RATIONAL DISCOVERY OF DENGUE TYPE 2 NON-COMPETITIVE INHIBITORS

HEH CHOON HAN

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF MEDICINE

UNIVERSITY OF MALAYA

KUALA LUMPUR

2012

Abstrak

Pelbagai kajian telah dijalankan dalam pembangunan terapeutik terhadap jangkitan dengue. Malah, kini, vaksin mahupun ejen anti-dengue yang berkesan masih belum ditemui. Dalam kajian ini, kami memberi tumpuan pada penemuan drug yang rasional terhadap ejen anti-dengue yang berpotensi dengan merujuk pada perencatan protease DEN-2 NS2B-NS3 yang tidak kompetitif. Sebuah model homologi, DH-1, untuk DEN-2 NS2B-NS3 (yang menggunakan kompleks protease West Nile Virus NS2B-NS3, 2FP7, sebagai rujukan) telah digunakan sebagai reseptor sasaran untuk reka bentuk tersebut. Pinostrobin, suatu flavanone, telah digunakan sebagai ligan piawai dalam kajian ini. Sejumlah 13,341 sebatian kecil, berstruktur yang mengandungi tulang belakang chalcone, flavanone and flavone, yang terdapat daripada pangkalan data yang bernama ZINC, telah digunakan untuk penyaringan maya. Keputusan penyaringan menghasilkan molekul yang mempunyai daya ikatan lebin kuat terhadap reseptor berbanding dengan ligan piawai. Cerakin perencatan terhadap aktiviti proteolitik DEN-2 NS2B-NS3 daripada sebatian terpilih yang berkedudukan tinggi dalam keputusan penyaringan menunjukkan perencatan lebin baik yang signifikan berbanding dengan ligan piawai, pinostrobin. Kesimpulannya, melalui cara rasional, kami telah berjaya menemui perencat yang berpotensi sebagai langkah pertama penemuan ejen anti-dengue. Sebatian 1 telah didapati menunjukkan aktiviti perencatan yang terbaik dengan $K_{i\,exp}$ *in vitro* bernilai 69 ± 9 μM. Kami juga mencadangkan satu sebatian petunjuk atau farmakofor yang berpotensi untuk perencat tidak kompetitif terhadap protease DEN-2 NS2B-NS3.

Abstract

Various works have been carried out in developing therapeutics against dengue infections. However, to date, no effective vaccine or anti-dengue agent has yet been discovered. In this study, we focused on rational drug discovery of potential anti-dengue agents based on non-competitive inhibition of DEN-2 NS2B-NS3 protease. A suitable DEN-2 NS2B-NS3 homology model DH-1 (using West Nile Virus NS2B-NS3 protease complex, 2FP7, as template) was used as the target receptor for the design. Pinostrobin, a flavanone, was used as the standard ligand in this study. A total of 13,341 small compounds, with the backbone structures of chalcone, flavanone and flavone, available from ZINC database were used in the virtual screening performed. Ranking of resulting compounds yielded those with higher binding affinities compared to the standard ligand. Inhibition assay of selected top ranking compounds against DEN-2 NS2B-NS3 proteolytic activity resulted in significantly better inhibition compared to the standard, pinostrobin. In conclusion, through rational approach, we have been able to discover potential inhibitors in our early step towards discovering anti-dengue agents. Compound 1 was found to exhibit the best inhibition activity with *in vitro* $K_{i\,exp}$ value of 69 ± 9 μM. We also suggested a potential lead structure or pharmacophore for non-competitive inhibitor against DEN-2 NS2B-NS3 protease.

Table of Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

ΔG_{dock}: Estimated Mean Free Energy of Binding

µL: Microliter

1D: One Dimensional

3D: Three Dimensional

A: Adenine

AMC: 7-Amino-4-methylcoumarin

APS: Ammonium Persulfate

Asp: Aspartic Acid

bp: Base pair

BSA: Bovine Serum Albumin

C: Cytosine

CBB R-250: Coomassie Brilliant Blue R-250

cDNA: Complimentary Deoxyribonucleic Acid

CF40.gly(T).NS3pro: His-tagged Recombinant Soluble Functional NS2B-NS3 Protease with $Gly_4ThrGly_4$ as flexible glycine linker

CF40: Hydrophilic NS2B

cm: Centimeter

DEN-1: Dengue Virus Type 1

DEN-2: Dengue Virus Type 2

DEN-3: Dengue Virus Type 3

DEN-4: Dengue Virus Type 4

dH_2O: Distilled Water

DMSO: Dimethylsulfoxide

DNA: Deoxyribonucleic Acid

dNTP: Deoxyribonucleotide Triphosphate (dNTP)

EDTA: Ethylenediaminetetraacetic Acid

g: Gram

G: Guanine

$Gly_4SerGly_4$:

Glycine-Glycine-Glycine-Glycine-Serine-Glycine-Glycine-Glycine-Glycine

$Gly_4ThrGly_4$:

Glycine-Glycine-Glycine-Glycine-Threonine-Glycine-Glycine-Glycine-Glycine

GT_F1: Forward Primer with Glycine Linker (Thr) at 5' for NS3 -

   5'-GGGGGCGGA<u>GGTACC</u>GGTGGAGGCGGGGCTGGAGTATTGTGGGA-

   T-3'

GT_R1: Reverse Primer with Glycine Linker (Thr) at 3' for NS2B -

   5'-CCCGCCTCCACC<u>GGTACC</u>TCCGCCCCCCAGTGTTTGTTCTTCCTC-3'

HCl: Hydrochloric acid

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid

His: Histidine

IF: Interaction Frequency

IPTG: Isopropyl-β-D-thiogalactopyranose

KCl: Potassium Chloride

$K_{i\,dock}$: Estimated Inhibition Constant

$K_{i\,exp}$: Experimental Inhibition Constant

L: Liter

LB: Lysogeny Broth

mg: Miligram

$MgCl_2$: Magnesium Chloride

mL: Milliliter

NaCl: Sodium Chloride

NaOH: Sodium Hydroxide

Ni^{2+}: Nickel ion (2+)

NTA: Nitrilotriacetic Acid

NS2B: Non-structural Protein 2B

NS2B_F1: Forward Primer for NS2B - 5'-CCG<u>GGATCC</u>GCCGATTTGGAACTG-3'

NS3: Non-structural Protein 3

NS3_R1: Reverse Primer for NS3 - 5'-CCC<u>AAGCTT</u>CAATTTTCTCTTTCG-3'

NS3pro: NS3 Serine Protease

NumCl: Number in Cluster

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

RAM: Random Access Memory

RFU: Relative Fluorescence Units

RMSD: Root Mean Square Deviation

RNA: Ribonucleic Acid

RNase: Ribonuclease

RT-PCR: Reverse Transcription Polymerase Chain Reaction

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Ser: Serine

T: Thymine

TBE: Tris/Borate/EDTA

TBS: Tris-Buffered Saline

TBST: Tris-Buffered Saline-Tween

TEMED: *N,N,N',N'*-Tetramethylethylenediamine

Thr: Threonine

UCLA: University of California, Los Angeles

UV: Ultraviolet

WNV: West Nile Virus

CHAPTER ONE

INTRODUCTION

1.1 General Introduction

Dengue virus causes diseases such as dengue fever, dengue haemorrhagic fever and dengue shock syndrome. It is among the major causes of morbidity and mortality, especially in children in many endemic Asian and South American countries (Gubler, 1998; Guha-Sapir & Schimmer, 2005). However, to date, there is no effective vaccine or anti-viral drug available in the market to protect against dengue diseases.

The protease, NS2B-NS3 (protease complex), of dengue virus type 2 (DEN-2) was reported to be involved in the cleavage of most of the non-structural proteins needed in viral replication. Inhibition of NS2B-NS3 protease complex is believed to suppress viral infections. It was previously reported that cardamonin (a chalcone) and pinostrobin (a flavanone) showed non-competitive inhibition towards DEN-2 NS2B-NS3 proteolytic activities, while panduratin A and 4-hydroxypanduratin A (both cyclohexenyl chalcone derivatives) showed competitive inhibition activities (Kiat *et al.*, 2006).

In this study, the models of DEN-2 NS2B-NS3 were studied computationally (*in silico*) using cardamonin, *R*-pinostrobin and *S*-pinostrobin to verify the suitability of the models as target receptors for non-competitive inhibition studies. Virtual screening of a series of small compounds from the ZINC database (Irwin & Shoichet, 2005) with backbone structures similar to chalcone, flavanone and flavone were then performed towards the suitable DEN-2 NS2B-NS3 model in an attempt to discover potential non-competitive inhibitors. The selected compounds were then submitted to DEN-2 NS2B-NS3 protease cleavage inhibition assay to validate their activities *in vitro*. A novel anti-dengue candidate was then obtained from the *in silico* and *in vitro* results.

1.2   Problem Statement

Dengue fever has been one of the most common diseases in the tropical and subtropical countries. It is reported to be one of the most fatal diseases in Malaysia (KPKM, 2012). To date, there is still no licensed vaccine for prevention, or anti-viral drug for treatment of dengue infections. The computational molecular modelling approach for new compound development is still in its early stage in Malaysia. It is a tool that can be used to design new compounds and simulate their potential activities on particular proteins by referring to their inter-molecular interactions. Thus, in this study, *in silico* molecular modelling and docking approaches were used to verify the suitability of DEN-2 NS2B-NS3 protease models for non-competitive inhibition studies. Screening of compounds using both *in silico* and *in vitro* methods were also performed, in an effort to search for compounds with activities against viral infections. The whole study design involved in this research can be considered as a holistic approach towards drug discovery of new compounds with expected therapeutic activities. The involvements of computational and experimental techniques in this study are essential in our effort towards the search for therapeutic drugs against dengue virus infections. Hence, it is believed that the approaches used in this study will be beneficial to the advancement of drug development area of research in the country.

1.3   Study Aim and Objectives

The aim of this study is to obtain a suitable model for non-competitive inhibition study towards DEN-2 NS2B-NS3 protease, and use this model in the rational discovery of new potential anti-dengue agents.

1.3.1 Study Objectives

i. To obtain the suitable DEN-2 NS2B-NS3 protease models for *in silico* docking study in search of hits for non-competitive inhibition activity.

ii. To perform virtual screening of chalcones, flavanones, flavones and other compounds which are structurally similar towards DEN-2 NS2B-NS3 protease.

iii. To clone, express and purify recombinant DEN-2 NS2B-NS3 protease and perform DEN-2 NS2B-NS3 protease assay on selected top ranking compounds from virtual screening experiment.

iv. To perform structure-activity relationship (SAR) studies for the biologically screened potential non-competitive inhibitors.

1.4 Research Project Workflow
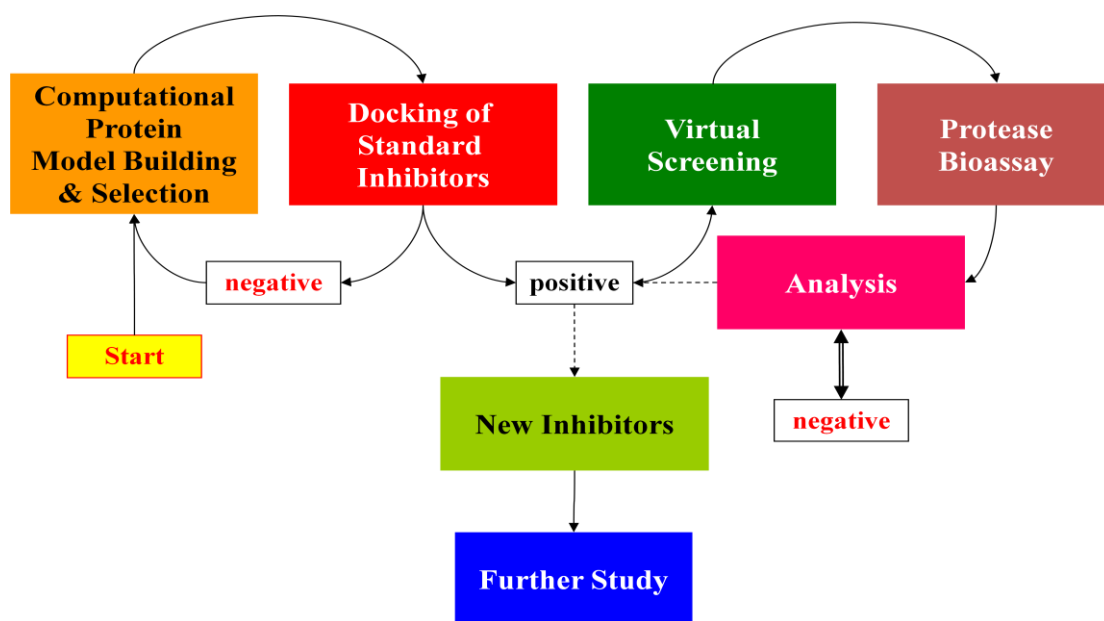
Figure 1.1 Workflow of overall research project.

CHAPTER TWO

BACKGROUND / LITERATURE REVIEW

2.1 Dengue

2.1.1 Dengue Epidemiology in the World and Malaysia

Every year, the World Health Organization (WHO) estimates that there are 50 - 100 million cases of dengue infection worldwide. Over 2.5 billion people (over 40% of the world's population) who live in tropical or sub-tropical regions of the world (as shown in Figure 2.1) are at risk from dengue infections, and about 1.8 billion of these people (more than 70%) reside in Asia Pacific countries (WHO, 2008; WHO, 2012a).
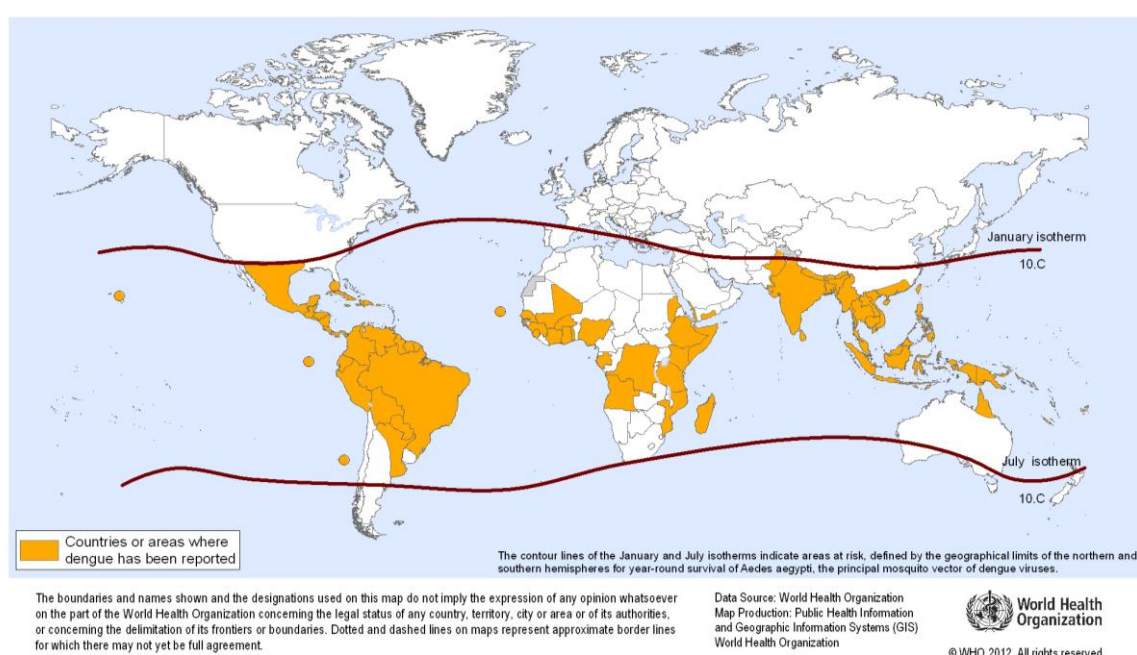
Figure 2.1 Countries or areas of the world that are at risk of dengue infection (WHO, 2011b). The two dark red lines indicate latitudes around 35 ^{o}N and 35 ^{o}S.

Based on Figure 2.2, in 2010, it was reported that there were 354,009 cases with 1,075 deaths (CFR 0.30%) in the Western Pacific Region. It is the highest reported number of cases since the large dengue outbreaks in 1998 (WHO, 2012b).
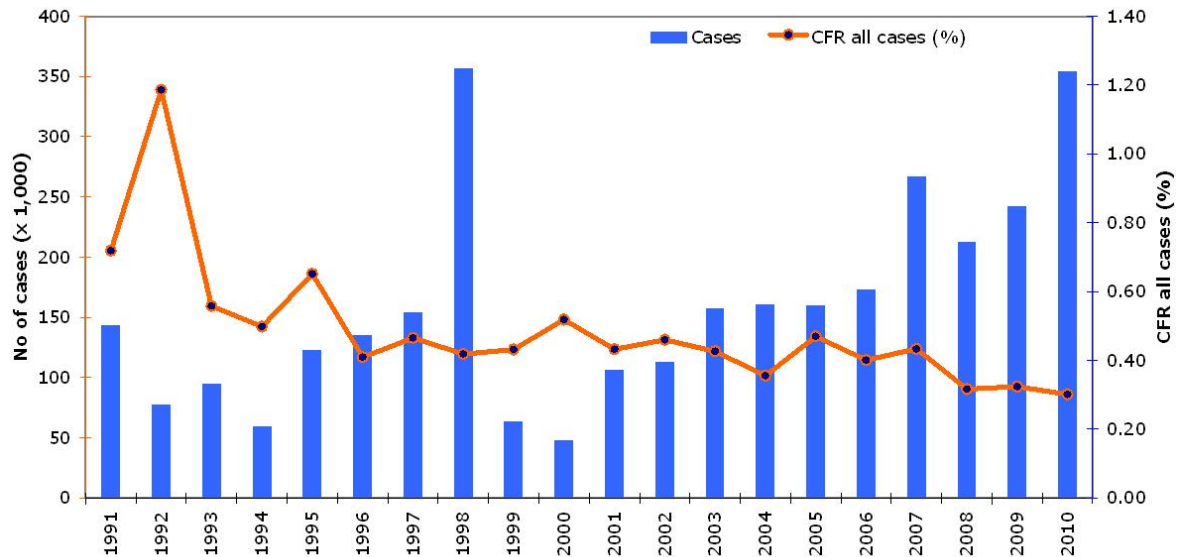
Figure 2.2 The number of reported dengue cases (DF/DHF cases) and Case Fatality Rates (CFR) in the Western Pacific Region from 1991 - 2010 (WHO, 2011a).

In the year 2009, Malaysia was reported to have 41,486 dengue cases with 88 deaths (CFR 0.21%); while in 2010 and 2011, 46,171 cases with 134 deaths (CFR 0.29%) and 19,884 cases with 36 deaths (CFR 0.18%) were reported respectively (KPKM, 2011b; KPKM, 2011a). From the beginning of the year 2012 up to the 5th week of 2012, there were already 2,413 dengue cases with 10 deaths (CFR 0.41%) reported in Malaysia (KPKM, 2012), which marks the highest CFR in recent years and reaches a worrying situation. In 2011, ASEAN Health Ministers had even declared 15 of June to be ASEAN Dengue Day as an important annual event that allows the ASEAN members, in coordination with the WHO, to reinforce dengue prevention and control measures (WHO, 2011a).

2.1.2 Dengue Transmission

There are two predominant arthropod vectors, *Aedes aegypti* and *Aedes albopictus* mosquitoes (Figure 2.3), that are implicated in the transmission of dengue virus (Simmons *et al.*, 1931; Hammon *et al.*, 1960; Gould *et al.*, 1968; Rosen, 1983). People in tropical and subtropical regions of the world are at risk of dengue infection because the mosquitoes are widely in such regions, mostly between latitudes 35°N and 35°S, corresponding approximately to a winter isotherm of 10°C (Figure 2.1) (WHO, 2009). After virus incubation for 4 to 10 days, the infected mosquitoes are capable of transmitting the virus to humans during probing and blood feeding for the rest of their lives (WHO, 2012a).

Meanwhile, infected humans are the main carriers, multipliers and sources of the virus for uninfected mosquitoes (WHO, 2012a). Various methods involving environmental management and usage of pesticide had been introduced for vector control and dengue prevention (WHO, 2009). The types of environmental management include environmental modification by installation of reliable piped water supply to communities to reduce vector larval habitats; environmental manipulation by frequent emptying and cleaning of water containers to manipulate the vector habitats; and changes to human habitation or behaviour by installing mosquito screens on windows, doors and other entry points, and using mosquito nets while sleeping to reduce human-vector contact (WHO, 2009).

Aedes aegypti   *Aedes albopictus*

Figure 2.3   Images of the two predominant arthropod vectors of dengue virus, *Aedes aegypti* and *Aedes albopictus* (images are adopted from the Centers for Disease Control and Prevention (CDC)'s Public Health Image Library (PHIL), with identification number, 9258 and 2165, respectively).

2.1.3   Dengue Characteristic

Symptomatic dengue virus infection can be categorized into three categories, which are undifferentiated fever, dengue fever (DF) and dengue haemorrhagic fever (DHF), whilst DHF, can be further classified into four severity grades, with grades III and IV being defined as dengue shock syndrome (DSS) (WHO, 2009). DF is a flu-like illness with a variety of nonspecific signs and symptoms and should be suspected when a high fever (40°C) concurrently appearing with two of the following symptoms: severe headache, pain behind the eyes, nausea, vomiting, muscle and joint pains, swollen glands or rashes (Gubler, 1998; WHO, 2012a). These symptoms usually occur following an incubation period of 3 - 14 days after the infective mosquito bite (WHO, 2012a). On the other hand, DHF is a potentially deadly complication that is characterized by high fever and can cause haemorrhagic manifestations, which may lead to DSS (Gubler, 1998; WHO, 2012a). The warning signs such as severe abdominal pain, rapid breathing, bleeding gums, restlessness, fatigue, persistent vomiting and/or blood

in vomit appear 3 - 7 days after the first symptoms along with a decrease in temperature (below 38 ^{o}C) (WHO, 2012a). Patients in shock may die within 8 to 24 hours but usually recover following anti-shock therapy (Gubler, 1998). As a result, various studies have been performed for the development of therapeutics against dengue, including the use of anti-viral inhibitors (Leung *et al.*, 2001; Hrobowski *et al.*, 2005; Whitby *et al.*, 2005; Kiat *et al.*, 2006; Yin *et al.*, 2006a; Yin *et al.*, 2006b; Lee *et al.*, 2007; Lescar *et al.*, 2008; Othman *et al.*, 2008; Frecer & Miertus, 2010; Muhamad *et al.*, 2010; Frimayanti *et al.*, 2011) and vaccines (McKee *et al.*, 1987; Durbin *et al.*, 2001; Whitehead *et al.*, 2003; Hanley *et al.*, 2004; Durbin *et al.*, 2005; Robert Putnak *et al.*, 2005; Edelman, 2007).

However, to date, there is no effective vaccine or anti-viral drug available in the market to protect against dengue (Monath, 1994; Kautner *et al.*, 1997). Nevertheless, early detection and access to proper medical care, such as maintenance of the patient's body fluid volume, could lower fatality rates from more than 20% to less than 1% (WHO, 2012a).

2.1.4 Dengue Virus Taxonomy

Dengue viruses belong to the Flavivirus genus, member of the Flaviviridae family (Westaway *et al.*, 1985). In fact, there are more than 70 viruses belonging to Flavivirus genus and many of them are arthropod-borne human pathogens, such as dengue virus, Japanese encephalitis virus (JEV), West Nile virus (WNV), and yellow fever virus (YFV) (Lindenbach *et al.*, 2007).

The dengue viruses can be categorized into four serotypes based on the antigens on the surface of the virus, which are dengue virus type 1, 2, 3 and 4 (DEN-1 to DEN-4) (Lanciotti *et al.*, 1992; Siqueira *et al.*, 2005; Lindenbach *et al.*, 2007). As members of the Flaviviridae family, all the dengue viruses share common morphological characteristics, genome structures, and replication and translation strategies (Westaway *et al.*, 1985; Kautner *et al.*, 1997; Gubler, 1998; Lindenbach *et al.*, 2007).

2.1.5 Dengue Virus Morphology and Life Cycle

The virus particle (virion) is about 50 nm in diameter, surrounded by structural proteins - envelope (E) protein and membrane (M) protein on the surface, with nucleocapsid, which consists of capsid (C) protein and genomic RNA, in the inner part (Figure 2.4) (Lindenbach *et al.*, 2007).
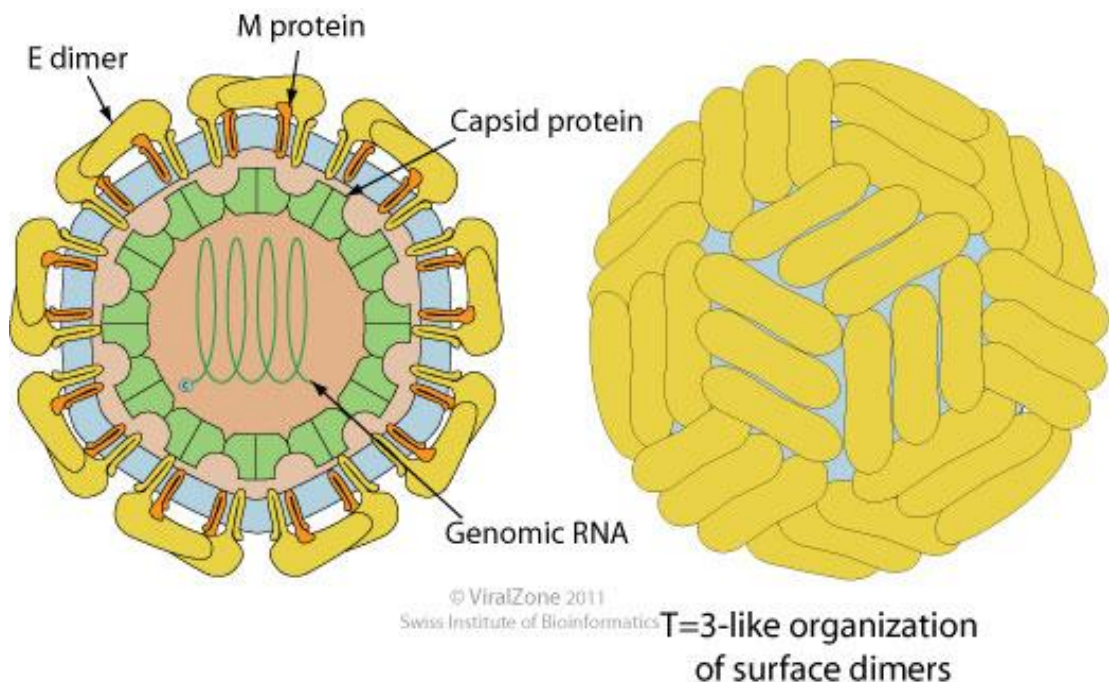
Figure 2.4   Illustrated virions of flavivirus by ViralZone (http://viralzone.expasy.org) based on cryo-electron microscopy images by Zhang *et al*. (2003) (images are adopted from http://viralzone.expasy.org/all_by_species/43.html).

The dengue virus life cycle is shown in Figure 2.5B. Dengue virus, as well as other flaviviruses, uses cell surface receptor-mediated endocytosis for cell entry (Rodenhuis-Zybert *et al.*, 2010). The process is followed by membrane fusion of the viral membrane with the host cell membrane, which is catalyzed by acidic pH of the environment, to uncoat the nucleocapsid and release the viral RNA genome into the cell cytosol. Subsequently, the RNA genome is translated as a single polyprotein by the host ribosomes, which translocate across the ER membrane (Figure 2.5) (Lindenbach *et al.*, 2007; Rodenhuis-Zybert *et al.*, 2010). The polyprotein is then processed co- and post-translationally by the cellular (host) and virus-derived proteases into three structural proteins and seven nonstructural proteins (Svitkin *et al.*, 1984; Markoff, 1989; Lindenbach *et al.*, 2007; Rodenhuis-Zybert *et al.*, 2010). Then, the nonstructural proteins initiate the replication of viral RNA genome right after the protein translation and folding of the individual proteins (Clyde *et al.*, 2006). The replication of RNA is catalyzed by virus replicase, which associates with membranes through interactions involving nonstructural proteins, viral RNA and probably some host factor (Lindenbach *et al.*, 2007). After this, the newly synthesized RNA is packaged by the C protein to form a nucleocapsid, while the prM and E proteins form heterodimers that are oriented into the lumen of the ER. Immature virion budding subsequently takes place through encapsulation of nucleocapsid by the prM/E heterodimers. However, the engulfment mechanism of the nucleocapsid by prM/E proteins is still unclear (Rodenhuis-Zybert *et al.*, 2010). Then, the immature virions formed in the ER will be transported and released through the host secretory pathway by travelling to the Golgi compartment and maturing in the secretion (Clyde *et al.*, 2006; Lindenbach *et al.*, 2007; Rodenhuis-Zybert *et al.*, 2010). Mature virion are formed and able to infect new cells when prM protein is cleaved into soluble pr peptide and M protein by host protease, furin (Stadler

et al., 1997; Keelapang *et al.*, 2004; Clyde *et al.*, 2006; Lindenbach *et al.*, 2007; Rodenhuis-Zybert *et al.*, 2010).
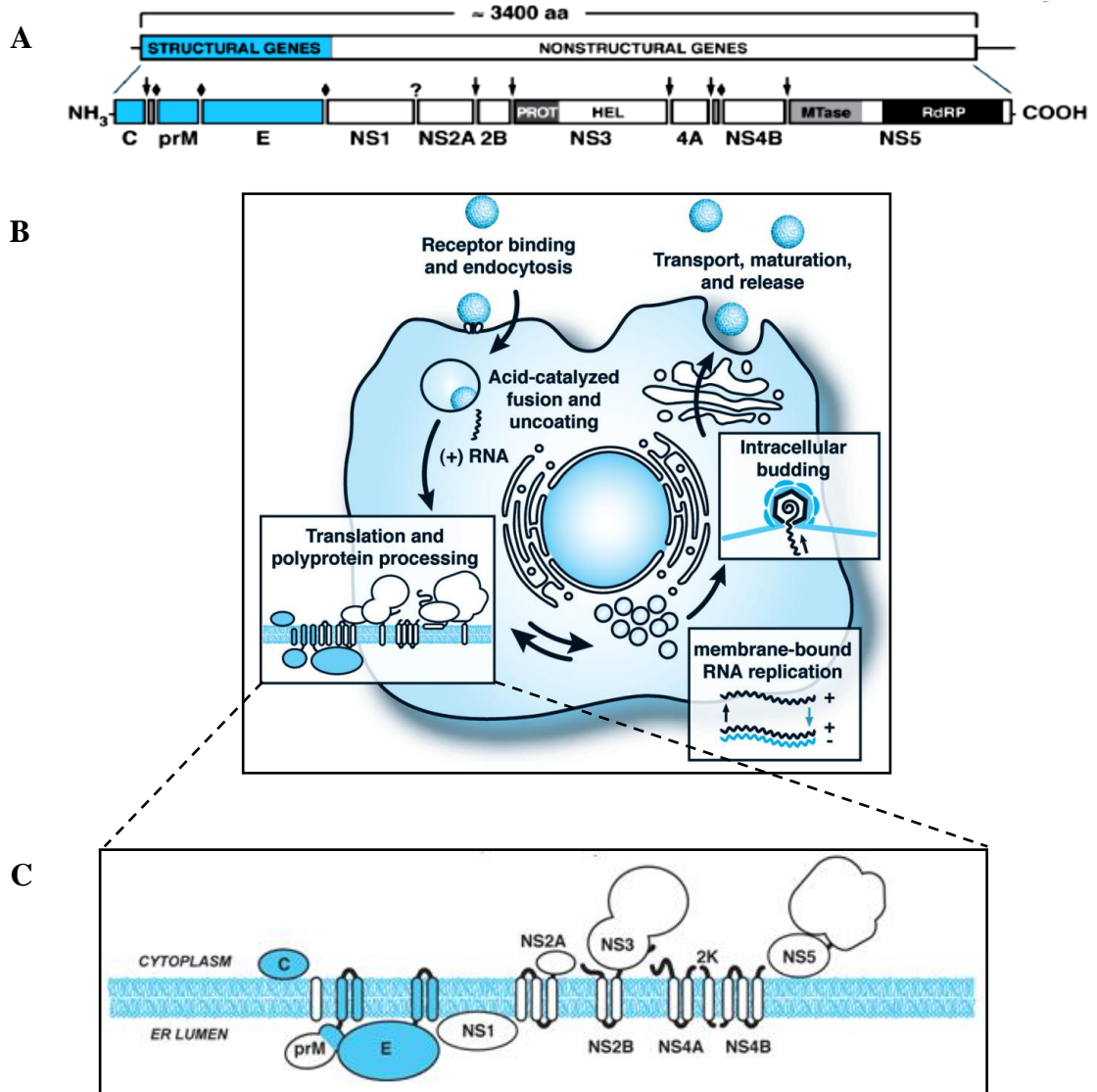
Figure 2.5   Polyprotein of dengue virus. **A:** Polyprotein processing and cleavage products of dengue virus. Structural proteins are coloured in cyan and nonstructural proteins are coloured in white. ♦ indicates cleavage sites for host signal peptidase, ↓ indicates cleavage sites for viral serine protease and ? indicates cleavage site for unknown protease. **B:** Dengue virus life cycle. **C:** The proposed topology of the flavivirus polyprotein cleavage products with respect to the endoplasmic reticulum (ER) membrane (Lindenbach *et al.* 2007).

2.1.5.1 Dengue Virus RNA

The viral genomic RNA is a single-stranded RNA of positive polarity with about 11 kb and type 1 cap at the 5'-end but lacking the poly A tail at the 3'-end (Russell *et al.*, 1980). The RNA gene order for a single polyprotein precursor of 3,391 amino acids is C-prM(M)-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5, consisting of three structural (C, prM and E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) as shown in Figure 2.5A (Irie *et al.*, 1989).

2.1.5.2 Structural Proteins

The C protein is a highly basic protein of about 11 kDa that serves to encapsulate the viral RNA genome for protecting and disseminating the viral RNA to suitable hosts (Lucas, 2001; Lindenbach *et al.*, 2007). The prM protein, about 18 kDa, is the glycoprotein precursor of M protein with a major function of preventing E protein from undergoing acid-catalyzed rearrangement to the T=3 fusogenic form during transport through the host secretory pathway, that would lead to premature membrane fusion of virus particle towards the Golgi compartment before release from the cell as secretion (Stadler *et al.*, 1997; Keelapang *et al.*, 2004; Clyde *et al.*, 2006; Lindenbach *et al.*, 2007; Rodenhuis-Zybert *et al.*, 2010). The M protein, about 8 kDa, is a small proteolytic fragment of the precursor prM protein, produced after being cleaved by host protease, furin, during maturation of nascent virus particles within the host secretory pathway after the soluble pr peptide is cleaved from prM (Stadler *et al.*, 1997; Keelapang *et al.*, 2004; Lindenbach *et al.*, 2007). Meanwhile, the functions of E protein (about 50 kDa) are to mediate binding and fusion during virus entry, and it is the main antigenic determinant on the virus particle as the target for immune system (Lindenbach *et al.*, 2007).

2.1.5.3 Nonstructural Proteins

The NS1 glycoprotein, is about 46 kDa in size (Lindenbach *et al.*, 2007). It was reported to be an important antigen along with E protein in immunization of mice for human vaccine development (Zhang *et al.*, 1988; Srivastava *et al.*, 1995). However, in recent year, it had been used for the early detection of dengue virus infection (Xu *et al.*, 2006; Hang *et al.*, 2009; Datta & Wattal, 2010; Lima Mda *et al.*, 2011). One of the examples of commercially available dengue NS1 detection kits is Platelia Dengue NS1 Ag (Bio-Rad).

The NS2A protein is a hydrophobic protein with size of about 22 kDa (Lindenbach *et al.*, 2007). It plays an important role in virus assembly (Leung *et al.*, 2008). On the other hand, the NS2B protein, which is about 14 kDa, is a membrane-associated protein. It forms a stable complex with NS3 and also acts as a cofactor for the NS2B-NS3 serine protease (Lindenbach *et al.*, 2007).

The NS3 protein is about 70 kDa, a large multifunctional protein, containing several functions required for polyprotein processing and RNA replication (Lindenbach *et al.*, 2007). As referred to Figure 2.5A, it consists of a trypsin-like serine protease domain within the N-terminal 180 residues (Bazan & Fletterick, 1989) and a domain with NTPase/helicase acitivity at the C-terminal (Li *et al.*, 1999; Luo *et al.*, 2008). The active site of NS3 serine protease carries the catalytic triad, comprising of three amino acid residues, namely His51, Asp75 and Ser135, with NS2B acting as a cofactor of NS3 protease for optimal catalytic activity (Preugschat *et al.*, 1990; Arias *et al.*, 1993). The NS2B-NS3 protease complex was reported to be responsible for cleaving the NS2A/NS2B, NS2B/NS3, NS3/NS4A and NS4B/NS5 junctions, preferentially with adjacent basic residues (Preugschat *et al.*, 1990; Falgout *et al.*, 1991; Cahour *et al.*, 1992;

Zhang *et al.*, 1992). Besides, the protease also generates the C-termini of mature C protein (Amberg *et al.*, 1994; Yamshchikov & Compans, 1994) and NS4A (Lin *et al.*, 1993). Meanwhile, the role for NS3 NTPase/helicase activity in flavivirus life cycle has been shown in the genomic RNA replication in an unwinding step, where the NS3 NTPase/helicase catalyzes the hydrolysis of ATP that is required for the unwinding of the double-stranded RNA for RNA replication (Li *et al.*, 1999).

NS4A and NS4B are both hydrophobic proteins of about 16 kDa and 27 kDa in sizes (Lindenbach *et al.*, 2007). The two proteins are membrane-associated and were reported to play an important role in RNA replication by colocalization with replication complexes, which involved NS3 (Preugschat & Strauss, 1991; Lin *et al.*, 1993; Umareddy *et al.*, 2006). As shown in Figure 2.5A and 2.5C, there is a signal peptide, designated 2k fragment, located at the C-terminal region of NS4A, serving as a signal sequence for the translocation of the adjacent NS4B into the ER lumen (Miller *et al.*, 2007).

NS5 is a multifunctional protein, about 103 kDa in size, with methyltransferase (MTase) and RNA-dependent RNA polymerase (RdRp) activities (Figure 2.5A) (Lindenbach *et al.*, 2007). The NS5 MTase is responsible for methylating the viral RNA cap structure to cap-1 structure, which further is being recognized for polyprotein translation (Zhou *et al.*, 2007). On the other hand, the NS5 RdRp catalyzed the viral replication by synthesizing a transient double-stranded replicative RNA intermediate which consists viral plus- and minus-strand RNAs (Bartholomeusz & Thompson, 1999; Yap *et al.*, 2007). The newly synthesized minus strand is subsequently used as a template for synthesizing additional plus-strand RNAs (Bartholomeusz & Thompson, 1999; Yap *et al.*, 2007).

2.1.6   The DEN-2 NS2B-NS3 Protease

Until now, many studies on DEN-2 had been carried out for better understanding of its replication mechanism. Amongst these studies, DEN-2 nonstructural proteins were some of the common research targets (Falgout *et al.*, 1991; Brinkworth *et al.*, 1999; Yusof *et al.*, 2000). As mentioned in section 2.1.5.3, the NS2B-NS3 virus protease complex was reported to play a very important role in cleaving most of dengue virus nonstructural proteins and some structural proteins (Falgout *et al.*, 1991; Yusof *et al.*, 2000) which will further complete the virus replication cycle (Lindenbach *et al.*, 2007). Thus, in this study, the inhibition of DEN-2 NS2B-NS3 protease was targeted for drug discovery.

2.2   Plant Extracts Against Dengue Infection

Dengue is a famous disease among the tropical and subtropical regions of the world and the native people of these regions have used various traditional herbal remedies to treat the disease. Since 1960s, it was reported that aqueous agar extracts give inhibition effects on infectious and hemagglutinating properties of DEN-2 (Schulze & Schlesinger, 1963). In recent years, various studies of plant extracts against dengue infection had been carried out, which included the use of plant extracts as mosquito larvicidal agents (Wandscheer *et al.*, 2004; Chowdhury *et al.*, 2008; Kumar *et al.*, 2010; Kalaivani *et al.*, 2012; Kovendan *et al.*, 2012; Mahesh Kumar *et al.*, 2012; Marimuthu *et al.*, 2012), mosquito repellents (Rajkumar & Jebanesan, 2010), plant-derived vaccines (Malabadi *et al.*, 2011), treatment for dengue fever (Ahmad *et al.*, 2011) and also anti-viral agents (Kiat *et al.*, 2006; Muliawan *et al.*, 2006; Jain *et al.*, 2008; Muhamad *et al.*, 2010; Tang *et al.*, 2012).

2.2.1 *Boesenbergia rotunda* Extract

Boesenbergia rotunda (L.) Mansf. (synonym of *Boesenbergia pandurata*), which is known as fingerroot, Chinese ginger (China and Southeast Asia) or "temu kunci" (Malaysia and Indonesia) (Porcher, 2003), is a common spice and herb belonging to the Zingiberaceae (ginger) family. Its extract has been reported to contain essential oils (Ultee, 1957) and various small compounds such as boesenbergin, cardamonin, pinostrobin, pinocembrin, alpinetin, panduratin A, 4-hydroxypanduratin A, 5,7-dimethoxyflavone and 1,8-cineole (Jaipetch *et al.*, 1982; Pancharoen *et al.*, 1987; Kiat *et al.*, 2006). These small compounds exhibit anti-bacterial (Ungsurungsie *et al.*, 1982), anti-fungal (Taweechaisupapong *et al.*, 2010), anti-inflammatory, anti-oxidant (Panthong *et al.*, 1989; Tuchinda *et al.*, 2002; Shindo *et al.*, 2006; Isa *et al.*, 2012), anti-ulcerogenic (Abdelwahab *et al.*, 2011), anti-tumor (Morikawa *et al.*, 2008), anti-HIV (Tewtrakul *et al.*, 2003; Cheenpracha *et al.*, 2006) and anti-dengue activities (Kiat *et al.*, 2006). As DEN-2 NS2B-NS3 protease inhibitors, cardamonin, pinostrobin, panduratin A and 4-hydroxypanduratin A have shown good inhibition activities (Kiat *et al.*, 2006).

2.2.2 Competitive and Non-competitive Inhibitors of DEN-2 NS2B-NS3 Protease

Competitive inhibitors competed with the active substrates for the same binding site (the active site). Thus, the concentration of the inhibitors is a variable that needs to be monitored since too high a concentration may be toxic to the human body. For this reason, our study focuses on the non-competitive inhibition of DEN-2 NS2B-NS3 protease, where the inhibition of ligand happens at binding site other than the active site. In this case, the inhibitor will not have to compete with the existing substrate for the binding site.

Cardamonin (chalcone) and pinostrobin (flavanone) showed non-competitive inhibition towards DEN-2 NS2B-NS3 proteolytic activities, while panduratin A and 4-hydroxypanduratin A (both cyclohexenyl chalcone derivatives) showed competitive inhibition activities (Figure 2.6) (Kiat *et al.*, 2006). As non-competitive inhibitors, cardamonin and pinostrobin were reported to show experimental inhibition constants ($K_{i\ exp}$) of 377 ± 77 μM and 345 ± 70 μM, respectively. Whereas the competitive inhibitors, panduratin A and 4-hydroxypanduratin A demonstrated $K_{i\ exp}$ values of 25 ± 8 μM and 21 ± 6 μM, respectively (Kiat *et al.*, 2006). In this study, the reported small compounds, cardamonin and pinostrobin (with non-competitive inhibition activities) were used as standards in our effort to develop anti-dengue agents.

2.3   *In Silico* (Computational) Study

Computational molecular modelling was first developed in the 1960's and since then it has become increasingly popular and is now frequently used in the molecular design field. Previous studies have reported the use of computational molecular modelling softwares for developing new molecule models (Ooms, 2000) in the development of new drugs such as local anesthetic drugs (Lipkind & Fozzard, 2005), anti-cancer agents (Bartulewic *et al.*, 2000), potential anti-malarials (Portela *et al.*, 2003), and anti-fungal drugs (Baginski *et al.*, 2005).
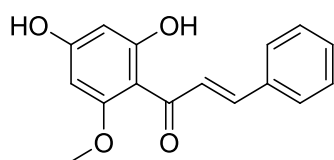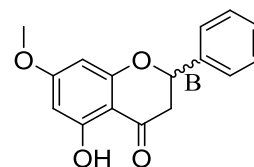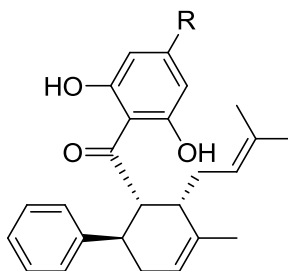
R = OMe   : Panduratin A

R = OH   : 4-hydroxypanduratin A

Figure 2.6 Structures of the small compounds which showed competitive and non-competitive inhibition activities towards DEN-2 NS2B-NS3 proteolytic activities (Kiat *et al.*, 2006).

2.3.1 Automated Docking

Automated docking is a computational simulation method for predicting an ideal binding site and orientation of a ligand towards a macromolecule target (protein receptor) via specific ligand-protein interaction formula and followed by scoring function to indicate the binding affinity (Lengauer & Rarey, 1996; Morris *et al.*, 1998). AutoDock 4.2 is an automated molecular docking software for calculating the free energy of binding (ΔG) in a implicit water environment using a semiempirical free energy force field, which includes six pair-wise evaluations (V) and an estimation of the

conformational entropy lost upon binding (ΔS_{conf}) (Huey *et al.*, 2007). ΔG is given by the following equation:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P})$$
$$+ \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}\right) \qquad \text{...(1)}$$

where L is ligand and P is protein. It is assumed that the ligand and the protein are sufficiently distant from one another in the unbound state that $V_{unbound}^{P-L}$ is zero. The disallowed motion in the protein results in the bound state of the protein (V_{bound}^{P-P}) being identical with the protein in the unbound state ($V_{unbound}^{P-P}$), which causes the difference in their intramolecular energy to be zero (Huey *et al.*, 2007). Thus, the formula for ΔG becomes:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + \left(V_{bound}^{P-L} + \Delta S_{conf}\right) \qquad \text{...(2)}$$

As for V, the pair-wise atomic terms include evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation (Huey *et al.*, 2007):

$$V = W_{vdw} \sum_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond} \sum_{ij} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right)$$
$$+ W_{elec} \sum_{ij} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{ij} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \qquad \text{...(3)}$$

where W_{vdw}, W_{hbond}, W_{elec} and W_{sol} are the weighting constants, which are also known as the free energy coefficients, and are optimized for each pair-wise terms to calibrate the empirical free energy based on a set of experimentally characterized complexes. For V^{L-L}, i and j are all pairs of atoms in the ligand that are separated by

three or more bonds while for V^{P-L}, i and j are pairs of ligand and protein atoms, respectively, with r_{ij} being the distance between atoms i and j (Morris *et al.*, 1998). A and B are parameters taken from the Amber force field (Weiner *et al.*, 1984), used for optimizing the first term based on the Lennard-Jones 12-6 potential for dispersion/repulsion interactions (Jones, 1924). C and D are parameters assigned to give a maximal well depth of 5 kcal/mol at 1.9 Å for O-H and N-H and a depth of 1 kcal/mol at 2.5 Å for S-H for the second term, a directional 12-10 hydrogen bonding term (Goodford, 1985) with directionality $E(t)$ dependent on the angle t away from ideal bonding geometry (Boobbyer *et al.*, 1989; Morris *et al.*, 1998). The third term in equation (3) is for electrostatic interactions, which is based on a screened Coulomb potential, where q is the partial atomic charge of an atom and $\varepsilon(r_{ij})$ is a distant-dependent dielectric constant (Mehler & Solmajer, 1991). The final term is a desolvation potential based on the volume (V) of the atoms surrounding a given atom, weighted by a solvation parameter (S) and an exponential term based on the distance, with the distance weighting factor σ set to 3.5 Å (Stouten *et al.*, 1993).

On the other hand, the term for the loss of conformational entropy upon binding (ΔS_{conf}) is directly proportional to the number of rotatable bonds in the molecule (N_{tors}) (Huey *et al.*, 2007):

$$\Delta S_{conf} = W_{conf} N_{tors} \qquad \text{...(4)}$$

where W_{conf} is the weighting constant for torsional term, and hence, equation (2) becomes:

$$\Delta G = V_{bound}^{L-L} - V_{unbound}^{L-L} + V_{bound}^{P-L} + W_{conf} N_{tors} \qquad \text{...(5)}$$

So, the estimated free energy of binding (ΔG) can now be summarized as:

Estimated free energy of binding (ΔG) = *Final Total Internal Energy – Unbound System's*

Energy + Final Intermolecular Energy + Torsional Free Energy ...(6)

which is the expression used in the AutoDock 4 series.

The estimated K_i between protein and ligand is calculated by:

$$K_i = e^{\Delta G / RT}$$ ...(7)

where R is the gas constant, 1.987 cal K^{-1} mol^{-1}, and T is the absolute temperature of body temperature, 310.15 K (Morris *et al.*, 1998).

2.3.2   The Lamarckian Genetic Algorithm in AutoDock

The Lamarckian genetic algorithm had been well described by Morris *et al.* (1998). It is a search algorithm available in AutoDock 3.0 software and later versions, for estimating the optimum free energy of binding of ligands with proteins in a given number of cycles (known as number of evaluations in AutoDock software). The Lamarckian genetic algorithm uses the concept of a genetic algorithm for a wider global search. The genetic algorithm is based on Darwinian evolution theory, where crossover with random pairs of individuals to produce new offspring, and random mutation of offspring will occur. The fitness of a particular generation is determined by the free energy of binding, in which the gene with better fitness will be selected for inheritance. In the genetic algorithm of molecular docking, the ligand's translation, orientation and conformation with respect to the protein correspond to the genotype (genetic

constitution), whereas the ligand's atomic coordinates correspond to the phenotype (characteristic as a result of genetic or environmental factor).

Besides the genetic algorithm for global search, Lamarckian genetic algorithm refines its search by implementing a more specific local search based on Lamarckian inheritance theory, where the characteristic of an individual resulting from an environmental factor can be passed on to its offspring. The concept of the Lamarckian genetic algorithm is illustrated in Figure 2.7. However, in molecular docking, a local search is carried on by continuously converting from the genotype to the phenotype, and thus, inverse mapping (as shown in Figure 2.7) is not required. Nevertheless, the parent's genotype will be replaced by the resulting genotype with better fitness that follows the Lamarckian inheritance theory.
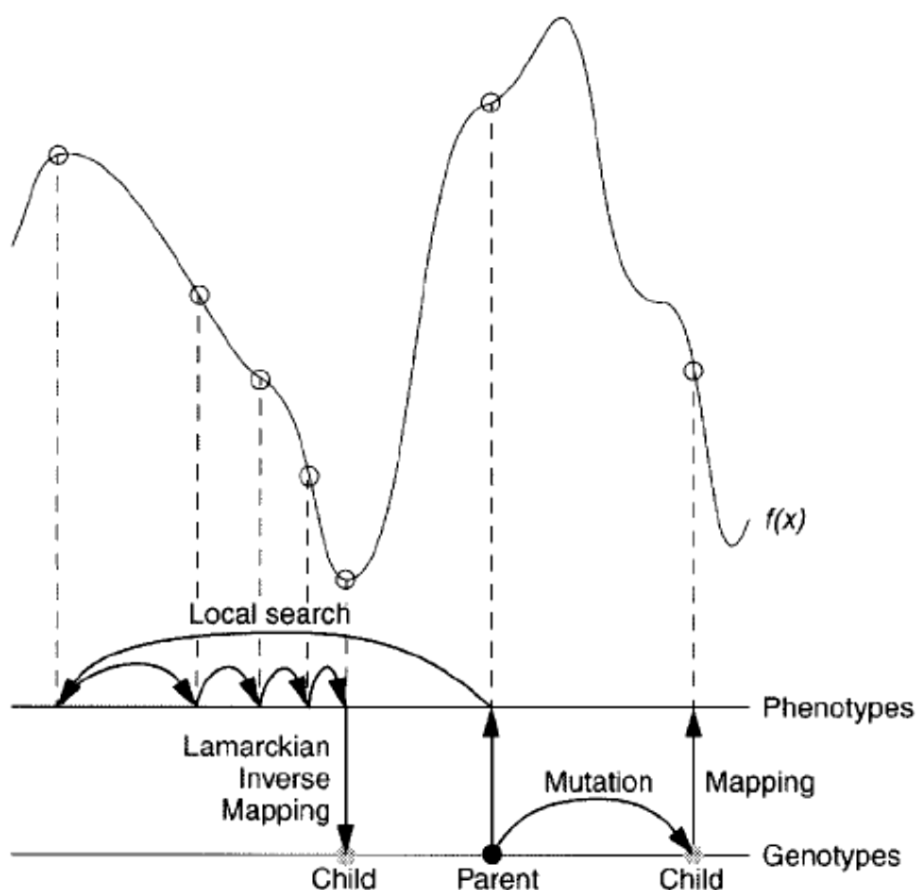
Figure 2.7 Genotypic and phenotypic search. It is based on global search with Darwinian evolution theory (right-hand side of the figure) and local search with Lamarckian inheritance theory (left-hand side) (Morris *et al.*, 1998). The result of the search is the fitness function, f(x).

2.3.3 AutoDock Virtual Screening and Zinc Database

Virtual screening, which is the use of computational tools to identify a reduced number of compounds from a large group with increased potential for bioactivity (Alvarez, 2004), has recently been contributing greatly to the drug discovery process (Schneider *et al.*, 2000; Bailey & Brown, 2001). AutoDock is a common and suitable tool for performing virtual screening (Li *et al.*, 2004; Park *et al.*, 2006). In virtual screening, large libraries of known or even unknown compounds (database of compounds) were usually being screened. Hence, ZINC, a free database of

commercially available compounds (Irwin & Shoichet, 2005) is a suitable database of compounds for virtual screening.

2.3.4 DEN-2 NS2B-NS3 Protease Models

For the *in silico* study, especially in molecular docking, a good DEN-2 NS2B-NS3 protease model that is suitable for the study of a specific type of inhibition (competitive or non-competitive) is really necessary. Currently, there is only one DEN-2 NS2B-NS3 protease crystal (PDB id: 2FOM) (Erbel *et al.*, 2006) available in the Protein Data Bank (PDB). However, other NS2B-NS3 protease crystals from West Nile Virus (WNV) (PDB id: 2FP7; 2GGV; 2IJO; 3E90) (Erbel *et al.*, 2006; Aleshin *et al.*, 2007; Robin *et al.*, 2009), DEN-1 (PDB id: 3L6P; 3LKW) (Chandramouli *et al.*, 2010) and DEN-3 (PDB id: 3U1I; 3U1J) (Noble *et al.*, 2012) have been reported. A few homology modelling studies of DEN-2 NS2B-NS3 have been performed using Hepatitis C Virus (HCV) NS3-NS4A (PDB id: 1JXP) (Yan *et al.*, 1998; Lee *et al.*, 2006), a mixture of NS2B from 2FP7, NS3 from 2FOM, 2FP7 and whole 2IJO (Wichapong *et al.*, 2010) as templates. Modeller software (Sali & Blundell, 1993) was used in these studies for homology modelling. Nevertheless, these studies focused on the active site (for competitive inhibition) of the protease.

In this study, nine models of DEN-2 NS2B-NS3 protease, namely 2FOM and eight homology models (DH-1 to DH-8) generated using 2FP7, 2GGV, 2IJO, 3E90, 3L6P, 3LKW, 3U1I and 3U1J as the templates were evaluated in order to obtain a suitable DEN-2 NS2B-NS3 protease model for non-competitive inhibition study.

2.3.5 Evaluation of Homology Model

The quality of homology model can be determined by using Ramachandran plot (Ramachandran *et al.*, 1963) and Verify3D software (Bowie *et al.*, 1991; Luthy *et al.*, 1992). Ramachandran plot (Figure 2.8) can be used to visualize backbone dihedral angels (Figure 2.9), ψ (psi) against φ (phi) of amino acid residues in protein structure. Based on Figure 2.8, the regions drawn with solid lines are regions allowed at standard full radii, dashed lines at reduced radii and dotted lines at relaxed tau angle (N-C^{α}-C). The edges of the Ramachandran plot continued right to left and bottom to top as dihedral angle values are circular and 0° is the same as 360°. Protein structure can then be predicted or validated according to the regions the amino acid residues fall into.

Procheck software (Laskowski *et al.*, 1993) was programed to generate Ramachandran plot for a specific protein. It summarizes the result of Ramachandran plot into favoured, allowed, generously allowed and disallowed region for the involving amino acid residues of the protein model. Protein model with more amino acid residues in the favoured region show better stereochemical quality while more amino acid residues in the disallowed region indicate more stereochemical errors of a model and thus low in quality.

Verify3D software analyzes the compatibility of folded 3D protein structure with its own 1D amino acid sequences by finding sequences that are most compatible with the environments of the residues in the 3D structure (Bowie *et al.*, 1991). These environments are categorized into, i) the area of the residue buried in the protein and inaccessible to solvent, ii) the fraction of side-chain area that is covered by polar atoms (O and N) and iii) the local secondary structure. By determining the environment class of a given position in a protein structure, a score is assigned for finding each of the 20

amino acid types at that position in some related protein structure. These scores are called 3D-1D scores and can be used in a sequence alignment algorithm to find the best alignment of amino acid sequences to the environment string. A sequence database is then derived from a collection of good structures and used as a reference for alignment and compatibility validation of new 3D structures according to the 3D-1D scores.
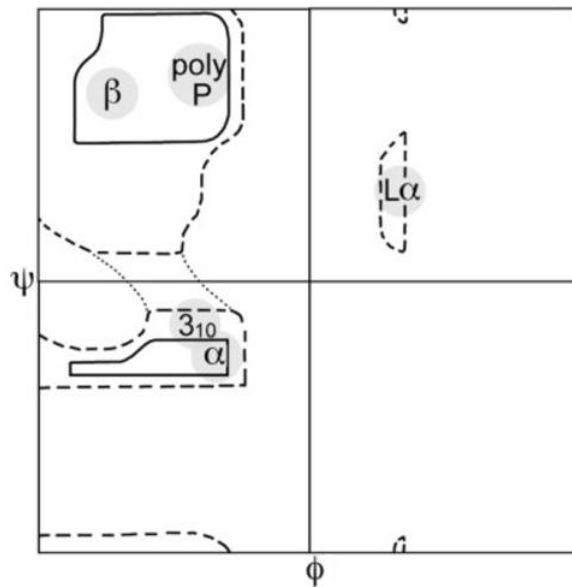
Figure 2.8   Outline example of original Ramachandran plot. Regions labeled with β for β-sheet, polyP for polyproline helix, $3_{10}\alpha$ for 3-10 helix and Lα for left-handed α helix.
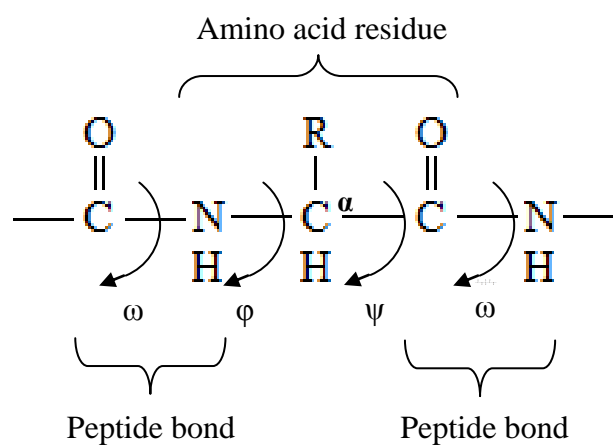
Figure 2.9   Dihedral angles (ψ, φ and ω) of a protein.

2.4   *In Vitro* Study

2.4.1   Cloning, Expression and Purification of DEN-2 NS2B-NS3 Protease

One of the earliest cloning experiments of the NS2B-NS3 protease was carried out by Chambers *et al.* (1990), for yellow fever virus (YFV). Subsequently, recombinant DEN-4 NS2B-NS3 protease was constructed and activated as two separate protein chains, NS2B cofactor and NS3 serine protease, after expression in mammalian cells (Falgout *et al.*, 1991). This was followed by recombinant DEN-2 NS2B-NS3 protease, which was expressed in mammalian cells (Zhang *et al.*, 1992; Clum *et al.*, 1997) or in bacterial cells (Yusof *et al.*, 2000) and purified into two protein chains as the active form (Clum *et al.*, 1997; Yusof *et al.*, 2000). Purification was conducted either by recognition of soluble fractions through inclusion of a FLAG monoclonal antibody (Hopp *et al.*, 1988) into the recombinant DEN-2 NS2B-NS3 protease (Clum *et al.*, 1997), or by applying Ni^{2+}-NTA-agarose His-tagged protein binding through the inclusion of hexahistidine tag into the protease (Yusof *et al.*, 2000). On the other hand, the deletion of hydrophobic regions of the NS2B leaving just the hydrophilic region with 40 amino acid residues relieved the membrane requirement for the activation of NS3 serine protease, while keeping the original activity of the recombinant DEN-2 NS2B-NS3 protease leaving very little precursor unprocessed (Clum *et al.*, 1997). Single chain recombinant DEN-2 NS2B-NS3 protease was then created by genetically fusing the hydrophilic NS2B cofactor region with the NS3 serine protease domain via a glycine linker between the two proteins (Leung *et al.*, 2001). This single chain recombinant DEN-2 NS2B-NS3 protease showed good protease activity as the glycine linker provided enough flexibility for the NS2B cofactor to interact with NS3 serine protease to fold into active form and was able to be purified in its native form as the NS2B-NS3 cleavage site was removed.

2.4.2 DEN-2 NS2B-NS3 Protease Assay (Activity and Inhibition Assays)

As the DEN-2 NS2B-NS3 protease was reported to be responsible for cleaving the NS2A/NS2B, NS2B/NS3, NS3/NS4A and NS4B/NS5 junctions, preferentially with adjacent basic residues as the recognition site (Preugschat *et al.*, 1990; Falgout *et al.*, 1991; Cahour *et al.*, 1992; Zhang *et al.*, 1992), the substrates for the protease assay were designed to have an amino acid sequence which resembled that of the cleavage site, containing dibasic amino acid residues (Clum *et al.*, 1997; Yusof *et al.*, 2000; Leung *et al.*, 2001; Chanprapaph *et al.*, 2005; Ganesh *et al.*, 2005; Kiat *et al.*, 2006; Yin *et al.*, 2006b; Yang *et al.*, 2011). However, different approaches such as radioisotope *trans*-[^{35}S]methionine labeled (Clum *et al.*, 1997), fluorogenic with 7-amino-4-methylcoumarin (AMC) attached (Yusof *et al.*, 2000; Ganesh *et al.*, 2005; Kiat *et al.*, 2006; Yin *et al.*, 2006b) and chromogenic with *para*-nitroaniline attached (Leung *et al.*, 2001; Chanprapaph *et al.*, 2005; Yang *et al.*, 2011) substrates were used in various studies. In this study, our protease assay was based on the assays from studies previously conducted and reported by our group, using Boc-Gly-Arg-Arg-AMC as a fluorogenic peptide substrate (Yusof *et al.*, 2000; Kiat *et al.*, 2006). Like other fluorophores, AMC needs to be excited by absorbing a certain wavelength of light or electromagnetic radiation in order to emit various wavelengths of fluorescent radiation (Harris, 2004).

2.4.3 Determination of Mechanisms of Enzyme Inhibition through Enzyme Kinetics

In order to determine the mechanisms of enzyme-substrate-inhibitor reactions (competitive or non-competitive inhibition), a study of the enzyme kinetics is necessary. Enzyme kinetics involves the measurement of the velocity of an enzyme-catalyzed reaction (rate of reaction) under varying conditions. A very common model is the Michaelis-Menten kinetics model, which consists of the parameters, V_{max}, the maximum enzyme-catalyzed reaction velocity without inhibitor, and K_m (Michaelis constant), the substrate concentration that provides a reaction velocity that is half of the V_{max} obtained under saturating substrate conditions (Figure 2.10) (Copeland, 2002b). Furthermore, the strength of an inhibitor can also be determined by its inhibition constant, K_i through the Michaelis-Menten kinetics model.

Generally, linear regression which uses the Lineweaver-Burk plot for linearizing enzyme kinetics data, and nonlinear regression which directly uses Michaelis-Menten equation, are the two methods that can be used to determine enzyme kinetics parameters (Motulsky & Christopoulos, 2004). However, the use of linear regression should be avoided as it transforms nonlinear data to create a linear relationship *(Y = slope ·X + intercept)* where the transformation distorts the experimental error. This is because the linear regression assumes that the scatter of points around the line follows a Gaussian distribution and that the standard deviation is the same at every value of X, and these assumptions are rarely true after transformation of data (Motulsky & Christopoulos, 2004). Thus, the values derived from the slope and the intercept of the linear regression line are also less accurate. Nevertheless, Lineweaver-Burk plot is still widely used as it can easily display the type of inhibitions (competitive or non-competitive) as shown in Figure 2.11.

On the other hand, nonlinear regression fits data to any equation that defines Y as a function of X and one or more parameters. This requires a computationally intensive and iterative approach, hence, it produces enzyme kinetics parameters more accurately (Motulsky & Christopoulos, 2004). The Michaelis-Menten equation (Copeland, 2002a) is defined as:

$$Velocity = k_2[ES] = \frac{V_{max}[S]}{[S] + K_m} \qquad \ldots(8)$$
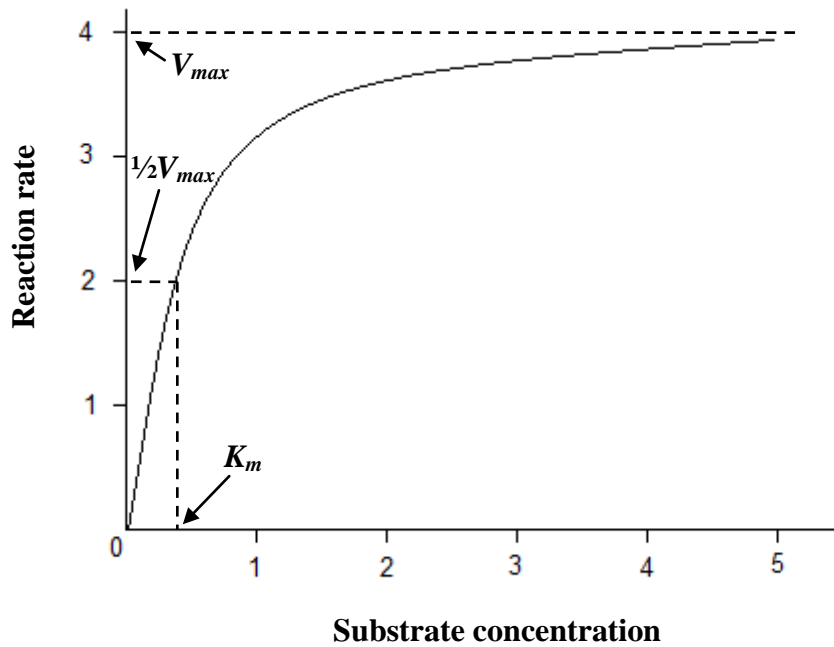
Figure 2.10   Michaelis-Menten saturation curve of an enzyme reaction

.

$$\text{Intercept} = -\frac{1}{K_M}\left(\frac{K_I}{K_I + [I]}\right)$$

$$\frac{1}{V_{max}}$$

$$\frac{1}{v_0}$$

$$\frac{1}{[S]}$$

(a)

$$\text{Intercept} = \frac{1}{V_{max}}\left(1 + \frac{[I]}{K_I}\right)$$

$$\frac{1}{v_0}$$

$$-\frac{1}{K_M}$$

$$\frac{1}{[S]}$$

(b)
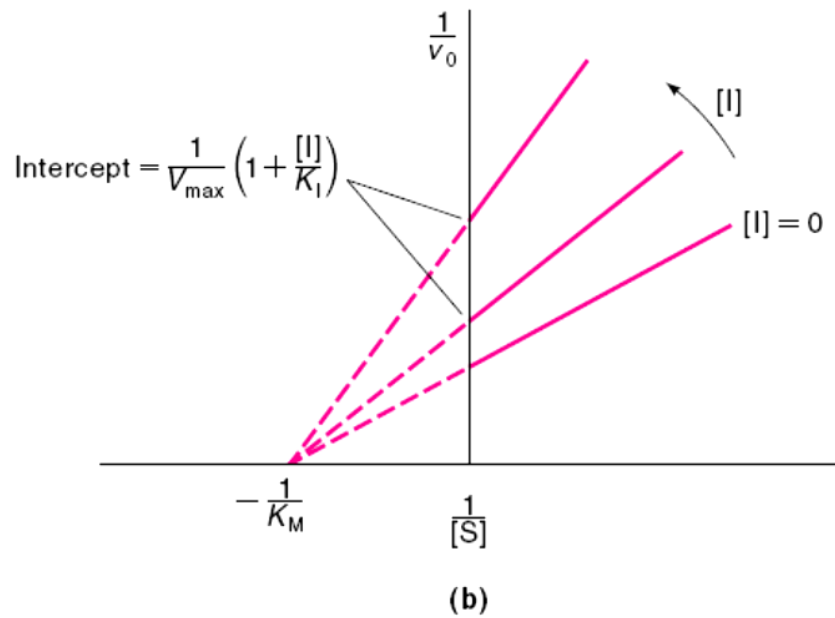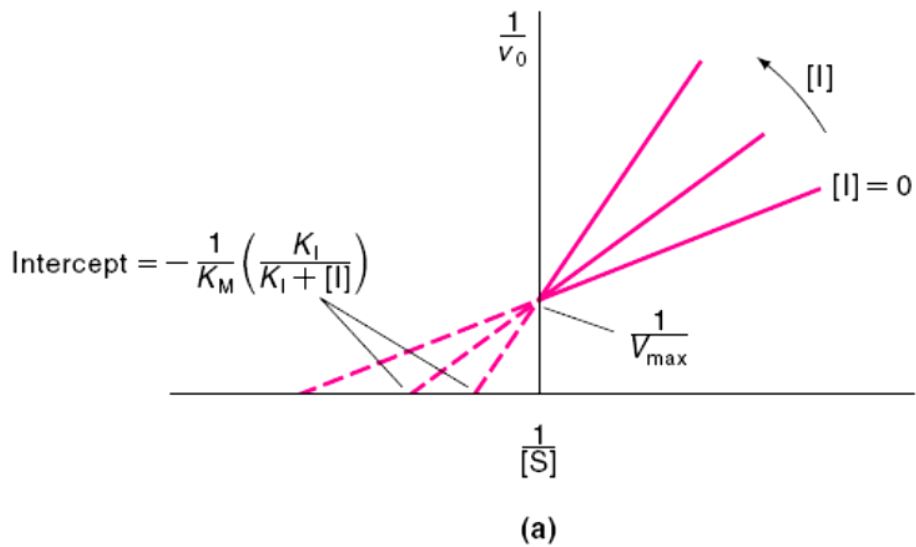
Figure 2.11   Lineweaver-Burk plots showing: (a) competitive inhibition, and (b) non-competitive inhibition (Chang, 2005). The arrow indicates the effect of the inhibitor, I, with increasing concentrations.

2.4.3.1 Competitive and Non-competitive Inhibition

The enzyme kinetics of competitive and non-competitive inhibitions has been well explained by Copeland (2002). The reaction schemes and simplified mechanisms for both competitive and non-competitive inhibitions are shown in Figure 2.12 and 2.13, respectively. As shown in Figure 2.12A, a competitive inhibitor competes with the

substrate for the same binding site (the active site) in the free enzyme, and not at all in the *ES* complex. As a result, by referring to Figure 2.13, complete competitive inhibition is characterized by values of $\alpha = \infty$ and $\beta = 0$. The effect of a competitive inhibitor is to increase the concentration of substrate that is required to reach half-maximal velocity (K_m of enzyme for its substrate will increase) without affecting the value of V_{max}. However in competitive inhibition, a high concentration of substrate can easily displace the inhibitor. The formula of the reaction velocity for competitive inhibition is given by:

$$V = \frac{V_{max} \cdot [S]}{[S] + K_m \left(1 + \frac{[I]}{K_i}\right)} \qquad \qquad \ldots(9)$$

On the other hand, in the case of non-competitive inhibition, the inhibitor binds at an allosteric site of the enzyme and does not compete with the substrate for binding towards the enzyme at the active site (Figure 2.12B). Thus, the inhibitor shows a binding affinity for both the free enzyme and the *ES* complex. By referring to Figure 2.13, complete non-competitive inhibition is then characterized by values of $\alpha = 1$ and $\beta = 0$. As the non-competitive inhibitor binds at the allosteric site and not the active site of the enzyme, it cannot be overcome by increasing the concentration of substrate. Hence, the effect of a non-competitive inhibitor is to decrease the value of V_{max} without affecting the K_m. The formula of the reaction velocity for non-competitive inhibition is given by:

$$V = \frac{V_{max} \cdot [S]}{([S] + K_m)\left(1 + \frac{[I]}{K_i}\right)} \qquad \qquad \ldots(10)$$
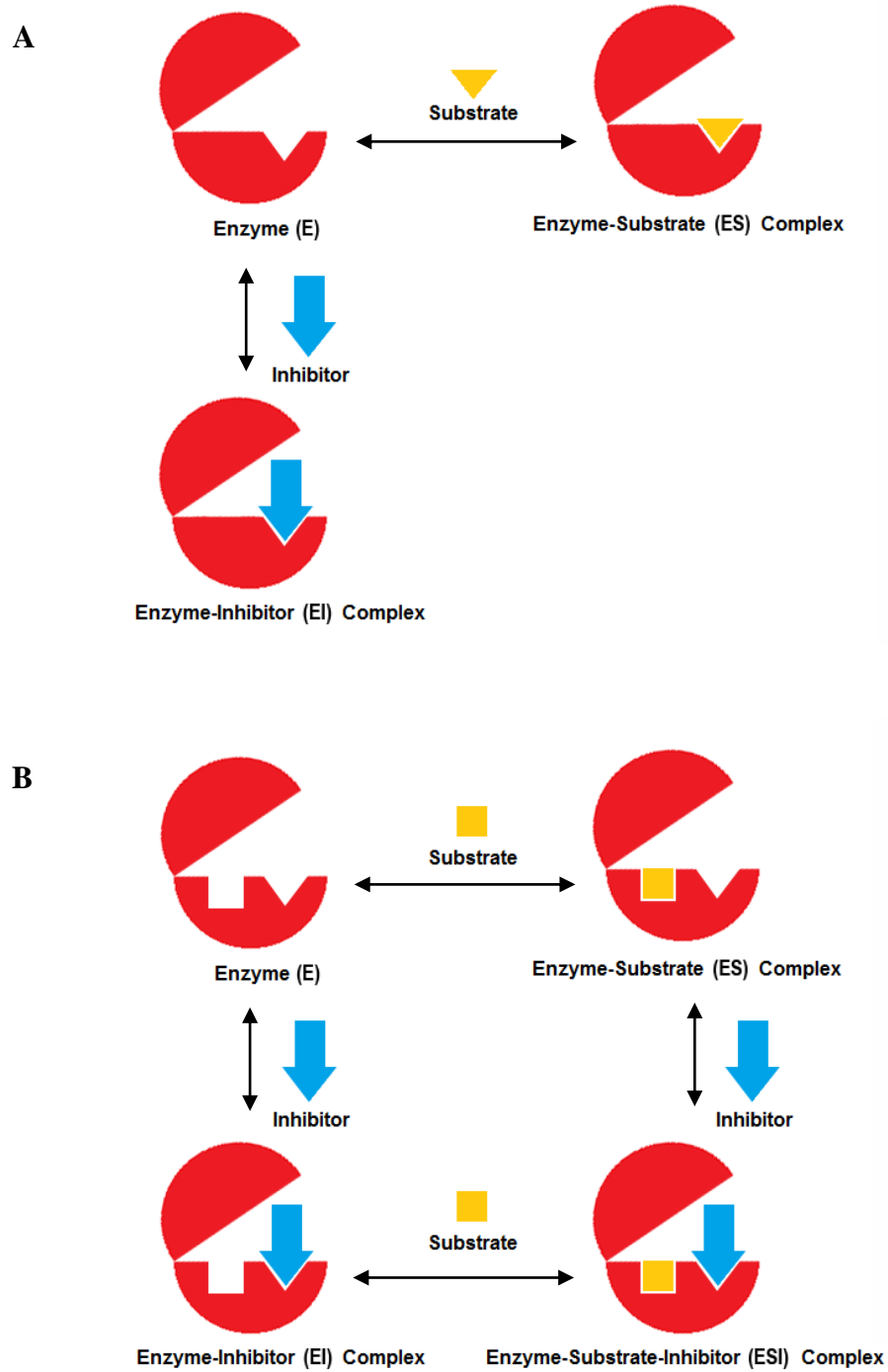
Figure 2.12   Simplified mechanism of, **A:** competitive inhibition. **B:** non-competitive inhibition.

$$E + S \quad \underset{}{\overset{K_s}{\rightleftharpoons}} \quad ES \quad \xrightarrow{k_p} \quad E + P$$

$$+ \qquad\qquad\qquad +$$

$$I \qquad\qquad\qquad I$$

$$\Big\updownarrow K_i \qquad\qquad \Big\updownarrow \alpha K_i$$

$$EI + S \quad \xrightarrow{\alpha K_s} \quad ESI \quad \xrightarrow{\beta k_p} \quad EI + P$$
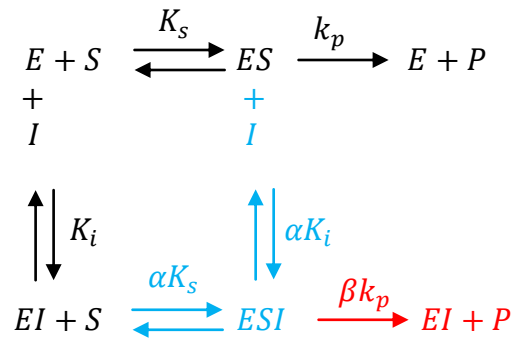
Figure 2.13   Equilibrium scheme to illustrate the enzyme reactions in the presence and absence of an inhibitor. E = enzyme, S = substrate, I = inhibitor, ES = enzyme-substrate complex, EI = enzyme-inhibitor complex, ESI = enzyme-substrate-inhibitor complex, P = product, K_s = equilibrium constant for dissociation of ES, K_i = dissociation constant for the EI and k_p = forward rate constant for product formation from ES or ESI. α = effect of inhibitor on the affinity of the enzyme for its substrate as well as the effect of substrate on the affinity of the enzyme for the inhibitor; β = change of the rate of product formation caused by the inhibitor. Competitive inhibition = black reaction pathways, non-competitive inhibition = black and light blue reaction pathways. (Copeland, 2002b).

2.4.3.2 Nonlinear Regression Mixed Model Inhibition Equation in GraphPad Prism 5.0 Software

GraphPad Prism 5.0 software can be used for both linear and nonlinear regression curve fitting from enzymatic assay data. For determining the mechanism of inhibition accurately, the use of nonlinear regression mixed model inhibition equation is suggested (Motulsky & Christopoulos, 2004). The equation defines Y (velocity of enzyme reactions) as a function of X (concentration of substrate) and it is shown by the following formula:

$$Y = \frac{\dfrac{V_{max}}{\left(1 + \dfrac{[I]}{\alpha K_i}\right)} \cdot X}{\left(\dfrac{K_m \left(1 + \dfrac{[I]}{K_i}\right)}{1 + \dfrac{[I]}{\alpha K_i}}\right) + X} \qquad \text{...(11)}$$

As mentioned in section 2.4.3.2, α determines the mechanism where its value determines the degree to which the binding of inhibitor changes the affinity of the enzyme for substrate. Its value should always be greater than zero. When $\alpha = 1$, the mixed-model is identical to non-competitive inhibition (equation 20). When α is very large, the mixed-model becomes identical to competitive inhibition (equation 21). When α is very small (but greater than zero), binding of the inhibitor enhances substrate binding to the enzyme, and the mixed model becomes nearly identical to an uncompetitive model, where uncompetitive inhibitors bind exclusively to the enzyme-substrate complex rather than to the free enzyme form (Copeland, 2002b). The formula of the reaction velocity for uncompetitive inhibition is given by:

$$V = \frac{\dfrac{V_{max}}{\left(1 + \dfrac{[I]}{\alpha K_i}\right)} \cdot [S]}{\dfrac{K_m}{\left(1 + \dfrac{[I]}{\alpha K_i}\right)} + [S]} \qquad \text{...(12)}$$

Hence, the nonlinear regression mixed model inhibition equation in GraphPad Prism 5.0 software was used for the determination of inhibition mechanism from enzymatic assay data in this study.

CHAPTER THREE

IN SILICO STUDIES

3.1 Methodology

In this part of the study, a workstation with four Intel Core 2 Duo E6850 3.00 GHz microprocessors, 8 GigaBytes of RAM and Ubuntu 10.04 Linux operating system was used.

3.1.1 Structure Verification of DEN-2 NS2B-NS3 Protease Models

The structures of DEN-2 NS2B-NS3 protease were first verified for their suitability to be the correct model as non-competitive docking target receptor. Then the suitable structure was used for virtual screening.

3.1.1.1 Homology Modelling

The 2FP7, 2GGV, 2IJO, 3E90, 3L6P, 3LKW, 3U1I and 3U1J crystal structures were obtained from the PDB. After removing water molecules and the substrates (if present), these crystal structures were used as a template for automatically generating homology models of the DEN-2 NS2B-NS3 protease using Modeller 9.11 software (Sali & Blundell, 1993). The homology models were built according to the amino acid sequence identical to that of 2FOM. The sequences were aligned using ClustalX 2.0 software (Thompson *et al.*, 1997; Larkin *et al.*, 2007) as shown in Figure 3.1. The sequence alignment in Figure 3.1 also showed that there are more than 50% identical amino acids between the crystal structures. Hence, the crystal structures are suitable to be used as template for DEN-2 protease homology model building. This is because a protein model that shares more than 30% sequence identity with another protein is indicative of an accurate structure for homology modelling (Eswar *et al.*, 2006). The homology modellings were carried out by referring to the Modeller online manual (Webb *et al.*, 2011), and run by using the command line:

mod9.11 model-default.py

Ten homology models were created using multiple templates and were then analyzed using Ramachandran plots generated by Procheck software for checking the stereochemical quality of a protein (Laskowski *et al.*, 1993), and Verify3D software for determining the compatibility of the 3D atomic models with their own 1D amino acid sequence (Luthy *et al.*, 1992). These softwares are available in their online server versions at the Structural Analysis and Verification Server of UCLA (University of California, Los Angeles; http://nihserver.mbi.ucla.edu/SAVES/). The homology model for each template that produced the best scores in structural analyses, namely DH-1 to DH-8, were then selected for docking studies.

3.1.1.2 Preparation of Macromolecule for Blind Docking

Blind docking that allowed the ligands to be docked freely in the whole structure of the macromolecule was performed for the DEN-2 apo protease (2FOM), and the homology models using AutoDock 4.2 software. The 3D crystal structure of DEN-2 NS2B-NS3 apo protease (2FOM) (Erbel *et al.*, 2006) was retrieved from the PDB, and chlorine atoms, water and glycerol molecules were removed. AutoDockTools 1.5.4 software was then used to add all hydrogen atoms, merging nonpolar hydrogen atoms, checking and repairing missing atoms, adding Gasteiger charges, checking and fixing total charges on residues, and assigning atom types to the protein structure (Figure 3.2A). A grid box of the protein structure was then generated using AutoGrid 4 software with default atom types (carbon, hydrogen, oxygen and nitrogen), grid spacing of 0.41 Å, dimension of 126 x 126 x 126 points along the x, y and z axes, and centered on the protein, covering the whole protein for the blind docking (Figure 3.2B).

As for the homology models, DH-1 to DH-8, docking parameters were set following those described above for 2FOM using the AutoDockTools 1.5.4 software.

3.1.1.3 Preparation of Flexible Ligands

Cardamonin, *R*-pinostrobin and *S*-pinostrobin were used as standard ligands. Structures (3D) of these ligands were constructed using Hyperchem Pro 8.0 software. The energies of all the ligands were also minimized using Hyperchem Pro 8.0 software, employing the steepest descent and conjugate gradient methods (termination conditions set to a maximum of 500 cycles or 0.1 kcal/Å mol rms gradient) (Othman *et al.*, 2008). The minimized structures were subsequently prepared with detected root of torsion and number of torsions for flexible-ligand docking using AutodockTools 1.5.4 software and saved as "ligand's name".pdbqt (e.g. cardamonin.pdbqt).

3.1.1.4 Preparation of AutoDock Parameters for Cardamonin, *R*-pinostrobin and *S*-pinostrobin

Parameters for blind docking of flexible ligands to DEN-2 protease were set to a population size of 150 individuals and 10,000,000 number of energy evaluations for 100 runs to produce 100 distinct conformations using the Lamarckian genetic algorithm search function (Morris *et al.*, 1998). The resulting 100 distinct conformations were set to be clustered in the same group with RMSD of not more than 0.5 Å, for the ease of analysis. The flexible-ligand (blind) dockings for each of the small compound (ligand) were performed using the AutoDock 4.2 software by applying all the parameters stored in the docking parameter file – "ligand's name".dpf.

Figure 3.1   The sequence alignment of NS2B and NS3 of 2FOM, 2FP7, 2GGV, 2IJO, 3E90, 3L6P, 3LKW, 3U1I and 3U1J using ClustalX 2.0 software. The grey histogram represents the level of convergence between two amino acids. Amino acids that were identical in both sequences will have a full grey bar and an asterisk symbol (*) on top of the specific amino acids, followed by colon symbol (:) for high similarity, period symbol (.) for low similarity and blank ( ) for non convergence.

A   **B**

Figure 3.2   Example of 3D structure of DEN-2 NS2B-NS3 crystal (2FOM). **A:** Cyan
– NS2B cofactor; Orange – NS3 serine protease; Green – catalytic triad residues of the
active site, His-51, Asp-75 and Ser135. **B:** Grid box was set to cover the whole protease.

3.1.1.5 Running AutoGrid 4 and AutoDock 4.2

AutoGrid 4 and AutoDock 4.2 softwares were installed in Ubuntu 10.04 Linux
operating system of the workstation. AutoGrid 4 was run following instructions in the
"AutoDock Version 4.2" user's manual (Morris *et al.*, 2010) using the command line:

autogrid4 -p protein.gpf -l protein.glg

and AutoDock 4.2 was run using the command line:

autodock4 -p ligand.dpf -l ligand.dlg

3.1.1.6 Analysis of Results

After the docking jobs were completed, the compounds were ranked based on the lowest estimated mean free energy of binding (ΔG_{dock}) coupled with the largest NumCl. ΔG_{dock} (equation (6) in section 2.3.1) was calculated using Autodock 4.2 software, while the estimated inhibition constant ($K_{i\ dock}$) was calculated using the formula (Morris *et al.*, 1998):

$$K_{i\ dock} = e^{\Delta Gdock/RT}$$

where R is the gas constant, 1.987 cal K^{-1} mol^{-1}, and T is the reaction or body temperature, 310.15 K (37 ^{o}C).

The number of distinct conformations that were grouped into the same cluster based on RMSD (NumCl) was used as a measure of the probability of a particular conformer to interact with the macromolecule target, where the higher NumCl number is proportionate to increased probability of interaction. All of the docked conformers were then subjected for interaction analyses using Ligplot 4.5.3 software. The hydrogen bonding distance was set to the range of 2.7 to 3.35 Å, and the hydrophobic interaction distance was set to the range of 2.9 to 3.9 Å (Wallace *et al.*, 1995). Interaction Frequency (*IF*) is determined as follows:

$$IF(\text{Acceptor})_{\text{Res X}}, IF(\text{Donor})_{\text{Res X}} \text{ and } IF(\text{Hydrophobic})_{\text{Res X}}$$

where $IF(\text{Acceptor})_{\text{Res X}}$ is the number of interactions as hydrogen bond acceptor in Residue X, $IF(\text{Donor})_{\text{Res X}}$ is the number of interactions as hydrogen bond donor in Residue X, and $IF(\text{Hydrophobic})_{\text{Res X}}$ is the number of hydrophobic interactions for Residue X.

The conformation from the docked conformational cluster, having the largest NumCl and exhibiting interaction with Lys74 from NS3, was selected as the best binding conformation. In cases where there were two or more clusters with similar NumCl (difference in NumCl ≤ 10), the cluster that showed lower ΔG_{dock} was chosen. Further interaction analyses using Discovery Studio Visualizer 3.1 (Accelrys Software Inc.) were performed for better insight.

The most suitable DEN-2 NS2B-NS3 protease model for non-competitive inhibition study was then identified following the completion of analysis of results.

3.1.2 Virtual Screening for Potential Non-competitive Inhibitors

3.1.2.1 Preparation of Parameters for Virtual Screening Parameters

Autodock 4.2 software was used for virtual screening. The suitable model of DEN-2 NS2B-NS3 protease for non-competitive inhibition and its docking parameters were prepared according to the method described in section 3.1.1.2. The docking parameters of the small compounds for virtual screening were modified to a population size of 150 individuals, and 1,750,000 number of energy evaluations for 20 runs using the Lamarckian genetic algorithm search function. Twenty distinct conformations that were produced were further clustered into the same group (NumCl) with RMSD of not more than 2.0 Ǻ. The docking parameters were modified to reduce the duration for running the large number of docking calculations in the virtual screening process. Pinostrobin was used as the standard as its reported $K_{i\ exp}$ value (345 ± 70 μM) was smaller than cardamonin's (377 ± 77 μM) (Kiat *et al.*, 2006) and was also rerun in docking with the same parameters as the rest of the small compounds in virtual screening.

A series of small compounds with structures having more than 50% similarity to chalcone (3,458), flavanone (4,886) and flavone (4,997) (Figure 3.3) were downloaded from the ZINC database (Irwin & Shoichet, 2005). Raccoon 1.0 software (Forli, 2010) was used for the preparation of all the input files for the virtual screening.

Chalcone   Flavanone   Flavone
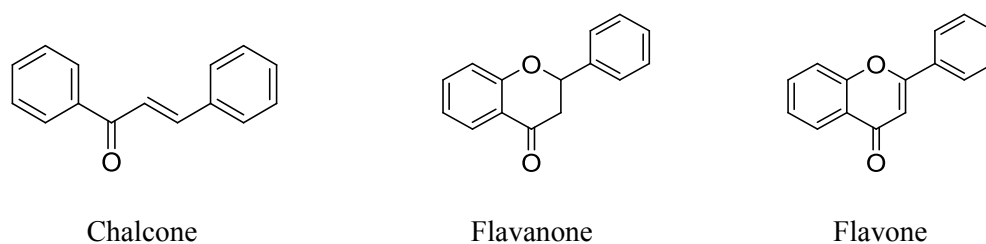
Figure 3.3   Structures of the chalcone, flavanone and flavone.

3.1.2.2 Running Virtual Screening

Virtual screening was run by using a script file generated by Raccoon 1.0, which sequentially and automatically runs all the docking of the compounds towards the suitable model of DEN-2 NS2B-NS3 protease using AutoDock 4.2.   After the virtual screening is completed, the compounds were ranked based on the lowest estimated binding energy with largest NumCl for ease of analysis.

3.1.2.3 High-throughput Analysis of Virtual Screening Results

Bash 4.1 software (Fox, 1989) was used to program automated sequential data submission, extraction and identification for high-throughput analysis. Those small compounds with ΔG_{dock} lower than the ΔG_{dock} for both of the standards, cardamonin and pinostrobin, and with NumCl more than 10, were further subjected to interaction analysis using Ligplot 4.5.3 software using the same parameters as described in section 3.1.1.6 (determination of IF). Data about the group, name, NumCl, ΔG_{dock}, $K_{i\ dock}$ value,

and interaction properties (IF(Acceptor)$_{\text{Res X}}$, IF(Donor)$_{\text{Res X}}$ and IF(Hydrophobic)$_{\text{Res X}}$) corresponding to these small compounds was then extracted. Further interaction analyses using Discovery Studio Visualizer 3.1 were also performed.

Further selection of the small compounds was then carried out to identify potential non-competitive inhibitors, based on the lowest ΔG_{dock} and interaction with Lys74 from NS3 (Othman *et al.*, 2008). The programming scripts are attached in the Appendix section. The selected small compounds were then traced from the ZINC database for their availability for purchase. The purchased compounds were then subjected to DEN-2 NS2B-NS3 protease cleavage inhibition assay.

3.2   Results and Discussions

3.2.1   Quality Verification of DEN-2 Protease Homology Models

From Ramachandran plots (Figure 3.4 – 3.11), it could be seen that more than 89% of residues of the homology models, DH-1 to DH-8, were located in the most favoured regions, and no residue was detected in the disallowed regions (Laskowski *et al.*, 1993). Thus, these models are of adequate stereochemical quality. Analyses with Verify3D (Figure 3.12 – 3.19) revealed that DH-1 to DH-8 contained adequate 3D atomic models that were compatible with their 1D amino acid sequences, with more than 73% of of the residues with an averaged 3D to 1D score of more than 0.2 (Luthy *et al.*, 1992), respectively.
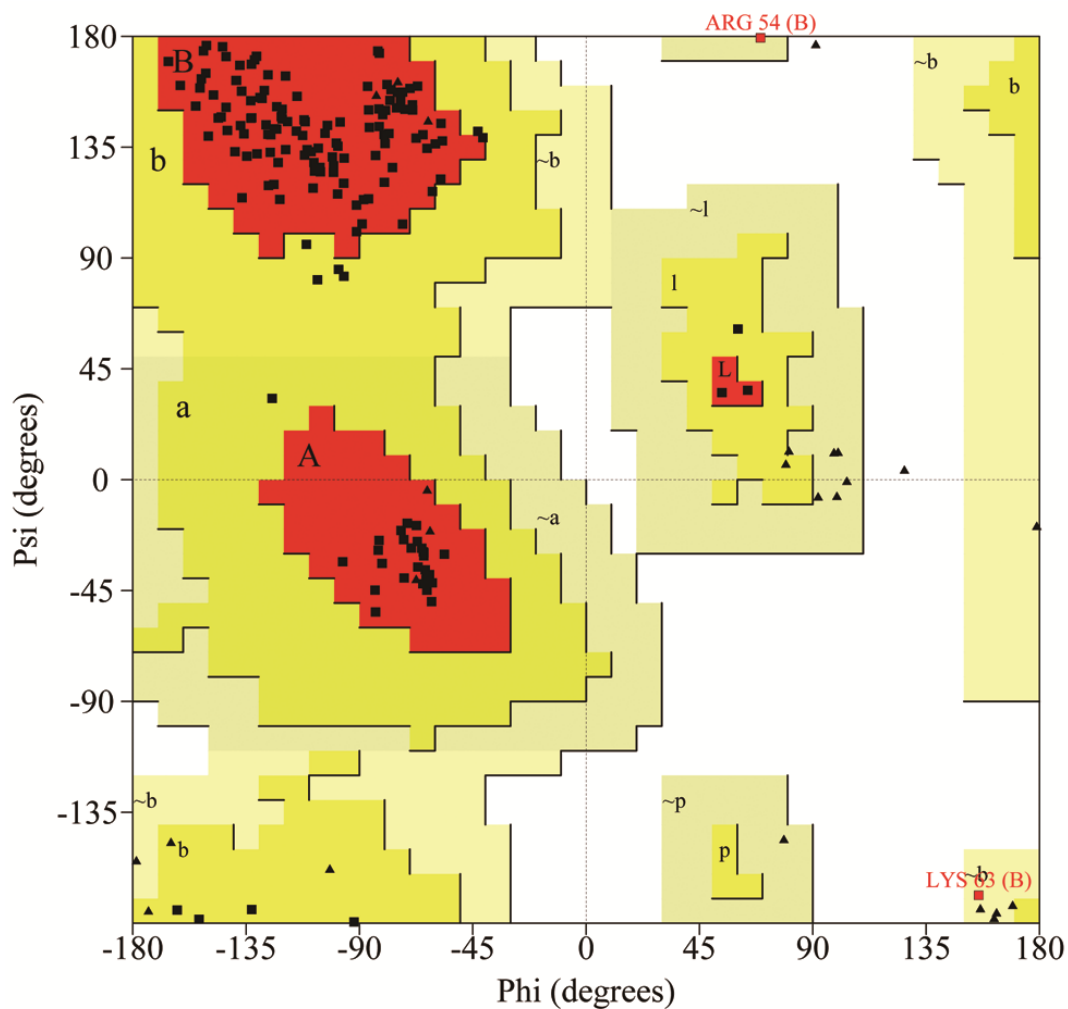
Plot statistics

Residues in most favoured regions [A,B,L]	142	91.6%
Residues in additional allowed regions [a,b,l,p]	11	7.1%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.3%
Residues in disallowed regions	0	0.0%
	----	-------
Number of non-glycine and non-proline residues	155	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	7	

Total number of residues	192	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 3.4   Ramachandran plot generated by Procheck software for homology model DH-1.
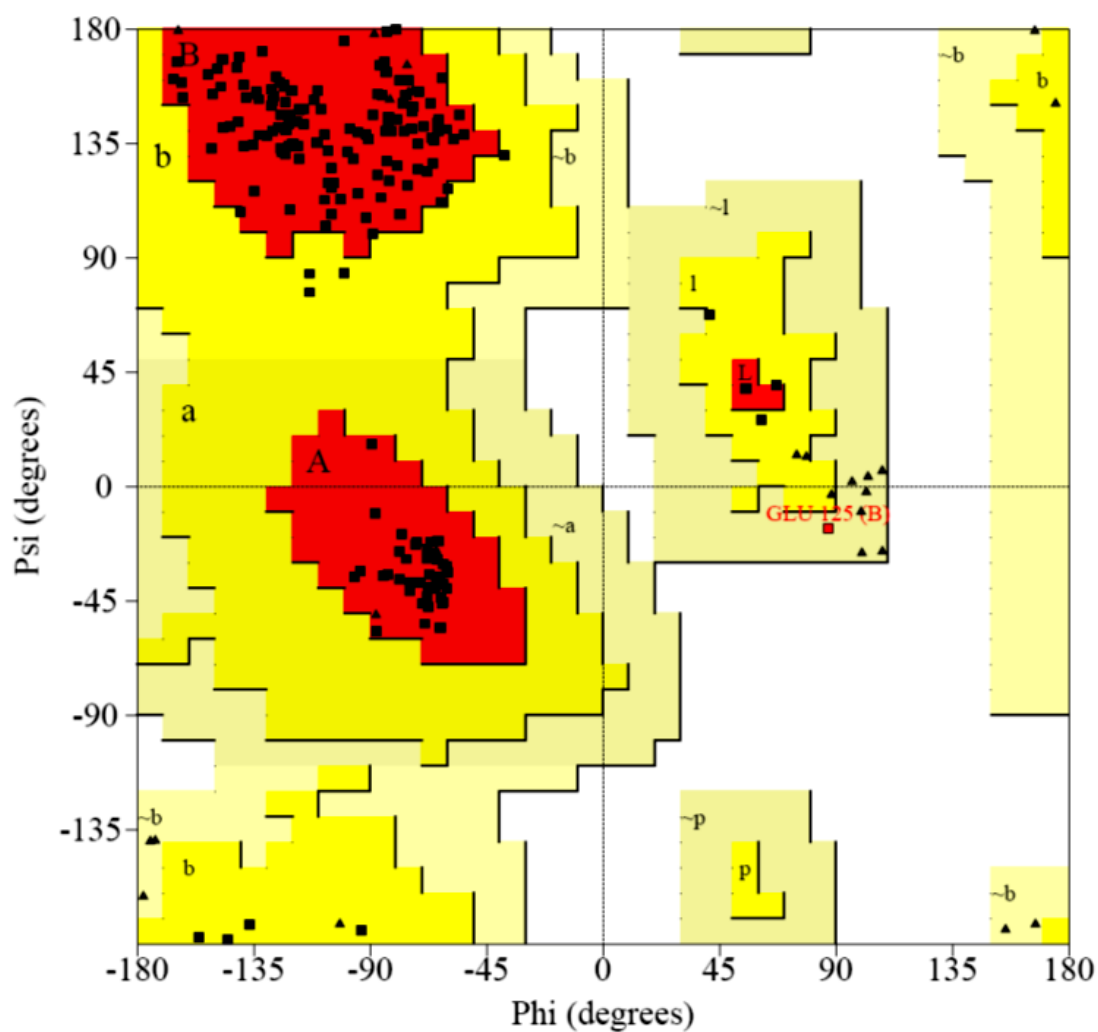
Plot statistics

Residues in most favoured regions [A,B,L]	161	92.5%
Residues in additional allowed regions [a,b,l,p]	12	6.9%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.6%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	174	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	11	

Total number of residues	215	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 3.5   Ramachandran plot generated by Procheck software for homology model
DH-2.
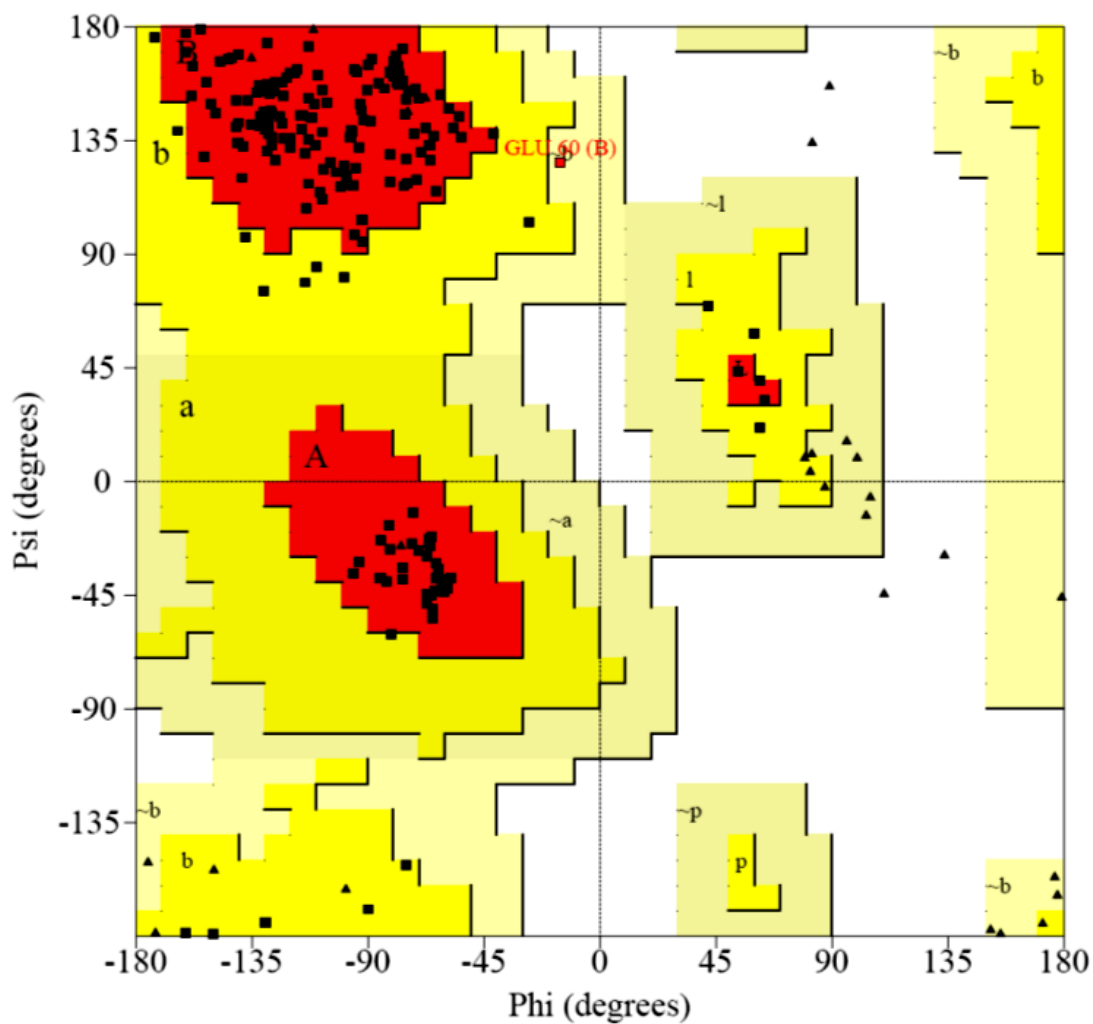
Plot statistics

Residues in most favoured regions [A,B,L]	156	89.7%
Residues in additional allowed regions [a,b,l,p]	17	9.8%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.6%
Residues in disallowed regions	0	0.0%
	----	-------
Number of non-glycine and non-proline residues	174	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	12	

Total number of residues	217	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 3.6   Ramachandran plot generated by Procheck software for homology model
DH-3.
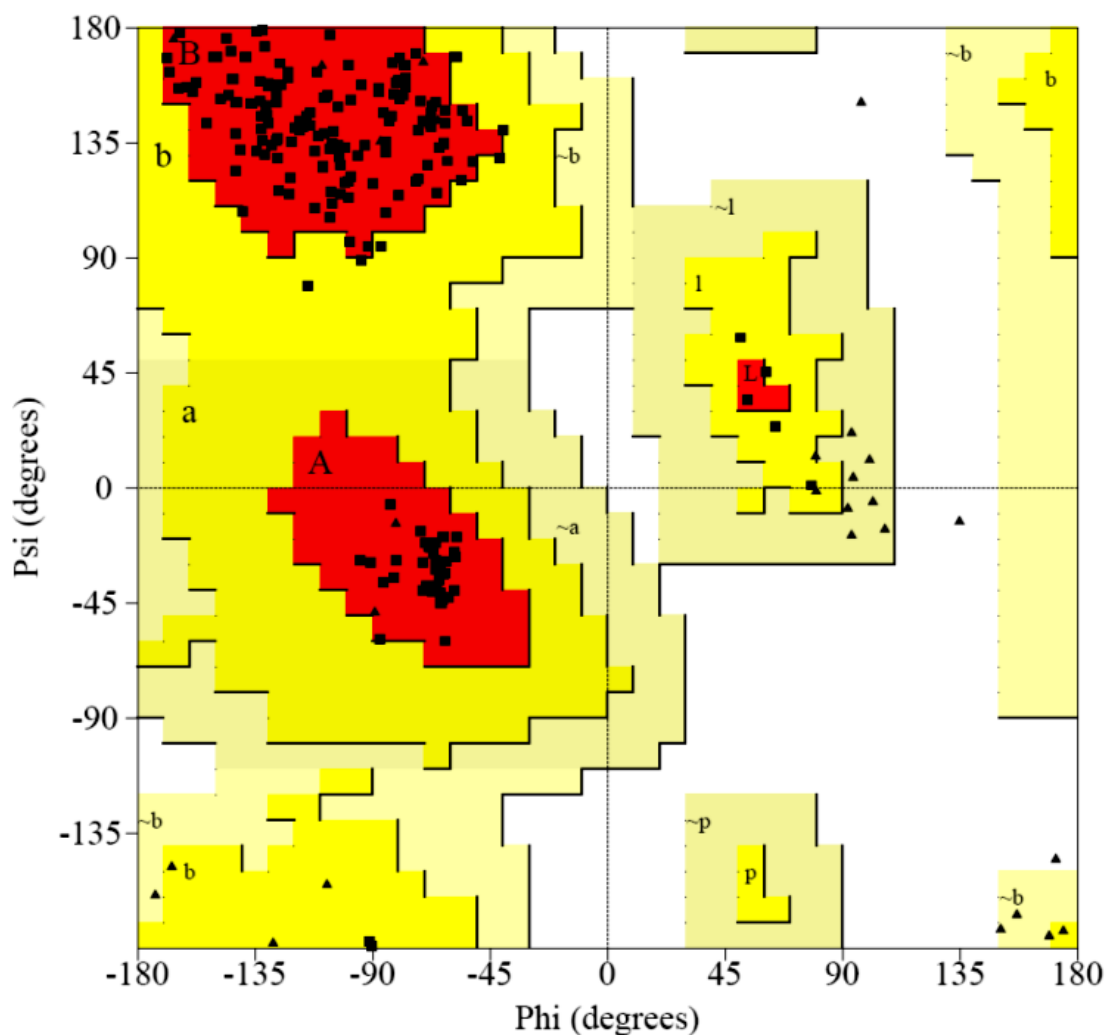
Plot statistics

Residues in most favoured regions [A,B,L]	162	93.1%
Residues in additional allowed regions [a,b,l,p]	12	6.9%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
	----	-------
Number of non-glycine and non-proline residues	174	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	11	

Total number of residues	216	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 3.7 Ramachandran plot generated by Procheck software for homology model DH-4.
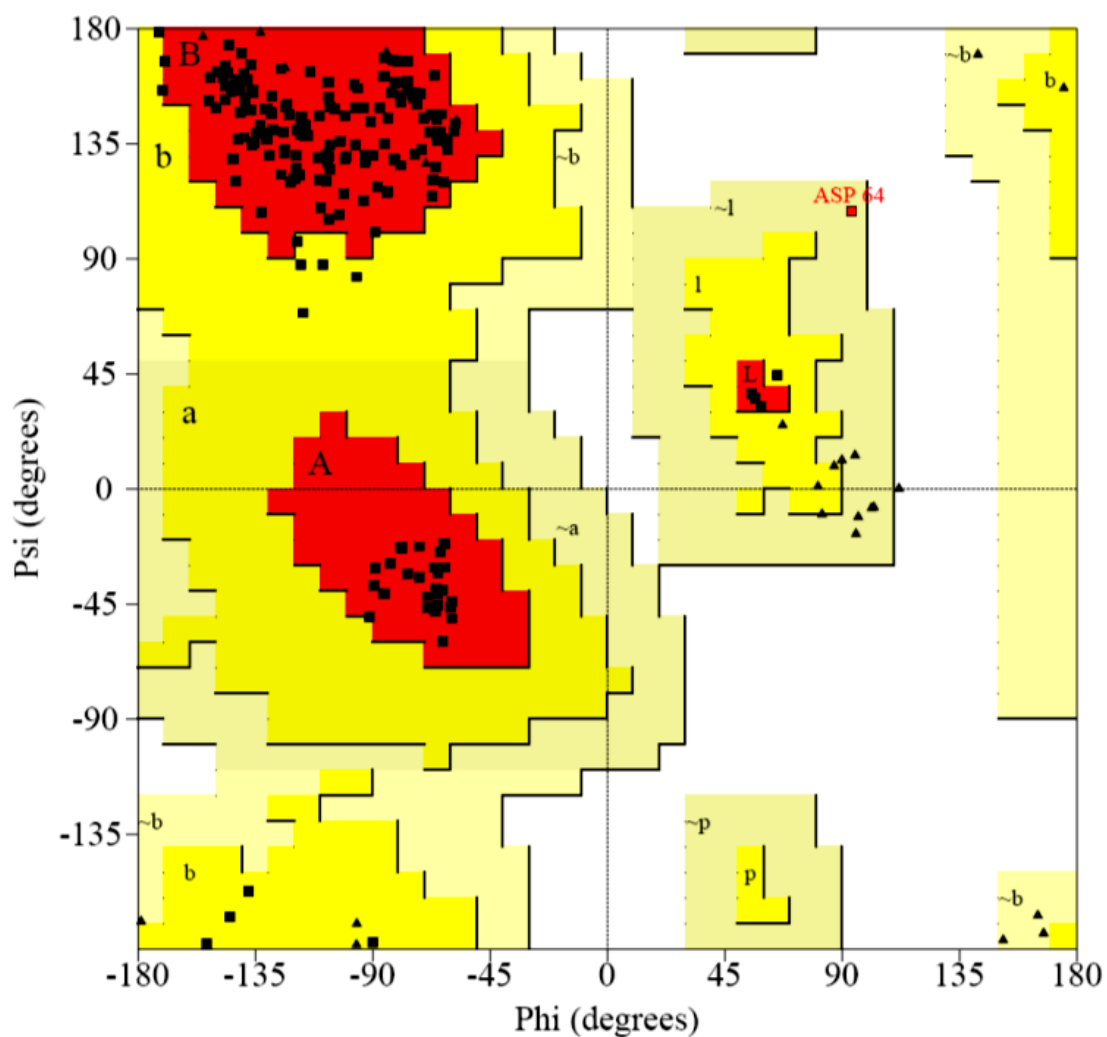
Plot statistics

Residues in most favoured regions [A,B,L]	163	92.1%
Residues in additional allowed regions [a,b,l,p]	13	7.3%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.6%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	177	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	11	

Total number of residues	217	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 3.8   Ramachandran plot generated by Procheck software for homology model DH-5.

Figure 3.9 Ramachandran plot generated by Procheck software for homology model DH-6.
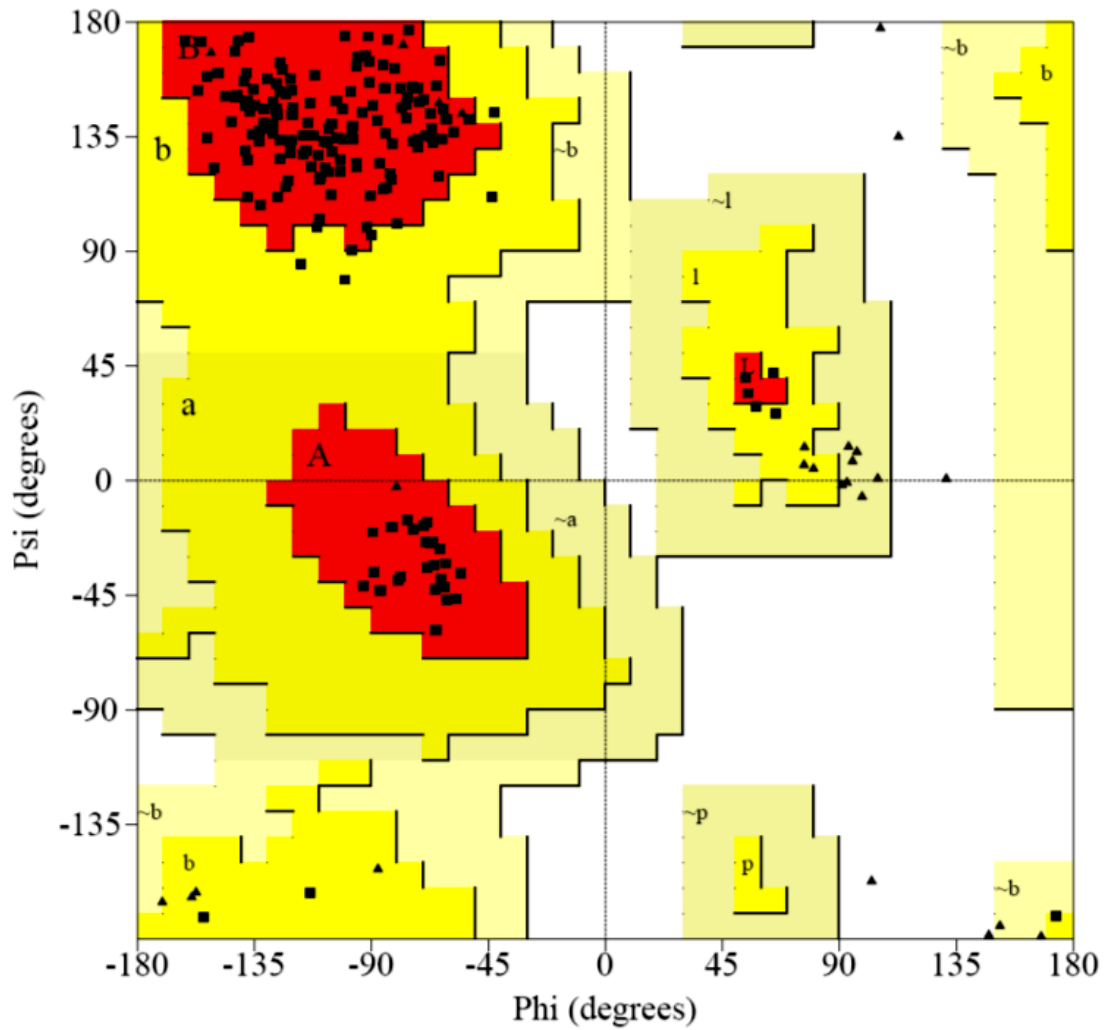
Plot statistics

Residues in most favoured regions [A,B,L]	156	93.4%
Residues in additional allowed regions [a,b,l,p]	11	6.6%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
	----	-------
Number of non-glycine and non-proline residues	167	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	11	

Total number of residues	209	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 3.10   Ramachandran plot generated by Procheck software for homology model DH-7.
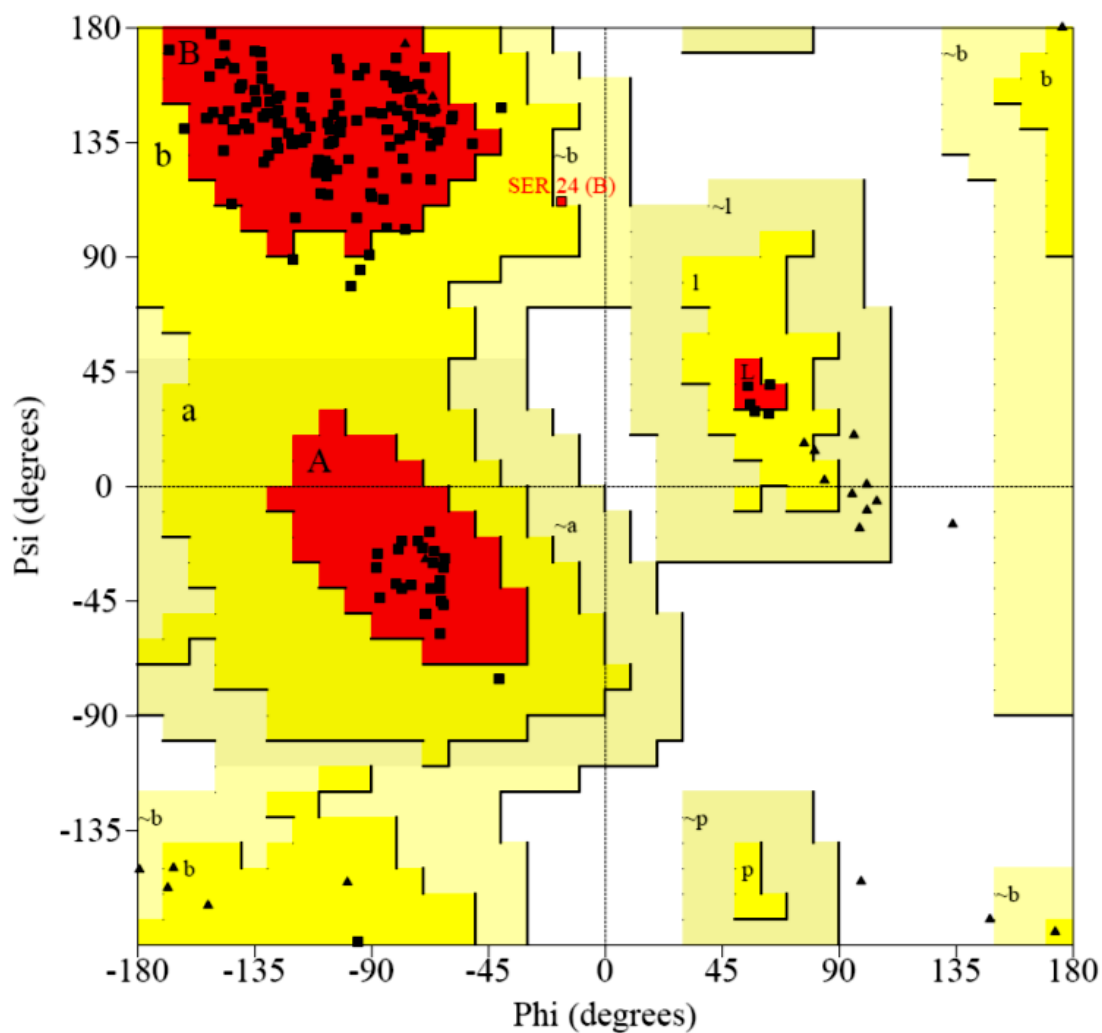
Plot statistics

Residues in most favoured regions [A,B,L]	138	93.9%
Residues in additional allowed regions [a,b,l,p]	8	5.4%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.7%
Residues in disallowed regions	0	0.0%
	----	-------
Number of non-glycine and non-proline residues	147	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	10	

Total number of residues	186	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 3.11   Ramachandran plot generated by Procheck software for homology model DH-8.
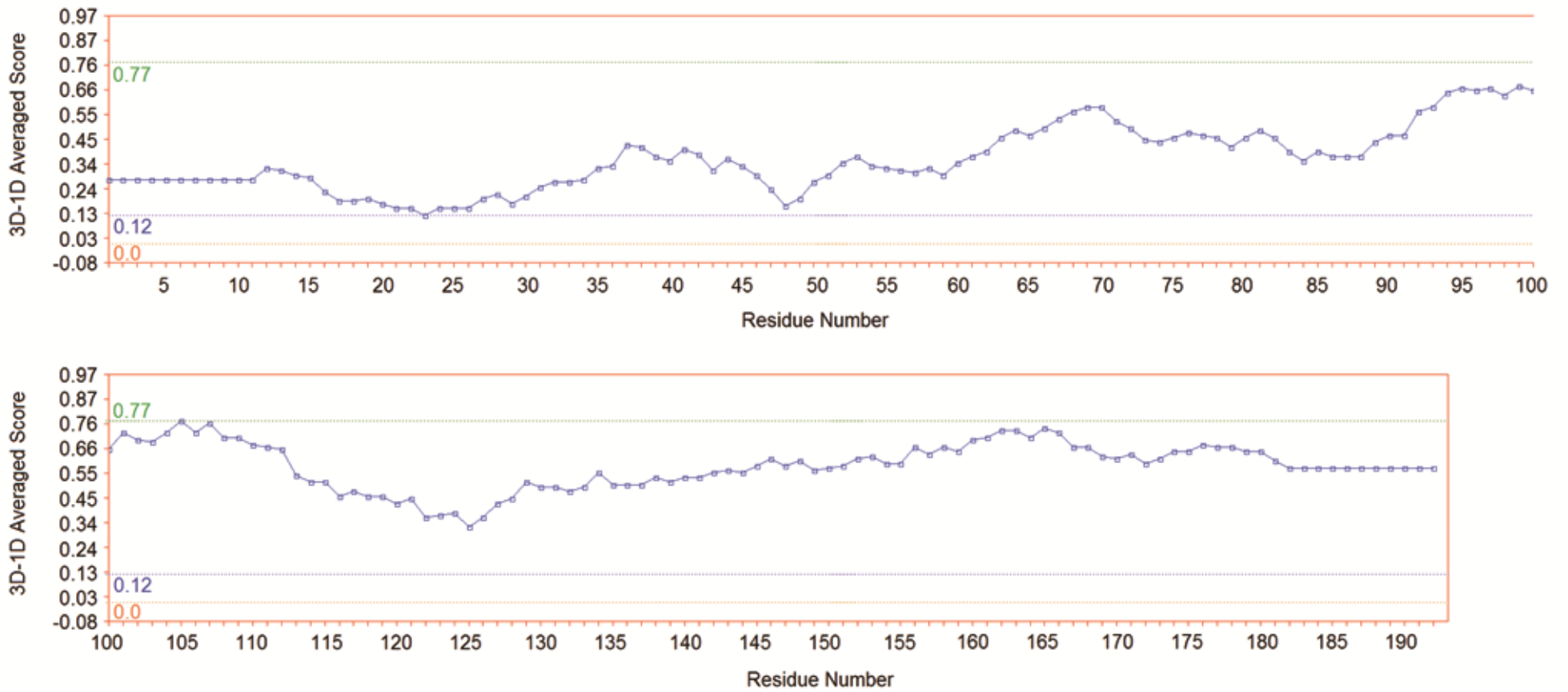
Figure 3.12   Verify3D plot for homology model DH-1. 92.8% of the residues had an averaged 3D-1D score of more than 0.2.
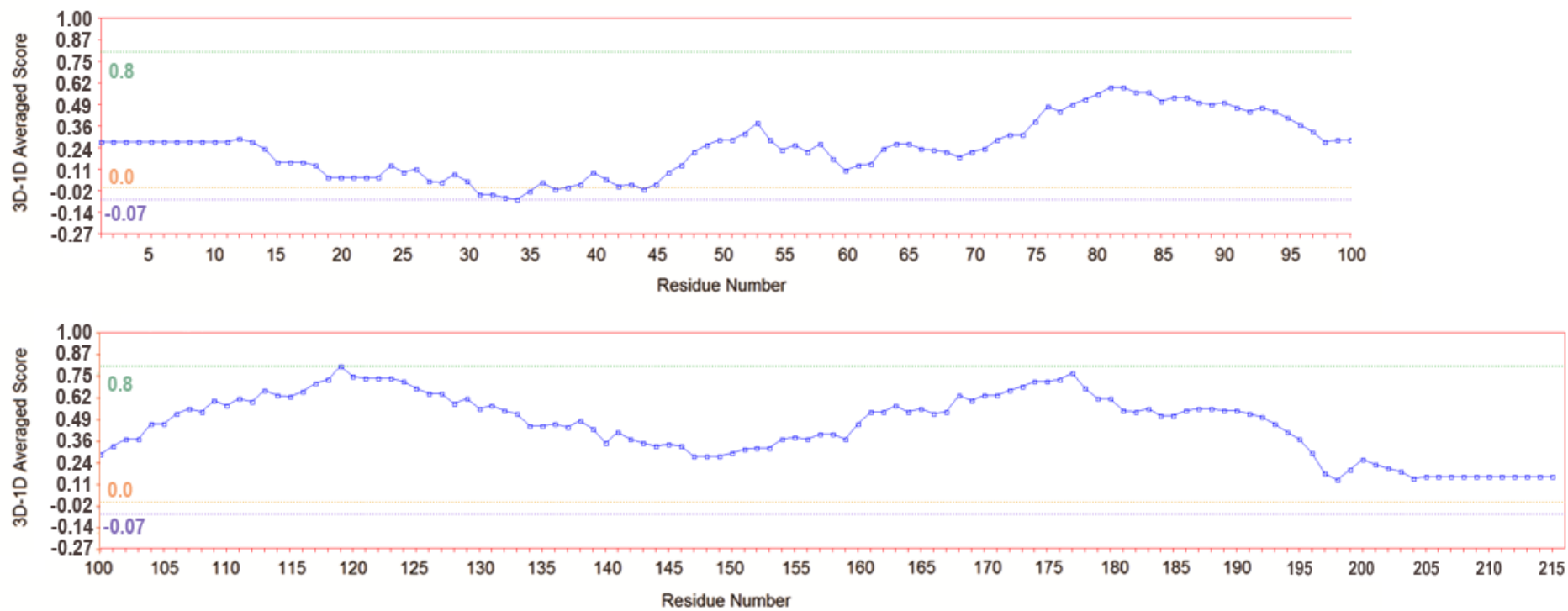
Figure 3.13 Verify3D plot for homology model DH-2. 74.5% of the residues had an averaged 3D-1D score of more than 0.2.
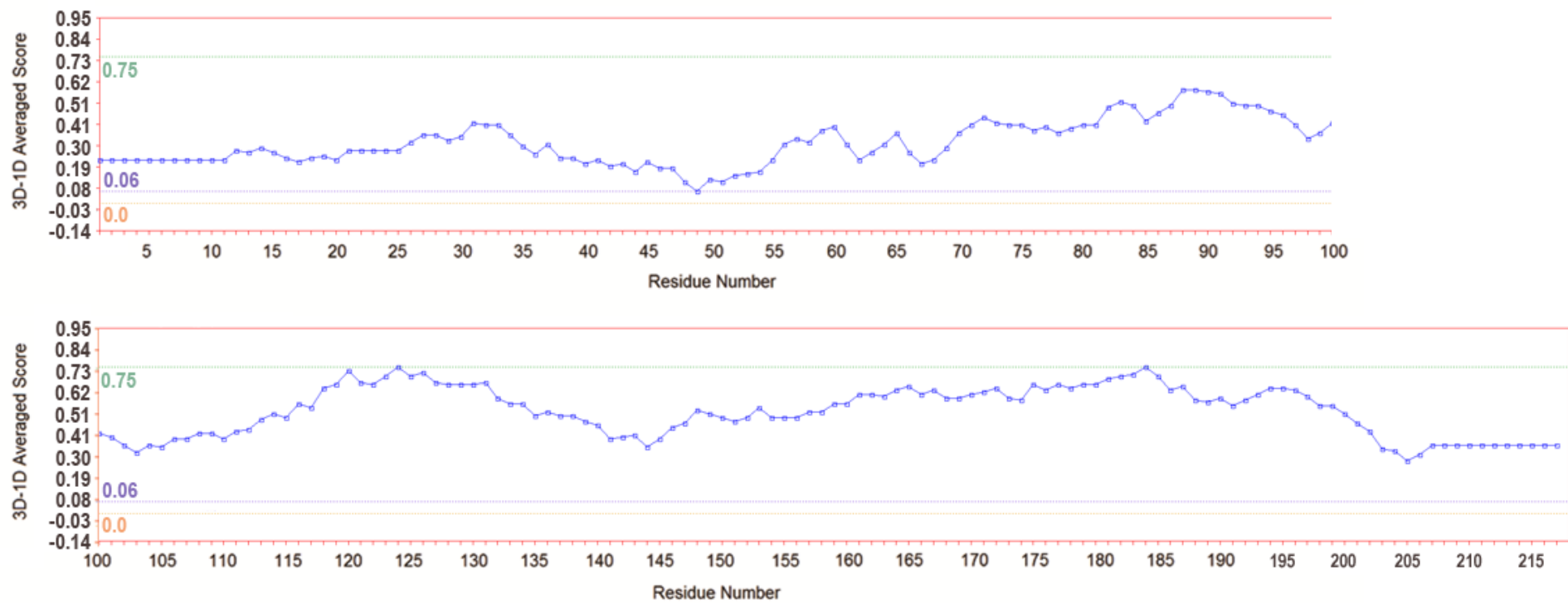
Figure 3.14 Verify3D plot for homology model DH-3. 95.0% of the residues had an averaged 3D-1D score of more than 0.2.
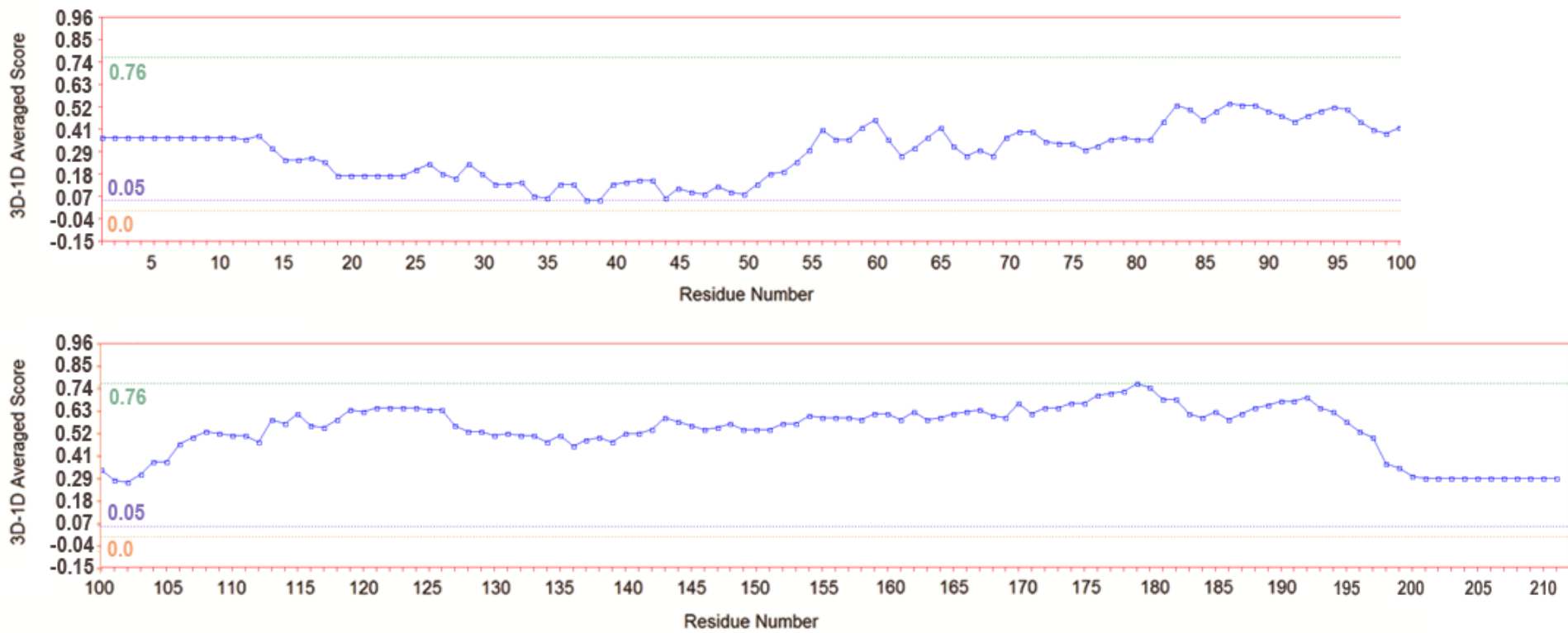
Figure 3.15   Verify3D plot for homology model DH-4. 85.3% of the residues had an averaged 3D-1D score of more than 0.2.
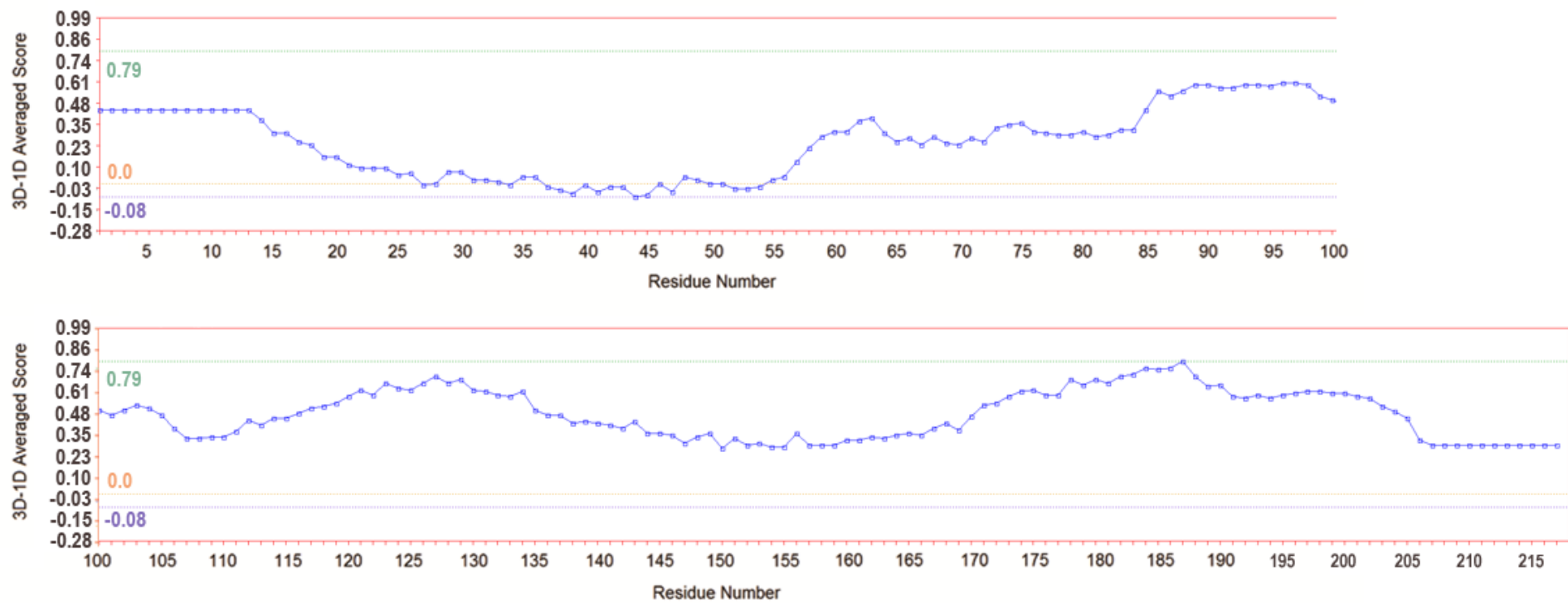
Figure 3.16 Verify3D plot for homology model DH-5. 82.1% of the residues had an averaged 3D-1D score of more than 0.2.
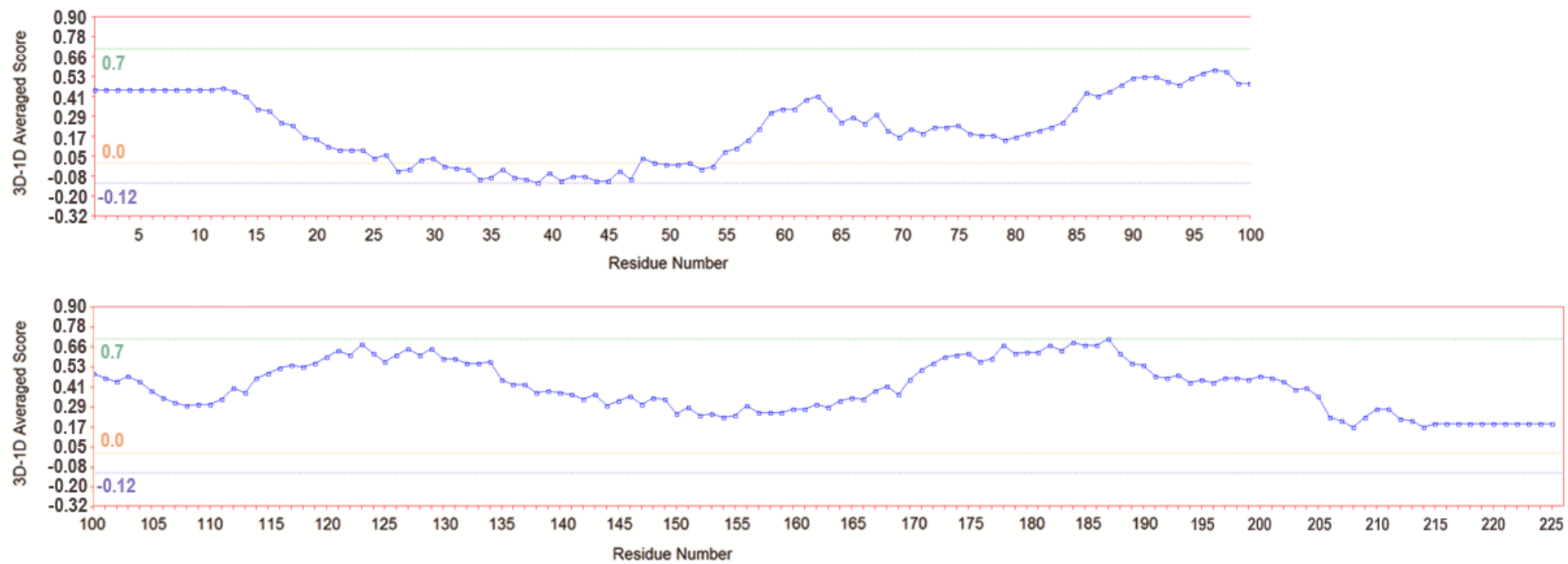
Figure 3.17   Verify3D plot for homology model DH-6. 73.0% of the residues had an averaged 3D-1D score of more than 0.2.
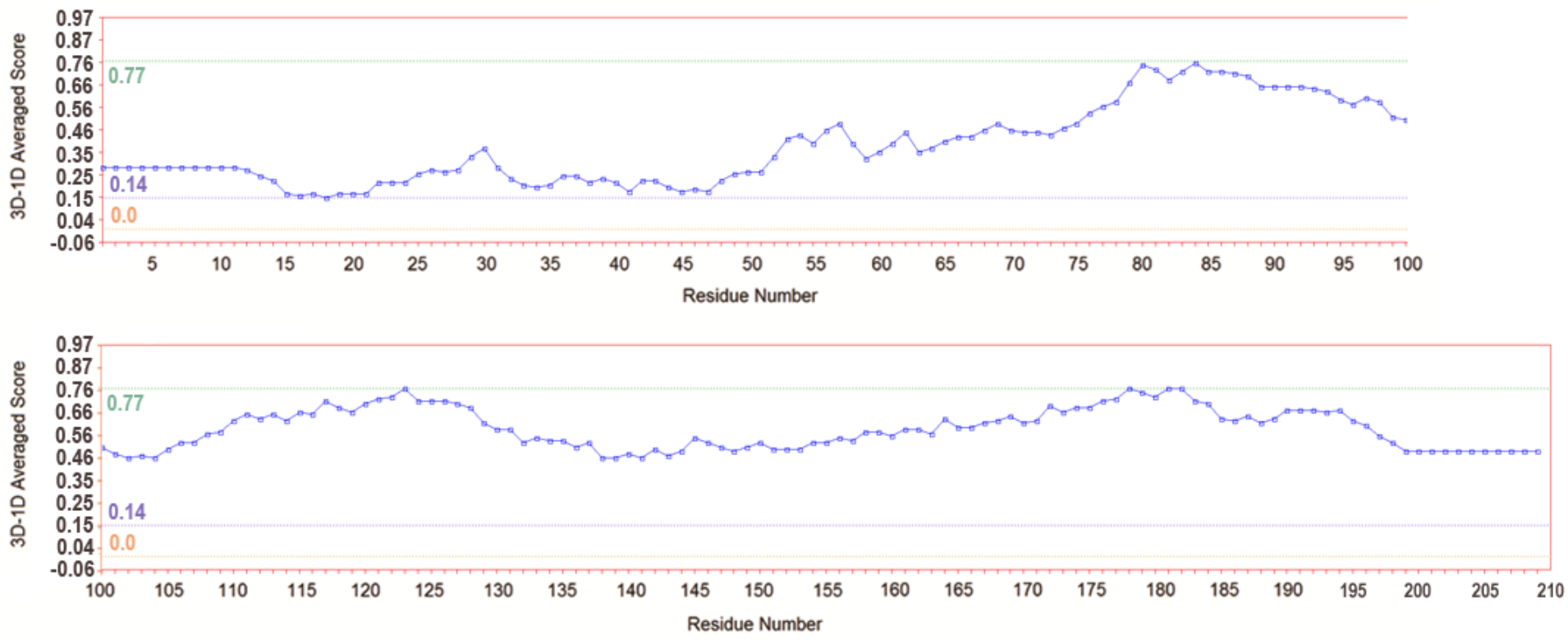
Figure 3.18   Verify3D plot for homology model DH-7. 93.8% of the residues had an averaged 3D-1D score of more than 0.2.
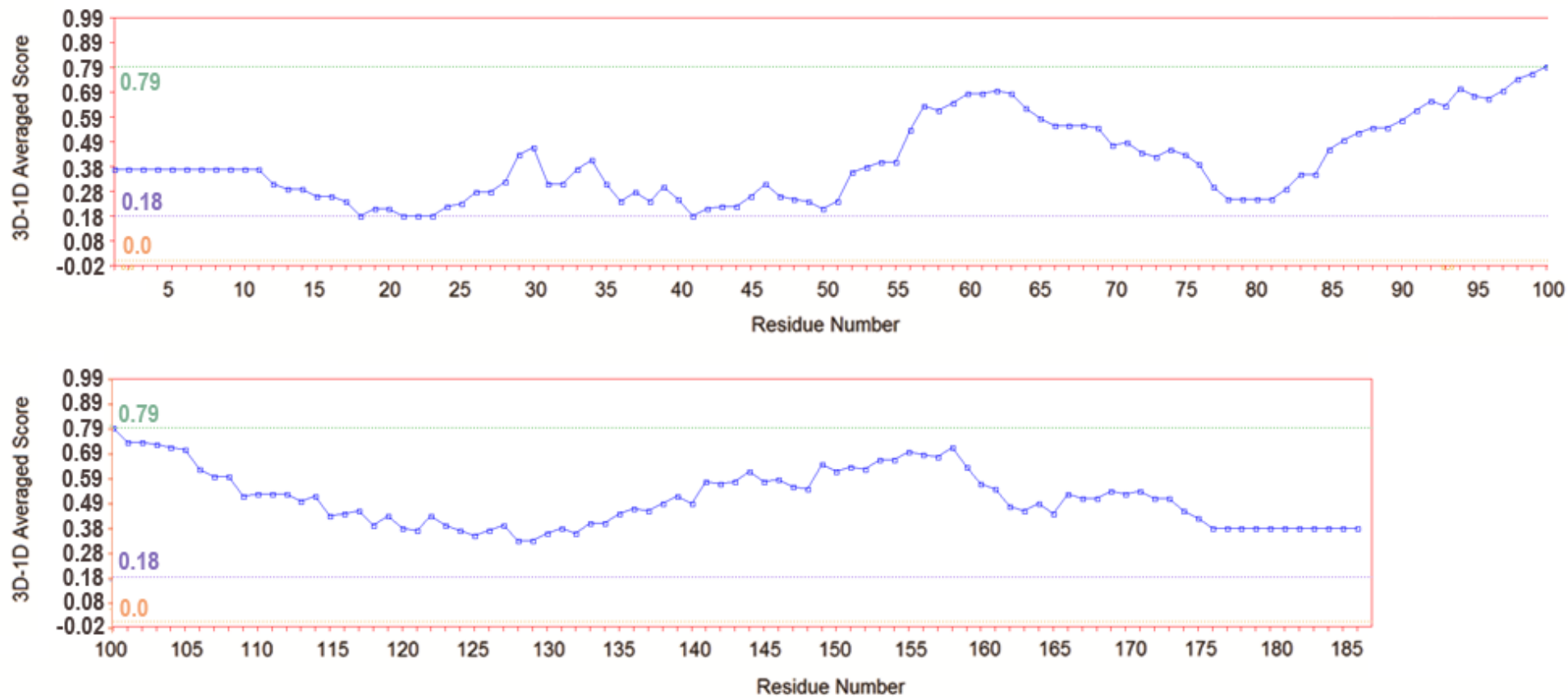
Figure 3.19 Verify3D plot for homology model DH-8. 97.3% of the residues had an averaged 3D-1D score of more than 0.2.

3.2.2 Docking of Standards towards DEN-2 NS2B-NS3 Models

An example of clustering histograms and estimated free energy of binding calculations for the docking results of standards – cardamonin, *R*-pinostrobin and *S*-pinostrobin docked towards 2FOM are shown in the Appendix section. From the histograms, the value of "Mean Binding Energy" was referred to instead of "Lowest Binding Energy" as "Mean Binding Energy" represents the free energy of binding for a group of conformations in a particular cluster while "Lowest Binding Energy" only represents one conformation with the lowest binding energy in a cluster. However, the conformation generated from each cluster is the conformation with the lowest binding energy; and this conformation is used for subsequent studies. The corresponding atomic coordinates of the standards were then merged with the atomic coordinates of 2FOM (same case for the other homology models, DH-1 to DH-8) for analysis using Ligplot 4.5.3 software. An example of Ligplot results are also shown in the Appendix section. These data were extracted computationally (following methods described in section 3.1.2.3) and are summarized in Table 3.1.

Table 3.1 : Docking output of the best binding conformations of the standard ligands towards DEN-2 NS2B-NS3 protease models which interacted with Lys74 from NS3. The $K_{i\,exp}$ values were obtained from a previously reported study (Kiat *et al.*, 2006).

Models	Compound Identity	NumCl	ΔG_{dock} (kcal mol^{-1})	$K_{i\,exp}$ (µM) from previous study[9] / $K_{i\,dock}$ (µM)
	Cardamonin			377 ±77
	Pinostrobin			345 ±70
		NumCl	ΔG_{dock} (kcal mol^{-1})	$K_{i\,dock}$ (µM)
2FOM	Cardamonin	10/100	-6.90	**13**
	R-pinostrobin	15/100	-7.55	**3.8**
	S-pinostrobin	14/100	-7.69	**3.8**
DH-1	Cardamonin	13/100	-4.75	**449**
	R-pinostrobin	28/100	-5.13	**243**
(2FP7 as template)	*S*-pinostrobin	11/100	-4.80	**414**
DH-2	Cardamonin	1/100	-4.29	**948**
	R-pinostrobin	58/100	-4.55	**622**
(2GGV as template)	*S*-pinostrobin	59/100	-4.66	**520**
DH-3	Cardamonin	1/100	-4.66	**520**
	R-pinostrobin	1/100	-4.84	**388**
(2IJO as template)	*S*-pinostrobin	5/100	-4.45	**731**

Table 3.1, continued

		NumCl	ΔG_{dock}	$K_{i\,dock}$
DH-4	Cardamonin	5/100	-5.41	**154**
(3E90 as template)	*R*-pinostrobin	29/100	-5.36	**167**
	S-pinostrobin	23/100	-4.78	**428**
DH-5	Cardamonin	2/100	-5.68	**99**
(3L6P as template)	*R*-pinostrobin	0/100	-	**-**
	S-pinostrobin	6/100	-4.74	**457**
DH-6	Cardamonin	0/100	-	**-**
(3LKW as template)	*R*-pinostrobin	0/100	-	**-**
	S-pinostrobin	0/100	-	**-**
DH-7	Cardamonin	2/100	-4.95	**325**
(3U1I as template)	*R*-pinostrobin	1/100	-4.24	**1028**
	S-pinostrobin	12/100	-4.41	**780**
DH-8	Cardamonin	1/100	-3.44	**3765**
(3U1J as template)	*R*-pinostrobin	6/100	-4.34	**874**
	S-pinostrobin	1/100	-3.81	**2066**

Bold and black = estimation of Ki about the same as Ki exp.

Bold and red = overestimation of Ki.

Bold and blue = underestimation of Ki.

NumCl = the number of conformations with RMSD < 0.5.

ΔG_{dock} = free energy of binding estimated from AutoDock 4.2 software.

$K_{i\,dock}$ = inhibition constant derived from ΔG_{dock}.

$K_{i\,exp}$ = inhibition constant calculated from *in vitro* experiment.

3.2.3 Selection of Model for Non-competitive Inhibition Study

Docking of the standard ligands, cardamonin, *R*-pinostrobin and *S*-pinostrobin, yielded best binding conformations with $K_{i\ dock}$ values that were in range with the reported $K_{i\ exp}$ values (Kiat *et al.*, 2006), and interated with Lys74 from NS3 only when DH-1 was used as the docking target (Tables 3.1). Thus, we suggest that the conformation of the allosteric binding pocket should resemble closely to that of DH-1. However, *R*-pinostrobin docked into DH-1 with a lower $K_{i\ dock}$ value than that of *S*-pinostrobin and the previously reported $K_{i\ exp}$ value. Future interaction analyses suggested that the better ΔG_{dock} and $K_{i\ dock}$ values of *R*-pinostrobin compared to *S*-pinostrobin might be due to the extra pi-cation interations between the phenyl side chain of *R*-pinostrobin and the ammonium groups of residues Lys73 and Lys74 (Figure 3.20). Hence, DH-1 and *R*-pinostrobin were used as the target receptor and standard for virtual screening, respectively.
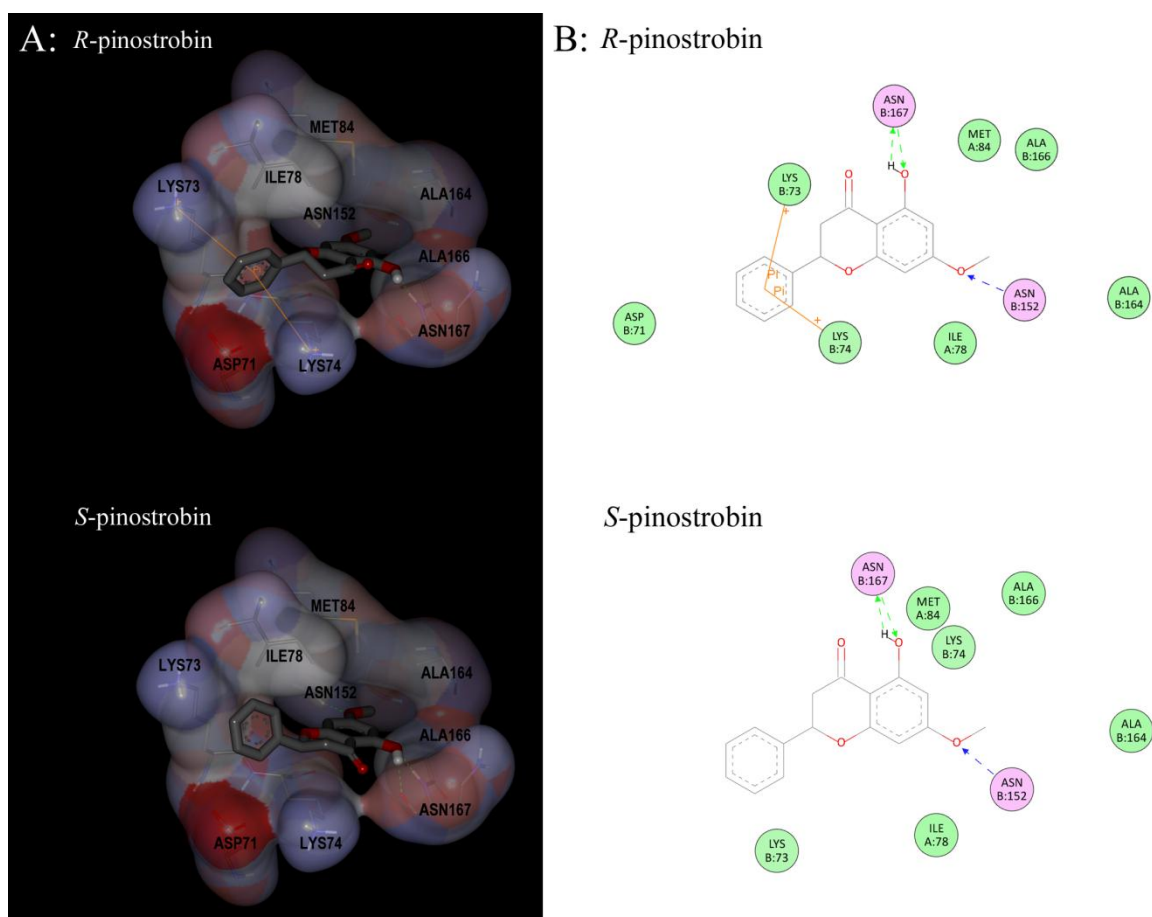
Figure 3.20 *R*-pinostrobin and *S*-pinostrobin docked at the allosteric binding site of DH-1. A: 3D representations of the binding poses. Ile78 and Met84 are from NS2B (chain A) while the rest of the residues are from NS3 (chain B). B: 2D representations of the binding modes. Residues with hydrogen bond interactions are coloured in pink, residues with hydrophobic or van der Waals interactions are coloured in green and pi-interactions are coloured in orange. The arrows are pointing from hydrogen bond donors towards hydrogen bond acceptors.

3.2.4　Virtual Screening

From the virtual screening of a total of 13,341 small compounds, 34 were identified to fulfill all the criteria of having ΔG_{dock} lower than that of the *R*-pinostrobin (standard), with NumCl of more than 10, and interacting with Lys74 of the NS3 protease in the allosteric binding site. However, only four of the small compounds (namely compounds 1 - 4 in this study) were available for purchase (Figure 3.21 and Table 3.2). The interactions of these 4 compounds (compound 1-4) towards the allosteric binding pocket of DH-1 model are shown in Figure 3.22, which were summarized in Table 3.2. The results from Ligplot analysis show there are hydrophobic or van der Waals interactions between compounds 1-4 and Lys74 from NS3 (Table 3.2). Further interaction analyses using Discovery Studio Visualizer 3.1 suggest the presence of pi-cation interactions between Lys74 and compounds 1, 2 and 3 (Figure 3.22). In addition, there were also hydrogen bonding interactions between Glu88, Gly124, Asn167 from NS3, and compounds 1, 2 and 4. The role of these amino acid residues (Lys74, Glu88, Gly124 and Asn167) is worth to be subjected for further study through site directed mutagenesis or protease-ligand interactions study by using x-ray crystallography.

Compounds 1 - 4 were then procured and tested for inhibition antivities.
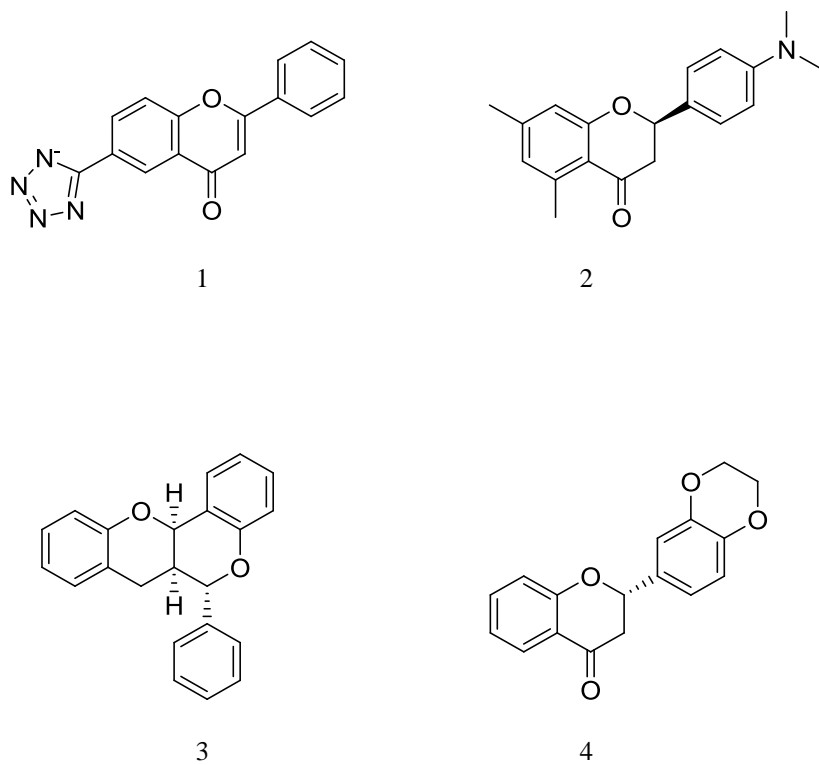
Figure 3.21 Structures of the small compounds identified from virtual screening against the DH-1 homology model, and were purchased due to their availability.
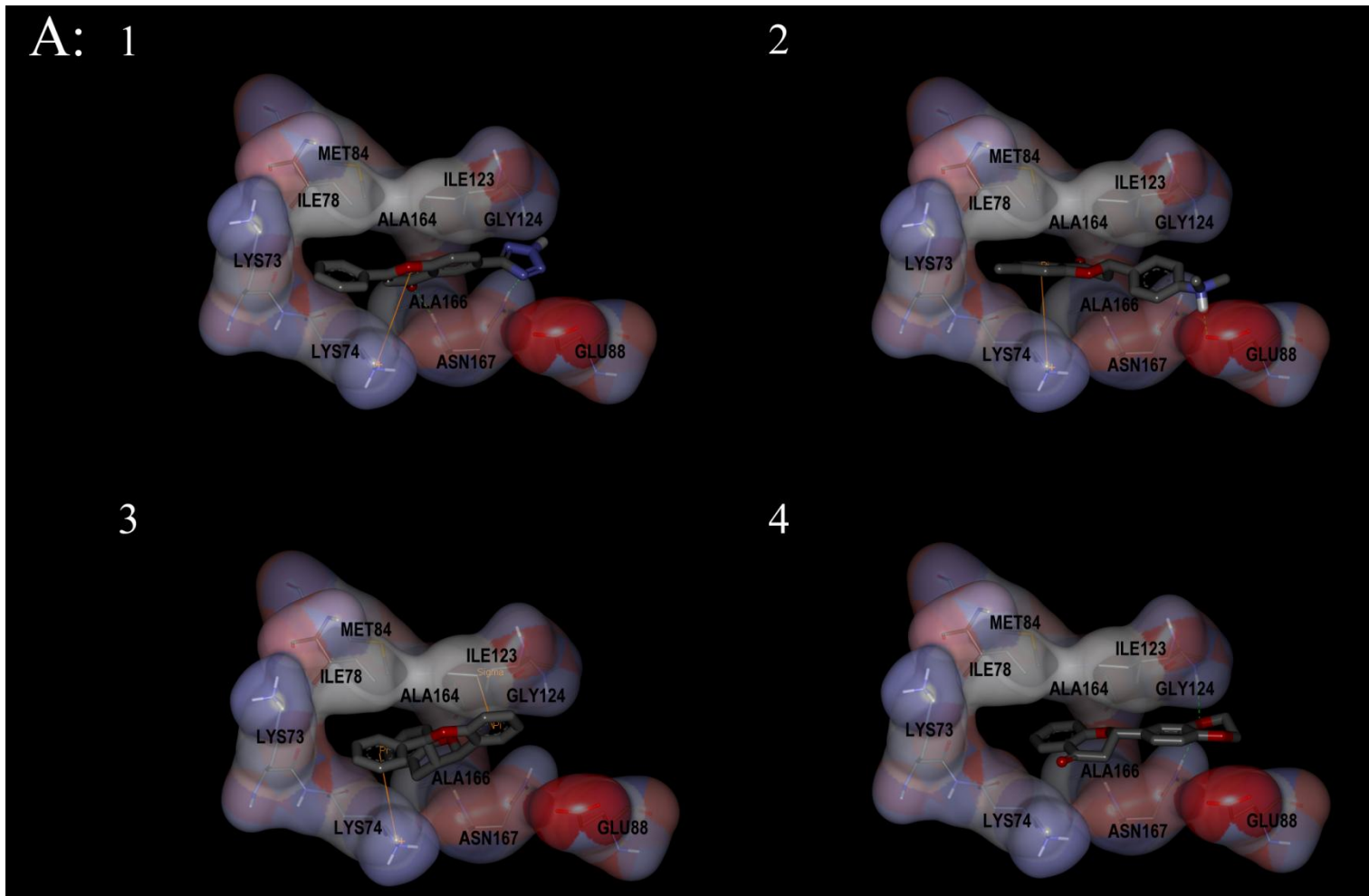
Figure 3.22A.

B: 1
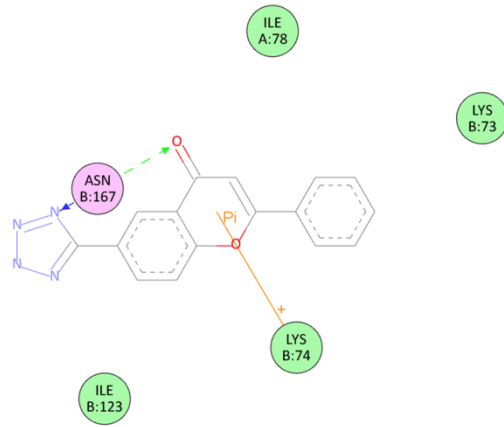
2
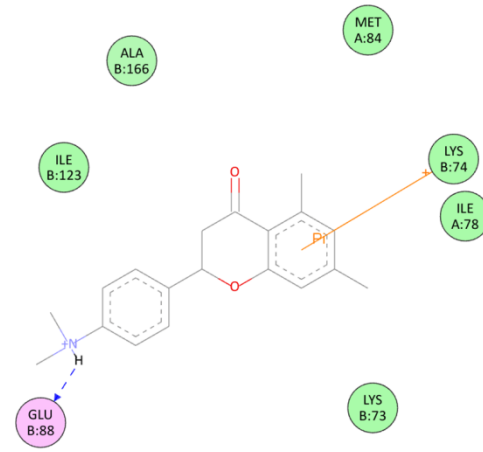
3
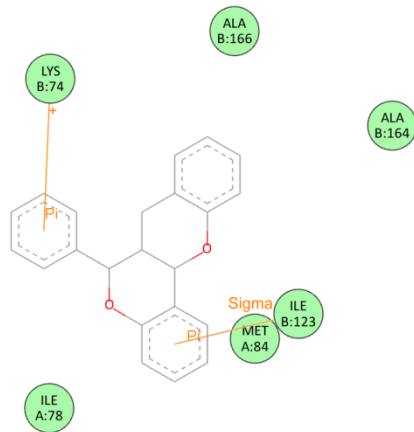
4
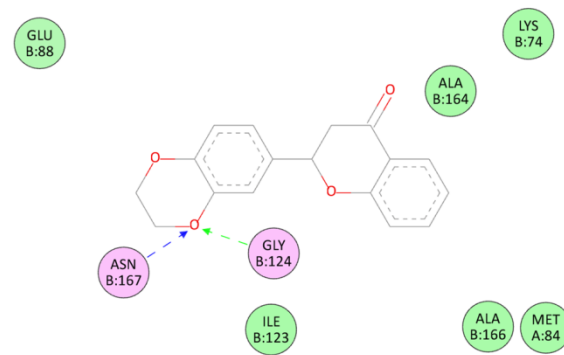
Figure 3.22B.

Figure 3.22   Compounds 1-4 docked at the allosteric binding site of DH-1. A: 3D representations of the binding poses. Ile78 and Met84 are from NS2B (chain A) while the rest of the residues are from NS3 (chain B). B: 2D representations of the binding modes. Residues with hydrogen bond interactions are coloured in pink, residues with hydrophobic or van der Waals interactions are coloured in green and pi-interactions are coloured in orange. The arrows are pointing from hydrogen bond donors towards hydrogen bond acceptors.

Table 3.2 : Small compounds identified from virtual screening against the DH-1 homology model which were available for purchase, with ΔG_{dock} lower than that of the standard *R*-pinostrohin, NumCl more than 10 and interacted with Lys74 from NS3.

Cpd	Zinc Codename / Chemical Name / (Source)	NumCl	ΔG_{dock} (kcal mol^{-1})	$K_{i\,dock}$ (μM)	Acceptor#	Donor#	Hydrophobic/ VDW#
1	ZINC00161709 2-phenyl-6-(1H-1,2,3,4-tetraazol-5-yl)-4H-chromen-4-one (Maybridge)	12/20	-6.17	45	B-ASN-167:1	B-ASN-167:1	A-ILE-78:2, B-LYS-73:5, ***B-LYS-74**:5, B-ILE-123:8, B-ASN-167:3
2	ZINC00148638 2-[4-(dimethylamino)phenyl]-5,7-dimethyl-3,4-dihydro-2H-1-benzopyran-4-one (Life Chemicals)	13/20	-5.77	86	B-GLU-88:1		A-ILE-78:5, A-MET-84:1, B-LYS-73:1, *** B-LYS-74**:4, B-GLU-88:1, B-ILE-123:9, B-ALA-166:2
3	ZINC16133933 6-phenyl-6a,12a-dihydro-6H,7H-chromeno[4,3-b]chromene (ChemBridge)	15/20	-5.33	175			A-ILE-78:1, A-MET-84:2, ***B-LYS-74**:7, B-ILE-123:9, B-ALA-164:2, B-ALA-166:8
4	ZINC00064430 2-(2,3-dihydro-1,4-benzodioxin-6-yl)-3,4-dihydro-2H-1-benzopyran-4-one (Szintekon)	11/20	-5.29	187	B-GLY-124:1, B-ASN-167:1	A-MET-84:2, ***B-LYS-74**:1, B-GLU-88:1, B-ILE-123:9, B-ALA-164:2, B-ALA-166:4, B-ASN-167:1	

NumCl = the number of conformations with RMSD < 2.0.

ΔG_{dock} = free energy of binding estimated from AutoDock 4.2 software.

$K_{i\,dock}$ = inhibition constant derived from ΔG_{dock}.

Acceptor = Hydrogen bond acceptor; Donor = Hydrogen bond donor; Hydrophobic/VDW = Hydrophobic or van der Waals interaction.

IF =Interaction frequency (number of interactions from the residues in binding site with the small compounds).

#Format of interaction indication = (Chain identity)-(Residue name)-(Residue number) : (*IF*).

Chain identity: A = NS2B; B = NS3.

*, bold and underlined: NS3 Lys74 (suggested key residue found in the allosteric site (Othman *et al.*, 2008)).

CHAPTER FOUR

IN VITRO STUDIES

4.1   Methodology

4.1.1   Cloning, Expression and Purification of Soluble DEN-2 NS2B-NS3 Protease

The following Table 4.1 shows the list of recipes or contents for the reaction mixtures or kits which were used for the *in vitro* studies:-

Table 4.1 :   Reaction mixtures or kits used in this study.

Code	Reaction Mixture/Kit	Recipe/Content	Amount
RMK01	10x PCR Buffer (1.0 mL) (200 mM Tris, 500 mM KCl, pH 8.4)	- Tris Base	24.0 mg
		- KCl	37.0 mg
		- pH 8.4	adjusted with HCl (and KOH)*
		- RNase-free Water	topped up to 1.0 mL
RMK02	PCR Master Mix (45.0 μL)	- 10x PCR Buffer (RMK01)	5.0 μL
		- 25 mM MgCl$_2$	2.0 μL
		- 10 mM dNTP Mix	5.0 μL
		- Forward Primer NS2B_F1 or GT_F1	0.5 μL
		- Reserve Primer NS3_R1 or GT_R1	0.5 μL
		- RNase-free Water	topped up to 45.0 μL
RMK03	QIAquick PCR Purification Kit (Qiagen)	- QIAquick Spin Columns	50
		- Buffer PBI	30.0 mL
		- Buffer PE	2 x 6.0 mL
		- Buffer EB	15.0 mL
		- Collection Tubes (2 mL)	50
RMK04	QIAquick Gel Extraction Kit (Qiagen)	- QIAquick Spin Columns	50
		- Buffer QG	2 x 50.0 mL
		- Buffer PE	2 x 10.0 mL
		- Buffer EB	15.0 mL
		- Collection Tubes (2 mL)	50
RMK05	LB Medium (1.0 L)	- Yeast Extract	5.0 g
		- Bacto-tryptone	10.0 g
		- NaCl	10.0 g
		- dH$_2$O	topped up to 1.0 L
RMK06	TransformAid Bacterial Transformation Kit (Fermentas)	- C-Medium	35.0 mL
		- T-Solution (A)	2 x 1.25 mL
		- T-Solution (B)	2 x 1.25 mL
RMK07	LB Agar (200.0 mL)	- Yeast Extract	1.0 g
		- Bacto-tryptone	2.0 g
		- NaCl	2.0 g
		- Agar	3.0 g
		- dH$_2$O	topped up to 200.0 mL

Table 4.1, continued

RMK08	QIAprep Spin Miniprep Kit (Qiagen)	- QIAprep Spin Columns	50
		- Buffer P1 (with RNase A added)	20.0 mL
		- Buffer P2	20.0 mL
		- Buffer N3	30.0 mL
		- Buffer PB	30.0 mL
		- Buffer PE	2 x 6.0 mL
		- Buffer EB	15.0 mL
		- Collection Tubes (2 mL)	50
RMK09	10x TBE Buffer (1.0 L) (890 mM Tris, 890 mM Boric Acid, 20 mM EDTA, pH 8.0)	- Tris Base	108.0 g
		- Boric Acid	55.0 g
		- EDTA	40.0 mL
		- pH 8.0	adjusted with HCl (and NaOH)*
		- dH$_2$O	topped up to 1.0 L
RMK10	Lysis Buffer (50.0 mL) (50mM HEPES, 300 mM NaCl, 20 mM Imidazole, 5% Glycerol, pH 7.5)	- HEPES	600.0 mg
		- NaCl	875.0 mg
		- Imidazole	68.0 mg
		- Glycerol	2.5 mL
		- Lysozyme	50.0 mg
		- pH 7.5	adjusted with HCl (and NaOH)*
		- dH$_2$O	top up to 50.0 mL
RMK11	Column Buffer (1.0 L) (50 mM HEPES, 300 mM NaCl, 20 mM Imidazole, pH 7.5)	- HEPES	12.0 g
		- NaCl	17.5 g
		- Imidazole	1.4 g
		- pH 7.5	adjusted with HCl (and NaOH)*
		- dH$_2$O	top up to 1.0 L
RMK12	Elution Buffer (10.0 mL) (50 mM HEPES, 300 mM NaCl, 100 mM Imidazole, pH 7.5)	- HEPES	120.0 mg
		- NaCl	175.0 mg
		- Imidazole	68.0 mg
		- pH 7.5	adjusted with HCl (and NaOH)*
		- dH$_2$O	top up to 10.0 mL
RMK13	Monomer Stock Solution (250.0 mL) (30% Acrylamide, 0.8% *N*,*N*'-methylenebisacrylamide)	- Acrylamide	75.0 g
		- *N*,*N*'-methylenebisacrylamide	2.0 g
		- dH$_2$O	top up to 250.0 mL
RMK14	4x Resolving Gel Buffer (500.0 mL) (1.5 M Tris, 0.4% SDS, pH 8.8)	- Tris Base	91.0 g
		- SDS	2.0 g
		- pH 8.8	adjusted with HCl (and NaOH)*
		- dH$_2$O	top up to 500.0 mL
RMK15	4x Stacking Gel Buffer (100.0 mL) (0.5 M Tris, 0.4% SDS, pH 6.8)	- Tris Base	6.1 g
		- SDS	0.4 g
		- pH 6.8	adjusted with HCl (and NaOH)*
		- dH$_2$O	top up to 100.0 mL
RMK16	12% Acrylamide in Resolving Gel Solution (1 gel)	- Monomer Stock Solution (RMK13)	3000.0 μL
		- 4x Resolving Gel Buffer (RMK14)	1875.0 μL
		- dH$_2$O	2625.0 μL
		- TEMED	5.0 μL
		- 10% APS (100 mg/mL)	25.0 μL

Table 4.1, continued

RMK17	Stacking Gel Solution (1 gel)	- Monomer Stock Solution (RMK13)	325.0 μL
		- 4x Resolving Gel Buffer (RMK15)	625.0 μL
		- dH$_2$O	1502.5 μL
		- TEMED	2.5 μL
		- 10% APS (100 mg/mL)	12.5 μL
RMK18	SDS Electrophoresis Buffer (1.0 L) (25 mM Tris, 192 mM Glycine, 0.1% SDS)	- Tris Base	3.0 g
		- Glycine	14.4 g
		- SDS	1.0 g
		- dH$_2$O	topped up to 1.0 L
RMK19	2x SDS-PAGE Loading Buffer (10.0 mL) (125 mM Tris, 20% Glycerol, 0.5% SDS, 2% 2-mercaptoethanol, 0.001% Bromophemol Blue)	- 4x Stacking Gel Buffer (RMK15)	2.5 mL
		- Glycerol	2.0 mL
		- SDS	0.4 g
		- 2-mercaptoethanol	0.2 mL
		- Bromophenol Blue	0.1 mg
		- dH$_2$O	topped up to 10.0 mL
RMK20	Fixing Solution and Destaining Solution (1.0 L) (50% Methanol, 10% Acetic Acid)	- Methanol	500.0 mL
		- Acetic Acid	100.0 mL
		- dH$_2$O	topped up to 1.0 L
RMK21	Staining Solution (1.0 L) (0.1% CBB R-250, 50% Methanol, 10% Acetic Acid)	- CBB R-250	1.0 g
		- Methanol	500.0 mL
		- Acetic Acid	100.0 mL
		- dH$_2$O	topped up to 1.0 L
RMK22	Storage Solution (1.0 L) (5% Acetic Acid)	- Acetic Acid	50.0 mL
		- dH$_2$O	topped up to 1.0 L
RMK23	Transfer Buffer (1.0 L) (25 mM Tris, 192 mM Glycine, 10% Methanol)	- Tris Base	3.0 g
		- Glycine	14.4 g
		- Methanol	100.0 mL
		- dH$_2$O	topped up to 1.0 L
RMK24	Blocking Solution (50.0 mL) (5% Non-fat Dry Milk in TBS)	- Non-fat Dry Milk	2.5 g
		- TBS (RMK25)	50.0 mL
RMK25	TBS Solution (500.0 mL) (50 mM Tris, 150 mM NaCl, pH 7.6)	- Tris Base	3029.0 mg
		- NaCl	4383.0 mg
		- pH 7.6	adjusted with HCl (and NaOH)*
		- dH$_2$O	topped up to 500.0 mL
RMK26	TBST Solution (500.0 mL) (50 mM Tris, 150 mM NaCl, 0.05% Tween 20, pH 7.6)	- Tris Base	3029.0 mg
		- NaCl	4383.0 mg
		- Tween 20	250.0 μL
		- pH 7.6	adjusted with HCl (and NaOH)*
		- dH$_2$O	topped up to 500.0 mL
RMK27	Tris-HCl Buffer (250.0 mL) (200 mM Tris, pH 8.5)	- Tris Base	6.0 g
		- pH 8.5	adjusted with HCl (and NaOH)*
		- dH$_2$O	topped up to 250.0 mL

*If pH was decreased below the favourable value by HCl, KOH or NaOH would be used to increase the pH for readjustment.

All buffers were stored at 4 ℃ before further usage, with a maximum of 6 months storage period.

4.1.1.1 Designing of Soluble DEN-2 NS2B-NS3 Protease Gene Insert

The NS2B-NS3 serine protease nucleotide sequence begins from nucleotide 4132 to nucleotide 5031 in the full length nucleotide sequence of DEN-2 gene, New Guinea C strain (Irie *et al.*, 1989). Designing of the nucleotide sequence for His-tagged recombinant soluble functional NS2B-NS3 protease (CF40.gly(T).NS3pro) is based on previous studies on insoluble DEN-2 NS2B-NS3 protease (Yusof *et al.*, 2000), soluble DEN-2 NS2B-NS3 protease with truncated NS2B region (CF40.gly.NS3pro) (Leung *et al.*, 2001) and DEN-2 NS2B-NS3 protease crystal structure, 2FOM (Erbel *et al.*, 2006). According to these studies, a glycine linker should be constructed and inserted between the hydrophilic NS2B gene (CF40: nucleotide 4276-4416) and NS3 serine protease gene (NS3pro: nucleotide 4522-5076) in order to get a single soluble DEN-2 NS2B-NS3 virus protease. This protease is flexible enough to be active without cleaving itself into individual NS2B cofactor and NS3 serine protease. However, in constrast to the previous works, in this study, a glycine linker with the peptide sequence $Gly_4ThrGly_4$ (nucleotide sequence – GGGGGCGGAGGTACCGGTGGAGGCGGG) was designed, instead of $Gly_4SerGly_4$. This was mainly because, by using Threonine (Thr) instead of Serine (Ser) would be more beneficial by giving an extra *Kpn*I restriction site, with GGTACC recognition nucleotide sequence in both CF40 and NS3pro genes for easier digestion and subsequent ligation. An additional nine nucleotides were also added to the NS3pro gene in the design of the reverse primer. This is to prevent primer dimerization of the reverse primer used, as reported previously (Leung *et al.*, 2001). Nevertheless, same to the study by Leung *et al.* (2001), the insert's forward primer was designed to have the *Bam*HI restriction site, with recognition sequence (GGATCC), while the reverse primer was designed to have *Hind*III restriction site, with recognition sequence (AAGCTT). Hence, this design allowed the protease gene to be cloned into the vector, pQE30 (Qiagen) (Figure 4.1), which also contains the two restriction sites. In addition,

pQE30 vector also contains the ampicilin resistant gene, which is required in subsequent culture work. These features of the designed insert are shown in Figure 4.2.
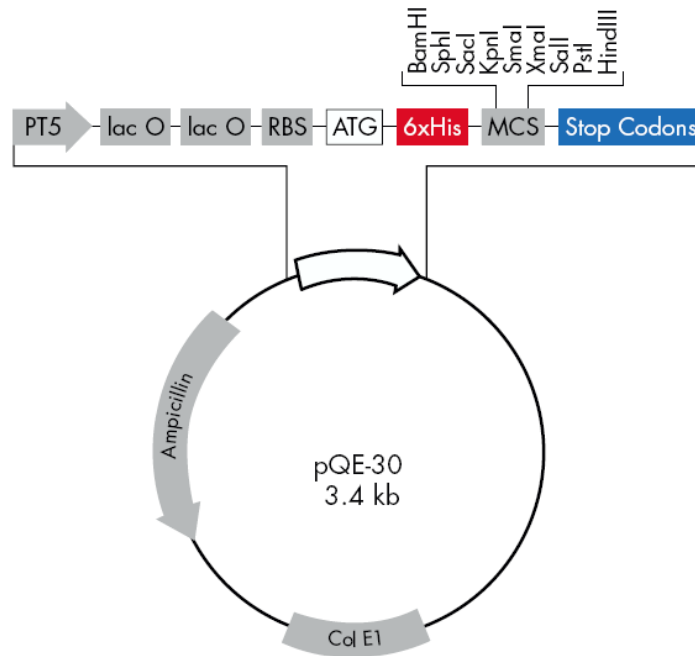
pQE-30 Vector

Positions of elements in bases	pQE-30
Vector size (bp)	3461
Start of numbering at *Xho*I (CTCGAG)	1–6
T5 promoter/lac operator element	7–87
T5 transcription start	61
6xHis-tag coding sequence	127–144
Multiple cloning site	145–192
Lambda t_0 transcriptional termination region	208–302
rrnB T1 transcriptional termination region	1064–1162
ColE1 origin of replication	1638
β-lactamase coding sequence	3256–2396

Figure 4.1   Sequence map of pQE30 vector from Qiagen. The image was adopted from http://www.qiagen.com/literature/pqesequences/pqe3x.pdf.

```
                                                        BamHI restriction enzyme
                          ccg gga tcc                   will act here

4041                                                                agctggcca
4141 ctaaatgagg ctatcatggc agtcgggatg gtgagcattt tggccagttc actcctaaag
4201 aatgacattc ccatgacagg accattagtg gctggagggc tcctcactgt gtgctacgtg
4261 ctcactggac gatcggccga tttggaactg gagagagccg ccgatgtcaa atgggaagat
4321 caggcagaga tatcaggaag cagtccaatc ctgtcaataa caatatcaga agatggtagc
4381 atgtcgataa aaaacgaaga ggaagaacaa acactgacca tactcatcag aacaggattg

                                           ggg ggc gga ggt acc
                        KpnI restriction    ggt gga ggc ggg
                        enzyme will act
   ggg ggc gga ggt acc  here followed
   ggt gga ggc ggg      by DNA ligase

4441 ctggtgatct caggactttt tcctgtatca ataccaatca cggcagcagc atggtacctg
4501 tgggaagtga agaaacaacg ggctggagta ttgtgggatg tcccttcacc cccacccgtg
4561 ggaaaggctg aactggaaga tggagcctat agaatcaagc aaaaagggat tcttggatat
4621 tcccagatcg gagccggagt ttacaaagaa ggaacattcc atacaatgtg gcatgtcaca
4681 cgcggcgctg ttctaatgca taaaggaaag aggattgaac catcatgggc ggacgttaag
4741 aaagacctaa tatcatatgg aggaggctgg aagctagaag gagaatggaa ggaaggagaa
4801 gaagtccagg tcttggcatt ggagcctgga aaaaatccaa gagccgtcca aacaaaacct
4861 ggtcttttca aaaccaacgc cggaaccata ggtgccgtat ctctggactt ttctcctgga
4921 acctcaggat cgccaatcat cgacaaaaaa ggaaaagttg tgggtcttta tggtaatggt
4981 gttgttacaa ggagtggagc atatgtgagt gctatagccc agactgaaaa aagtattgaa
5041 gacaatccag agatcgaaga tgacattttt cgaaagagaa aattg

                                                    aag ctt ggg
                        HindIII restriction
                        enzyme will act here
```

Figure 4.2   Nucleotide sequence of NS2B and NS3 genes. NS2B nucleotide sequence is from 4132 to 4521 while NS3 serine protease nucleotide sequence is from 4522 until 5031. The first sequence in red is the hydrophilic region of NS2B, CF40 (4276 - 4416), and the second sequence in red is the region of NS3 serine protease, NS3pro (4522 - 5076) that was used in previous study (Leung *et al.*, 2001). Our designed sequence of the NS3pro was chosen from 4522 to 5085 instead. The two sequences in red boxes were used for designing the forward primer, NS2B_F1 and reverse primer, NS3_R1 for insoluble and soluble DEN-2 NS2B-NS3 protease gene. *Bam*HI restriction site (GGATCC) is underlined in the upper red box and *Hind*III restriction site (AAGCTT) is underlined in the lower red box. The two sequences in blue boxes are the nucleotide sequence (GGGGGCGGAGGTACCGGTGGAGGCGGG)

for glycine linkers, which were used for designing the forward and reverse primer pairs, GT_F1 and GT_R1. *Kpn*I restriction site (GGTACC) is underlined in the blue boxes.

4.1.1.2 Primers' Design

All the primers were designed using an online PrimerAnalyser software (http://primerdigital.com/tools/PrimerAnalyser.html) from the PrimerDigital website (http://primerdigital.com), a web tool for calculating primer's melting temperature for suitable PCR annealing temperature. The website also predicts the efficiency of PCR by identifying primers' dimer formation (Kalendar, 2010) before the primers' sequences were sent for synthesis in a commercial company.

4.1.1.3 Construction of Soluble DEN-2 NS2B-NS3 Protease Gene Insert from Insoluble DEN-2 NS2B-NS3 Protease Gene

Reverse Transcription Polymerase Chain Reactions (RT-PCR) were conducted for DEN-2 virus RNA (obtained from Sungai Buloh Health Laboratory) with forward primer NS2B_F1, 5'-CCG GGATCC GCCGATTTGGAACTG-3' (*Bam*HI restriction site is underlined) and reverse primer NS3_R1, 5'-CCC AAGCTT CAATTTTCTCTTTCG-3' (*Hind*III restriction site is underlined) to get complimentary deoxyribonucleic acid (cDNA) of insoluble DEN-2 NS2B-NS3 protease gene with 828 base pairs (bp). The PCR Master Mix (Table 4.1:RMK02) was prepared in a 1.5 mL microcentrifuge. 2.5 μL of 40 μg/mL DEN-2 virus RNA and 2.5 μL of Taq polymerase (Fermentas) were then added into the PCR Master Mix just before the reaction mixture was subjected to RT-PCR. The heating steps for RT-PCR were followed as shown in Table 4.2 and were performed in a thermal cycler. The RT-PCR product was purified using QIAquick PCR Purification Kit (Qiagen) (Table 4.1:RMK03).

Table 4.2 : Heating steps for RT-PCR and PCR.

Reaction		Process	Number of cycle	Temperature	Duration
RT-PCR		cDNA synthesis	1	50 °C	30 minutes
	PCR	Denaturation	1	94 °C	2 minutes
		PCR amplification	40	94 °C 60 °C 72 °C	15 seconds (Denature) 30 seconds (Anneal) 1 minute (Extend)
		Final Extension	1	72 °C	10 minutes

For purification of PCR using QIAquick PCR Purification Kit, firstly, 5 volumes of Buffer PBI were added and mixed to one volume of the PCR reaction mixture. Then, a QIAquick spin column was placed in a 2 mL collection tube, and the sample was applied to the column, and centrifuged at 13,000 rpm at room temperature for 1 minute for DNA binding. The flow-through was then discarded and the column was placed back into the same tube. Subsequently, 0.75 mL of Buffer PE was added to the QIAquick spin column, and the column, along with the collection tube was centrifuged at 13,000 rpm at room temperature for 1 minute to wash out the unwanted primers and impurities, such as salts, enzymes, unincorporated nucleotides, agarose, dyes, ethidium bromide, oils and detergents. The flow-through was then discarded and the column was placed back again in the same tube. Following this, the column was centrifuged at 13,000 rpm at room temperature in the 2 mL collection tube for another 1 minute to remove residual ethanol from Buffer BE which may interfere with subsequent enzymatic reactions. The column was then placed in a clean 1.5 mL microcentrifuge tube. Next, 50 µl of Buffer EB was added to the center of the QIAquick membrane, and the column was let to stand for 1 minute before it was centrifuged at 13,000 rpm at room temperature for 1 minute to elute the purified DNA. The purified DNA was kept at -20 °C until further use.

PCR was then carried out for amplifying individual CF40 with 177 bp and NS3pro with 600 bp from insoluble DEN-2 NS2B-NS3 cDNA. The amplification for CF40 was carried out using the same reaction mixture of PCR Master Mix, but replaced with different forward and reverse primer pairs, NS2B_F1 / GT_R1 (5'-CCCGCCTCCACC<u>GGTACC</u>TCCGCCCCCCAGTGTTTGTTCTTCCTC-3'; *Kpn*I restriction site is underlined), and DEN-2 virus RNA was also replaced with insoluble DEN-2 NS2B-NS3 cDNA. On the other hand, the forward and reverse primer pairs, GT_F1 (5'-GGGGGCGGA<u>GGTACC</u>GGTGGAGGCGGGGCTGGAGTATTGTGGGA -T-3') / NS3_R1 were used for amplifying NS3pro and the rest of the reaction mixture was followed same as that of CF40. The heating steps for PCR were done according to Table 3.2 by using a thermal cycler. The PCR product was then purified using QIAquick PCR Purification Kit (Qiagen) with the same protocol mentioned in the previous paragraph. Calculations for number of base pairs in RT-PCR and PCR products are shown in Table 4.3.

Table 4.3 : Base pairs calculation of each gene.

Gene from RT-PCR / PCR	Base pairs from nucleotide sequence	Base pairs from primers / glycine linker	Total base pairs
Insoluble NS2B-NS3 (4276 - 5085)	5085 – 4276 + 1 = 810	2 x 9 =18	828
Hydrophillic NS2B, CF40 (4276 - 4416)	4416 – 4276 + 1 = 141	9 + 27 = 36	177
NS3 Serine Protease, NS3pro (4522 - 5085)	5085 – 4522 + 1 = 564	9 + 27 = 36	600
Soluble NS2B-NS3, CF40-gly(T).NS3pro	141 + 564 = 705	(2 x 9) + 27 = 45	750

Subsequently, digestions were performed for both genes of the hydrophilic NS2B and NS3 serine protease (CF40 and NS3pro) with glycine linker (Gly$_4$ThrGly$_4$) genes at the 3' and 5' ends, respectively, using *Kpn*I restriction enzyme at 37 ℃ for 4 hours. The two digested genes were then ligated with DNA ligase at 16 ℃ for 1 hour in the ratio of 1 to 1. The ligation of genes was verified using agarose gel electrophoresis (described in later section, 4.1.1.5) and the gene with the correct base pairs was cut out from the agarose gel and extracted by using a QIAquick Gel Extraction Kit (Qiagen) (Table 4.1:RMK04).

For gel extraction using the QIAquick Gel Extraction Kit, firstly, the DNA fragment was excised from the agarose gel with a clean and sharp scalpel. The size of the gel slice was minimized by removing extra agarose. The gel slice was then weighed in a colourless 1.5 mL microcentrifuge tube. Buffer QG (3 volumes) was added to 1 volume of gel (100 mg equivalent to approximately 100 μL) in the same tube. For gel slices which were more than 400 mg, more than one QIAquick spin columns were used since the maximum amount of gel slice per QIAquick spin column is 400 mg. The tube was then incubated at 50 ℃ for 10 minutes until the gel slice had completely dissolved. The reaction mixture was mixed by vortexing the tube every 2 - 3 minutes during the incubation to help dissolve the gel. After the gel slice had dissolved completely, the colour of the mixture was checked. This should be yellow to ensure that the pH ≤ 7.5 in order to maximize the adsorption of DNA to the QIAquick membrane. If pH > 7.5, the colour of Buffer QG turns to orange or violet, and the adsorption of DNA becomes less optimal. Next, the QIAquick spin column was placed in a 2 mL collection tube. The dissolved sample was applied to the QIAquick spin column and was centrifuged at 13,000 rpm at room temperature for 1 minute for DNA binding. As the maximum volume of the column reservoir was 800 μl, sample volumes of more than 800 μl were

loaded and spun again. The flow-through was discarded and the column was placed back in the same collection tube. Then, 0.5 mL of Buffer QG was added to the QIAquick spin column and was centrifuged at 13,000 rpm at room temperature for 1 minute to remove all traces of agarose. Buffer PE (0.75 mL) was then added to the column and it was centrifuged at room temperature for 1 minute to wash out the unwanted primers and impurities. The flow-through was discarded and the column was centrifuged at 13,000 rpm at room temperature for an additional 1 minute to remove residual ethanol from Buffer PE which could interfere with subsequent enzymatic reactions. QIAquick spin column was then placed into a clean 1.5 mL microcentrifuge tube. Subsequently, 50 µl of Buffer EB was added to the center of the QIAquick membrane, the column was let to stand for 1 minute and then was centrifuged at 13,000 rpm at room temperature for 1 minute to elute the DNA. The eluted DNA was kept at -20 ℃ until further use.

Subsequent PCR was then performed to amplify the ligated gene CF40.gly(T).NS3pro (CF40-Gly$_4$ThrGly$_4$-NS3pro) with primer pairs NS2B_F1/NS3_R1, so that the concentration was enough for it to become an insert for cloning. This was followed by *Bam*HI and *Hind*III restriction enzymes' digestions of both the insert and pQE30 vector for 4 hours at 37 ℃, for each of the digestion process. The products were then ligated with DNA ligase at 16 ℃ for 1 hour for the insert and pQE30 vector (ratio 1:1) with concentration of 20 µg/mL each.

4.1.1.4 Transformation of Ligated Insert and Vector

XL1-blue MRF' strain *Escherichia coli* bacteria, purchased from Stratagene company was cultured overnight for a maximum of 16 hours in Lysogeny Broth (LB) medium (Table 4.1:RMK05) on the day before transformation. The medium was then

autoclaved and 1 mL of 100 mg/mL of ampicillin was added after it was cooled to 50 ℃. The final ligation product, pQE30.CF40.gly(T).NS3pro was then transformed into XL1-blue MRF' bacteria using TransformAid Bacterial Transformation Kit (Fermentas) (Table 4.1:RMK06), followed by culture on LB agar plate containing 100 μg/mL of ampicilin with incubation of 16 hours at 37 ℃.

For preparation of LB agar plates, firstly, LB agar (Table 4.1:RMK07) was autoclaved and 200 μL of 100 mg/mL of ampicillin was added after it was cooled to 50 ℃. Then, the agar was poured into different Petri plates containing about 25 mL of agar, and the LB agar plates were then stored at 4 ℃ prior to usage.

For transformation of plasmid to the bacteria using TransformAid Bacterial Transformation Kit, firstly, 2 mL of C-Medium was inoculated with XL1-blue MRF' strain *Escherichia coli* bacteria (from frozen stock) in a culture tube. The culture was then incubated overnight in a shaker at 37 ℃. The next day, the culture tube containing 1.5 mL of new C-Medium, an amount sufficient for 2 transformations was pre-warmed at 37 ℃. Then, 1/10 volume of the overnight culture (0.15 mL) was added to the culture tube and the tube was incubated in a shaker at 37 ℃ for 20 minutes. Subsequently, 2 LB agar plates containing 100 μg/mL of ampicillin each, were pre-warmed at 37 ℃ for 20 minutes. TransformAid T-Solution was prepared by mixing 250 μl of T-Solution A and 250 μl of T-Solution B, to a total volume of 500 μl for 2 transformations. The T-Solution was then kept on ice. Fresh culture (1.5 mL) was dispensed into a tube and centrifuged at 13,000 rpm at room temperature for 1 minute. The supernatant was discarded and the pelleted cells were resuspended in 300 μl TransformAid T-Solution. The mixture was incubated on ice for 5 minutes. The cells were centrifuged at 13,000 rpm at room temperature again for 1 minute and then the supernatant was removed.

Next, the cells were resuspended in 120 μl of TransformAid T-Solution and incubated on ice for 5 minutes. After that, 2 sets of 2.5 μl of ligation products, pQE30.CF40.gly(T).NS3pro were dispensed into new microcentrifuge tubes and placed on ice for 2 minutes. Resuspended cells (50 μl) were added to each tube containing the pQE30.CF40.gly(T).NS3pro and incubated on ice for 5 minutes. Finally, the cells were plated on pre-warmed LB agar plates with 100 μg/mL of ampicillin and then incubated overnight at 37 ℃.

Ten bacteria colonies were then picked and cultured again, separately, in LB medium containing 100 μg/mL of ampicilin, with incubation for 16 hours at 37 ℃. Plasmids of the colonies' cultures were extracted using a QIAprep Spin Miniprep Kit (Qiagen) (Table 4.1:RMK08). For plasmid extraction from one colony of the bacterial culture using the kit, firstly, the 1 mL of bacteria culture was centrifuged at 13,000 rpm at room temperature for 1 minute and then the supernatant was removed. The pelleted bacterial cells were then resuspended in 250 μl of Buffer P1 and transferred to a microcentrifuge tube. Buffer P2 (250 μl) was added into the tube and the mixture was mixed thoroughly by inverting the tube 4 - 6 times. Buffer N3 (350 μl) was then added and the mixture was mixed immediately and thoroughly, by inverting the tube 4 - 6 times. The mixture was, then, centrifuged at 13,000 rpm at room temperature for 10 minutes. The supernatant obtained was then pipetted to the QIAprep spin column, placed in a 2 mL collection tube and centrifuged at 13,000 rpm at room temperature for 1 minute. Next, the flow-through was discarded. Subsequently, 0.5 mL of Buffer PB was added to the QIAprep spin column and the column along with the tube was centrifuged at 13,000 rpm at room temperature for 1 minute to wash and remove endonucleases. This was to ensure that the plasmid DNA was not degraded. Buffer PE (0.75 mL) was added to the QIAprep spin column and it was centrifuged at 13,000 rpm

at room temperature for 1 minute to wash and remove salts. The flow-through was then discarded and the tube was centrifuged at 13,000 rpm at room temperature for an additional 1 minute to remove residual ethanol from Buffer PE. Then, the QIAprep column was placed into a clean 1.5 mL microcentrifuge tube. Following this, 50 μl of Buffer EB was added to the center of the QIAprep membrane and the column was left to stand for 1 minute and centrifuged at 13,000 rpm at room temperature for 1 minute to elute the plasmid. The extracted plasmid was then kept at -20 ℃ for further use.

After the extraction process was completed, the plasmids were verified with digestion of the *Bam*HI and *Hind*III restriction enzymes. This was to confirm that the soluble DEN-2 NS2B-NS3 protease gene had been transformed into the vector prior to further nucleotide sequencing (section 4.1.1.5). The successfully transformed cultures were than kept as glycerol stock, in 15% of glycerol, and stored at -80 ℃ for future use.

4.1.1.5 Verification of Gene Products

All gene products from RT-PCR, PCR, digestions and ligations were verified using agarose gel electrophoresis (1% w/v agarose gel − 0.5 g agarose in 50 mL TBE buffer (Table 4.1:RMK09), containing 0.5 μg/mL ethidium bromide). For running of agarose gel electrophoresis, firstly, the agarose gel was cast with 8 wells in a horizontal gel box. After the gel had solidified, it was transferred into a gel tank and the gel was then submerged in TBE buffer. The loading samples were prepared as a mixture of 5 μL of DNA samples (RT-PCR and PCR products, ligated genes and plasmids) and 2 μL of DNA loading dye, and were loaded into the wells of the agarose gel, along with 5 μL of 1 kb DNA ladder in one of the wells. The electrophoresis experiment was run at 100 V constant voltages for 1 hour. DNA stained with ethidium bromide will produce fluorescence under ultraviolet (UV) light, and thus, its base pairs can be recognized

after separation using gel electrophoresis and compared with the DNA ladder. To validate the identity of the protease gene, nucleotide sequencing for the plasmids extracted from the successfully transformed cultures (as mentioned in section 4.1.1.4) was done at Medigene Sdn. Bhd.

4.1.1.6 Expression and Purification of Soluble DEN-2 NS2B-NS3 Protease

pQE30 vector was used for high level, inducible expression of amino-terminal hexahistidine-tagged recombinant proteins (Yusof *et al.*, 2000; Leung *et al.*, 2001). Cultures of XL1-blue bacteria, transformed with pQE30.CF40.gly(T).NS3pro (plasmid with soluble DEN-2 NS2B-NS3 protease gene), were grown in 1 liters of LB medium containing 100 μg/mL of ampicilin at 37 °C, until $A_{600\ nm}$ reached 0.6. One mL of the culture was pelleted by centrifugation at 13,000 rpm for 1 minute and kept as "Uninduced Protein" in -20 °C for further analysis. The bacteria cells were then induced for expression by the addition of IPTG to a final concentration of 1 mM and incubated for an additional 3 hours at 37 °C. After the incubation, 1 mL of the culture was also pelleted by centrifugation at 13,000 rpm for 1 minute and kept as "Induced Protein" in -20 °C for further analysis. The rest of the cells were harvested by centrifugation at 6,000 x rpm for 15 minutes and the pellets were stored at -80 °C until used.

For protein purification, the cell pellets were initially thawed and resuspended in 5.0 mL of lysis buffer (Table 4.1:RMK10) for every 1.0 g of cell pellet. The resuspended cells were subjected to probe sonication for five 30 seconds pulses on ice and then centrifuged at 10,000 rpm for 1 hour at 4 °C. Then, the supernatant was filtered through a 0.45 μm sterilized filter and 0.5 mL of the filtered supernatant was kept as "Total Protein" in -20 °C for further analysis. The rest of the filtrate was then, immediately incubated with 2 mL of Ni^{2+}-NTA-agarose (Qiagen), which was pre-

equilibrated with column buffer (Table 4.1:RMK11), in a 50 mL centrifuge tube for 1 hour. This step ensures the His-tagged protein binds to the Ni^{2+}-NTA-agarose. The mixture was then pelleted by centrifugation at 6,000 rpm for 30 minutes. Following this, 1.0 mL of the supernatant was collected and kept as "Non-Binding Protein" in -20 ℃ for further analysis. The pellet was then transferred into a 5 mL polypropylene column using column buffer and the column was extensively washed with 30 mL of column buffer. The first and last milliliters of the wash fractions were collected and kept as "Wash-First" and "Wash-Last" in -20 ℃ for further analysis. Finally, the protease was eluted from the column using elution buffer (Table 4.1:RMK12) into ten separate 1.5 mL microcentrifuge tubes. Each tube contained 1 mL of eluant and kept as "E1" to "E10" in -20 ℃ for further analysis and use.

4.1.1.7 Verification of Protease Using SDS-PAGE and Western Blot

All the proteins collected were analyzed using 12% SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) with CBB R-250 staining and Western Blot. To run 12% SDS-PAGE, firstly, the 12% acrylamide gel was cast with 10 wells in a gel casting chamber. Acrylamide (12%) in resolving gel solution (Table 4.1:RMK16) was pipetted into the gel casting chamber, until the solution level was just 0.5 cm below the wells. When the resolving gel had solidified, stacking gel solution (Table 4.1:RMK17) was poured on top of the resolving gel. A comb forming 10 wells was inserted into the chamber before the stacking gel solidified. Subsequently, the comb was removed and SDS electrophoresis buffer (Table 4.1:RMK18) was poured into the chamber to cover the gel. The loading samples were prepared by mixing 7 μL of protein samples and 7 μL of 2x SDS-PAGE loading buffer (Table 4.1:RMK19), and heated to 100 ℃ for 5 minutes. This denatures the proteins prior to loading into the wells along with a protein ladder. The electrophoresis experiment was run at 100 V constant

voltages for 2 hours. Next, the gel was fixed with a fixing solution (Table 4.1:RMK20) for 1 hour to prevent the protein from leaching out of the gel, followed by staining with CBB R-250 staining solution (Table 4.1:RMK21) for 30 minutes and then, destaining overnight with the destaining solution (Table 4.1:RMK20), shaking slowly on a shaker for each step. CBB R-250 binds to the protein and separated protein bands having different molecular weights will be detected as blue bands in the gel on a clear background. The gel was then stored in storage solution (Table 4.1:RMK22) for further use or reference.

For the western blot analysis, the gel preparation was the same as for SDS-PAGE analysis without any staining. A nitrocellulose membrane was first attached beside the gel with filter paper covered at both outer sides of the membrane and the gel. Another fiber pad was then attached to both outer sides of the filter paper. It was ensured that there was no bubble between the nitrocellulose membrane and the gel to avoid interference of result. Two cassette holders were then attached to the outer part of both the fiber pad and the whole attachment was fixed in the western blot apparatus with the nitrocellulose membrane facing the anode and the gel facing the cathode. An ice block was added, followed by the transfer buffer (Table 4.1:RMK23). The apparatus was run at 100 V constant voltages for 1 hour. Then, the nitrocellulose membrane was incubated overnight at 4 ℃ with blocking solution (Table 4.1:RMK24). Next, the membrane was transferred into the blocking solution with monoclonal anti-human adiponectin antibody (final concentration of 1 μg/mL) and incubated at room temperature for 1 hour on a shaker with low speed. Subsequently, the membrane was rinsed with TBS solution (Table 4.1:RMK25) and washed 3 times with TBST solution (Table 4.1:RMK26), 15 minutes each time, and then rinsed again with TBS solution. After soaking the membrane again with the blocking solution, a second antibody, mouse

IgG antibody (final concentration of 1 μg/mL) was added. The membrane was incubated again for 1 hour and washed again with TBS and TBST solutions, as mentioned in previously. Lastly, Western Blue Stabilized Substrate (Promega) was added to the membrane to catalyze the alkaline phosphatase for conjugation with the second antibody. The protein of interest was then detected as a purple band on a clear background in the membrane.

4.1.1.8 Protease Quantification using Bradford Protein Assay

Bradford protein assay was used in this study to quantitate the concentration of the protease. The assay was performed using a microplate reader. Reaction mixtures with a total volume of 200 μL per mixture were prepared in UV-transparent 96-well plates. At first, standard reaction mixtures containing 10 μL of 5 different concentrations of Bovine Serum Albumin (BSA) (0.2, 0.4, 0.6, 0.8 and 1.0 mg/mL) and without BSA were prepared in Quick Start Bradford Dye Reagent (Bio-Rad) with a total of 200 μL per mixture. The mixtures were then measured for UV absorption at a wavelength of 595 nm using a microplate reader. A BSA standard curve was then plotted and was later used for determination of the concentration of purified protease. The concentration of purified protease was determined by using a reaction mixture containing 10 μL of purified protease added in Quick Start Bradford Dye Reagent with a total volume of 200 μL per mixture.

4.1.1.9 Protease Activity and Inhibition Assays using Fluorogenic Peptide

The assays were performed using a Tecan Infinite M200 Pro fluorescence spectrophotometer. Reaction mixtures with a total volume of 200 μL per mixture were prepared in black 96-well plates. Firstly, standard solutions were prepared where reaction mixtures with 5 different concentrations of AMC (5, 10, 15, 20 and 25 μM) and

without AMC were mixed with 10 μL of DMSO each, buffered at pH 8.5 using 200 mM Tris-HCl (Table 4.1:RMK27). The standard solutions were used to plot AMC standard curve, which was used to calculate the amount of AMC released in subsequent assays. For the protease activity assay, the reaction mixtures were optimized at 2 μM of dengue protease CF40.gly(T).NS3pro and 100 μM of fluorogenic peptide substrate (Boc-Gly-Arg-Arg-AMC) (Peptide Institute, Inc), buffered at pH 8.5 using 200 mM Tris-HCl. The mixtures were then incubated at 37℃ for 30 minutes before measurements were taken. Substrate cleavage was optimized at 440 nm for emission and 350 nm for excitation. For the protease inhibition assay, reaction mixtures containing 100 μM of fluorogenic peptide substrate Boc-Gly-Arg-Arg-AMC, 2 μM dengue protease CF40.gly(T).NS3pro with or without potential inhibitors of varying concentrations, buffered at pH 8.5 using 200 mM Tris-HCl were prepared. The potential inhibitors were initially prepared in 10 μL of DMSO, and assayed at 5 different concentrations, which were 25, 50, 100, 200 and 400 mg/L. Kinetic studies were also performed by using 5 different concentrations (20, 40, 60, 80 and 100 μM) of the peptide substrate with the same reaction mixtures to determine the type of inhibition (competitive or non-competitive). For both the protease activity assay and protease inhibition assay, reaction mixtures without the addition of dengue protease CF40.gly(T).NS3pro were used as controls to correct the error obtained in measurement of the fluorescence intensity due to release of AMC by self-hydrolysis of the peptide substrate. For elimination of the DMSO solvent effect in the inhibition assay, those mixtures without potential inhibitors were tested with the addition of DMSO in the same amount as when the potential inhibitors were used. The reaction mixtures were incubated at 37℃ for 30 minutes before addition of the final fluorogenic peptide substrate. The mixture was then further incubated at 37℃ for 30 minutes before measurements. All measurements were carried out in triplicate.

The fluorescence intensity of AMC (released by the fluorogenic peptide substrate in the assay) was derived based on the fluorescence intensity detected after 30 minutes, using the following formula:

$$I_{sample} - I_{control} = I_{protease} - I_0$$

where I_{sample} is fluorescence intensity of the sample which is equivalent to $I_{protease}$, the fluorescence intensity of reaction mixture with the protease (CF40.gly(T).NS3pro), while $I_{control}$ is fluorescence intensity of the control which is equivalent to I_0, the fluorescence intensity of reaction mixture without protease.

Next, the velocity of enzyme-catalyzed reaction, which is equivalent to the rate of AMC released, was obtained from the AMC standard curve and the fluorescence intensity of AMC released after 30 minutes. The reaction velocities were then used to calculate the K_m and V_{max} for the protease activity assay, using nonlinear regression Michaelis-Menten equation in GraphPad Prism 5.0 software. The reaction velocities in the presence of potential inhibitors were also used to calculate the K_i values of the potential inhibitors and to verify the type of inhibitors, by using nonlinear regression mixed model inhibition (Copeland, 2002b) equation in GraphPad Prism 5.0 software. On the other hand, Lineweaver-Burk plots were used for better display of the type of inhibition. Unpaired t-tests were also performed by comparing the $K_{i\,exp}$ values of all the tested compounds with the standard, using GraphPad Prism 5.0 software to evaluate the statistical significance of the inhibition activities of the compounds.

4.1.2 Structure-Activity Relationship (SAR) Study

Structure-activity relationship (SAR) study was performed by comparing and relating the structure of the compounds with the inhibition activities in the *in vitro* study, based on the binding interactions observed from the docking experiments.

4.2 Results and Discussions

4.2.1 Cloning, Expression, Purification and Verification of Soluble DEN-2 NS2B-NS3 Protease

The RT-PCR result of DEN-2 virus RNA produced using designed primers showed that the correct insoluble DEN-2 NS2B-NS3 cDNA with 828 bp was being amplified (Figure 4.3A). The PCR result for the hydrophilic NS2B cofactor, CF40 gene with a glycine linker at the 3' end and NS3 serine protease, NS3pro gene with a glycine linker, gly-(T), at the 5' end also showed the desired length of 177 bp and 600 bp respectively (Figure 4.3B). Set 2 in Figure 4.28B was selected for digestion with KpnI restriction enzyme, followed by ligation with DNA ligase as fewer impurities from PCR products were detected.

The ligation of the NS2B cofactor and NS3 serine protease genes, with glycine linker in between them was successfully carried out. Two by-products were generated due to the ligation process, which were two NS2B cofactor genes ligated together with the glycine linker and two NS3 serine protease genes ligated together with the glycine linker. Two precursor genes were also present even after the ligation process was completed (Figure 4.4A). The 1% agarose gel electrophoresis for PCR amplified NS2B cofactor gene, ligated with NS3 serine protease gene via a glycine linker after sequential double digestions by BamHI and HindIII restriction enzymes, showed the desired gene length of about 750 bp (Figure 4.4B). The agarose gel electrophoresis profiles of the

DNA from the undigested, single digested (with only BamHI restriction enzyme) and double digested (with both BamHI and HindIII restriction enzymes in sequence) pQE30 vectors also showed that they had been digested correctly (Figure 4.4B).
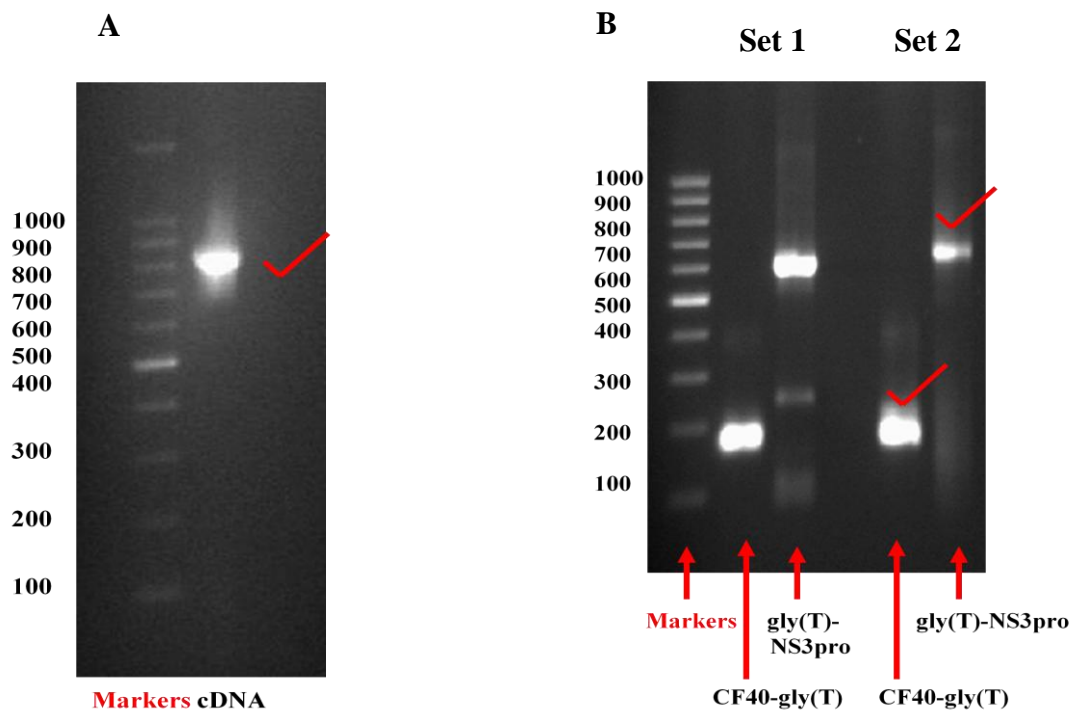
Figure 4.3 RT-PCR and PCR result. **A:** RT-PCR of insoluble DEN-2 NS2B-NS3 protease gene with 828 bp. **B:** PCR of 2 sets of CF40 gene with gly-(T) at the 3' end containing 177 bp and NS3pro gene with gly-(T) at the 5' end containing 600 bp. Tick signs show the bands wth the desired length of genes which were selected for further cloning work.
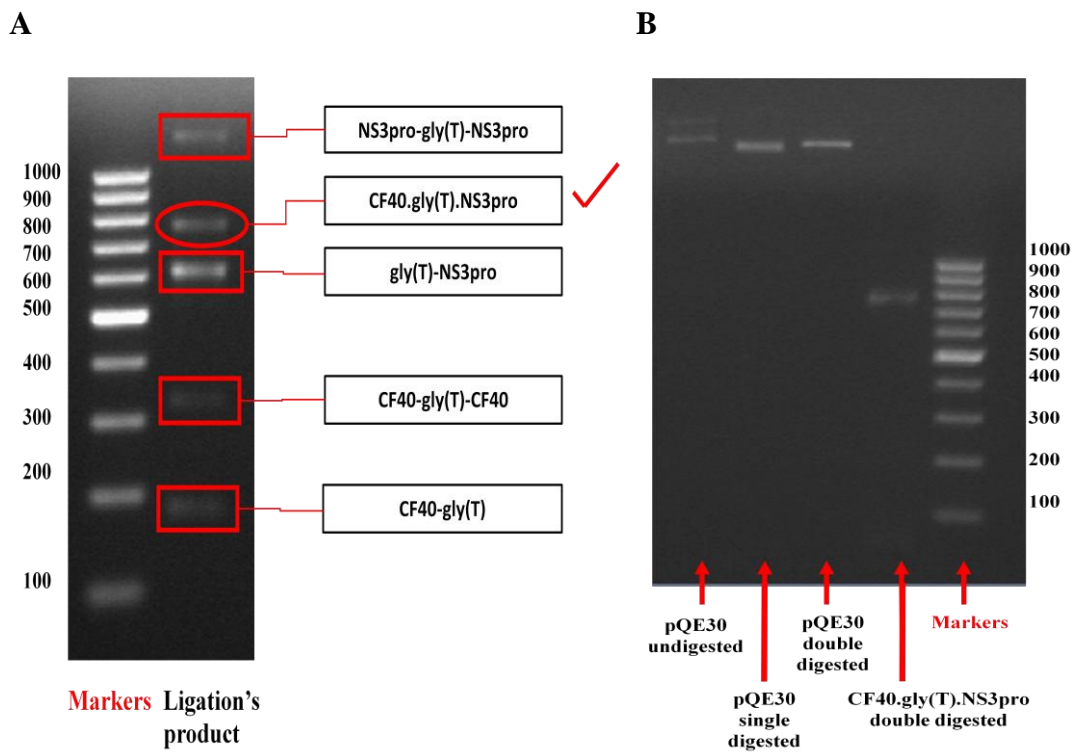
Figure 4.4 Ligation and digestion result. **A:** Ligation of CF40 gene with gly-(T) at 3' end and NS3pro gene with gly-(T) at 5' end (band shown with a tick). The agarose gel profile shows that there were two other by-products, CF40-gly(T)-CF40 and NS3pro-gly(T)-NS3pro genes. The two precursor genes of CF-gly(T) and gly(T)-NS3pro were also detected in the gel even after the reaction had ended. CF40.gly(T).NS3pro gene (ticked) was then extracted for further amplification with PCR. **B:** Differences in the DNA profiles in 1% agarose gel between the undigested, single digested, double digested and amplified CF40.gly(T).NS3pro gene after PCR and double digestion.

For the transformation of the pQE30 vector ligated with insert (soluble DEN-2 NS2B-NS3 protease, CF40.gly(T).NS3pro gene) into XL1-blue bacteria, only 1 out of a total of 24 selected bacteria colonies after 4 trials were shown to produce the desired gene length of about 750 bp from the sequential double digestion with BamHI and HindIII of their extracted plasmid (Figure 4.5).

Subsequently, the successfully transformed bacteria culture was used for protein expression and purification with the protein profile as shown in Figure 4.6. The presence of more than one band indicated that more than one protein were eluted out even after purification on Ni^{2+}-NTA-agarose (Lanes E1 and E2) was performed. In fact, when compared to the purification of insoluble protein under denaturing conditions, it is more likely for Ni^{2+}-NTA-agarose purification of soluble protein under native conditions to exhibit copurify associated proteins. These copurify associated proteins such as enzyme subunits and binding proteins presented in expressing cells, were added to the lysate before purification, or added to the Ni^{2+}-NTA matrix after the 6xHig-tagged protein was bound (Le Grice & GrÜNinger-Leitch, 1990; Flachmann & Kuhlbrandt, 1996; Qiagen, 2003). However, through western blot, the identity of our protein was confirmed to be the His-tagged soluble DEN-2 NS2B-NS3 protease, since the His-tagged protein will bind to the monoclonal anti-human adiponectin antibody used in western blot (Figure 4.7B). A previous study (Leung *et al.*, 2001) reported that the size of the soluble DEN-2 NS2B-NS3 protease was around 32 kDa and only a single protein band was observed after purification on Ni^{2+}-NTA-agarose column. In this study, we report the expression of soluble DEN-2 NS2B-NS3 protease (CF40.gly(T).NS3pro) with a molecular mass of about 38 kDa (Figure 4.6 and 4.7). The additional amino acids in this protease might be due to the additional nucleotides which were included in the design of the primers.
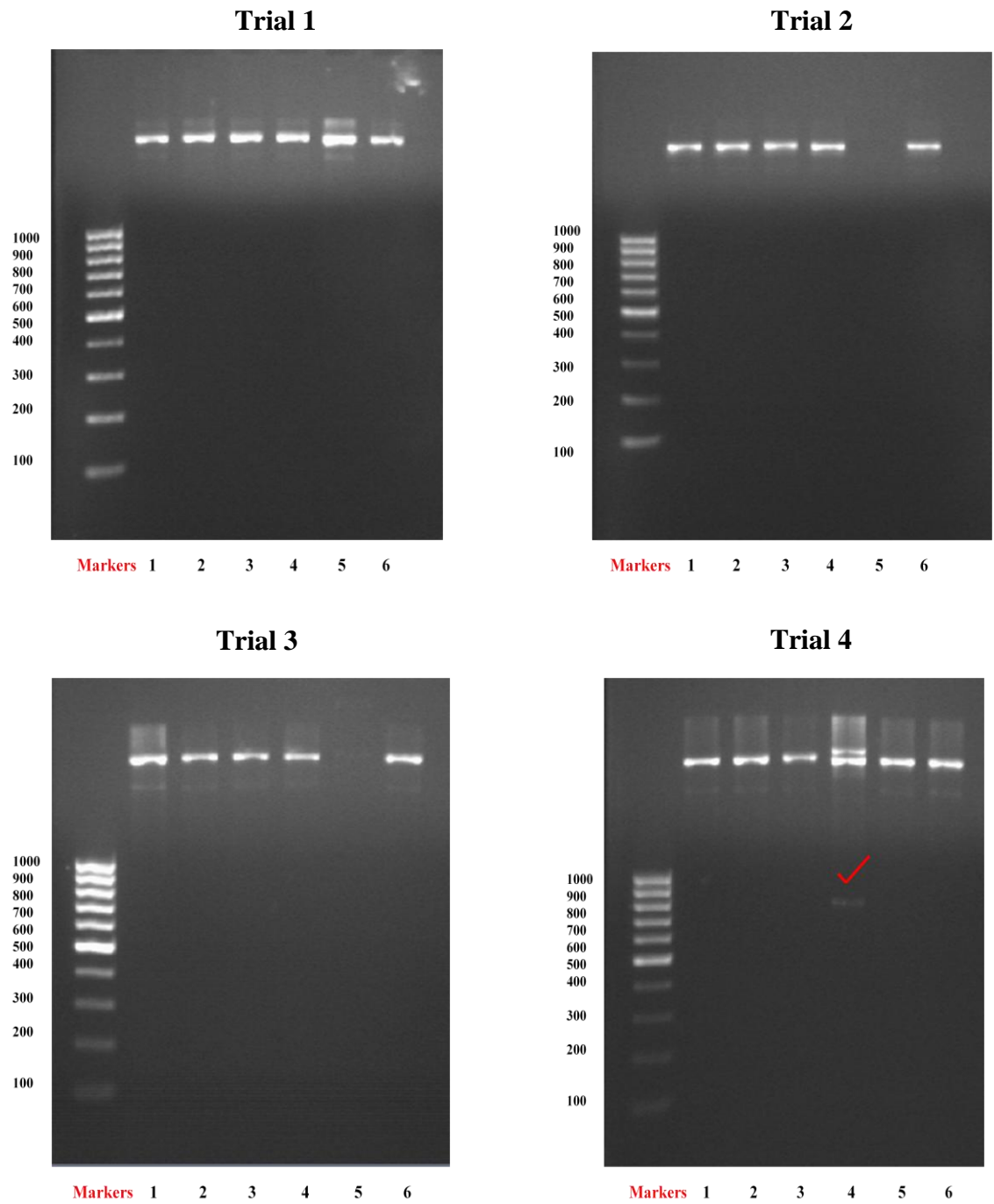
Figure 4.5 1% agarose gel electrophoresis profiles for 4 different trials (Trial 1 - 4) of LB agar cultures after transformations. 6 colonies picked from each set for LB medium cultures, followed by plasmids extraction and double digestion with BamHI and HindIII. Colony number 4 from Trial 4 (ticked) was the only culture with desired gene length of about 750 bp.
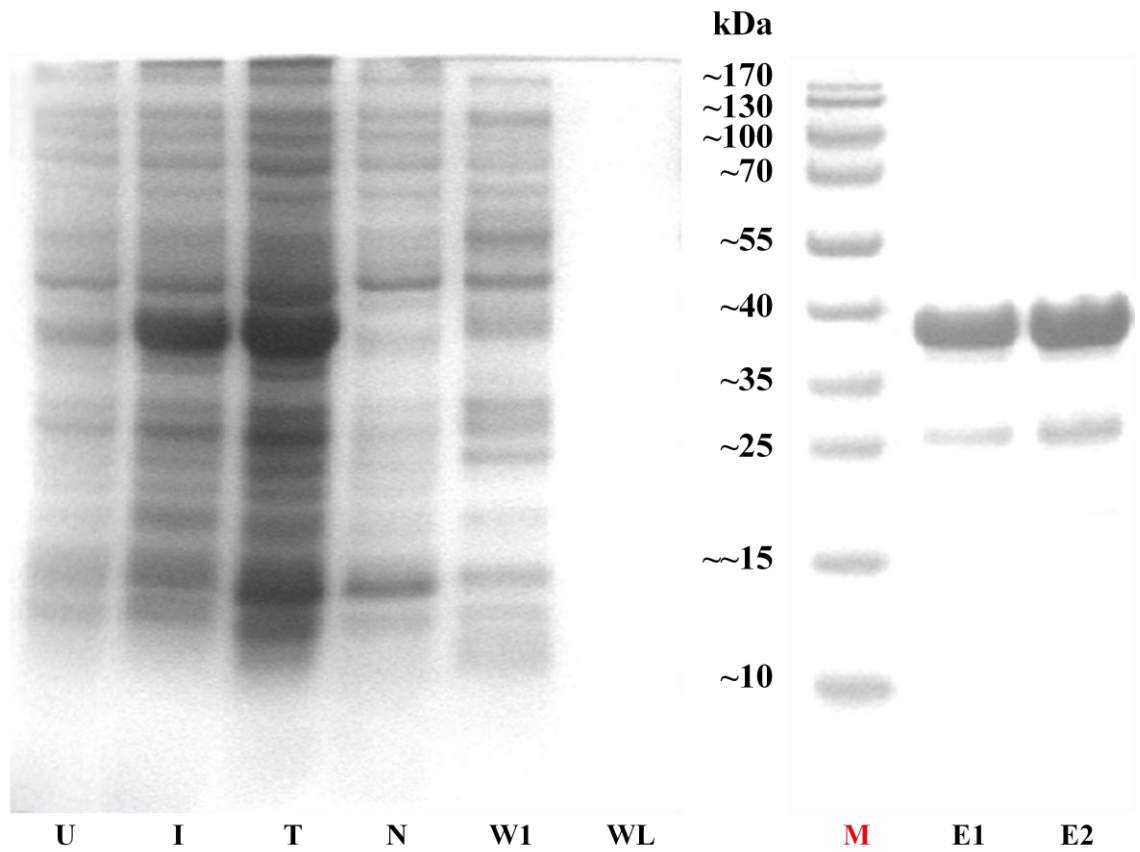
Figure 4.6   SDS-PAGE profile of "Uninduced Protein" (U), "Induced Protein" (I), "Total Protein" (T), "Non-Binding Protein" towards Ni^{2+}-NTA-agarose column (N), "Wash-First" (W1), "Wash-Last" (WL), protein molecular mass markers (M), first 1 ml of eluted elution buffer (E1) and followed by second ml (E2).
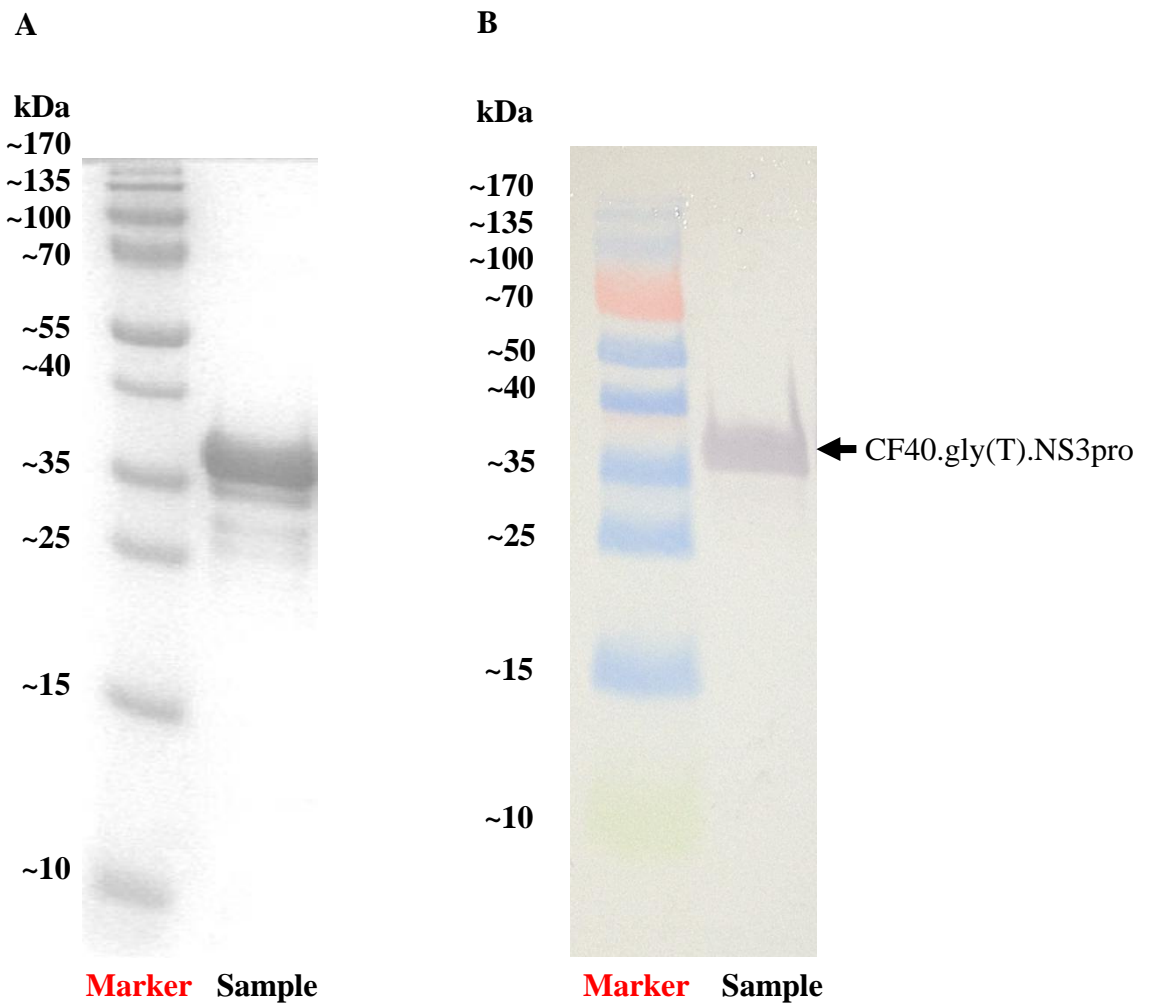
Figure 4.7   SDS-PAGE and Western blot profile. **A:** SDS-PAGE gel profile of Ni^{2+}-NTA-agarose purified CF40.gly(T).NS3pro. **B:** Western blot membrane profile of Ni^{2+}-NTA-agarose purified CF40.gly(T).NS3pro. Purple band indicates the binding of His-tagged CF40.gly(T).NS3pro, and it was observed at ~38 kDa.

Further purification using a Ni^{2+}-NTA-agarose column was also carried out with large amount of wash buffer (500 mL) in order to obtain pure CF40.gly(T).NS3pro (Figure 4.8).

For further verification, the translated amino acids sequence based on the nucleotide sequencing result of our rDNA had validated that there are more than 231/241 (95.9%) identity and 240/241 (99.6%) similarity (amino acids not identical but with similar physical properties) between our cloned protease and DEN-2 NS2B-NS3 protease from other strains of DEN-2 reported in NCBI (NCBI) through protein Blast (Altschul *et al.*, 1997) (Figure 4.9). Moreover, as shown in Figure 4.9, all the catalytic triad residues − His51, Asp75 and Ser135, and the suggested key residue found in the allosteric site − Lys74 (located in the NS3 serine protease domain) were aligned perfectly between our cloned protease and the DEN-2 NS2B-NS3 protease from the other strains. This shows that our protease is capable of representing the DEN-2 NS2B-NS3 protease generically for the development of non-specific protease inhibitor towards all the strains of DEN-2.
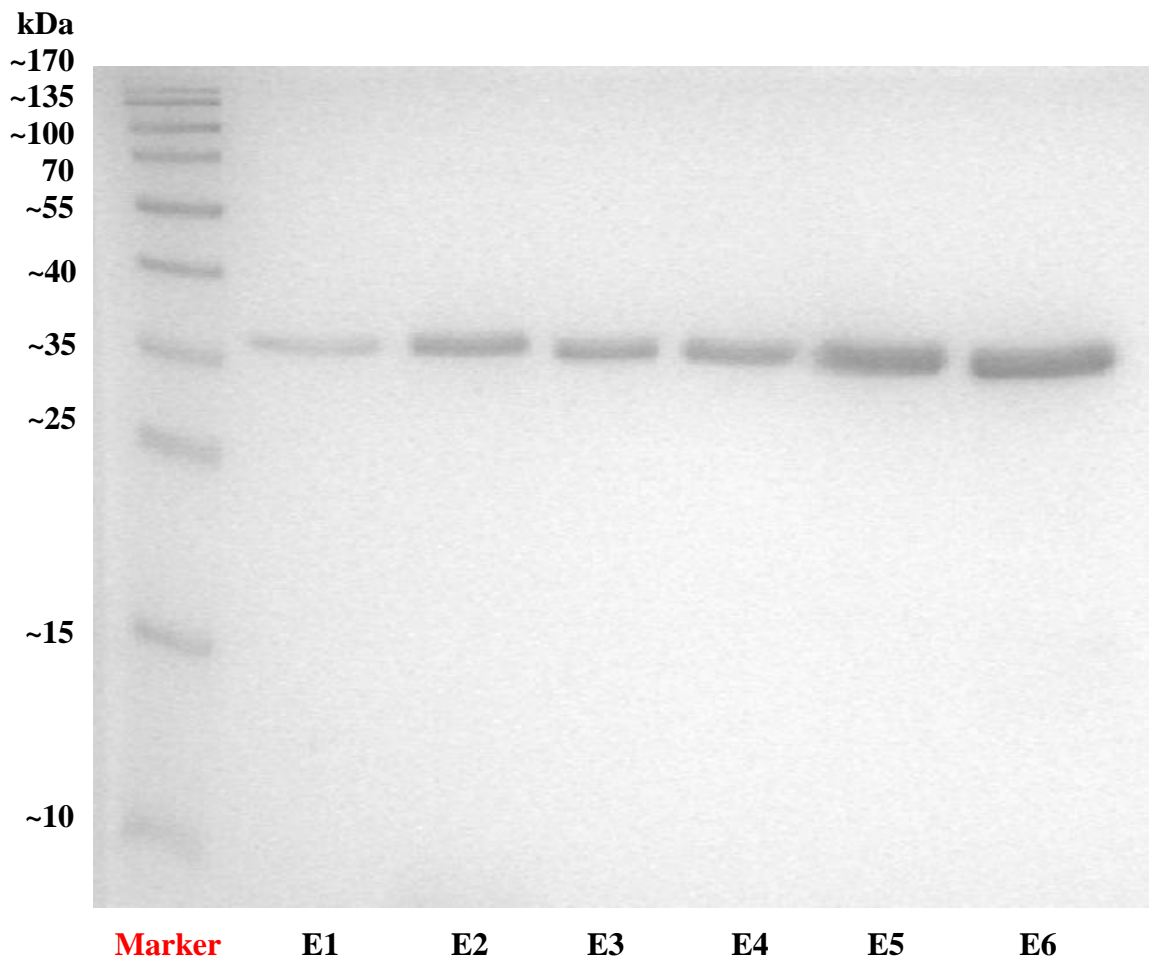
Figure 4.8   First elution (E1) to sixth elution (E6) of subsequent Ni^{2+}-NTA-agarose

purification using large amount of wash buffer (500 mL) before elution.
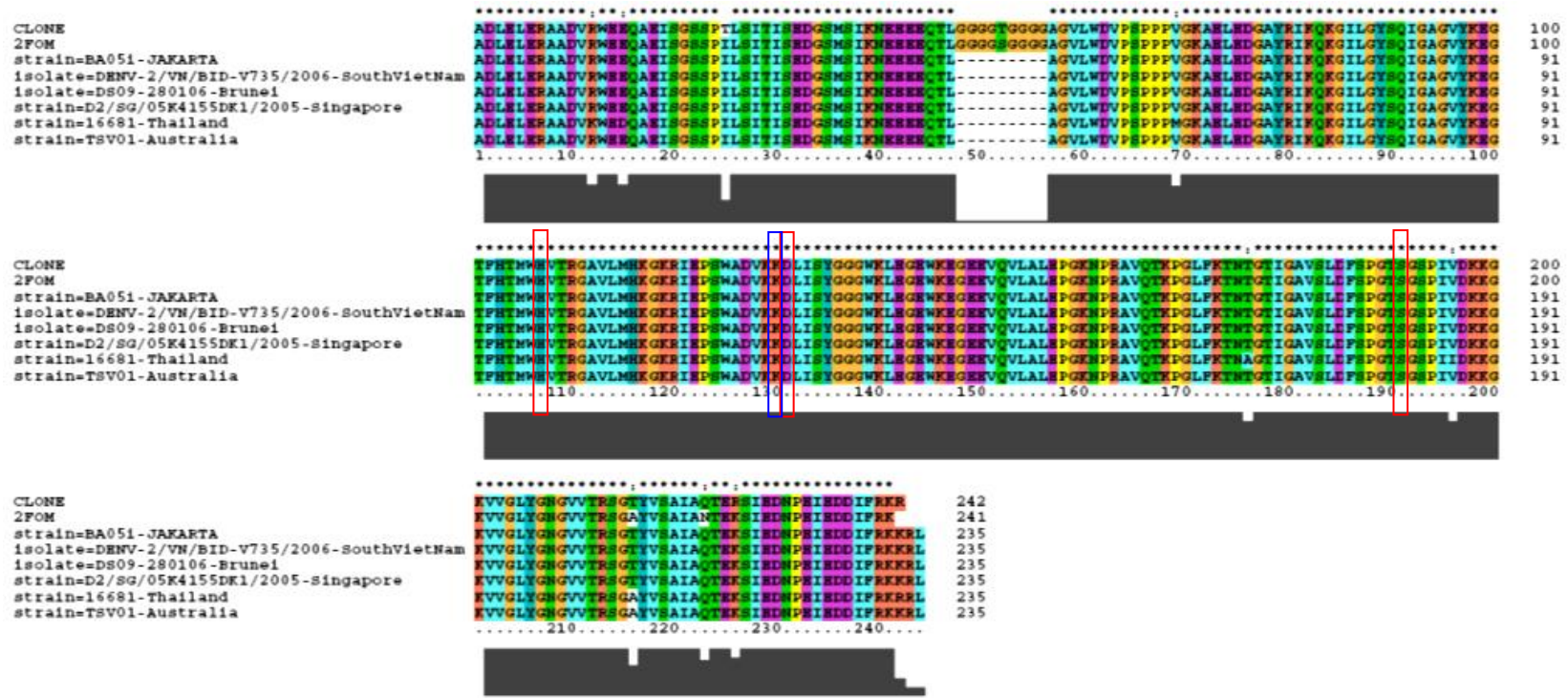
Figure 4.9 Sequence alignment of the protease cloned in this study and DEN-2 NS2B-NS3 protease from various strains of reported DEN-2. The grey histogram represents the level of convergence between two amino acids. Amino acids that were identical in both sequences will have a full grey bar and an asterisk symbol (*) on top of the specific amino acids, followed by colon symbol (:) for high similarity, period symbol (.) for low similarity and blank ( ) for non-convergence. Catalytic triad residues are highlighted in red-outlined boxes and NS3 Lys74 residues are highlighted in a blue-outlined box.

4.2.2 Protease Quantification, Protease Activity and Inhibition Assay

BSA standard curve was obtained from Bradford Protein Assay (Figure 4.10). The formula for the standard curve is $y = 0.445x$, which means "Absorbance" = 0.445 x "Concentration". Thus, "Concentration" = "Absorbance"/0.445. An example of the calculation for the concentration of our protease is shown as below:

Absorbance at 595 nm for E1 = 1.020 OD

Absorbance at 595 nm for blank (Quick Start Bradford Dye Reagent only) = 0.341 OD

Concentration of E1 = (1.020 − 0.341)/0.445

$\quad\quad\quad\quad\quad\quad\quad$ = 1.526 mg/mL

Our protease is about 38 kDa, thus about 38,000 g/mol.

Concentration of E1 $\quad \approx \quad \dfrac{1.526 \text{ mg/mL}}{38,000 \text{ g/mol}}$

$\quad\quad\quad\quad\quad\quad\quad \approx \quad \dfrac{1.526 \text{ mg/mL}}{38,000 \text{ mg/mmol}}$

$\quad\quad\quad\quad\quad\quad\quad \approx \quad 0.00004016 \text{ mol/L}$

$\quad\quad\quad\quad\quad\quad\quad \approx \quad 40.16 \text{ μM}$

Based on the dilution equation, $C_1V_1 \quad = \quad C_2V_2$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad C_1 \quad = \quad \dfrac{C_2V_2}{V_1}$

Thus, the concentration of 10 μL of E1 in 200 μL Tris-HCl Buffer (Table 4.1:RMK27)

$\approx \quad \dfrac{40.16 \text{ μM x 10 μL}}{200 \text{ μL}}$

$\approx \quad 2.01 \text{ μM}$

Absorbance Versus Concentrations

$y = 0.445x$
$R^2 = 0.9978$
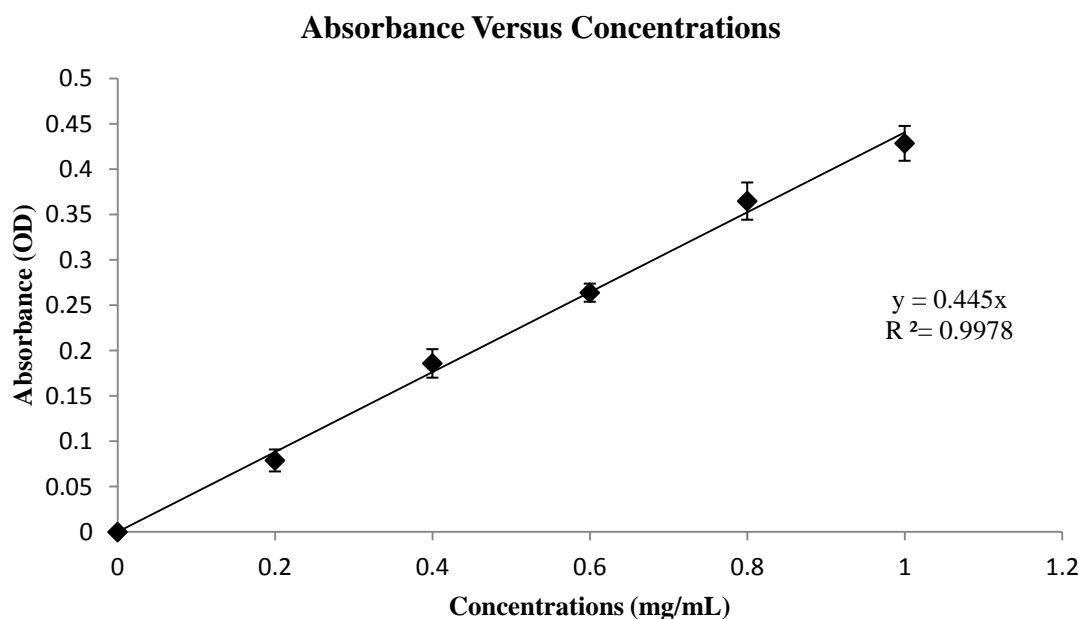
Figure 4.10 Standard curve for BSA concentration against absorbance.

Figure 4.11 shows the optimization of emission and excitation wavelengths for AMC's fluorescence intensity. These parameters were then applied in the protease activity and inhibition assays for the detection of fluorescence intensity of AMC released through substrate cleavage. Based on the methods described in section 4.1.1.9, AMC standard curve (Figure 4.12) was obtained. The formula for the AMC standard curve is y = 1420.7x, which means "Fluorescence Intensity of AMC" = 1420.7 x "Concentration of AMC". Thus, "Concentration of AMC" = "Fluorescence Intensity of AMC"/1420.7. The enzyme-catalyzed reaction velocity, which is equivalent to the rate of AMC released (μM/min), for protease activity and substrate optimization assays (Figures 4.13 and 4.14) as well as inhibition assays (Figure 4.15 - 4.20) was calculated following the example shown below:

Fluorescence intensity of 1.0 μM of dengue protease CF40.gly(T).NS3pro with 100 μM of fluorogenic peptide substrate (Boc-Gly-Arg-Arg-AMC) buffered at pH 8.5 by 200 mM Tris-HCl after 30 minutes

=   12857 RFU (Relative Fluorescence Units)

Fluorescence intensity of 100 μM fluorogenic peptide substrate (Boc-Gly-Arg-Arg-AMC) buffered at pH 8.5 by 200 mM Tris-HCl after 30 minutes

=   4055 RFU

Concentration of AMC released   = $\dfrac{(12857 - 4055)\ \mu M}{1420.7}$

                                    =   6.196 μM

Reaction velocity (Rate of AMC released)   = $\dfrac{6.196\ \mu M}{30\ min}$

                                    =   0.2065 μM/min

A
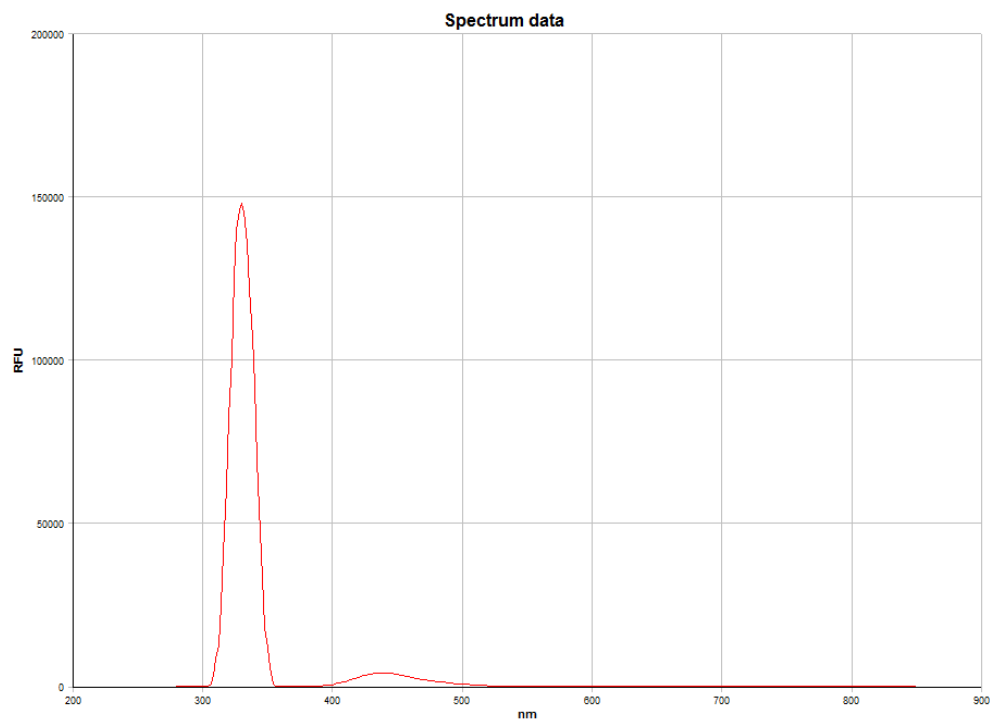
B
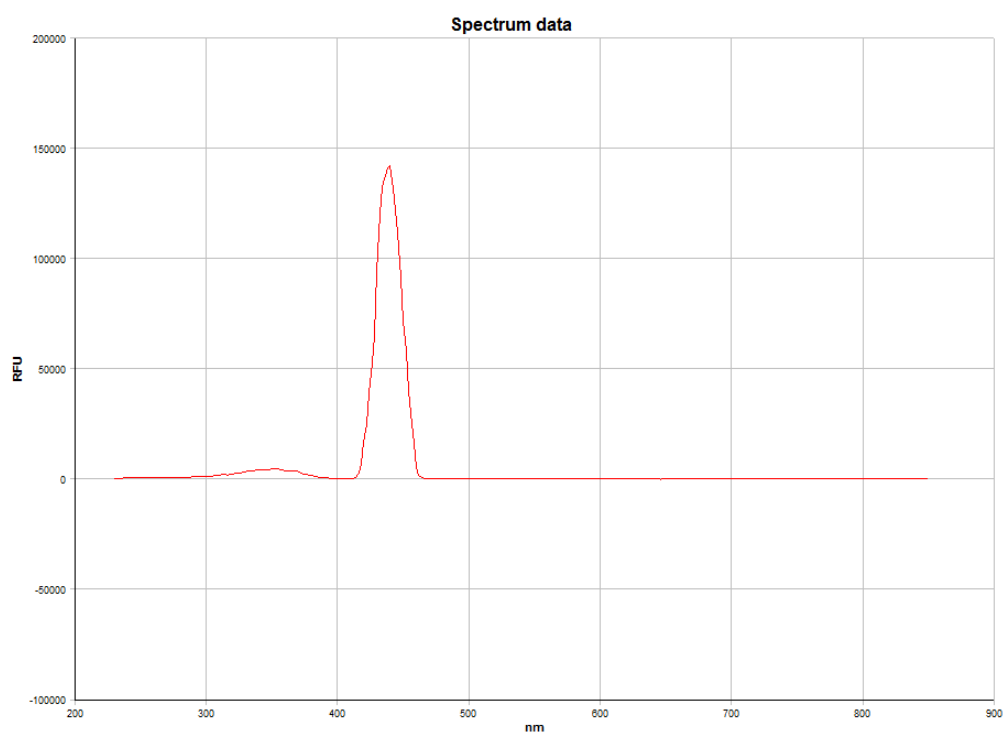
Figure 4.11 Optimization of AMC's fluorescence intensity. **A:** AMC's fluorescence intensity was optimum at 440 nm for emission scan. **B:** AMC's fluorescence intensity was optimum at 350 nm for excitation scan.
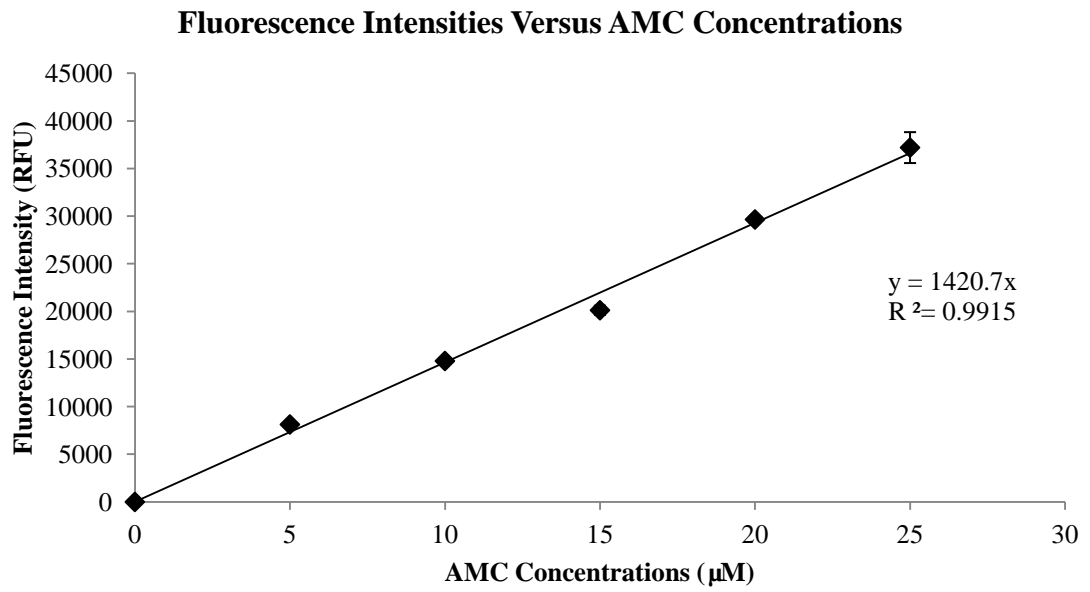
Figure 4.12   Standard curve for AMC concentration against absorbance.
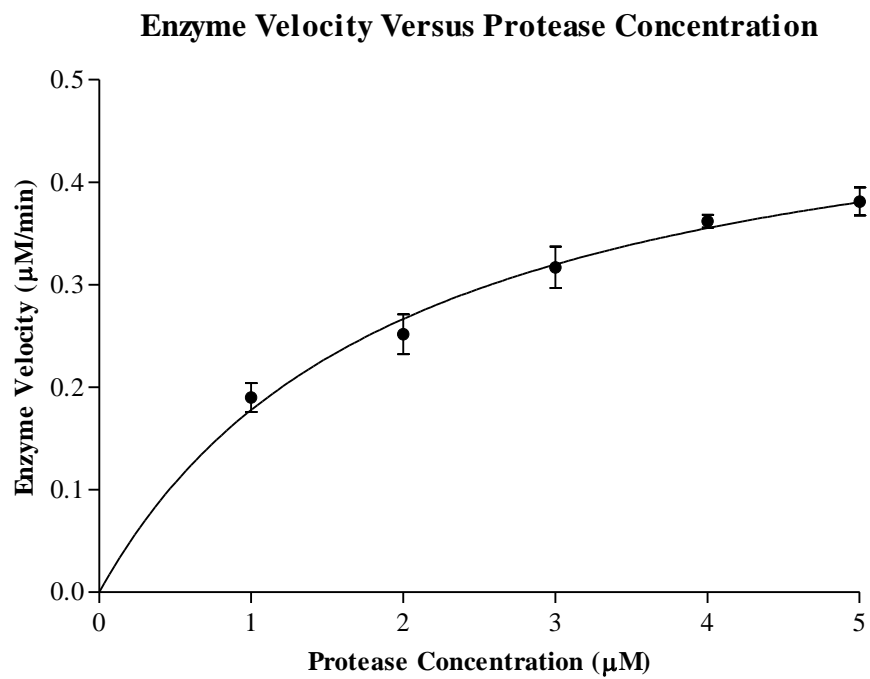
Figure 4.13   Protease activity optimization assay with 100 µM of fluorogenic peptide

substrate (Boc-Gly-Arg-Arg-AMC) buffered at pH 8.5 by 200 mM Tris-HCl.

Enzyme Velocity Versus Substrate Concentration
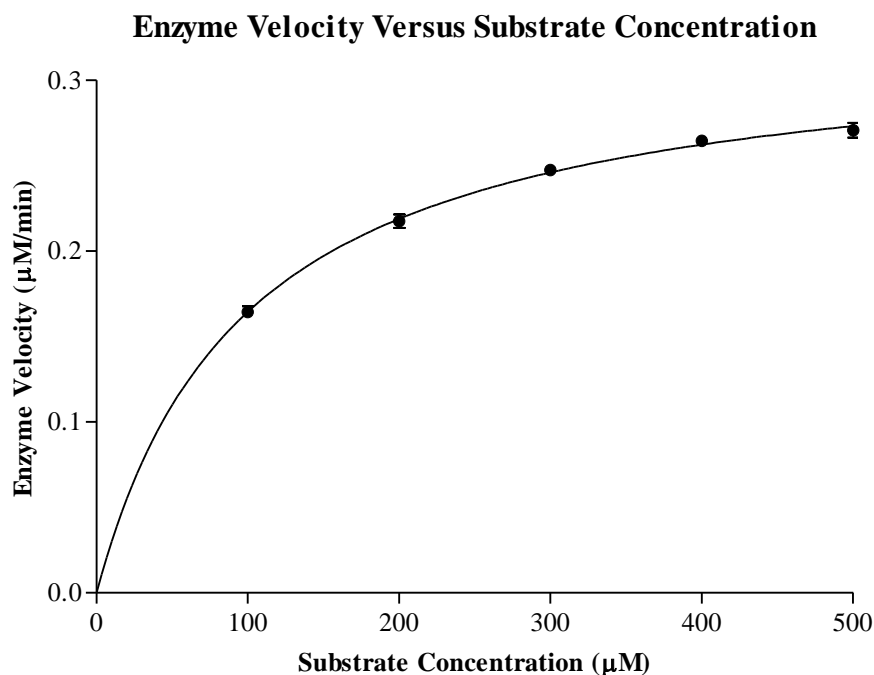
Figure 4.14 Fluorogenic peptide substrate optimization assay with 2 μM of dengue protease CF40.gly(T).NS3pro buffered at pH 8.5 by 200 mM Tris-HCl.

Figures 4.13 and 4.14 show the optimization of protease and substrate concentrations. By using the nonlinear regression Michaelis-Menten equation in GraphPad Prism 5.0 software, the maximum enzyme-catalyzed reaction velocity without inhibitor, V_{max}, of both protease and substrate were calculated as 0.53 ± 0.04 μM/min and 0.33 ± 0.01 μM/min, respectively, while the Michaelis-Menten constant, K_m, values for both were 2.0 ± 0.4 μM and 99 ± 4 μM, respectively (Table 4.4). As a result, based on the K_m values, 2 μM of dengue protease CF40.gly(T).NS3pro along with 20, 40, 60, 80 and 100 μM of fluorogenic peptide substrate (Boc-Gly-Arg-Arg-AMC) were chosen as optimum concentrations for subsequent inhibition assays.

The best-fit values of shared parameters for different concentrations of pinostrobin (standard), compounds 1, 2, 3 and 4, fitted using a nonlinear regression mixed model inhibition equation in GraphPad Prism 5.0 software, are shown in Tables 4.5 - 4.9 together with Lineweaver-Burk plots illustrating the type of protease inhibition (Figure 4.15 - 4.19). The value of alpha is a measure of the degree to which the binding of the inhibitor changes the affinity of the enzyme for its substrate. Its value should always be greater than zero. When alpha = 1, the inhibitor does not alter binding of substrate to the enzyme and the mixed-model is identical to noncompetitive inhibition. When alpha is very large, binding of the inhibitor prevents binding of the substrate and the mixed-model becomes identical to competitive inhibition. When alpha is very small (but greater than zero), binding of the inhibitor enhances substrate binding to the enzyme and the mixed model becomes nearly identical to an uncompetitive model (Copeland, 2002b; Motulsky & Christopoulos, 2004).

A review of the literature yielded no bioassay data for compounds 1, 2 and 3 while compound 4 had been reported to be active as an inhibitor in various bioassays in PubChem database (http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?sid= 112156252&viewopt=PubChem) (PubChem, 2005), including flavivirus genomic capping enzyme inhibition assay for DEN-2 (Geiss *et al.*, 2009). The $K_{i\,exp}$ values for all the tested compounds in this study are shown in Figure 4.20 and Table 4.10. Compounds 1, 2 and 4 showed better inhibition activities with significantly lower $K_{i\,exp}$ values compared to that of the standard pinostrobin, while compound 3 showed no significant difference (Figure 4.20). For most of the compounds, including pinostrobin, the $K_{i\,exp}$ obtained correlated well with the docking results, except for compound 3. The poor correlation in the case of compound 3 might be due to the lack of hydrogen bond interaction towards any of the protease's amino acid residues (as shown in Table 2).

This lack of a hydrogen bond interaction is mainly caused by the absence of a hydrophilic side chain in compound 3 as compared to the other compounds (Figure 1). The physiochemical properties of compound 3 may also lead to an overestimation by AutoDock 4.2 software in generating the estimated $K_{i\ dock}$ value (Table 3), where the algorithm in AutoDock takes into account both electronic and solvation properties of compounds (Morris *et al.*, 1998).

From the kinetic analysis using the GraphPad Prism 5.0 software and Lineweaver-Burk plots, all of the test compounds were also shown to be non-competitive inhibitors. This further supports our proposal that the structure of the allosteric binding site for non-competitive inhibition of dengue virus should resemble that of DH-1. Consequently, DH-1 is suggested to be suitable to be used for allosteric binding studies and virtual screening for non-competitive inhibitors. In this study, compound 1 proved to be the most potent inhibitor among all the tested compounds, with a computational $K_{i\ dock}$ value of 45 μM and an *in vitro* $K_{i\ exp}$ of 69 ± 9 μM. This compound could then be used as a lead for ligand-based drug design of anti-dengue agents.

Table 4.4 :   Best-fit values for V_{max} and K_m for protease activity and substrate optimization assays using nonlinear regression Michaelis-Menten equation in GraphPad Prism 5.0 software.

Optimization	Protease	Substrate
Best-fit values		
V_{max} (μM/min)	0.53	0.33
K_m (μM)	2.0	99
Std. Error		
V_{max} (μM/min)	0.04	0.01
K_m (μM)	0.4	4

Table 4.5 : Best-fit values of shared parameters, V_{max}, Alpha, $K_{i\ exp}$ and K_m for standard pinostrobin with different concentrations, I, fitted using nonlinear regression mixed model inhibition method in GraphPad Prism 5.0 software.

Concentrations (μM)	1480	740	370	185	92.5	0
Mixed model inhibition						
Best-fit values						
$Vmax$ (μM/min)	0.32	0.32	0.32	0.32	0.32	0.32
I (μM)	= 1480	= 740	= 370	= 185	= 92.5	= 0
Alpha	1.1	1.1	1.1	1.1	1.1	1.1
$Ki\ exp$ (μM)	415	415	415	415	415	415
Km (μM)	100	100	100	100	100	100
Std. Error						
$Vmax$ (μM/min)	0.03	0.03	0.03	0.03	0.03	0.03
Alpha	0.6	0.6	0.6	0.6	0.6	0.6
$Ki\ exp$ (μM)	85	85	85	85	85	85
Km (μM)	14	14	14	14	14	14

$Vmax$ = maximum enzyme velocity without inhibitor.

Km = Michaelis-Menten constant.

Alpha = 1, mixed-model = noncompetitive inhibition.

Alpha very large, mixed-model = competitive inhibition.

Alpha very small but > 0, mixed-model = uncompetitive inhibition.

$Ki\ exp$ = inhibition constant calculated from in vitro experiment.

Lineweaver-Burk Plot for *R*-Pinostrobin
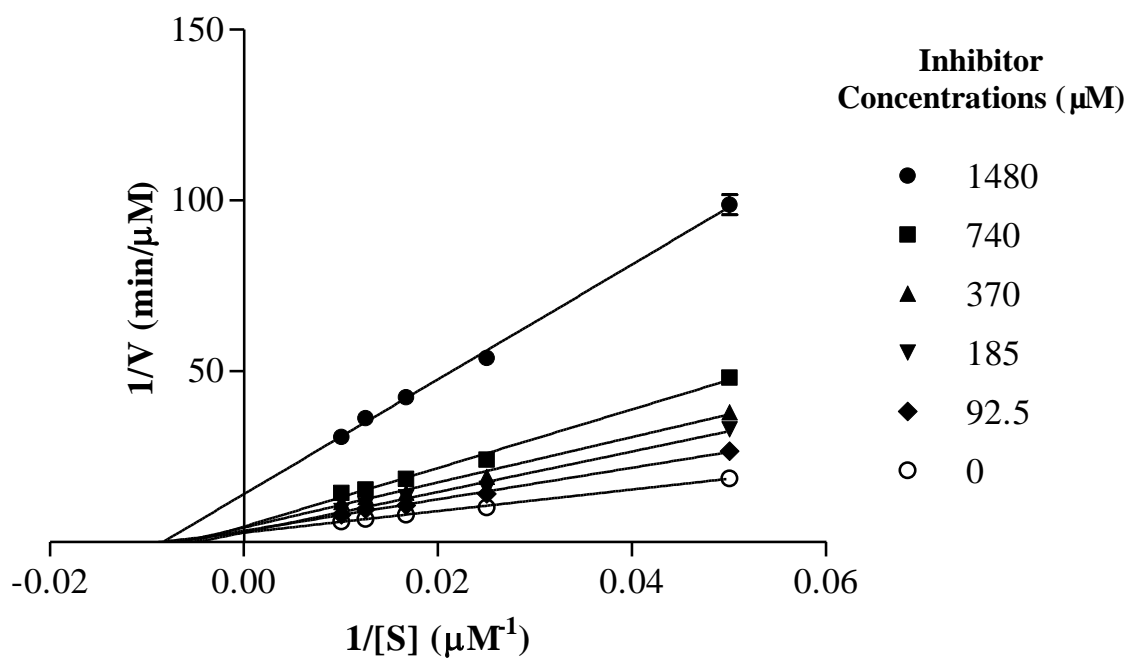
Figure 4.15 Protease inhibition assay with pinostrobin as inhibitor. All lines intercept at the same point at X-axis with negative value indicates that the reaction kinetics involves non-competitive inhibition.

Table 4.6 : Best-fit values of shared parameters, V_{max}, Alpha, $K_{i\ exp}$ and K_m for compound 1 with different concentrations, I, fitted using nonlinear regression mixed model inhibition method in GraphPad Prism 5.0 software.

Concentrations (μM)	1380	689	345	172	86.1	0
Mixed model inhibition						
Best-fit values						
$Vmax$ (μM/min)	0.31	0.31	0.31	0.31	0.31	0.31
I (μM)	= 1380	= 689	= 345	= 172	= 86.1	= 0
Alpha	1.2	1.2	1.2	1.2	1.2	1.2
$Ki\ exp$ (μM)	69	69	69	69	69	69
Km (μM)	86	86	86	86	86	86
Std. Error						
$Vmax$ (μM/min)	0.02	0.02	0.02	0.02	0.02	0.02
Alpha	0.3	0.3	0.3	0.3	0.3	0.3
$Ki\ exp$ (μM)	9	9	9	9	9	9
Km (μM)	8	8	8	8	8	8

$Vmax$ = maximum enzyme velocity without inhibitor.

Km = Michaelis-Menten constant.

Alpha = 1, mixed-model = noncompetitive inhibition.

Alpha very large, mixed-model = competitive inhibition.

Alpha very small but > 0, mixed-model = uncompetitive inhibition.

$Ki\ exp$ = inhibition constant calculated from in vitro experiment.

Lineweaver-Burk Plot for Compound 1
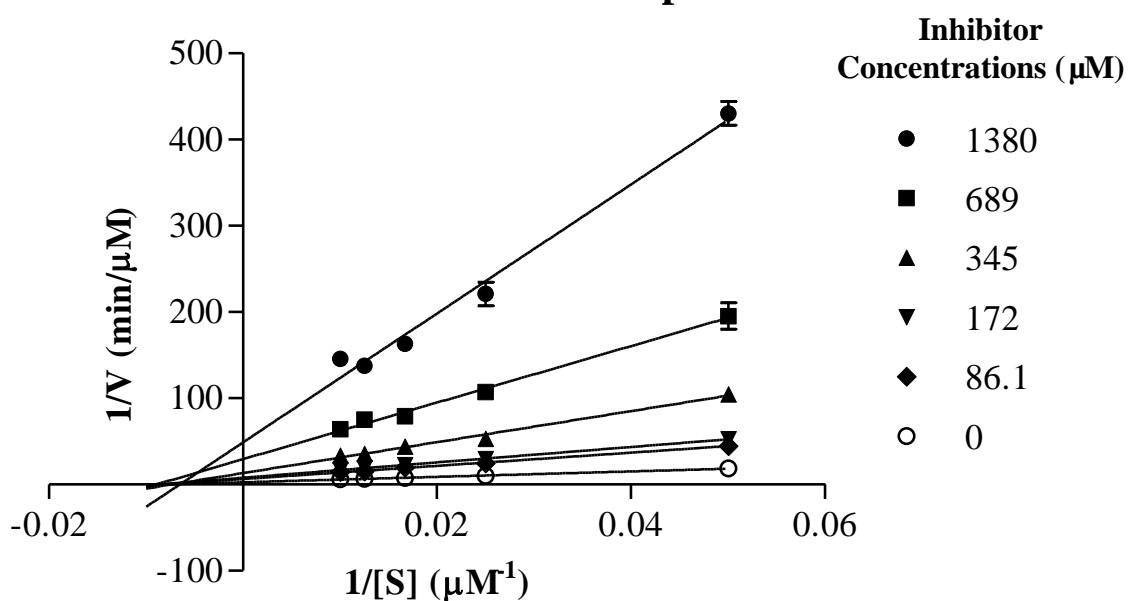
Figure 4.16   Protease inhibition assay with compound 1 as inhibitor. All lines intercept at the same point at X-axis with negative value indicates that the reaction kinetics involves non-competitive inhibition.

Table 4.7 : Best-fit values of shared parameters, V_{max}, Alpha, $K_{i\ exp}$ and K_m for compound 2 with different concentrations, I, fitted using nonlinear regression mixed model inhibition method in GraphPad Prism 5.0 software.

Concentrations (µM)	1350	677	339	169	84.6	0
Mixed model inhibition						
Best-fit values						
$Vmax$ (µM/min)	0.32	0.32	0.32	0.32	0.32	0.32
I (µM)	= 1350	= 677	= 339	= 169	= 84.6	= 0
Alpha	1.7	1.7	1.7	1.7	1.7	1.7
$Ki\ exp$ (µM)	121	121	121	121	121	121
Km (µM)	94	94	94	94	94	94
Std. Error						
$Vmax$ (µM/min)	0.02	0.02	0.02	0.02	0.02	0.02
Alpha	0.6	0.6	0.6	0.6	0.6	0.6
$Ki\ exp$ (µM)	14	14	14	14	14	14
Km (µM)	9	9	9	9	9	9

$Vmax$ = maximum enzyme velocity without inhibitor.

Km = Michaelis-Menten constant.

Alpha = 1, mixed-model = noncompetitive inhibition.

Alpha very large, mixed-model = competitive inhibition.

Alpha very small but > 0, mixed-model = uncompetitive inhibition.

$Ki\ exp$ = inhibition constant calculated from in vitro experiment.

Lineweaver-Burk Plot for Compound 2
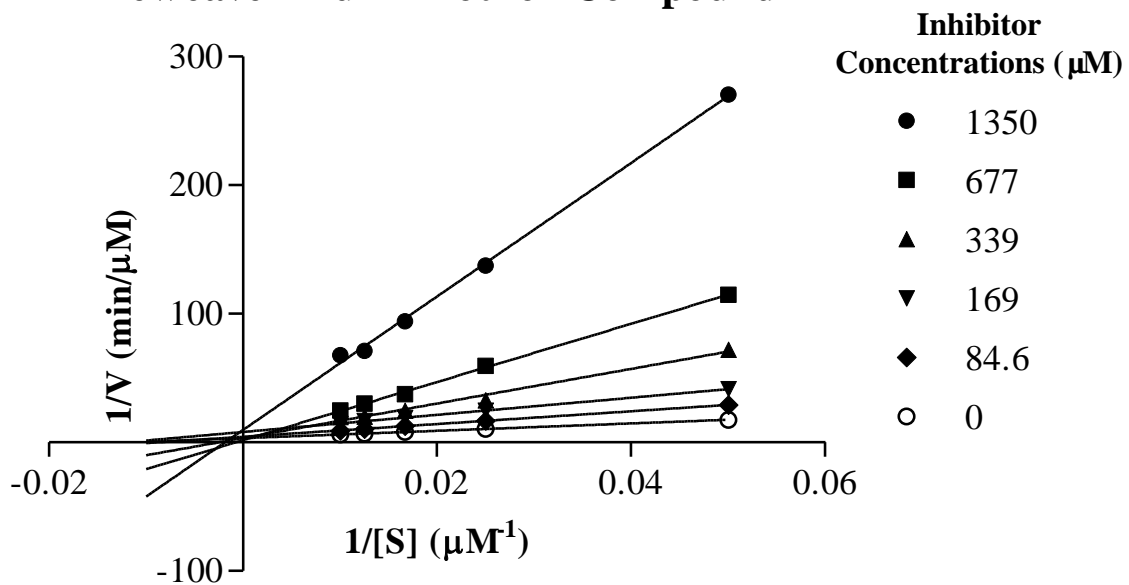
Figure 4.17 Protease inhibition assay with compound 2 as inhibitor. All lines intercept at the same point at X-axis with negative value indicates that the reaction kinetics involves non-competitive inhibition.

Table 4.8 :   Best-fit values of shared parameters, V_{max}, Alpha, $K_{i\ exp}$ and K_m for compound 3 with different concentrations, I, fitted using nonlinear regression mixed model inhibition method in GraphPad Prism 5.0 software.

Concentrations (μM)	1270	636	318	159	79.5	0
Mixed model inhibition						
Best-fit values						
$Vmax$ (μM/min)	0.30	0.30	0.30	0.30	0.30	0.30
I (μM)	= 1270	= 636	= 318	= 159	= 79.5	= 0
Alpha	0.9	0.9	0.9	0.9	0.9	0.9
$Ki\ exp$ (μM)	510	510	510	510	510	510
Km (μM)	93	93	93	93	93	93
Std. Error						
$Vmax$ (μM/min)	0.03	0.03	0.03	0.03	0.03	0.03
Alpha	0.5	0.5	0.5	0.5	0.5	0.5
$Ki\ exp$ (μM)	120	120	120	120	120	120
Km (μM)	14	14	14	14	14	14

$Vmax$ = maximum enzyme velocity without inhibitor.

Km = Michaelis-Menten constant.

Alpha = 1, mixed-model = noncompetitive inhibition.

Alpha very large, mixed-model = competitive inhibition.

Alpha very small but > 0, mixed-model = uncompetitive inhibition.

$Ki\ exp$ = inhibition constant calculated from in vitro experiment.

Lineweaver-Burk Plot for Compound 3
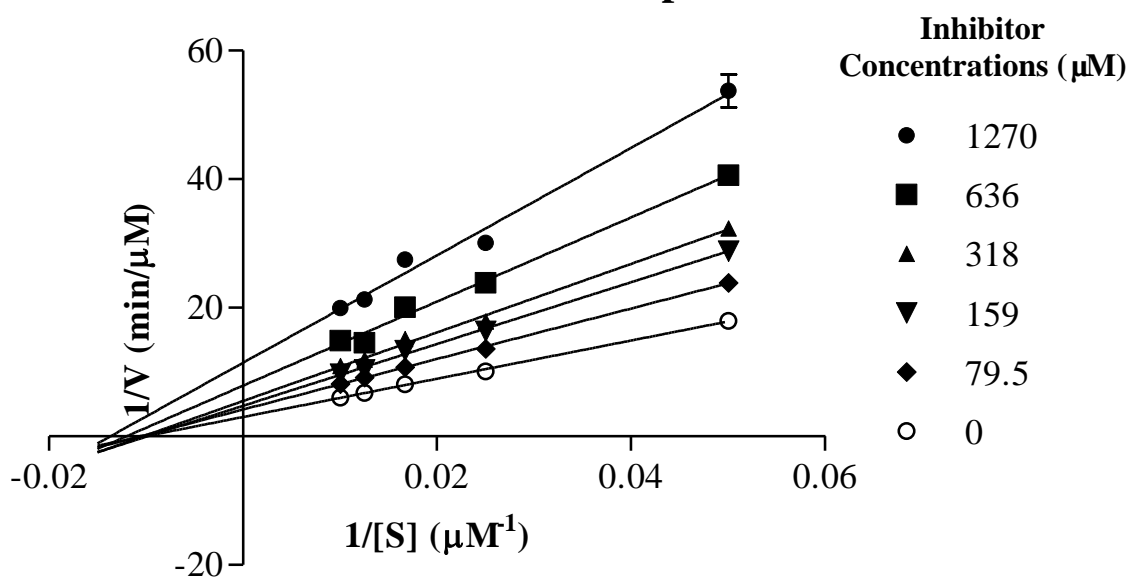
Figure 4.18   Protease inhibition assay with compound 3 as inhibitor. All lines intercept at the same point at X-axis with negative value indicates that the reaction kinetics involves non-competitive inhibition.

Table 4.9 :   Best-fit values of shared parameters, V_{max}, Alpha, $K_{i\ exp}$ and K_m for compound 4 with different concentrations, I, fitted using nonlinear regression mixed model inhibition method in GraphPad Prism 5.0 software.

Concentrations (μM)	1420	709	354	177	88.6	0
Mixed model inhibition						
Best-fit values						
Vmax	0.34	0.34	0.34	0.34	0.34	0.34
I	= 1420	= 709	= 354	= 177	= 88.6	= 0
Alpha	1.1	1.1	1.1	1.1	1.1	1.1
Ki exp	186	186	186	186	186	186
Km	98	98	98	98	98	98
Std. Error						
Vmax	0.03	0.03	0.03	0.03	0.03	0.03
Alpha	0.5	0.5	0.5	0.5	0.5	0.5
Ki exp	38	38	38	38	38	38
Km	15	15	15	15	15	15

Vmax = maximum enzyme velocity without inhibitor.

Km = Michaelis-Menten constant.

Alpha = 1, mixed-model = noncompetitive inhibition.

Alpha very large, mixed-model = competitive inhibition.

Alpha very small but > 0, mixed-model = uncompetitive inhibition.

Ki exp = inhibition constant calculated from in vitro experiment.
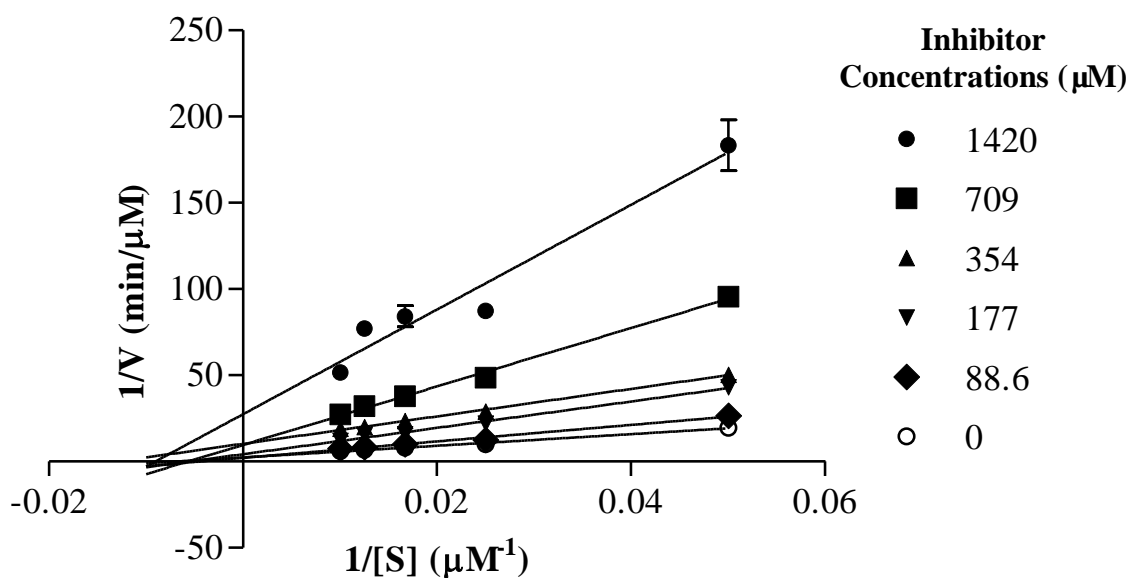
Figure 4.19 Protease inhibition assay with compound 4 as inhibitor. All lines intercept at the same point at X-axis with negative value indicates that the reaction kinetics involves non-competitive inhibition.
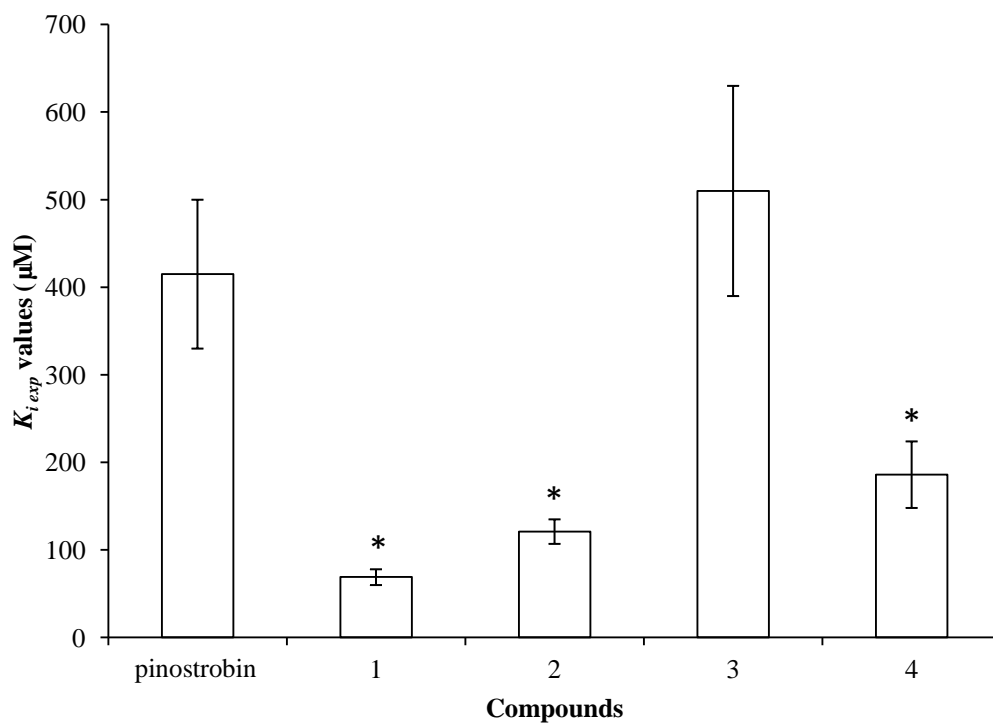
Figure 4.20   Unpaired t-tests for $K_{i\,exp}$ values of compounds 1 - 4 compared with $K_{i\,exp}$ value of standard pinostrobin. * indicates significant different with p value < 0.05.

Table 4.10 : NumCl, ΔG_{dock} and $K_{i\,dock}$ values of the best binding conformations of the small compounds from virtual screening towards DEN-2 NS2B-NS3 proteases, homology model DH-1 compared to the $K_{i\,exp}$ values obtained from protease bioassay in this study.

Compound	Compound Identity	NumCl	ΔG_{dock} (kcal mol^{-1})	$K_{i\,dock}$ (μM)	$K_{i\,exp}$ (μM) in this study
1	2-phenyl-6-(1H-1,2,3,4-tetraazol-5-yl)-4H-chromen-4-one	12/20	-6.17	45	69 ± 9 *
2	2-[4-(dimethylamino)phenyl]-5,7-dimethyl-3,4-dihydro-2H-1-benzopyran-4-one	13/20	-5.77	86	121 ± 14 *
3	6-phenyl-6a,12a-dihydro-6H,7H-chromeno[4,3-b]chromene	15/20	-5.33	175	510 ± 120
4	2-(2,3-dihydro-1,4-benzodioxin-6-yl)-3,4-dihydro-2H-1-benzopyran-4-one	11/20	-5.29	187	186 ± 38 *
Standard	***R*-pinostrobin**				**pinostrobin**
		6/20	**-4.89**	**358**	**415 ± 85**

NumCl = the number of conformations with RMSD < 2.0.

ΔG_{dock} = free energy of binding estimated from AutoDock 4.2 software.

$K_{i\,dock}$ = inhibition constant derived from ΔG_{dock}.

* indicates significant different (p value < 0.05) of unpaired t-tests for $K_{i\,exp}$ values of compounds 1 - 4 compared with $K_{i\,exp}$ value of standard pinostrobin.

ΔG_{dock} and $K_{i\,dock}$ values for *R*-pinostrobin are slightly different from the values in Table 3.1 due to rerun of docking in virtual screening using different parameters.

4.2.3   SAR Study

From the *in vitro* result, compound 1, which is a flavone showed the best DEN-2 NS2B-NS3 protease inhibition activity. It also contains a side chain with hydrogen bond acceptor and donor properties at position 6 of the structure that is lacking in the other compounds (Figure 3.21, p. 70; Figure 4.26). This side chain might play an important role in providing hydrogen bonding interactions for binding in the binding site (Figure 4.21). Conversely, *R*-pinostrobin (Figure 2.6, p.21) which has a side chain (methoxyl group) at position 5 which acts as a hydrogen acceptor, and a side chain (hydroxyl group) at position 7 with hydrogen bond donor property, showed weaker protease inhibition, and therefore implies that these substitutions are less important (Figure 4.25).

On the other hand, compounds 2 and 4 showed better protease inhibition activity than the standard. This might be due to the side chain at position 2 for compounds 2 and 4 having hydrogen bond donor and acceptor properties, respectively, as opposed to the standard.

Furthermore, based on docking results, amino acid residues Glu88, Gly124 and Ala167 located inside the binding pocket, play important role in forming hydrogen bond interactions as hydrogen bond acceptor (Glu88) and hydrogen bond donor (Gly124 and Ala167) (Figures 4.21 - 4.25). Thus, a side chain containing electronegative atom(s) such as nitrogen, oxygen or fluorine is necessary as hydrogen bond acceptor as well as -NH or -OH group as hydrogen bond donor. This side chain should be located either at position 2 or 6 of the structure.

From Figures 4.21 and 4.22, it is highlighted that the benzopyrone ring structure of the inhibitor is necessary for the formation of pi-cation interaction with the ammonium group of the residue Lys74. For compound 3, pi-cation and pi-sigma interactions could be observed between the two benzyl rings of the inhibitor with Lys74 and Ile122, respectively (Figure 4.23). However, this structure is not recommended as the *in vitro* result demonstrated low inhibition activity. As mentioned in section 4.2.2, this might due to the lack of hydrogen bond interactions with any of the residues Glu88, Glu124 or Ala167 in the other end of the binding site, which could be necessary in establishing binding stability in the binding pocket.

Figures 4.21 – 4.25 also highlights an important region (at the left side of the binding site) which is hydrophobic. It can be seen that hydrophobic side chain (ring) of the inhibitors fit comfortably into this hydrophobic pocket. In addition, it was postulated that the hydrophobic side chain of the residues is also necessary for fitting the structure towards the proper position for the mentioned pi-cation and hydrogen bonding interactions.

In addition to the interactions between the inhibitors and the binding site of the protease discussed above, shape complementarity seemed to also play its role in ensuring a stable binding complex in this study. Three dimensional inspection of the docking results highlighted compound 1 to be maximizing the spacial accomodation of the binding site.
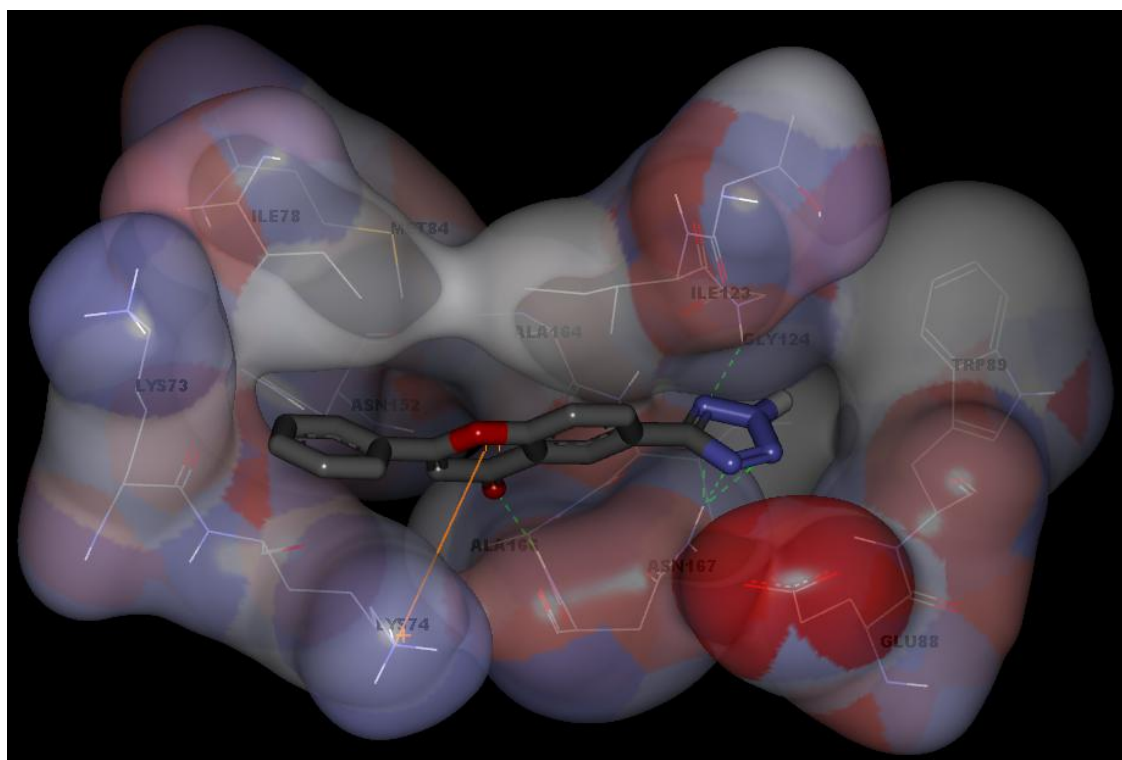
Figure 4.21 Interactions between compound 1 and the allosteric binding pocket. Pi-cation interaction (orange) and hydrogen bonding interactions (green) were observed.
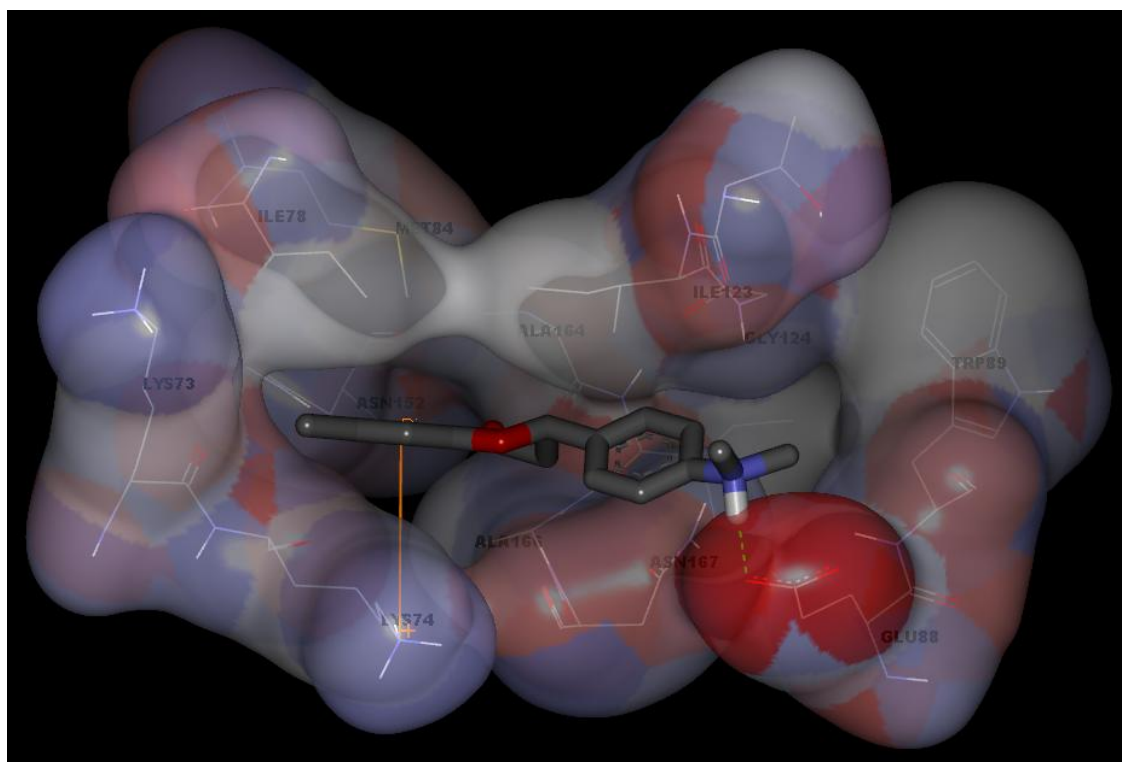
Figure 4.22 Interactions between compound 2 and the allosteric binding pocket. Pi-cation interaction (orange) and hydrogen bonding interactions (green) were observed.
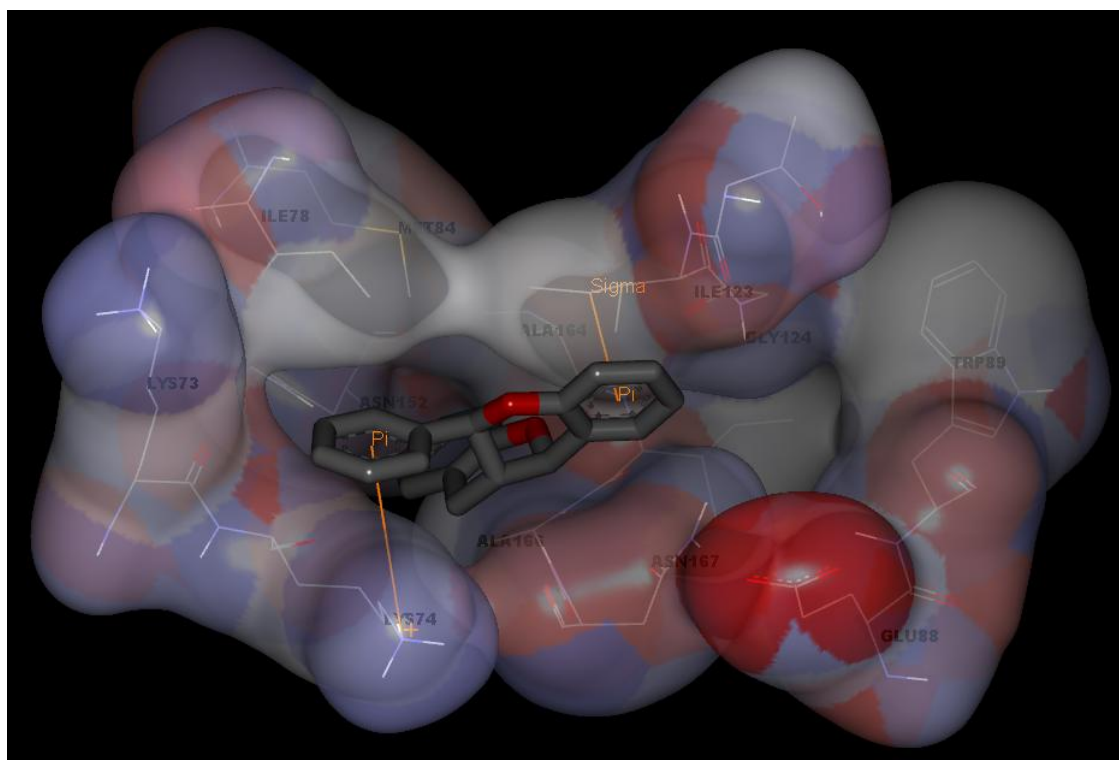
Figure 4.23 Interactions between compound 3 and the allosteric binding pocket. Pi-cation and pi-sigma interactions (orange) were observed.
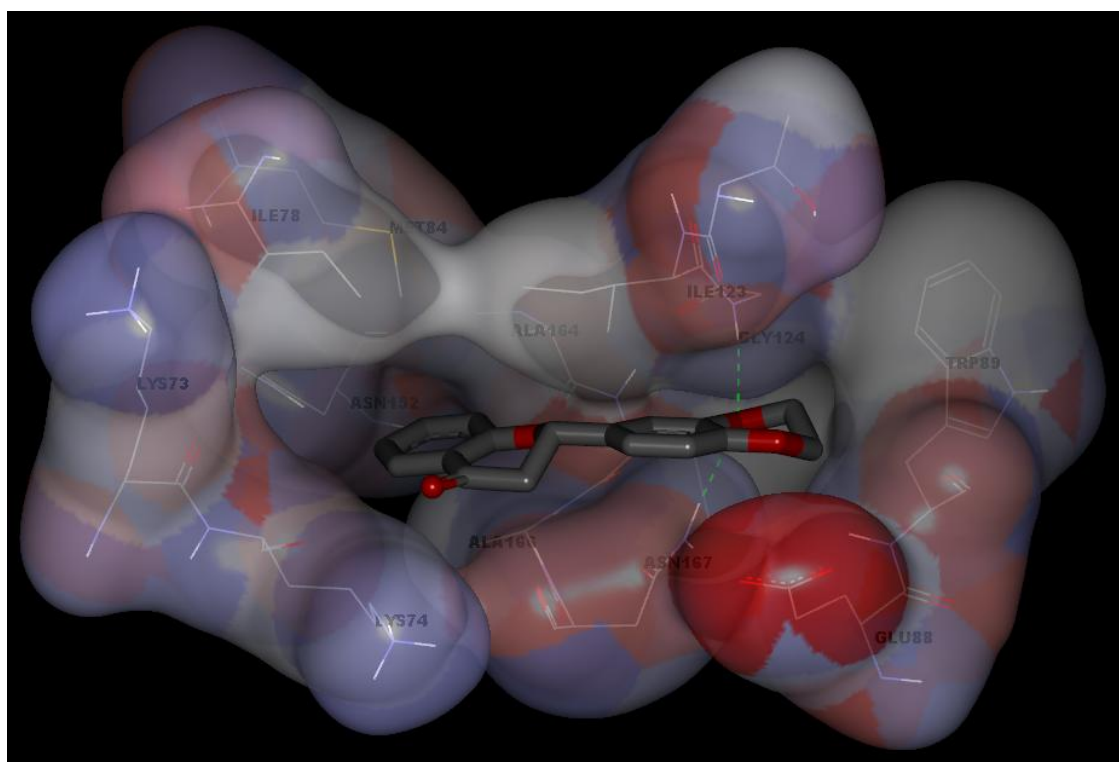
Figure 4.24 Interactions between compound 4 and the allosteric binding pocket. Hydrogen bonding interactions (green) were observed.
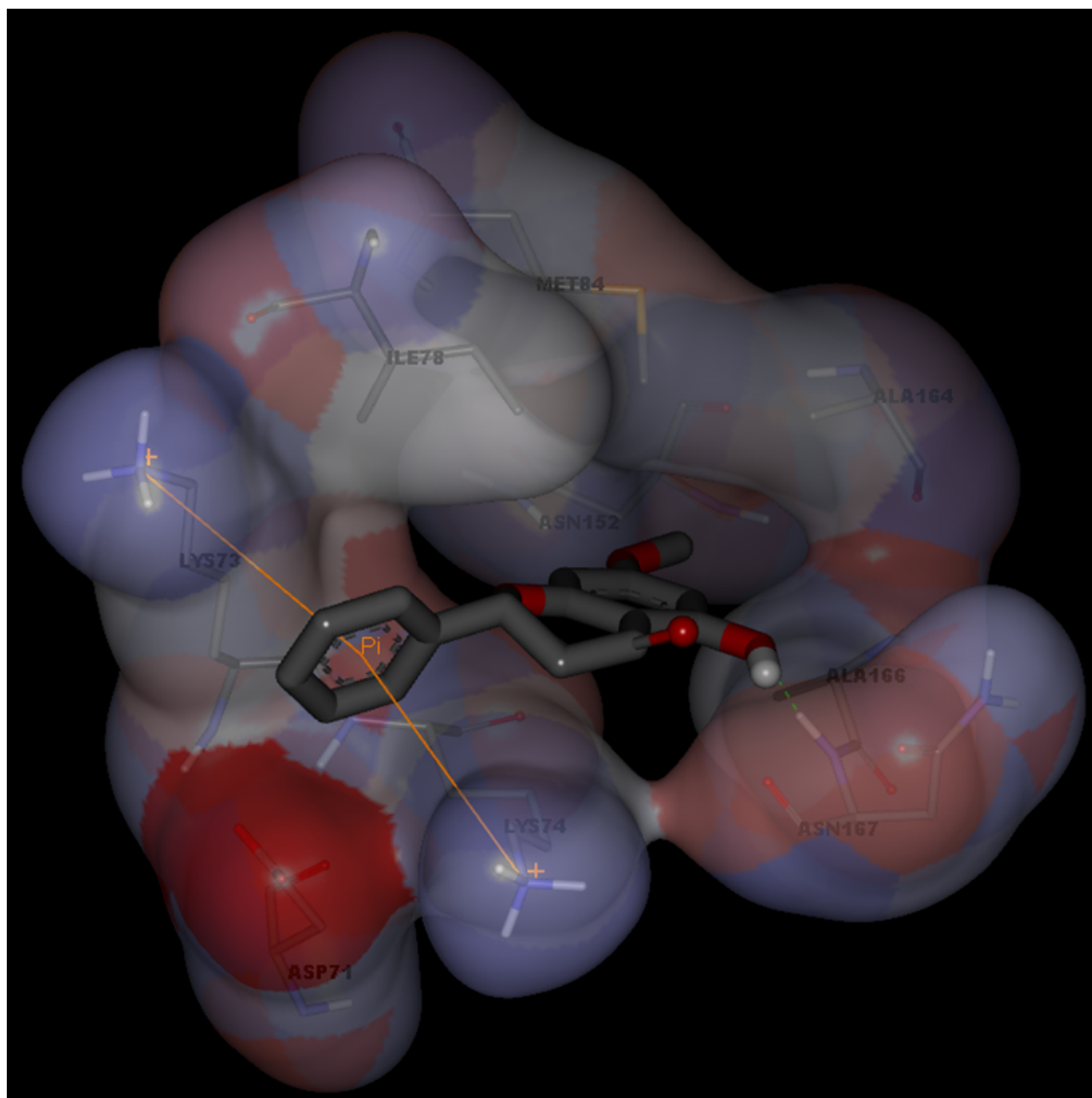
Figure 4.25 Interactions between *R*-pinostrobin (standard) and the allosteric binding pocket. Pi-cation interaction (orange) and hydrogen bonding interactions (green) were observed.

Hence, a potential lead structure for DEN-2 NS2B-NS3 protease inhibitor could have properties as shown in Figure 4.26.
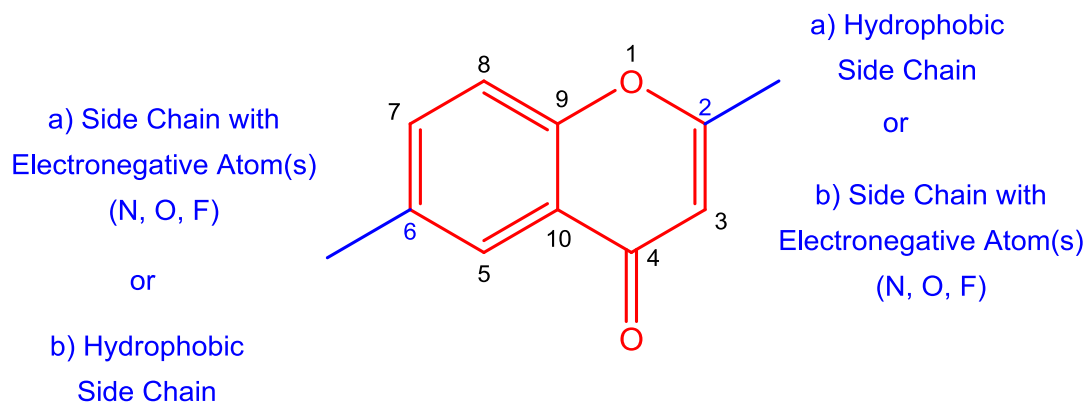
Figure 4.26 Suggested potential lead structure towards the design of DEN-2 NS2B-NS3 protease inhibitor. The important features are benzopyrone ring structure (coloured in red) and side chains with the specific functional groups at positions 2 and 6 (coloured in blue).

CHAPTER FIVE

CONCLUSION

5.1   Overall Conclusion

Based on chalcone, flavanone, and flavone scaffolds, virtual screening of compounds from the online ZINC database was performed in search of potential non-competitive inhibitors against DEN-2 NS2B-NS3 protease. A total of 34 small compounds were identified to have higher binding affinities compared to the standard ligand used in this study. However, only 4 of these small compounds were available for purchase. Modeling analysis suggests that the compounds form hydrogen bonding interactions with the amino acid residues Glu88, Gly124 and Asn167, as well as pi-cation interactions with Lys74. The role of these amino acid residues (Lys74, Glu88, Gly124 and Asn167) is worthy of further study.

Soluble DEN-2 NS2B-NS3 protease gene was successfully cloned and the corresponding protease could be expressed, even though with higher molecular mass. The results from *in vitro* inhibition assays supported the *in silico* results obtained. Compound 1 was found to be the best inhibitor of DEN-2.

It is proposed that for non-competitive inhibition studies on DEN-2 protease, an appropriate model should exhibit a conformation of the allosteric binding site that resembles the homology model, DH-1, built in this study. DH-1 could also be used for further virtual screening studies involving a larger database of compounds and for drug design studies of non-competitive inhibitors against DEN-2 NS2B-NS3 protease.

From this study, compound 1, 2 and 4 showed better inhibition against DEN-2 proteolytic activity compared to the pinostrobin (standard). Compound 1 exhibited the best inhibition activity with *in vitro* $K_{i\,exp}$ of 69 ± 9 μM. There were no previous reports on these non-competitve inhibition activities against DEN-2 NS2B-NS3 protease. On

the other hand, most reports described the competitive inhibition activities of compounds, even with lower $K_{i\ exp}$ values, which indicate better protease activity inhibition. However, these competitive inhibitors could be displaced when higher concentrations of the substrates are available. Hence, logically, a higher concentration of inhibitor will be needed to compete with the substrate for binding to active site. Consequently, this may lead to these competitive inhibitors reaching the toxic concentrations and becoming harmful to human. In view of this, compound 1 is currently the best non-competitive inhibitor to be reported.

We also suggested a potential lead structure or pharmacophore for non-competitive inhibitor against DEN-2 NS2B-NS3 protease based on SAR study.

In conclusion, the rational discovery method described here has potential for use in the discovery of lead compounds for the treatment of dengue, as well as other disease targets.

5.2   Future Studies

The potential lead structure or pharmacophore (Figure 4.26) could be used in future anti-dengue drug design, starting by synthesizing new compounds with the suggested features for *in vitro* assay verification. With a larger number of different types of compounds, DH-1 could also be used to perform virtual screening to obtain novel non-competitive inhibitors against DEN-2 NS2B-NS3 protease. Crystallization trials involving compound 1 could be carried out to enable a more accurate assessment of the binding mechanism of the non-competitive inhibitors found in this study and to assist in further work on the design and development of anti-dengue agents using small compounds. Furthermore, compounds 1, 2 and 4 could also be used in further *in vivo*

studies such as cell bioassay to verify their inhibition activities as well as toxicity at the cellular level.

5.3   Limitations of Study

One of the limitations of this study is that, the DEN-2 NS2B-NS3 protease model used for virtual screening is a homology model instead of actual crystal structure. An actual crystal structure of DEN-2 NS2B-NS3 protease might behave differently and might not even fold into the same conformation as that of the homology model. However, since the actual crystal structure of DEN-2 NS2B-NS3 protease complexed with non-competitive inhibitor has yet available, the homology model, DH-1 is currently the best model for non-competitive inhibition study.

Futuremore, the ΔG_{dock} calculated by AutoDock software cannot be used directly to represent the actual binding energy of a compound. This is mainly because the AutoDock software uses implicit water environment, with only parameters optimized for water environment, rather than uses explicit water environment where actual water molecules will be included for calculation. However, the latter method might be more accurate, it consumes a lot of time and computer power. Moreover, AutoDock software uses semiempirical force field for energy calculation instead of quantum mechanic force field with more accuracy. Nevertheless, the latter method also time and computer power consuming. As virtual screening consists of a very large number of compounds, using calculation with explicit water environment and quantum mechanic force field would be irrational. Thus, AutoDock software with faster calculating time was used in spite of it is less accurate.

Besides, not all the compounds identified in virtual screening were available for purchase. SAR study might be less accurate with *in vitro* result involving just a few compounds. Moreover, the purchasable compounds were expensive in cost.

Lastly, synthesis of the potential compounds resulted from virtual screening was unable to be performed due to insufficient time.

REFERENCES

Abdelwahab, S. I., Mohan, S., Abdulla, M. A., Sukari, M. A., Abdul, A. B., Taha, M. M., Syam, S., Ahmad, S., & Lee, K. H. (2011). The methanolic extract of Boesenbergia rotunda (L.) Mansf. and its major compound pinostrobin induces anti-ulcerogenic property in vivo: possible involvement of indirect antioxidant action. *J Ethnopharmacol, 137*(2), 963-970. doi: 10.1016/j.jep.2011.07.010

Ahmad, N., Fazal, H., Ayaz, M., Abbasi, B. H., Mohammad, I., & Fazal, L. (2011). Dengue fever treatment with Carica papaya leaves extracts. *Asian Pac J Trop Med, 1*(4), 330-333. doi: 10.1016/S2221-1691(11)60055-5

Aleshin, A. E., Shiryaev, S. A., Strongin, A. Y., & Liddington, R. C. (2007). Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. *Protein Sci, 16*(5), 795-806. doi: 10.1110/ps.072753207

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402.

Alvarez, J. C. (2004). High-throughput docking as a source of novel drug leads. *Curr Opin Chem Biol, 8*(4), 365-370. doi: 10.1016/j.cbpa.2004.05.001

Amberg, S. M., Nestorowicz, A., McCourt, D. W., & Rice, C. M. (1994). NS2B-3 proteinase-mediated processing in the yellow fever virus structural region: in vitro and in vivo studies. *J Virol, 68*(6), 3794-3802.

Arias, C. F., Preugschat, F., & Strauss, J. H. (1993). Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. *Virology, 193*(2), 888-899. doi: 10.1006/viro.1993.1198

Baginski, M., Sternal, K., Czub, J., & Borowski, E. (2005). Molecular modelling of membrane activity of amphotericin B, a polyene macrolide antifungal antibiotic. *Acta Biochim Pol, 52*(3), 655-658.

Bailey, D., & Brown, D. (2001). High-throughput chemistry and structure-based design: survival of the smartest. *Drug Discov Today, 6*(2), 57-59.

Bartholomeusz, A., & Thompson, P. (1999). Flaviviridae polymerase and RNA replication. *J Viral Hepat, 6*(4), 261-270.

Bartulewic, D., Markowska, A., Wolczynski, S., Dabrowska, M., & Rozanski, A. (2000). Molecular modelling, synthesis and antitumour activity of carbocyclic analogues of netropsin and distamycin--new carriers of alkylating elements. *Acta Biochim Pol, 47*(1), 23-35.

Bazan, J. F., & Fletterick, R. J. (1989). Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology, 171*(2), 637-639.

Boobbyer, D. N. A., Goodford, P. J., McWhinnie, P. M., & Wade, R. C. (1989). New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *Journal of Medicinal Chemistry, 32*(5), 1083-1094. doi: 10.1021/jm00125a025

Bowie, J. U., Luthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science, 253*(5016), 164-170.

Brinkworth, R. I., Fairlie, D. P., Leung, D., & Young, P. R. (1999). Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. *J Gen Virol, 80 (Pt 5)*, 1167-1177.

Cahour, A., Falgout, B., & Lai, C. J. (1992). Cleavage of the dengue virus polyprotein at the NS3/NS4A and NS4B/NS5 junctions is mediated by viral protease NS2B-NS3, whereas NS4A/NS4B may be processed by a cellular protease. *J Virol, 66*(3), 1535-1542.

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. *J Virol, 84*(6), 3059-3067. doi: 10.1128/JVI.02044-09

Chang, R. (2005). *Physical Chemistry for the Biosciences*: University Science.

Chanprapaph, S., Saparpakorn, P., Sangma, C., Niyomrattanakit, P., Hannongbua, S., Angsuthanasombat, C., & Katzenmeier, G. (2005). Competitive inhibition of the dengue virus NS3 serine protease by synthetic peptides representing polyprotein cleavage sites. *Biochem Biophys Res Commun, 330*(4), 1237-1246. doi: 10.1016/j.bbrc.2005.03.107

Cheenpracha, S., Karalai, C., Ponglimanont, C., Subhadhirasakul, S., & Tewtrakul, S. (2006). Anti-HIV-1 protease activity of compounds from Boesenbergia pandurata. *Bioorg Med Chem, 14*(6), 1710-1714. doi: 10.1016/j.bmc.2005.10.019

Chowdhury, N., Ghosh, A., & Chandra, G. (2008). Mosquito larvicidal activities of Solanum villosum berry extract against the dengue vector Stegomyia aegypti. *BMC Complement Altern Med, 8*, 10. doi: 10.1186/1472-6882-8-10

Clum, S., Ebner, K. E., & Padmanabhan, R. (1997). Cotranslational membrane insertion of the serine proteinase precursor NS2B-NS3(Pro) of dengue virus type 2 is required for efficient in vitro processing and is mediated through the hydrophobic regions of NS2B. *J Biol Chem, 272*(49), 30715-30723.

Clyde, K., Kyle, J. L., & Harris, E. (2006). Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *J Virol, 80*(23), 11418-11431. doi: 10.1128/JVI.01257-06

Connelly, P. R., Aldape, R. A., Bruzzese, F. J., Chambers, S. P., Fitzgibbon, M. J., Fleming, M. A., Itoh, S., Livingston, D. J., Navia, M. A., Thomson, J. A., & et al. (1994). Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. *Proc Natl Acad Sci U S A, 91*(5), 1964-1968.

Copeland, R. A. (2002a). Kinetics of Single-Substrate Enzyme Reactions *Enzymes* (pp. 109-145): John Wiley & Sons, Inc.

Copeland, R. A. (2002b). Reversible Inhibitors *Enzymes* (pp. 266-304): John Wiley & Sons, Inc.

Datta, S., & Wattal, C. (2010). Dengue NS1 antigen detection: a useful tool in early diagnosis of dengue virus infection. *Indian J Med Microbiol, 28*(2), 107-110. doi: 10.4103/0255-0857.62484

Durbin, A. P., Karron, R. A., Sun, W., Vaughn, D. W., Reynolds, M. J., Perreault, J. R., Thumar, B., Men, R., Lai, C. J., Elkins, W. R., Chanock, R. M., Murphy, B. R., & Whitehead, S. S. (2001). Attenuation and immunogenicity in humans of a live dengue virus type-4 vaccine candidate with a 30 nucleotide deletion in its 3'-untranslated region. *Am J Trop Med Hyg, 65*(5), 405-413.

Durbin, A. P., Whitehead, S. S., McArthur, J., Perreault, J. R., Blaney, J. E., Jr., Thumar, B., Murphy, B. R., & Karron, R. A. (2005). rDEN4delta30, a live attenuated dengue virus type 4 vaccine candidate, is safe, immunogenic, and highly infectious in healthy adult volunteers. *J Infect Dis, 191*(5), 710-718. doi: 10.1086/427780

Edelman, R. (2007). Dengue vaccines approach the finish line. *Clin Infect Dis, 45 Suppl 1*, S56-60. doi: 10.1086/518148

Erbel, P., Schiering, N., D'Arcy, A., Renatus, M., Kroemer, M., Lim, S. P., Yin, Z., Keller, T. H., Vasudevan, S. G., & Hommel, U. (2006). Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat Struct Mol Biol, 13*(4), 372-373. doi: 10.1038/nsmb1073

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics, Chapter 5*, Unit 5 6. doi: 10.1002/0471250953.bi0506s15

Falgout, B., Pethel, M., Zhang, Y. M., & Lai, C. J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J Virol, 65*(5), 2467-2475.

Flachmann, R., & Kuhlbrandt, W. (1996). Crystallization and identification of an assembly defect of recombinant antenna complexes produced in transgenic tobacco plants. *Proc Natl Acad Sci U S A, 93*(25), 14966-14971.

Forli, S. (2010). Raccoon | AutoDock VS: an automated tool for preparing AutoDock virtual screenings (Version 1.0). Retrieved from http://autodock.scripps.edu/resources/raccoon

Fox, B. (1989). Bash, the GNU Bourne-Again Shell (Version 4.1). Retrieved from http://www.gnu.org/software/bash/

Frecer, V., & Miertus, S. (2010). Design, structure-based focusing and in silico screening of combinatorial library of peptidomimetic inhibitors of Dengue virus NS2B-NS3 protease. *J Comput Aided Mol Des, 24*(3), 195-212. doi: 10.1007/s10822-010-9326-8

Frimayanti, N., Chee, C. F., Zain, S. M., & Rahman, N. A. (2011). Design of new competitive dengue ns2b/ns3 protease inhibitors-a computational approach. *Int J Mol Sci, 12*(2), 1089-1100. doi: 10.3390/ijms12021089

Ganesh, V. K., Muller, N., Judge, K., Luan, C.-H., Padmanabhan, R., & Murthy, K. H. M. (2005). Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. *Bioorganic & Medicinal Chemistry, 13*(1), 257-264. doi: 10.1016/j.bmc.2004.09.036

Geiss, B. J., Thompson, A. A., Andrews, A. J., Sons, R. L., Gari, H. H., Keenan, S. M., & Peersen, O. B. (2009). Analysis of Flavivirus NS5 Methyltransferase Cap Binding. *Journal of Molecular Biology, 385*(5), 1643-1654.

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry, 28*(7), 849-857. doi: 10.1021/jm00145a002

Gould, D. J., Yuill, T. M., Moussa, M. A., Simasathien, P., & Rutledge, L. C. (1968). An insular outbreak of dengue hemorrhagic fever. 3. Identification of vectors and observations on vector ecology. *Am J Trop Med Hyg, 17*(4), 609-618.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev, 11*(3), 480-496.

Guha-Sapir, D., & Schimmer, B. (2005). Dengue fever: new paradigms for a changing epidemiology. *Emerg Themes Epidemiol, 2*(1), 1. doi: 10.1186/1742-7622-2-1

Hammon, W. M., Rudnick, A., & Sather, G. E. (1960). Viruses associated with epidemic hemorrhagic fevers of the Philippines and Thailand. *Science, 131*, 1102-1103.

Hang, V. T., Nguyet, N. M., Trung, D. T., Tricou, V., Yoksan, S., Dung, N. M., Van Ngoc, T., Hien, T. T., Farrar, J., Wills, B., & Simmons, C. P. (2009). Diagnostic accuracy of NS1 ELISA and lateral flow rapid tests for dengue sensitivity, specificity and relationship to viraemia and antibody responses. *PLoS Negl Trop Dis, 3*(1), e360. doi: 10.1371/journal.pntd.0000360

Hanley, K. A., Manlucu, L. R., Manipon, G. G., Hanson, C. T., Whitehead, S. S., Murphy, B. R., & Blaney, J. E., Jr. (2004). Introduction of mutations into the non-structural genes or 3' untranslated region of an attenuated dengue virus type 4 vaccine candidate further decreases replication in rhesus monkeys while retaining protective immunity. *Vaccine, 22*(25-26), 3440-3448. doi: 10.1016/j.vaccine.2004.02.031

Harris, D. C. (2004). *Exploring Chemical Analysis*: W. H. Freeman.

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Pat Cerretti, D., Urdal, D. L., & Conlon, P. J. (1988). A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Nat Biotech, 6*(10), 1204-1210.

Hrobowski, Y. M., Garry, R. F., & Michael, S. F. (2005). Peptide inhibitors of dengue virus and West Nile virus infectivity. *Virol J, 2*, 49. doi: 10.1186/1743-422X-2-49

Huey, R., Morris, G. M., Olson, A. J., & Goodsell, D. S. (2007). A semiempirical free energy force field with charge-based desolvation. *J Comput Chem, 28*(6), 1145-1152. doi: 10.1002/jcc.20634

Irie, K., Mohan, P. M., Sasaguri, Y., Putnak, R., & Padmanabhan, R. (1989). Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene, 75*(2), 197-211.

Irwin, J. J., & Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model, 45*(1), 177-182. doi: 10.1021/ci049714+

Isa, N. M., Abdelwahab, S. I., Mohan, S., Abdul, A. B., Sukari, M. A., Taha, M. M., Syam, S., Narrima, P., Cheah, S. C., Ahmad, S., & Mustafa, M. R. (2012). In vitro anti-inflammatory, cytotoxic and antioxidant activities of boesenbergin A,

a chalcone isolated from Boesenbergia rotunda (L.) (fingerroot). *Braz J Med Biol Res.*

Jain, M., Ganju, L., Katiyal, A., Padwad, Y., Mishra, K. P., Chanda, S., Karan, D., Yogendra, K. M., & Sawhney, R. C. (2008). Effect of Hippophae rhamnoides leaf extract against Dengue virus infection in human blood-derived macrophages. *Phytomedicine, 15*(10), 793-799. doi: 10.1016/j.phymed.2008.04.017

Jaipetch, T., Kanghae, S., Pancharoen, O., Patrick, V., Reutrakul, V., Tuntiwachwuttikul, P., & White, A. (1982). Constituents of *Boesenbergia pandurata* (syn. *Kaempferia pandurata*): Isolation, Crystal Structure and Synthesis of (±)-Boesenbergin A. *Aust J Chem, 35*(2), 351-361. doi: http://dx.doi.org/10.1071/CH9820351

Jones, J. E. (1924). On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society of London. Series A, 106*(738), 463-477. doi: 10.1098/rspa.1924.0082

Kalaivani, K., Senthil-Nathan, S., & Murugesan, A. G. (2012). Biological activity of selected Lamiaceae and Zingiberaceae plant essential oils against the dengue vector Aedes aegypti L. (Diptera: Culicidae). *Parasitol Res, 110*(3), 1261-1268. doi: 10.1007/s00436-011-2623-x

Kalendar, R. (2010). Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analyses. Retrieved from http://primerdigital.com/tools/

Kautner, I., Robinson, M. J., & Kuhnle, U. (1997). Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis, and prevention. *J Pediatr, 131*(4), 516-524.

Keelapang, P., Sriburi, R., Supasa, S., Panyadee, N., Songjaeng, A., Jairungsri, A., Puttikhunt, C., Kasinrerk, W., Malasit, P., & Sittisombut, N. (2004). Alterations of pr-M cleavage and virus export in pr-M junction chimeric dengue viruses. *J Virol, 78*(5), 2367-2381.

Kiat, T. S., Pippen, R., Yusof, R., Ibrahim, H., Khalid, N., & Rahman, N. A. (2006). Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoids of fingerroot, Boesenbergia rotunda (L.), towards dengue-2 virus NS3 protease. *Bioorg Med Chem Lett, 16*(12), 3337-3340. doi: 10.1016/j.bmcl.2005.12.075

Kovendan, K., Murugan, K., & Vincent, S. (2012). Evaluation of larvicidal activity of Acalypha alnifolia Klein ex Willd. (Euphorbiaceae) leaf extract against the malarial vector, Anopheles stephensi, dengue vector, Aedes aegypti and

Bancroftian filariasis vector, Culex quinquefasciatus (Diptera: Culicidae). *Parasitol Res, 110*(2), 571-581. doi: 10.1007/s00436-011-2525-y

KPKM. (2011a). *"Situasi Semasa Demam Denggi Dan Chikungunya Di Malaysia Bagi Minggu 52/2010 (26 Disember 2010 hingga 1 Januari 2011)"*.

KPKM. (2011b). *"Situasi Semasa Demam Denggi Di Malaysia Bagi Minggu 52/2011 (25 hingga 31 Disember 2011)"*.

KPKM. (2012). *"Peningkatan Kes Dan Kematian Denggi Yang Tertinggi Pada Tahun 2012"*.

Kumar, S., Warikoo, R., & Wahab, N. (2010). Larvicidal potential of ethanolic extracts of dried fruits of three species of peppercorns against different instars of an indian strain of dengue fever mosquito, Aedes aegypti L. (Diptera: Culicidae). *Parasitol Res, 107*(4), 901-907. doi: 10.1007/s00436-010-1948-1

Lanciotti, R. S., Calisher, C. H., Gubler, D. J., Chang, G. J., & Vorndam, A. V. (1992). Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. *J Clin Microbiol, 30*(3), 545-551.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948. doi: 10.1093/bioinformatics/btm404

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*(2), 283-291. doi: 10.1107/s0021889892009944

Le Grice, S. F. J., & GrÜNinger-Leitch, F. (1990). Rapid purification of homodimer and heterodimer HIV-1 reverse transcriptase by metal chelate affinity chromatography. *European Journal of Biochemistry, 187*(2), 307-314. doi: 10.1111/j.1432-1033.1990.tb15306.x

Lee, Y. K., Rozana, O., Habibah, A. W., Rohana, Y., & Noorsaadah, A. R. (2006). A Revisit into the DEN2 NS2B/NS3 Virus Protease Homology Model: Structural Verification and Comparison with Crystal Structure of HCV NS3/4A and DEN2 NS3. *Malaysian Journal of Science, 25*(1), 15-22.

Lee, Y. K., Tan, S. K., Habibah, A. W., Rohana, Y., & Noorsaadah, A. R. (2007). Nonsubstrate Based Inhibitors of Dengue Virus Serine Protease: A Molecular Docking Approach to Study Binding Interactions between Protease and

Inhibitors. *Asia Pacific Journal of Molecular Biology and Biotechnology, 15*(2), 53-59.

Lengauer, T., & Rarey, M. (1996). Computational methods for biomolecular docking. *Curr Opin Struct Biol, 6*(3), 402-406.

Lescar, J., Luo, D., Xu, T., Sampath, A., Lim, S. P., Canard, B., & Vasudevan, S. G. (2008). Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. *Antiviral Res, 80*(2), 94-101. doi: 10.1016/j.antiviral.2008.07.001

Leung, D., Schroder, K., White, H., Fang, N. X., Stoermer, M. J., Abbenante, G., Martin, J. L., Young, P. R., & 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. *J Biol Chem, 276*(49), 45762-45771. doi: 10.1074/jbc.M107360200

Leung, J. Y., Pijlman, G. P., Kondratieva, N., Hyde, J., Mackenzie, J. M., & Khromykh, A. A. (2008). Role of nonstructural protein NS2A in flavivirus assembly. *J Virol, 82*(10), 4731-4741. doi: 10.1128/JVI.00002-08

Li, C., Xu, L., Wolan, D. W., Wilson, I. A., & Olson, A. J. (2004). Virtual screening of human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase against the NCI diversity set by use of AutoDock to identify novel nonfolate inhibitors. *J Med Chem, 47*(27), 6681-6690. doi: 10.1021/jm049504o

Li, H., Clum, S., You, S., Ebner, K. E., & Padmanabhan, R. (1999). The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. *J Virol, 73*(4), 3108-3116.

Lima Mda, R., Nogueira, R. M., Schatzmayr, H. G., de Filippis, A. M., Limonta, D., & dos Santos, F. B. (2011). A new approach to dengue fatal cases diagnosis: NS1 antigen capture in tissues. *PLoS Negl Trop Dis, 5*(5), e1147. doi: 10.1371/journal.pntd.0001147

Lin, C., Amberg, S. M., Chambers, T. J., & Rice, C. M. (1993). Cleavage at a novel site in the NS4A region by the yellow fever virus NS2B-3 proteinase is a prerequisite for processing at the downstream 4A/4B signalase site. *J Virol, 67*(4), 2327-2335.

Lindenbach, B. D., Thiel, H. J., & Rice, C. M. (2007). Flaviviridae: The Viruses and Their Replication. In D. M. Knipe & P. M. Howley (Eds.), *Fields Virology, 5th Edition* (pp. 1101-1152): Lippincott-Raven Publishers, Philadelphia.

Lipkind, G. M., & Fozzard, H. A. (2005). Molecular modeling of local anesthetic drug binding by voltage-gated sodium channels. *Mol Pharmacol, 68*(6), 1611-1622. doi: 10.1124/mol.105.014803

Lucas, W. (2001). Viral Capsids and Envelopes: Structure and Function *eLS*: John Wiley & Sons, Ltd.

Luo, D., Xu, T., Hunke, C., Gruber, G., Vasudevan, S. G., & Lescar, J. (2008). Crystal structure of the NS3 protease-helicase from dengue virus. *J Virol, 82*(1), 173-183. doi: 10.1128/JVI.01788-07

Luthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature, 356*(6364), 83-85.

Mahesh Kumar, P., Murugan, K., Kovendan, K., Panneerselvam, C., Prasanna Kumar, K., Amerasan, D., Subramaniam, J., Kalimuthu, K., & Nataraj, T. (2012). Mosquitocidal activity of Solanum xanthocarpum fruit extract and copepod Mesocyclops thermocyclopoides for the control of dengue vector Aedes aegypti. *Parasitol Res*. doi: 10.1007/s00436-012-2876-z

Malabadi, R. B., Ganguly, A., Silva, J. A., Parashar, A., Suresh, M. R., & Sunwoo, H. (2011). Overview of plant-derived vaccine antigens: Dengue virus. *J Pharm Pharm Sci, 14*(3), 400-413.

Marimuthu, G., Rajamohan, S., Mohan, R., & Krishnamoorthy, Y. (2012). Larvicidal and ovicidal properties of leaf and seed extracts of Delonix elata (L.) Gamble (Family: Fabaceae) against malaria (Anopheles stephensi Liston) and dengue (Aedes aegypti Linn.) (Diptera: Culicidae) vector mosquitoes. *Parasitol Res*. doi: 10.1007/s00436-011-2802-9

Markoff, L. (1989). In vitro processing of dengue virus structural proteins: cleavage of the pre-membrane protein. *J Virol, 63*(8), 3345-3352.

McKee, K. T., Jr., Bancroft, W. H., Eckels, K. H., Redfield, R. R., Summers, P. L., & Russell, P. K. (1987). Lack of attenuation of a candidate dengue 1 vaccine (45AZ5) in human volunteers. *Am J Trop Med Hyg, 36*(2), 435-442.

Mehler, E. L., & Solmajer, T. (1991). Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Eng, 4*(8), 903-910.

Miller, S., Kastner, S., Krijnse-Locker, J., Buhler, S., & Bartenschlager, R. (2007). The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner. *J Biol Chem, 282*(12), 8873-8882. doi: 10.1074/jbc.M609919200

Monath, T. P. (1994). Dengue: the risk to developed and developing countries. *Proc Natl Acad Sci U S A, 91*(7), 2395-2400.

Morikawa, T., Funakoshi, K., Ninomiya, K., Yasuda, D., Miyagawa, K., Matsuda, H., & Yoshikawa, M. (2008). Medicinal foodstuffs. XXXIV. Structures of new prenylchalcones and prenylflavanones with TNF-alpha and aminopeptidase N inhibitory activities from Boesenbergia rotunda. *Chem Pharm Bull (Tokyo), 56*(7), 956-962.

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry, 19*(14), 1639-1662. doi: 10.1002/(sici)1096-987x(19981115)19:14<1639::aid-jcc10>3.0.co;2-b

Morris, G. M., Goodsell, D. S., Pique, M. E., Lindstrom, W. L., Huey, R., Forli, S., Hart, W. E., Halliday, S., Belew, R., & J., O. A. (2010). AutoDock Version 4.2 - Automated Docking of Flexible Ligands to Flexible Receptors

Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression : a practical guide to curve fitting*: Oxford University Press.

Muhamad, M., Kee, L. Y., Rahman, N. A., & Yusof, R. (2010). Antiviral actions of flavanoid-derived compounds on dengue virus type-2. *Int J Biol Sci, 6*(3), 294-302.

Muliawan, S. Y., Kit, L. S., Devi, S., Hashim, O., & Yusof, R. (2006). Inhibitory potential of Quercus lusitanica extract on dengue virus type 2 replication. *Southeast Asian J Trop Med Public Health, 37 Suppl 3*, 132-135.

NCBI. National Center for Biotechnology Information, 2012, from www.ncbi.nlm.nih.gov

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. *J Virol, 86*(1), 438-446. doi: 10.1128/JVI.06225-11

Ooms, F. (2000). Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr Med Chem, 7*(2), 141-158.

Othman, R., Kiat, T. S., Khalid, N., Yusof, R., Newhouse, E. I., Newhouse, J. S., Alam, M., & Rahman, N. A. (2008). Docking of noncompetitive inhibitors into dengue virus type 2 protease: understanding the interactions with allosteric binding sites. *J Chem Inf Model, 48*(8), 1582-1591. doi: 10.1021/ci700388k

Pancharoen, O., Picker, K., Reutrakul, V., Taylor, W., & Tuntiwachwuttikul, P. (1987). Constituents of the Zingiberaceae. X. Diastereomers of [7-Hydroxy-5-Methoxy-2-Methyl-2-(4'-Methylpent-3'-Enyl)-2H-Chromen-8-yl] [3"-Methyl-2'-(3"'-Methylbut-2"'-Enyl]-6"-Phenylcyclohex-3"-Enyl]M Ethanone (Panduratin B), a Constituent of the Red Rhizomes of a Variety of *Boesenbergia pandurata*. *Aust J Chem, 40*(3), 455-459. doi: http://dx.doi.org/10.1071/CH9870455

Panthong, A., Tassaneeyakul, W., Kanjanapothi, D., Tantiwachwuttikul, P., & Reutrakul, V. (1989). Anti-inflammatory activity of 5,7-dimethoxyflavone. *Planta Med, 55*(2), 133-136. doi: 10.1055/s-2006-961905

Park, H., Lee, J., & Lee, S. (2006). Critical assessment of the automated AutoDock as a new docking tool for virtual screening. *Proteins, 65*(3), 549-554. doi: 10.1002/prot.21183

Porcher, M. H. (2003). Multilingual Multiscript Plant Name Database: Sorting Boesenbergia Names Retrieved 1st April, 2012, from http://www.plantnames.unimelb.edu.au/Sorting/Boesenbergia.html

Portela, C., Afonso, C. M., Pinto, M. M., & Ramos, M. J. (2003). Computational studies of new potential antimalarial compounds--stereoelectronic complementarity with the receptor. *J Comput Aided Mol Des, 17*(9), 583-595.

Preugschat, F., & Strauss, J. H. (1991). Processing of nonstructural proteins NS4A and NS4B of dengue 2 virus in vitro and in vivo. *Virology, 185*(2), 689-697.

Preugschat, F., Yao, C. W., & Strauss, J. H. (1990). In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. *J Virol, 64*(9), 4364-4374.

PubChem. (2005). AC1MW9BY - PubChem Public Chemical Database Retrieved 20 December 2011, from http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?sid=112156252&view opt=PubChem

Qiagen. (2003). *The QIAexpressionist - A handbook for high-level expression and purification of 6xHis-tagged proteins* (Fifth ed.).

Rajkumar, S., & Jebanesan, A. (2010). Prevention of Dengue fever through plant based mosquito repellent Clausena dentata (Willd.) M. Roem (Family: Rutaceae) essential oil against Aedes aegypti l. (Diptera: Culicidae) mosquito. *Eur Rev Med Pharmacol Sci, 14*(3), 231-234.

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol, 7*, 95-99.

Robert Putnak, J., Coller, B. A., Voss, G., Vaughn, D. W., Clements, D., Peters, I., Bignami, G., Houng, H. S., Chen, R. C., Barvir, D. A., Seriwatana, J., Cayphas, S., Garcon, N., Gheysen, D., Kanesa-Thasan, N., McDonell, M., Humphreys, T., Eckels, K. H., Prieels, J. P., & Innis, B. L. (2005). An evaluation of dengue type-2 inactivated, recombinant subunit, and live-attenuated vaccine candidates in the rhesus macaque model. *Vaccine, 23*(35), 4442-4452. doi: 10.1016/j.vaccine.2005.03.042

Robin, G., Chappell, K., Stoermer, M. J., Hu, S. H., Young, P. R., Fairlie, D. P., & Martin, J. L. (2009). Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. *J Mol Biol, 385*(5), 1568-1577. doi: 10.1016/j.jmb.2008.11.026

Rodenhuis-Zybert, I. A., Wilschut, J., & Smit, J. M. (2010). Dengue virus life cycle: viral and host factors modulating infectivity. *Cell Mol Life Sci, 67*(16), 2773-2786. doi: 10.1007/s00018-010-0357-z

Rosen, L. (1983). *The global importance of dengue infection and disease.* Paper presented at the Proceedings of the International Conference on DHF, Kuala Lampur, Malaysia: University of Malaysia.

Russell, P. K., Brandt, W. E., & Dalrymple, J. M. (1980). *Chemical and antigenic structure of flaviviruses.* Paper presented at the Academic Press, New York.

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol, 234*(3), 779-815. doi: 10.1006/jmbi.1993.1626

Schneider, G., Clement-Chomienne, O., Hilfiger, L., Schneider, P., Kirsch, S., Bohm, H. J., & Neidhart, W. (2000). Virtual Screening for Bioactive Molecules by Evolutionary De Novo Design Special thanks to Neil R. Taylor for his help in preparation of the manuscript. *Angew Chem Int Ed Engl, 39*(22), 4130-4133. doi: 10.1002/1521-3773(20001117)39:22<4130::AID-ANIE4130>3.0.CO;2-E

Schulze, I. T., & Schlesinger, R. W. (1963). Inhibition of infectious and hemagglutinating properties of type 2 dengue virus by aqueous Agar extracts. *Virology, 19*, 49-57.

Shindo, K., Kato, M., Kinoshita, A., Kobayashi, A., & Koike, Y. (2006). Analysis of antioxidant activities contained in the Boesenbergia pandurata Schult. Rhizome. *Biosci Biotechnol Biochem, 70*(9), 2281-2284.

Simmons, J. S., St. Johns, J. H., & Reynolds, F. H. K. (1931). Experimental studies of dengue. *Philippine Journal of Science, 44*, 1-251.

Siqueira, J. B., Jr., Martelli, C. M., Coelho, G. E., Simplicio, A. C., & Hatch, D. L. (2005). Dengue and dengue hemorrhagic fever, Brazil, 1981-2002. *Emerg Infect Dis, 11*(1), 48-53.

Srivastava, A. K., Putnak, J. R., Warren, R. L., & Hoke, C. H., Jr. (1995). Mice immunized with a dengue type 2 virus E and NS1 fusion protein made in Escherichia coli are protected against lethal dengue virus infection. *Vaccine, 13*(13), 1251-1258.

Stadler, K., Allison, S. L., Schalich, J., & Heinz, F. X. (1997). Proteolytic activation of tick-borne encephalitis virus by furin. *J Virol, 71*(11), 8475-8481.

Stouten, P. F. W., Frömmel, C., Nakamura, H., & Sander, C. (1993). An Effective Solvation Term Based on Atomic Occupancies for Use in Protein Simulations. *Molecular Simulation, 10*(2), 97 - 120.

Svitkin, Y. V., Lyapustin, V. N., Lashkevich, V. A., & Agol, V. I. (1984). Differences between translation products of tick-borne encephalitis virus RNA in cell-free systems from Krebs-2 cells and rabbit reticulocytes: involvement of membranes in the processing of nascent precursors of flavivirus structural proteins. *Virology, 135*(2), 536-541.

Tang, L. I., Ling, A. P., Koh, R. Y., Chye, S. M., & Voon, K. G. (2012). Screening of anti-dengue activity in methanolic extracts of medicinal plants. *BMC Complement Altern Med, 12*, 3. doi: 10.1186/1472-6882-12-3

Taweechaisupapong, S., Singhara, S., Lertsatitthanakorn, P., & Khunkitti, W. (2010). Antimicrobial effects of Boesenbergia pandurata and Piper sarmentosum leaf extracts on planktonic cells and biofilm of oral pathogens. *Pak J Pharm Sci, 23*(2), 224-231.

Tewtrakul, S., Subhadhirasakul, S., Puripattanavong, J., & Panphadung, T. (2003). HIV-1 protease inhibitory substances from the rhizomes of Boesenbergia pandurata Holtt. *Songklanakarin J. Sci. Technol., 25*(4), 503-508.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res, 25*(24), 4876-4882.

Tuchinda, P., Reutrakul, V., Claeson, P., Pongprayoon, U., Sematong, T., Santisuk, T., & Taylor, W. C. (2002). Anti-inflammatory cyclohexenyl chalcone derivatives in Boesenbergia pandurata. *Phytochemistry, 59*(2), 169-173.

Ultee, A. J. (1957). The ethereal oil of *Gastrochilus Panduratum. Ridl Verslag Akad Wetenschappen Amsterdam, 36*, 1262-1264.

Umareddy, I., Chao, A., Sampath, A., Gu, F., & Vasudevan, S. G. (2006). Dengue virus NS4B interacts with NS3 and dissociates it from single-stranded RNA. *J Gen Virol, 87*(Pt 9), 2605-2614. doi: 10.1099/vir.0.81844-0

Ungsurungsie, M., Suthienkul, O., & Paovalo, C. (1982). Mutagenicity screening of popular Thai spices. *Food Chem Toxicol, 20*(5), 527-530.

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng, 8*(2), 127-134.

Wandscheer, C. B., Duque, J. E., da Silva, M. A., Fukuyama, Y., Wohlke, J. L., Adelmann, J., & Fontana, J. D. (2004). Larvicidal action of ethanolic extracts from fruit endocarps of Melia azedarach and Azadirachta indica against the dengue mosquito Aedes aegypti. *Toxicon, 44*(8), 829-835. doi: 10.1016/j.toxicon.2004.07.009

Webb, B., Madhusudhan, M. S., Shen, M.-Y., Marti-Renom, M. A., Eswar, N., Alber, F., Topf, M., Oliva, B., Fiser, A., Sánchez, R., Yerkovich, B., Badretdinov, A., Melo, F., Overington, J. P., & Feyfant, E. (2011). MODELLER A Program for Protein Structure Modeling Release 9.10, r8346, 2011, from http://salilab.org/modeller/manual/

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society, 106*(3), 765-784. doi: 10.1021/ja00315a051

Westaway, E. G., Brinton, M. A., Gaidamovich, S., Horzinek, M. C., Igarashi, A., Kaariainen, L., Lvov, D. K., Porterfield, J. S., Russell, P. K., & Trent, D. W. (1985). Flaviviridae. *Intervirology, 24*(4), 183-192.

Whitby, K., Pierson, T. C., Geiss, B., Lane, K., Engle, M., Zhou, Y., Doms, R. W., & Diamond, M. S. (2005). Castanospermine, a potent inhibitor of dengue virus infection in vitro and in vivo. *J Virol, 79*(14), 8698-8706. doi: 10.1128/JVI.79.14.8698-8706.2005

Whitehead, S. S., Falgout, B., Hanley, K. A., Blaney Jr, J. E., Jr., Markoff, L., & Murphy, B. R. (2003). A live, attenuated dengue virus type 1 vaccine candidate with a 30-nucleotide deletion in the 3' untranslated region is highly attenuated and immunogenic in monkeys. *J Virol, 77*(2), 1653-1657.

WHO. (2008). The Dengue Strategic Plan for the Asia Pacific Region, 2008-2015: World Health Organization.

WHO. (2009). Dengue - Guidelines for Diagnosis, Treatment, Prevention and Control: World Health Organization.

WHO. (2011a). Action Against Dengue: Dengue Day Campaigns Across Asia: World Health Organization.

WHO. (2011b). International Travel and Health 2011. from http://www.who.int/ith/en

WHO. (2012a). Dengue and severe dengue fact sheet N°117: World Health Organization.

WHO. (2012b). Dengue in the Western Pacific Region 2011. from http://www.wpro.who.int/health_topics/dengue/

Wichapong, K., Pianwanit, S., Sippl, W., & Kokpol, S. (2010). Homology modeling and molecular dynamics simulations of Dengue virus NS2B/NS3 protease: insight into molecular interaction. *J Mol Recognit, 23*(3), 283-300. doi: 10.1002/jmr.977

Xu, H., Di, B., Pan, Y. X., Qiu, L. W., Wang, Y. D., Hao, W., He, L. J., Yuen, K. Y., & Che, X. Y. (2006). Serotype 1-specific monoclonal antibody-based antigen capture immunoassay for detection of circulating nonstructural protein NS1: Implications for early diagnosis and serotyping of dengue virus infections. *J Clin Microbiol, 44*(8), 2872-2878. doi: 10.1128/JCM.00777-06

Yamshchikov, V. F., & Compans, R. W. (1994). Processing of the intracellular form of the west Nile virus capsid protein by the viral NS2B-NS3 protease: an in vitro study. *J Virol, 68*(9), 5765-5771.

Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J. L., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L. C., & Chen, Z. (1998). Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 A resolution structure in a hexagonal crystal form. *Protein Sci, 7*(4), 837-847. doi: 10.1002/pro.5560070402

Yang, C. C., Hsieh, Y. C., Lee, S. J., Wu, S. H., Liao, C. L., Tsao, C. H., Chao, Y. S., Chern, J. H., Wu, C. P., & Yueh, A. (2011). Novel dengue virus-specific NS2B/NS3 protease inhibitor, BP2109, discovered by a high-throughput screening assay. *Antimicrob Agents Chemother, 55*(1), 229-238. doi: 10.1128/AAC.00855-10

Yap, T. L., Xu, T., Chen, Y. L., Malet, H., Egloff, M. P., Canard, B., Vasudevan, S. G., & Lescar, J. (2007). Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J Virol, 81*(9), 4753-4765. doi: 10.1128/JVI.02283-06

Yin, Z., Patel, S. J., Wang, W. L., Chan, W. L., Ranga Rao, K. R., Wang, G., Ngew, X., Patel, V., Beer, D., Knox, J. E., Ma, N. L., Ehrhardt, C., Lim, S. P., Vasudevan, S. G., & Keller, T. H. (2006a). Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. *Bioorg Med Chem Lett, 16*(1), 40-43. doi: 10.1016/j.bmcl.2005.09.049

Yin, Z., Patel, S. J., Wang, W. L., Wang, G., Chan, W. L., Rao, K. R., Alam, J., Jeyaraj, D. A., Ngew, X., Patel, V., Beer, D., Lim, S. P., Vasudevan, S. G., & Keller, T. H. (2006b). Peptide inhibitors of Dengue virus NS3 protease. Part 1: Warhead. *Bioorg Med Chem Lett, 16*(1), 36-39. doi: 10.1016/j.bmcl.2005.09.062

Yusof, R., Clum, S., Wetzel, M., Murthy, H. M., & Padmanabhan, R. (2000). Purified NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro. *J Biol Chem, 275*(14), 9963-9969.

Zhang, L., Mohan, P. M., & Padmanabhan, R. (1992). Processing and localization of Dengue virus type 2 polyprotein precursor NS3-NS4A-NS4B-NS5. *J Virol, 66*(12), 7549-7554.

Zhang, Y. M., Hayes, E. P., McCarty, T. C., Dubois, D. R., Summers, P. L., Eckels, K. H., Chanock, R. M., & Lai, C. J. (1988). Immunization of mice with dengue structural proteins and nonstructural protein NS1 expressed by baculovirus recombinant induces resistance to dengue virus encephalitis. *J Virol, 62*(8), 3027-3031.

Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., Guo, Y., Bernard, K. A., Shi, P. Y., & Li, H. (2007). Structure and function of flavivirus NS5 methyltransferase. *J Virol, 81*(8), 3891-3903. doi: 10.1128/JVI.02704-06

APPENDICES

Publication and Presentation

Poster Presentation:

Choon Han Heh, Rozana Othman and Noorsaadah Abd. Rahman (2008) Homology modelling of Dengue Type 1, 3 and 4 NS2B-NS3 virus proteases based on Dengue Type 2 NS2B-NS3 virus protease. The Second International Conference on Dengue and Dengue Haemorrhagic Fever, Phuket, Thailand.

Oral Presentation:

C. H. Heh, R. Othman, N. Abd. Rahman, M. J. C. Buckle and R. Yusof (2010) Virtual screening of chalcones and flavanones into Dengue Virus Type 2 NS2B-NS3 protease using AutoDock 4.2. The 6th Mathematics and Physical Sciences Graduate Congress 2010, Faculty of Science, University of Malaya.

Journal:

1.   Eng-Chong, T., Yean-Kee, L., Chin-Fei, C., Choon-Han, H., Sher-Ming, W., Li-Ping, C. T., . . . Yusof, R. (2012). Boesenbergia rotunda: From Ethnomedicine to Drug Discovery. *Evidence-Based Complementary and Alternative Medicine, 2012, 25. doi: 10.1155/2012/473637 (ISI-Cited Publication)*

2.   Rothan, H., Han, H. C., Ramasamy, T. S., Othman, S., Rahman, N. A., & Yusof, R. (2012). Inhibition of dengue NS2B-NS3 protease and viral replication in Vero cells by recombinant retrocyclin-1. *BMC Infect Dis, 12(1), 314. (ISI-Cited Publication)*

3.	Heh, C. H., Othman, R., Buckle, M. J. C., Sharifuddin, Y., Yusof, R., & Rahman, N. A. (2013). Rational Discovery of Dengue Type 2 Non-competitive Inhibitors. *Chemical Biology & Drug Design, n/a-n/a. doi: 10.1111/cbdd.12122 (Accepted ISI-Cited Publication)*

HOMOLOGY MODELING OF DENGUE TYPE 1, 3 AND 4 NS2B-NS3 VIRUS PROTEASES BASED ON DENGUE TYPE 2 NS2B-NS3 VIRUS PROTEASE

Choon Han, Heh[1], Rozana, Othamn[1], Noorsaadah and Abdul Rahman[1]

[1]Faculty of Science, Department of Chemistry, University of Malaya, Kuala Lumpur 52100, Malaysia

BACKGROUND: Dengue, including dengue fever, dengue haemorrhagic fever and dengue shock syndrome is among the major causes of morbidity and mortality in children in many endemic Asian and South American countries. The viruses that cause dengue are members of the Flaviviridae family and they can be categorized into four serotypes, which are Dengue virus type 1, 2, 3 and 4 (DEN-1 to DEN-4). The dengue virus protease (protease complex), NS2B-NS3 was reported required for the cleavage of virus non-structural proteins, and play an important role in viral replication. However, to date, there is only DEN-2 NS2B-NS3 virus protease crystal structure available in the protein data bank.

OBJECTIVES: The objectives of this study are to obtain homology models of DEN-1, 3 and 4, and study the similarity and differences among the models, which could be useful in the development of antiviral agents against dengue virus infections.

METHODS: In this study, the homology models were built using DEN-2 NS2B-NS3 virus protease crystal structure as template with the aid of computer software Modeller 9.3 and the models were then sent for structure verification through UCLA-DOE web server.

RESULTS: The results showed that all the homology models are qualify enough for further usage. For NS2B structures, there were only slight differences among DEN-1, DEN-2 and DEN-3, while significant differences were shown for DEN-4. All the models showed only slight differences for their NS3 structures.

CONCLUSIONS: Since all the catalytic triad regions of the NS3 structures exhibit similarities, inhibitor that binds to catalytic triad in DEN 2 NS3 virus protease is believed to have the same inhibition activities on catalytic triad in DEN 1, DEN 3 and DEN 4 NS3 proteases .

Virtual Screening of Chalcones and Flavanones into Dengue Virus Type 2 NS2B-NS3 Protease Using AutoDock 4.2

C. H. Heh[a], R. Othman[a], N. Abd. Rahman[b], M. J. C. Buckle[a] and R. Yusof[c]

[a]Department of Pharmacy, Faculty of Medicine, Univeristy of Malaya.
[b]Department of Chemistry, Faculty of Science, Univeristy of Malaya.
[c]Department of Molecular Medicine, Faculty of Science, Univeristy of Malaya.

Abstract. Virus protease, NS2B-NS3 of dengue virus type 2 (DEN-2) was reported to be involved in the cleavage of most of the non-structural proteins needed in viral replication. Inhibition of the protease is believed to suppress virus infections. A previous bioassay study showed that cardamonin (a chalcone) and pinostrobin (a flavanone) showed non-competitive inhibition activities towards DEN-2 NS2B-NS3 proteolytic activities. Thus, in this study, virtual screening of a library of chalcones and flavanones, obtained from the Zinc database, against DEN-2 protease was performed using the AutoDock 4.2 program. The results obtained showed that flavanones could be better inhibitors compared to chalcones. Top five flavanones, ranked based on binding-affinity, docked into the same allosteric region as reported by Othman et al. (2008). It can thus be suggested that the binding pocket be used for further development of new anti-dengue agent. Further study is in progress to validate the results obbtained.

Keywords: Dengue virus type 2, DEN-2, NS2B-NS3, Virtual Screening, AutoDock, Chalcones, Flavanones.

Bash 4.1 software scripts for high-throughput analysis of virtual screening results:

```
for d in *2FOM;

do echo $d;

cd $d;

csplit -k -s -n 3 -f "$d"- "$d".dlg '%^MODEL%' '/^MODEL/' '{*}';

    for z in $d-*;

    do echo $z;

    filename="$z".pdb;

    echo $filename;

    mv $z $filename;

    done;

cd ..;

done;
```

```
echo "Group | Compound | Runs | Rank | NumCl | BEnergy | BEmean | Ki |
Atoms | Torsions" > newsummary.txt;

for d in *;

do echo $d;

cd $d;

cd 2FOM;

    for z in *-2FOM;

    do echo $z;

    cd $z;

    co=0;

    count=1;

        for f in *-???.pdb;

        do echo $f;

        echo $count;

        n=$(grep Number $f);

        cn=${n//[!0-9]/};

            if [ $cn -gt $co ] && [ $count -le 10 ];
```

```
        then co=$cn;

        cf=$f;

        echo $co;

        echo $cf;

        else echo $co;

        echo $cf;

        fi;

    let count++;

    done;

echo $co;

echo $cf;

cr=$(grep "Cluster Rank" $cf);

r=${cr//[!0-9]/};

echo $r;

s=$(grep "Estimated Free" $cf);

l=${s#*=};

v=${l% *};

e=${v//[!0-9.-]/};

echo $e;

runs=$(grep "ga_run" $z.dlg);

runs=${runs//[!0-9]/};

dc=$(grep "distinct" $z.dlg);

dc=${dc%, *};

dc=${dc//[!0-9]/};

grep -- "$e |" $z.dlg > etemp.txt;

emin=$(grep -w "$r |    " etemp.txt);

emin=${emin%$co *};

emin=${emin%| *};

emin=$(Connelly et al., *|);

emin=$(Connelly et al., *|);

emin=$(Connelly et al., *|);

emin=$(echo $emin | tr " " "\n" | awk NR==1);

ki=$(grep "Ki" $cf);
```

```
    ki=${ki#*=};

    ki=${ki% *};

    ki=${ki% *};

    ki=${ki% *};

    ki=${ki% *};

    ki=${ki% *};

    ki=${ki% *};

    atms=$(grep "Total number of atoms" $z.dlg);

    atms=${atms//[!0-9]/};

    tors=$(grep "Rotatable Bonds" $z.dlg);

    tors=${tors//[!0-9]/}; group=$(echo $d | sed 's/-d//g');

    name=$(echo $z | sed 's/-dock//g'); echo $group "|" $name "|" $runs
    "|" $r "|" $co "|" $e "|" $emin "|" $ki "|" $atms "|"
    $tors >> ../../../newsummary.txt;

    cd ..;

    done;

cd ../..;

done;

echo "Group | Compound | Runs | Rank | NumCl | BEnergy | BEmean | Ki |
Atoms | Torsions" > newsummary.sort;

cat newsummary.txt | sort -k7n -t"|" >> newsummary.sort;
```

##Script for extracting the group, name, NumCl, *ΔG,* rank and coordinates
of the conformation with largest NumCl along with the macromolecule
coordinates as preparation file for running Ligplot:

```
for d in *;

do echo $d;

cd $d;

cd 2FOM;

    for z in *-2FOM;

    do echo $z;

    cd $z;

    rm *lelc.pdbqt;

    co=0;

        for f in *-???.pdb;

        do echo $f;
```

```
        n=$(grep Number $f);

        cn=${n//[!0-9]/};

            if [ $cn -gt $co ];

            then co=$cn;

            cf=$f;

            echo $co;

            echo $cf;

            else echo $co;

            echo $cf;

            fi;

        done;

    echo $co;

    echo $cf;

    cr=$(grep "Cluster Rank" $cf);

    r=${cr//[!0-9]/};

    echo $r;

    s=$(grep "Estimated Free" $cf);

    l=${s#*=};

    v=${l% *};

    e=${v//[!0-9.-]/};

    echo $e;

    group=$(echo $d | sed 's/-d//g');

    name=$(echo $z | sed 's/-dock//g');

    grep ATOM "$cf" > "$name"_"$r"_"$e"_"$co"_hetatmlelc.pdb;

    grep TER "$cf" >> "$name"_"$r"_"$e"_"$co"_hetatmlelc.pdb;

    sed -i 's/ATOM  /HETATM/g' "$name"_"$r"_"$e"_"$co"_hetatmlelc.pdb;

    sed -i 's/<1> d     /NEW D 180/g'
    "$name"_"$r"_"$e"_"$co"_hetatmlelc.pdb;

    cp "$z".pdbqt "$group"_"$name"_"$r"_"$e"_"$co"_lelc.pdbqt;

    grep HETATM "$name"_"$r"_"$e"_"$co"_hetatmlelc.pdb >>
    "$group"_"$name"_"$r"_"$e"_"$co"_lelc.pdbqt;

    grep TER "$name"_"$r"_"$e"_"$co"_hetatmlelc.pdb >>
    "$group"_"$name"_"$r"_"$e"_"$co"_lelc.pdbqt;

    cd ..;

done;
```

```
cd ../..;

done;
```

Script for copying all the Ligplot preparation files into a another new directory "newlelc" for the ease of access:

```
mkdir newlelc;

for d in *;

do echo $d;

cd $d;

cd 2FOM;

    for f in *-2FOM;

    do echo $f;

    cd ../../newlelc;

    mkdir $f;

    cd ../$d/2FOM;

    cd $f;

        for z in *_lelc.pdbqt;

        do cp $z ../../../newlelc/$f;

        done;

    cd ..;

    done;

cd ../..;

done;
```

Script for running sequential Ligplot for all the compounds in newlelc directory:

```
cd newlelc;

for d in *2fom-dock;

do echo $d;

cd $d;

ligplot *.pdbqt NEW 180 NEW 180 D;

cd ..;

done;
```

```
cd ..;
```

Script for extracting hydrogen bonding interactions data from Ligplot result:

```
echo Chain "|" ResName "|" ResID "|" Atom "|" FullID "|" A-D Distance >
hb-acceptor.txt;

echo Chain "|" ResName "|" ResID "|" Atom "|" FullID "|" D-A Distance >
hb-donor.txt;

cd newlelc;

for f in *;

do echo $f;

cd $f;

n=$(grep -c "NEW" ligplot.hhb);

        for d in $(seq 1 $n);

        do grep "NEW" ligplot.hhb | awk NR==$d > lighhb$d.log;

        done;

                for b in lighhb*;

                do echo $b;

                s=$(grep "NEW" $b);

                res1=$(echo $s | tr " " "\n"| awk NR==1);

                id1=$(echo $s | tr " " "\n"| awk NR==3);

                atm1=$(echo $s | tr " " "\n"| awk NR==4);

                ch1=$(echo $s | tr " " "\n"| awk NR==2);

                res2=$(echo $s | tr " " "\n"| awk NR==5);

                id2=$(echo $s | tr " " "\n"| awk NR==7);

                atm2=$(echo $s | tr " " "\n"| awk NR==8);

                ch2=$(echo $s | tr " " "\n"| awk NR==6);

                dis=$(echo $s | tr " " "\n" | awk NR==9);

                s=${s%% *};

                echo $s;

                        if [ $s == NEW ];

                then acp=$(echo $ch2 "|" $res2 "|" $id2 "|" $atm2 "|" $ch2-
                $res2-$id2-$atm2 "|" $dis);

                echo $acp >> ../../hb-acceptor.txt; else don=$(echo $ch1 "|"
                $res1 "|" $id1 "|" $atm1 "|" $ch1-$res1-$id1-$atm1 "|" $dis);

                echo $don >> ../../hb-donor.txt;
```

```
            fi;

         done;

       cd ..;

       done;

cd ..;

cat hb-acceptor.txt | sort -k3n -t"|" > hb-acceptor.presort;

echo Chain "|" ResName "|" ResID "|" Atom "|" FullID "|" A-D Distance >
hb-acceptor.sort;

grep    ^A    hb-acceptor.presort    >>    hb-acceptor.sort;    grep    ^B    hb-
acceptor.presort >> hb-acceptor.sort;

cat hb-donor.txt | sort -k3n -t"|" > hb-donor.presort; echo Chain "|"
ResName "|" ResID "|" Atom "|" FullID "|" A-D Distance > hb-donor.sort;

grep ^A hb-donor.presort >> hb-donor.sort; grep ^B hb-donor.presort >> hb-
donor.sort;
```

Script for extracting hydrophobic interactions data from Ligplot result:

```
echo Chain "|" ResName "|" ResID "|" Atom "|" FullID "|" Distance >
phobic.txt;

cd new1st;

for f in *;

do echo $f;

cd $f;

n=$(grep -c "NEW" ligplot.nnb);

      for d in $(seq 1 $n);

   do grep "NEW" ligplot.nnb | awk NR==$d > lignnb$d.log;

   done;

      for b in lignnb*;

      do echo $b; s=$(grep "NEW" $b);

      res=$(echo $s | tr " " "\n"| awk NR==5);

      id=$(echo $s | tr " " "\n"| awk NR==7);

      atm=$(echo $s | tr " " "\n"| awk NR==8);

      ch=$(echo $s | tr " " "\n"| awk NR==6);

      dis=$(echo $s | tr " " "\n" | awk NR==9);

       echo $ch "|" $res "|" $id "|" $atm "|" $ch-$res-$id-$atm "|"
       $dis >> ../../phobic.txt;

       done;
```

```
        cd ..;

        done;
cd ..;

cat phobic.txt | sort -k3n -t"|" > phobic.presort;

echo Chain "|" ResName "|" ResID "|" Atom "|" FullID "|" A-D Distance >
phobic.sort;

grep ^A phobic.presort >> phobic.sort; grep ^B phobic.presort >>
phobic.sort;



##Script for hydrogen bond acceptor interactions analysis:

file="acceptor-analysis.txt";

echo ResFullID "|" ResFreq "|" Atom "|" AtomFeq > $file;

x=hb-acceptor.sort;

c=$(grep -c "|" $x);

molo= ;

t=1;

atmo= ;

a=1;

for n in $(seq 2 $c);

do ch=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==1);

res=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==3);

id=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==7);

moln=$ch-$res-$id;

    if [[ "$moln" == "$molo" || "$molo" == "" ]] && [[ "$atmo" == "$atm"
    || "$atmo" == "" ]];

    then molo=$moln;

    atmo=$atm;

    echo $molo $t $atmo $a;

    let t++ a++;

    elif [ "$moln" == "$molo" ] && [ "$atmo" != "$atm" ];

    then tn=$(($t-1));

    an=$(($a-1));

    echo $molo $tn $atmo $an;

    echo $molo "|" $tn "|" $atmo "|" $an >> $file;
```

```
        a=1;

        atmo=$atm;

        echo $molo $t $atmo $a;

        let t++ a++;

        else tn=$(($t-1));

        an=$(($a-1));

        echo $molo $tn $atmo $an;

        echo $molo "|" $tn "|" $atmo "|" $an >> $file;

        t=1;

        a=1;

        molo=$moln;

        atmo=$atm;

        echo $molo $t $atmo $a;

        let t++ a++;

        fi;

done;

tn=$(($t-1));

an=$(($a-1));

echo $molo $tn $atmo $an;

echo $molo "|" $tn "|" $atmo "|" $an >> $file;

cat $file | sort -k3 -t"|" > atm-acceptor-analysis.sort;


##Script for hydrogen bond donor interactions analysis:

file="donor-analysis.txt";

echo ResFullID "|" ResFreq "|" Atom "|" AtomFeq > $file;

x=hb-donor.sort;

c=$(grep -c "|" $x);

molo= ;

t=1;

atmo= ;

a=1;

for n in $(seq 2 $c);

do ch=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==1);
```

```
res=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==3);

id=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==7);

moln=$ch-$res-$id;

    if [[ "$moln" == "$molo" || "$molo" == "" ]] && [[ "$atmo" == "$atm"
    || "$atmo" == "" ]];

    then molo=$moln;

    atmo=$atm;

    echo $molo $t $atmo $a;

    let t++ a++;

    elif [ "$moln" == "$molo" ] && [ "$atmo" != "$atm" ];

    then tn=$(($t-1));

    an=$(($a-1));

    echo $molo $tn $atmo $an;

    echo $molo "|" $tn "|" $atmo "|" $an >> $file;

    a=1;

    atmo=$atm;

    echo $molo $t $atmo $a;

    let t++ a++;

    else tn=$(($t-1));

    an=$(($a-1));

    echo $molo $tn $atmo $an;

    echo $molo "|" $tn "|" $atmo "|" $an >> $file;

    t=1;

    a=1;

    molo=$moln;

    atmo=$atm;

    echo $molo $t $atmo $a;

    let t++ a++;

    fi;

done;

tn=$(($t-1));

an=$(($a-1));

echo $molo $tn $atmo $an;
```

```
echo $molo "|" $tn "|" $atmo "|" $an >> $file;

cat $file | sort -k3 -t"|" > atm-donor-analysis.sort;
```

##Script for hydrophobic interactions analysis:

```
file="phobic-analysis.txt";

echo ResFullID "|" ResFreq "|" Atom "|" AtomFeq > $file;

x=phobic.sort;

c=$(grep -c "|" $x);

molo= ;

t=1;

atmo= ;

a=1;

for n in $(seq 2 $c);

do ch=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==1);

res=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==3);

id=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==7);

moln=$ch-$res-$id;

    if [[ "$moln" == "$molo" || "$molo" == "" ]] && [[ "$atmo" == "$atm"
    || "$atmo" == "" ]];

    then molo=$moln;

    atmo=$atm;

    echo $molo $t $atmo $a;

    let t++ a++;

    elif [ "$moln" == "$molo" ] && [ "$atmo" != "$atm" ];

    then tn=$(($t-1));

    an=$(($a-1));

    echo $molo $tn $atmo $an;

    echo $molo "|" $tn "|" $atmo "|" $an >> $file;

    a=1;

    atmo=$atm;

    echo $molo $t $atmo $a;

    let t++ a++;
```

```
    else tn=$(($t-1));

    an=$(($a-1));

    echo $molo $tn $atmo $an;

    echo $molo "|" $tn "|" $atmo "|" $an >> $file;

    t=1;

    a=1;

    molo=$moln;

    atmo=$atm;

    echo $molo $t $atmo $a;

    let t++ a++;

    fi;

done;

tn=$(($t-1));

an=$(($a-1));

echo $molo $tn $atmo $an;

echo $molo "|" $tn "|" $atmo "|" $an >> $file;

cat $file | sort -k3 -t"|" > atm-phobic-analysis.sort;
```

##Script for generating interaction frequency for hydrogen bond acceptor interactions:

```
file="freq-hba.txt";

echo -n "ResFullID | ResFreq " > $file;

x=atm-acceptor-analysis.sort;

c=$(grep -c "|" $x);

for n in $(seq 2 $c);

do atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

echo $atm;

num=$(cat $file | awk NR==1 | tr " " "\n" | awk 'END { print NR }');

numo=1;

a=0;

    while [ $numo -le $num ];

    do echo $numo;

    temp=$(cat $file | awk NR==1 | tr " " "\n" | awk NR==$numo);

    echo $temp;
```

```
                    if [ "$atm" != "$temp" ];

                    then let a++;

                    elif [ "$atm" == "$temp" ];

                    then let a=a+100;

                    fi;

           let numo=numo+2;

           done;

           if [ $a -lt 100 ];

           then echo yeah;

           echo -n "| $atm "  >> $file;

           fi;

done;

file="freq-hba.txt";

x=acceptor-analysis.txt;

c=$(grep -c "|" $x);

resx=;

for n in $(seq 2 $c);

do res=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==1);

atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

echo $atm $res;

    if [ "$res" != "$resx" ];

    then resx=$res;

    echo $resx;

    echo -ne "\n$resx | xxx " >> $file;

    num=$(cat $file | awk NR==1 | tr " " "\n" | awk 'END { print NR }');

    numo=5;

        while [ $numo -le $num ];

        do atmx=$(cat $file | awk NR==1 | tr " " "\n" | awk NR==$numo);

        check=$(cat $x | grep $resx | grep -w $atmx);

        if [ "$check" == "" ];

        then freq=0;

        else freq=$(cat $x | grep $resx | grep -w $atmx | tr " " "\n" | awk
        NR==7);

        fi;
```

174

```
        echo -n "| $freq " >> $file;

        let numo=numo+2;

        done;

    line=$(grep -c "|" $file);

    numx=$(cat $file | awk NR==$line | tr " " "\n" | awk 'END { print
    NR }');

    numy=5; tfreq=0;

        while [ $numy -le $numx ];

        do atmfreq=$(cat $file | awk NR==$line | tr " " "\n" | awk
        NR==$numy);

        echo $atmfreq;

        let tfreq=tfreq+atmfreq;

        echo $tfreq;

        let numy=numy+2;

        done;

    echo $tfreq;

    sed -i s/xxx/"$tfreq"/g $file;

    else echo "done already";

    fi;

done;

grep ^R freq-hba.txt > freq-hba.sort;

grep ^A freq-hba.txt >> freq-hba.sort;

grep ^B freq-hba.txt >> freq-hba.sort;
```

##Script for generating interaction frequency for hydrogen bond donor interactions:

```
file="freq-hbd.txt";

echo -n "ResFullID | ResFreq " > $file;

x=atm-donor-analysis.sort;

c=$(grep -c "|" $x);

for n in $(seq 2 $c);

do atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

echo $atm;

num=$(cat $file | awk NR==1 | tr " " "\n" | awk 'END { print NR }');
```

```
numo=1;

a=0;

    while [ $numo -le $num ];

    do echo $numo;

    temp=$(cat $file | awk NR==1 | tr " " "\n" | awk NR==$numo);

    echo $temp;

        if [ "$atm" != "$temp" ];

        then let a++;

        elif [ "$atm" == "$temp" ];

        then let a=a+100;

        fi;

    let numo=numo+2;

    done;

    if [ $a -lt 100 ];

    then echo yeah;

    echo -n "| $atm "  >> $file;

    fi;

done;

file="freq-hbd.txt";

x=donor-analysis.txt;

c=$(grep -c "|" $x);

resx=;

for n in $(seq 2 $c);

do res=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==1);

atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

echo $atm $res;

    if [ "$res" != "$resx" ];

    then resx=$res;

    echo $resx;

    echo -ne "\n$resx | xxx " >> $file;

    num=$(cat $file | awk NR==1 | tr " " "\n" | awk 'END { print NR }');

    numo=5;

        while [ $numo -le $num ];

        do atmx=$(cat $file | awk NR==1 | tr " " "\n" | awk NR==$numo);
```

```
        check=$(cat $x | grep $resx | grep -w $atmx);

        if [ "$check" == "" ]; then freq=0;

        else freq=$(cat $x | grep $resx | grep -w $atmx | tr " " "\n" | awk
        NR==7);

        fi;

        echo -n "| $freq " >> $file;

        let numo=numo+2;

        done;

    line=$(grep -c "|" $file);

    numx=$(cat $file | awk NR==$line | tr " " "\n" | awk 'END { print
    NR }');

    numy=5;

    tfreq=0;

        while [ $numy -le $numx ];

        do atmfreq=$(cat $file | awk NR==$line | tr " " "\n" | awk
        NR==$numy);

        echo $atmfreq;

        let tfreq=tfreq+atmfreq;

        echo $tfreq;

        let numy=numy+2;

        done;

    echo $tfreq;

    sed -i s/xxx/"$tfreq"/g $file;

    else echo "done already";

    fi;

done;

grep ^R freq-hbd.txt > freq-hbd.sort;

grep ^A freq-hbd.txt >> freq-hbd.sort;

grep ^B freq-hbd.txt >> freq-hbd.sort;


##Script for generating interaction frequency for hydrophobic interactions:

file="freq-phobic.txt";

echo -n "ResFullID | ResFreq " > $file;

x=atm-phobic-analysis.sort;
```

```
c=$(grep -c "|" $x);

for n in $(seq 2 $c);

do atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

echo $atm;

num=$(cat $file | awk NR==1 | tr " " "\n" | awk 'END { print NR }');

numo=1;

a=0;

    while [ $numo -le $num ];

    do echo $numo;

    temp=$(cat $file | awk NR==1 | tr " " "\n" | awk NR==$numo);

    echo $temp;

        if [ "$atm" != "$temp" ]; then let a++; elif [ "$atm" == "$temp" ];

        then let a=a+100;

        fi;

    let numo=numo+2;

    done;

    if [ $a -lt 100 ];

    then echo yeah; echo -n "| $atm "  >> $file;

    fi;

done;

file="freq-phobic.txt";

x=phobic-analysis.txt;

c=$(grep -c "|" $x);

resx=;

for n in $(seq 2 $c);

do res=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==1);

atm=$(cat $x | awk NR==$n | tr " " "\n" | awk NR==5);

echo $atm $res;

    if [ "$res" != "$resx" ];

    then resx=$res;

    echo $resx;

    echo -ne "\n$resx | xxx " >> $file;

    num=$(cat $file | awk NR==1 | tr " " "\n" | awk 'END { print NR }');

    numo=5;
```

```
        while [ $numo -le $num ];

        do atmx=$(cat $file | awk NR==1 | tr " " "\n" | awk NR==$numo);

        check=$(cat $x | grep $resx | grep -w $atmx);

            if [ "$check" == "" ];

            then freq=0;

            else freq=$(cat $x | grep $resx | grep -w $atmx | tr " " "\n" |
            awk NR==7);

            fi;

        echo -n "| $freq " >> $file;

        let numo=numo+2;

        done;

    line=$(grep -c "|" $file);

    numx=$(cat $file | awk NR==$line | tr " " "\n" | awk 'END { print
    NR }');

    numy=5;

    tfreq=0;

        while [ $numy -le $numx ];

        do atmfreq=$(cat $file | awk NR==$line | tr " " "\n" | awk
        NR==$numy);

        echo $atmfreq;

        let tfreq=tfreq+atmfreq;

        echo $tfreq;

        let numy=numy+2;

        done;

    echo $tfreq;

    sed -i s/xxx/"$tfreq"/g $file;

    else echo "done already";

    fi;

done;

grep ^R freq-phobic.txt > freq-phobic.sort;

grep ^A freq-phobic.txt >> freq-phobic.sort;

grep ^B freq-phobic.txt >> freq-phobic.sort;
```

An example of clustering histogram from docking result of *S*-pinostrobin docked towards 2FOM by AutoDock 4.2 software. All conformations were chosen for further analysis.

```
_____|_____|_____|_____|_____|_____
       |          |     |          |     |
Clus   | Lowest   | Run | Mean     | Num | Histogram
-ter   | Binding  |     | Binding  | in  |
Rank   | Energy   |     | Energy   | Clus|    5   10   15   20   25   30   35
_____|_____|_____|_____|_____|___:____|____:____|____:____|____:____|____:___
   1  |    -7.60 |  91 |    -7.58 |  63
|############################################################
   2  |    -7.32 | 100 |    -7.28 |  12 |############
   3  |    -7.18 |  28 |    -7.18 |   1 |#
   4  |    -6.50 |  63 |    -6.49 |   6 |######
   5  |    -6.42 |  24 |    -6.39 |   7 |#######
   6  |    -6.41 |  58 |    -6.34 |   2 |##
   7  |    -6.28 |   1 |    -6.26 |   2 |##
   8  |    -6.20 |  42 |    -6.20 |   1 |#
   9  |    -6.07 |  95 |    -6.07 |   1 |#
  10  |    -5.95 |  15 |    -5.91 |   2 |##
  11  |    -5.86 |  61 |    -5.86 |   1 |#
  12  |    -5.76 |  62 |    -5.76 |   1 |#
  13  |    -5.76 |  40 |    -5.76 |   1 |#
_____|_____|_____|_____|_____|_____
```

An example of the calculation of estimated free energy of binding from docking result (towards 2FOM) for one of the *S*-pinostrobin's conformation. Its atomic coordinates are coloured in blue.

```
MODEL        91
USER     Run = 91
USER     Cluster Rank = 1
USER     Number of conformations in this cluster = 63
USER
USER     RMSD from reference structure       = 8.698 A
USER
USER     Estimated Free Energy of Binding    = -7.60 kcal/mol   [=(1)+(2)+(3)-(4)]
USER     Estimated Inhibition Constant, Ki   = 2.70 uM (micromolar) [Temperature = 298.15
K]
USER
USER     (1) Final Intermolecular Energy     = -8.48 kcal/mol
USER         vdW + Hbond + desolv Energy     = -8.40 kcal/mol
USER         Electrostatic Energy            = -0.08 kcal/mol
USER     (2) Final Total Internal Energy     = -0.30 kcal/mol
USER     (3) Torsional Free Energy           = +0.89 kcal/mol
USER     (4) Unbound System's Energy         = -0.28 kcal/mol
USER
USER
USER
USER     DPF = spinos_2fom-dock.dpf
USER     NEWDPF move    spinos.pdbqt
USER     NEWDPF about   -1.782400 -6.455800 -0.583200
USER     NEWDPF tran0   -5.059249 -13.273567 6.082521
USER     NEWDPF axisangle0    0.125805 0.295535 -0.947012 110.017891
USER     NEWDPF quaternion0   0.103065 0.242115 -0.775832 0.573449
USER     NEWDPF dihe0  -121.94 173.96 -74.96
USER
USER                              x        y        z    vdW    Elec        q     RMS
ATOM     1  C7_ <1> d         -3.983 -13.265   4.869 -0.36 -0.04   +0.169   8.698
ATOM     2  C8_ <1> d         -3.348 -14.177   5.952 -0.47 -0.02   +0.121   8.698
ATOM     3  C9_ <1> d         -3.647 -13.583   7.307 -0.40 -0.02   +0.167   8.698
ATOM     4  C10 <1> d         -5.004 -13.053   7.546 -0.43 -0.01   +0.076   8.698
ATOM     5  C11 <1> d         -5.805 -12.804   6.413 -0.35 -0.02   +0.089   8.698
ATOM     6  O12 <1> d         -5.337 -13.053   5.158 -0.17 +0.08   -0.340   8.698
ATOM     7  O13 <1> d         -2.708 -13.280   8.044 -0.28 +0.05   -0.292   8.698
ATOM     8  C14 <1> d         -5.529 -12.801   8.831 -0.46 -0.02   +0.085   8.698
ATOM     9  C15 <1> d         -6.834 -12.291   8.964 -0.48 -0.03   +0.074   8.698
ATOM    10  C16 <1> d         -7.634 -12.042   7.835 -0.43 -0.02   +0.073   8.698
ATOM    11  C17 <1> d         -7.112 -12.310   6.559 -0.34 -0.02   +0.074   8.698
ATOM    12  C5_ <1> d         -3.831 -13.868   3.489 -0.38 +0.01   -0.029   8.698
ATOM    13  C4_ <1> d         -4.206 -13.129   2.350 -0.24 -0.00   +0.010   8.698
ATOM    14  C3_ <1> d         -4.036 -13.667   1.064 -0.25 -0.00   +0.001   8.698
ATOM    15  C1_ <1> d         -3.480 -14.945   0.903 -0.41 -0.00   +0.000   8.698
ATOM    16  C2_ <1> d         -3.088 -15.683   2.032 -0.53 -0.00   +0.001   8.698
ATOM    17  C6_ <1> d         -3.257 -15.144   3.318 -0.58 -0.00   +0.010   8.698
ATOM    18  O18 <1> d         -4.855 -13.053   9.982 -0.60 +0.13   -0.360   8.698
ATOM    19  _H19 <1> d        -5.218 -13.551  10.689 -0.24 -0.17   +0.217   8.698
ATOM    20  O20 <1> d         -8.903 -11.557   7.907 -0.56 +0.05   -0.356   8.698
ATOM    21  C21 <1> d         -9.710 -11.181   6.832 -0.44 -0.04   +0.210   8.698
TER
ENDMDL
```

An example of graphical result generated by Ligplot 4.5.3 software. It illustrates the hydrogen bond and hydrophobic interactions between the *S*-pinostrobin's conformation and amino acid residues of 2FOM. New 180(D) is the temporary code name for ligand *S*-pinostrobin.
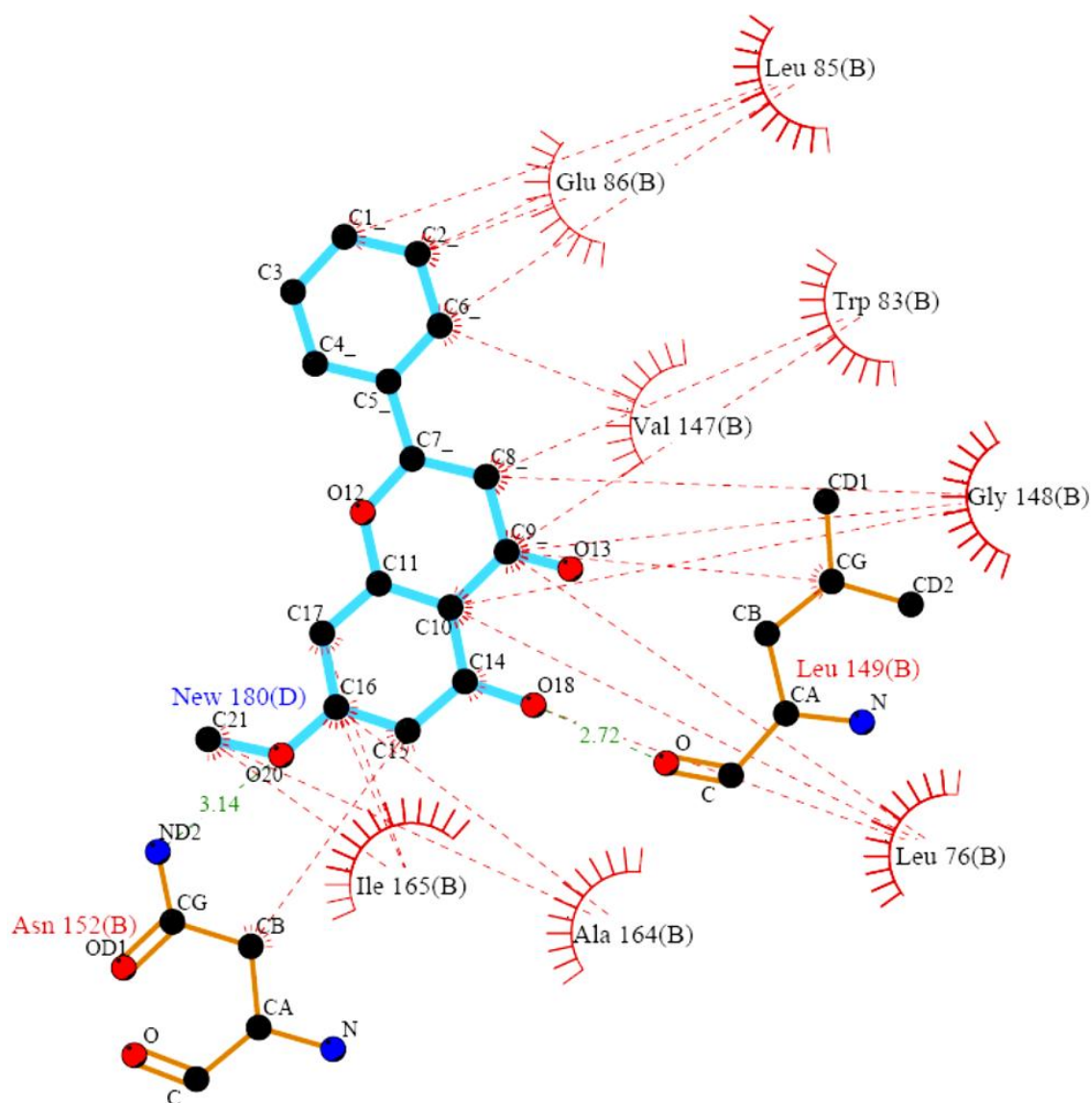
The text version of hydrogen bond and hydrophobic interactions between the *S*-pinostrobin's conformation and amino acid residues of 2FOM. New 180(D) is the temporary code name for ligand *S*-pinostrobin.

```
Hydrogen Bond Interactions:

Donor                  Acceptor               Distance
ASN B  152   ND2       NEW D  180   O20       3.14
NEW D  180   O18       LEU B  149   O         2.72

Hydrophobic Interactions:

Atom 1                 Atom 2              Distance
NEW D  180   C21       ILE B  165   C       3.71
NEW D  180   C17       ILE B  165   C       3.63
NEW D  180   C16       ILE B  165   C       3.87
NEW D  180   C16       ILE B  165   CA      3.81
NEW D  180   C21       ALA B  164   CB      3.84
NEW D  180   C16       ALA B  164   C       3.77
NEW D  180   C15       ASN B  152   CB      3.74
NEW D  180   C9_       LEU B  149   CG      3.65
NEW D  180   C9_       GLY B  148   C       3.56
NEW D  180   C8_       GLY B  148   C       3.80
NEW D  180   C10       GLY B  148   CA      3.62
NEW D  180   C9_       GLY B  148   CA      3.29
NEW D  180   C8_       GLY B  148   CA      3.14
NEW D  180   C6_       VAL B  147   C       3.86
NEW D  180   C2_       GLU B   86   C       3.85
NEW D  180   C6_       LEU B   85   CB      3.69
NEW D  180   C2_       LEU B   85   CB      3.79
NEW D  180   C2_       LEU B   85   C       3.32
NEW D  180   C1_       LEU B   85   C       3.78
NEW D  180   C9_       TRP B   83   CZ2     3.69
NEW D  180   C8_       TRP B   83   CZ2     3.62
NEW D  180   C14       LEU B   76   CD2     3.40
NEW D  180   C10       LEU B   76   CD2     3.50
NEW D  180   C9_       LEU B   76   CD2     3.56
NEW D  180   C10       LEU B   76   CD1     3.90
```