CHAPTER 1
INTRODUCTION

1.1 Dengue Fever (DF) and Dengue Haemorrhagic Fever (DHF)

1.1.1 Symptoms and prevalence

Dengue is a mosquito-borne viral infectious disease that has become a major public health concern in the recent years. Human infected by this disease will exhibit fever of 3 to 5 days, intense headache, myalgia, anthralgic retro-orbital pain, anorexia, GI disturbances, rash and leucopenia as symptoms.

Dengue is usually found in tropical and sub-tropical regions around the world, mainly in urban and semi-urban areas. In the recent years, dengue become as a serious disease that is endemic in over 100 countries (Figure 1.1), with more than 2.5 billion people at risk for epidemic transmission (Gubler, 1996). About 100 million cases of Dengue Fever (DF) and 500,000 cases of Dengue Haemorrhagic Fever (DHF) have been reported globally with this figure is still rising (http://www.searo.who.int/en/Section10/Section332_1103htm, 14 August 2006).

Dengue Haemorrhagic Fever (DHF) epidemic was first reported in 1953 in the Philippines. This disease has greatly expanded to most Asian countries and has become the top ten leading causes of hospitalisation and death among children (WHO, 1997).
In Malaysia, a recent report revealed 38 deaths in the first 10 weeks of 2010 amongst 10,462 cases of dengue fever. In comparison, in 2009 statistics showed 41,486 dengue cases with 88 deaths reported. The death cases in the first 10 weeks of 2010 already comprised 43% of the death cases in 2009. (http://www.moh.gov.my/MohPortal/newsFull.jsp?action=load&id=432, 14 March 2010). 40% of the world's populations are now at risk from dengue without effective treatment, vaccine or drug (Kautner et al., 1997; Monath, 1994).


**Figure 1.1:** World distribution of Dengue in year 2008 (source: http://www.cdc.gov/dengue/resources/Dengue&DFH%20Information%20for%20Health%20Care%20Practitioners_2009.pdf, 15 May 2010)

### 1.1.2 Diagnosis and treatment

Thrombocytopenia and haemoconcentration are the constant indication for a patient infected with dengue virus. A person infected by the dengue virus would typically exhibit a drop of platelet count to below 100 000 per mm$^3$ between the 3$^{rd}$ and 8$^{th}$ day of the fever. A rise in haematocrit level indicating plasma leakage is always observed in the blood stream of DHF patients. Other common observations such as hypoproteinaemia caused by the loss of albumin, hyponatraemia, and increased level of
serum aspartate aminotransferase are also commonly observed. Pleural effusion on the right side of chest is also visible under X-ray in dengue patients. (WHO, 1997)

Patients with dengue infection could only be treated at the early stage of the infection by relieving the symptoms to prevent complications and death. Appropriate, intensive and supportive therapy by maintaining the circulating fluid volume of the patient is given to reduce the mortality to less than 1%. Aspirin and ibuprofen are avoided due to its ability to increase bleeding tendency and stomach pain. Painkillers such as paracetamol are often prescribed on medical advice and patients will be hospitalised immediately after being diagnosed with DHF disease. (http://www.searo.who.int/EN/Section10/Section332/Section1631.htm, 4 August 2006)

1.2 Dengue Virus, the Genome and Lifecycles
1.2.1 Transmission

Dengue fever and dengue haemorrhagic fever are caused by dengue virus of the flavivirus family. There are 4 serotypes of dengue viruses; denoted as DEN1, DEN2 DEN3 and DEN4. The DEN2 serotype is the most prevalent amongst the four serotypes. The recovery from the infection by one serotype will provide lifelong immunity against that particular serotype, but only partial and transient protection against subsequent infections by the other three serotypes. There have been clinical evidences showing that sequential infection will increase the risk of more serious disease resulting in DHF (http://www.searo.who.int/en/Section10/Section332_1103.htm, 14 August 2006).
The main dengue vector is the female mosquito of the species *Aedes aegypti* and *Aedes albopictus*. An infected mosquito can remain infected for life. The virus is transmitted to a person when the mosquito feeds on a dengue patient during the first to fifth days of illness. Following the virus incubation period for 8-10 days in the vector, the virus can then be transmitted by these *Aedes* mosquitoes to susceptible individuals through blood feeding (http://www.who.int/mediacentre/factsheets/fs117/en/, 13 March 2009).

### 1.2.2 Polyprotein processing

Dengue viruses are single-stranded, positive-sense RNA that has approximately 10,723 nucleotides. The genomic RNA has a single open reading frame that encodes a polyprotein of 3,391 amino acids. These amino acids are processed into 3 structural proteins (core protein, C; membrane-associated protein, prM; and envelope protein, E) and seven non-structural proteins; NS1 to NS5 that are further assembled into the virion. The virion is approximately 50 nm in diameter. These proteins are expressed in the infected cells (Irie *et al.*, 1989), as depicted in Figure 1.2.
Figure 1.2: Structural and non-structural polyprotein assembly of DEN2 virus. (a) Schematic representation of dengue virus structure and morphology. (b) Arrangement of viral proteins in the single and positive stranded dengue RNA genome-encoded precursor polyprotein with their respective cleaving enzyme.

The envelope protein, which is a part of structural protein, is responsible for neutralization, fusion and interactions with virus receptors on the target cell (Klasse et al., 1998). Through receptor mediated endocytosis or direct fusion, target cells can be penetrated by the virus. Jahn and co-workers (Jahn et al., 2003) reported that the fusion process of the virus envelope protein could be controlled by specific fusion proteins which are complexed with lipids and other proteins at the fusion site. “Fusion peptide”, a term given to the special segment of a polypeptide chain of fusion proteins, is exposed and inserted into the target cell during the fusion process. There are at least two different classes of structural viral fusion proteins, termed Class I and Class II.

Dengue virus envelope polyprotein, like other flaviviruses and alphaviruses, is classified as Class II fusion protein (Rey et al., 1995). The fusion protein appears as an
internal loop between two β-strands and it is buried under the protein interface. At low pH, the protein is exposed to conformational changes (Rey et al., 1995; Lescare et al., 2001). Modis and co-workers (Modis et al., 2003) have reported the crystallisation of the E protein from Dengue virus type 2 which has a ectodomain soluble fragment (residue 1-394) that is similar to tick-borne encephalitis (TBE) virus. The hydrophobic ligand lined by the residues in this dimeric crystal structureis reported to affect the pH threshold for fusion (Modis et al., 2003). In the crystal structure of the E protein, the detergent, n-octyl-β-D-glucoside used in the crystallisation process was found to be in its binding pocket which indicated the hydrophobic nature of the binding site (Modis et al., 2003). This information could then be used to aid in the design of small molecules that will fit into the binding pocket of the E protein which may inhibit the fusion process and subsequently, inhibit the entry of the pathogenic microbes into the host cells.

The non-structural protein, NS1, is a glycoprotein which was found to be important for virus viability. In vitro infection has revealed the translocation of NS1 into endoplasmic reticulum (ER) through a hydrophobic sequence that is localized at the C-terminal of E protein (Falgout and Markoff, 1995). A homodimer of NS1 was formed and interacted with membrane in the ER (Winkler et al., 1989). A fraction of NS1 protein was found to be enrolled in the early stage of viral replication by association with intracellular organelles (Mackenzie et al., 1996; Muylaert et al., 1997). The NS1 protein was also found to be exported along the secretory pathway of the plasma membrane by remaining anchored to the glycophsphosphatidylinositol group (Jacobs et al., 1992) or as a soluble hexamer (sNS1) (Flamand et al., 1995; Crooks, 1994).
The NS3 of DEN2 has multifunctional protein fragments that contain serine protease, NS2B/NS3 as well as helicase. Optimal activity of the NS3 serine protease is required for the maturation of the virus. In addition, the presence of the NS2B co-factor is found to be a pre-requisite for the optimal catalytic activity of NS3 (Bianchi and Pessi, 2002 and references therein). Studies has revealed that this second largest protein contained a serine protease catalytic triad within the terminal region of 180 amino acid residues which require 40 amino acid residues of NS2B for protease activity (Falgout et al., 1991; Arias et al. 1993; Jan et al., 1995). The polyprotein precursor processing occurs co-translationally as well as post-translationally and is performed by either the host cell proteases, furin or signalases in association with the membranes of the endoplasmic reticulum or the viral protease. The protease cleaves the viral polyprotein at four junctions, NS2A-NS2B (Arg-Ser), NS2B-NS3 (Arg-Ala), NS3-NS4A (Lys-Ser), and NS4B-NS5 (Arg-Gly) where a pair of dibasic amino acids at the P2 and P1 positions followed by a small, non-branched amino acid (Gly, Ala or Ser) at P1’ was found as the consensus of substrate cleavage motif (see Figure 1.2 for more details) (Yusof et al., 2000 and references therein). In addition, the viral protease has been found to cleave internally within NS2A (Nestorowicz et al., 1994) and NS3 (Falgout et al., 1991). The NS3 residues from 180 to 618 contain conserved motifs that were found in several NTPase and the DEXH family of RNA helicases. The mutational and competitive experiments using ATP and its analog as substrate suggested that both RTPase and NTPase activities share the same active sites (Bartelma and Padmanathan, 2002). Here, the impaired helicase of dengue viruses was unable to replicate, implying the important role of NS3 protein in the flavivirus life cycle. Two crystal forms of dengue virus type 2 NTPase/helicases at 2.4Å and 2.8Å, respectively, were reported by Xu and co-workers (Xu et al., 2005). The crystal structure comprises 3 domains with an asymmetric distribution of charges on its surface and a tunnel that is enough to fit in a
single-stranded RNA. Its catalytic mechanism was assisted by the presence of the
divalent metal ion, when a sulfate ion was found at the NTPase active site. (Xu et al.,
2005).

NS5, the largest non-structural protein of DEN2 (predicted to be of molecular
weight 103-104 kD), is the most highly conserved protein among the flavivirus protein
(Mandl et al. 1988). It was found to have two enzymes, 5” RNA O-methyltransferase
and RNA-dependant RNA polymerase, and that O-methyltransferase is involved in 5’
capping, whereas the latter is involved in viral replication in infected cells. The crystal
structure of NS5 that containing guanyltransferrase /methyltransferase (Egloff et al.,
2002) has been reported. Kapoor and co-workers found NS3 and NS5 proteins to be
present in infected cells as a stable complex (Kapoor et al. 1995), and that NS5 is able
to trigger the NTPase activities of NS3 (Yon et al., 2005). These findings may resemble
the role of heterodimeric NS3-NS5 complex in unwinding double stranded RNA during
replication process (Wahab et al., 2007).

1.3 Serine Proteases

At present, over 155,000 peptidase gene sequences have been classified into 52
clans and 208 families. Based on the MEROPS database, over thirty percent of the
proteolytic enzymes are classified as serine proteases, withmore than 55,000 serine
proteases grouped into 16 clans and 46 families (http://merops.sanger.ac.uk/cgi-bin/
statistics_index?type=P, 15 May 2010). The serine protease was named for its
nucleophilic serine which plays a vital role in the hydrolysis of peptide substrates in the
active sites. In the serine protease, this proteolytic mechanism is distinguished by the
appearance of a catalytic triad that formed a proton shuttling relay (Hedstrom, 2002).
example, the catalytic triad in the serine protease, chymotrypsin, comprises Ser-195, His-57 and Asp-102. The proteolytic mechanism of the serine protease will be described further in the section below.

Serine proteases are widely distributed in all form of cellular life including viral genomes. Chymotrypsin-like proteases that classified under Clan PA are the most abundant serine proteases (Rawlings et al., 2010). There are many important mammalian physiological processes that involve chymotrypsin-like proteases, such as digestion, hemostasis, reproduction, signal transduction, apoptosis, and the immunity responses (Hedstrom, 2002 and references therein).

1.3.1 Dengue Virus NS2B/NS3 Serine Protease

When the NS2B/NS3 serine protease was found to be important in the polyprotein processing in dengue virus, many approaches and efforts were made to understand the mechanism, structure and molecular interaction between the serine protease of NS2B/NS3 complex and its substrates. The minimum domain size required for protease activity of the 69-kD NS3 protein has been mapped to 167 residues at the N terminus (Li et al., 1999). The virus sequence alignments analysed revealed that structural motifs as well as the characteristic catalytic triad (His-Asp-Ser) of mammalian serine proteases are conserved in all flaviviruses (Bazan and Fletterick, 1989; Gorbalenya et al., 1989).

Sequence comparison among the serine proteases and mutational analysis verified that a catalytic triad of NS2B/NS3 comprised of the residues His51, Asp75, and Ser135, and that replacement of the catalytic Ser135 residue by alanine resulted in an
enzymatically inactive NS3 protease (Yan et al., 1998). The presence of a peptide co-factor is essential for optimal catalytic activity of the flaviviral proteases with natural polyprotein substrates (Bartenschlager et al., 1995; Chambers et al., 1991).

Although the dengue virus NS3 protease exhibits NS2B-independent activity with model substrates for serine proteases such as N-α-benzoyl-L-arginine-p-nitroanilide, the enzymatic cleavage of dibasic peptides is markedly enhanced in NS2B/NS3 complex. In addition, the presence of the NS2B co-factor has been shown to be an absolute requirement for trans-cleavage of a cloned polyprotein substrate (Yusof et al, 2000). A genetically engineered NS2B(H)-NS3pro protease containing a non-cleavable nonamer glycine linker between the NS2B activation sequence and the protease moiety exhibited higher specific activity with para-nitroanilide peptide substrates than the NS2B(H)-NS3pro molecule (Leung et al., 2001). The NS2B-NS3pro protease incorporating a full-length NS2B cofactor sequence could catalyze the cleavage of 12-mer peptide substrates representing native polyprotein junctions (Khumthong et al., 2002; Khumthong et al., 2003).

A model of the NS2B/NS3 dengue virus protease was first constructed through homology modeling using the crystal structure of HCV NS3/4A complex as template by Brinkworth and co-workers, with the suggestion that the 40 amino acid residues of dengue virus NS2B co-factor could be reduced to 12 hydrophobic residues (Brinkworth et al., 1999). Experimental data on hepatitis C virus protease showed some structural and mechanistic explanations for the protease activation by its co-factor, where the NS4A co-factor was found to affect the folding of the NS3 protease. This resulted in the conformational rearrangements of the N-terminal 28 residues of the protease and a strand displacement that lead to the formation of a well-ordered array of three β-sheets
with the co-factor as an integral part of the protease fold (Kim et al., 1996; Yan et al., 1998). These conformational changes reorient the residues of the catalytic triad making it more favourable for proton shuttling during proteolytic process.

Mutational analyses revealed the importance of several amino acid residues that are highly conserved among the flaviviruses, where 5 putative substrate binding residues (Asp-129, Phe-130, Tyr-150, Asn-152 and Gly-153) were proposed (Valle and Falgout, 1998). Computer modeling study of a substrate binding at the catalytic triad of the crystal structure of NS3 without its NS2B cofactor revealed that Gly-133 and Ser-135 to be the most likely to form the oxyanion hole (Murthy et al., 1999). Hydrogen bonding interactions have been observed between the main chain of P1 and P2 residues with appropriate main chain atoms of Gly-153 and Asn-152 to generate the short section of β-sheet common in serine protease-inhibitor interactions (Read and James, 1986). Three residues, Ser-131, Tyr-150, and Ser-163, are within the S1 pocket. A serine side chain at P1’ fits into the S1’ pocket formed by the catalytic His-51 and Ser-135 and residues Gly-35, Ile-36, and Val-52. The Oε1 atom of Asn-152 forms a salt bridge/hydrogen bond with Ne of the P2 Arg in the modeled complex (Murthy et al., 1999).

Although the crystal structure of DEN2 NS3 has been reported (Murthy et al., 1999), the absence of NS2B cofactor therein makes the mechanism of proteolytic process activation unclear. The orientation of the carboxyl side chain of Asp-75 away from His-51 in the catalytic triad of NS3 crystals formed an open conformation that may lead to the inefficiency of proteolytic activity.
1.3.2 Mechanism of action

Proteases, or proteinases are enzymes that recognise protein or peptide as their substrates and cleave these substrates by hydrolysing their amide bonds. Classification of the proteases (serine, aspartic, cysteine and metallo protease) is made after their critical amino acid residue used in the hydrolysis process. For dengue virus, the NS2B/NS3 complex is classified as chymotrypsin-like serine protease that has 3 critical amino acid residues (Ser135, His51, and Asp75) in the active site for catalytic process.

Using chymotrypsin as an example, the mechanism of proteolysis by serine protease is illustrated in Figure 1.3 (Murrays et al., 2003). From the carboxyl group of Asp-102, the electron-rich group is transferred through the cyclic amine of His-57 via a hydrogen bond, causing the hydroxyl group of Ser195 to become more nucleophilic and ready for catalytic proteolytic process. When the peptide substrate moves into the active site, the nucleophilic hydroxyl of Ser-195 attacks the scissile electron-deficient carbonyl of the amide bond in the substrate, and subsequently forming the hemiketal tetrahedral intermediate which is stabilised by an oxyanion hole. Proton transfer from the charged His-57 to the substrate caused the amide bond to break and thereafter, releasing the C-terminal product (the amine). An acyl-enzyme complex with the N-terminal product bound covalently with hydroxyl of Ser-195 is also formed. Water then act as a nucleophile to attack the carbonyl of the ester forming the hemiketal tetrahedral intermediate which is stabilised by an oxyanion hole. The breakdown of the hemiketal intermediate was then initiated by the proton transfer from His-57, yielding the N-terminal product (carboxylic acid) and regeneration of catalytic triad that is ready for the next substrate cleavage. All the catalytic process is promoted under acidic or basic conditions.
catalysis of His-57 and Asp-102, with the help of the backbone N-H groups of the Ser-195 and Gly-193 in the oxyanion hole during the formation of hemiketal tetrahedral intermediate.

Figure 1.3: Proteolytic process at the catalytic triad of serine protease

1.4 Approaches towards Dengue Virus Inhibition

1.4.1 Attenuated vaccine

Vaccine development is a very popular and effective method for human to fight against diseases, especially when the disease is endemic. Several developed vaccine such as hepatitis B, rubella, tetanus are shown to be effectively able to control these diseases.

Initiative such as the production of a live-attenuated vaccine in suckling mouse brain (Hotta, 1957; Sabin and Schlesinger, 1945; Schlesinger et al., 1956; Wiseman et
1.4.2 Therapeutic agents: virus inhibitor

HIV serine protease inhibitor was amongst some examples of commercially available medication for human administration (West and Fairlie, 1995). In other examples, two heptapeptides containing amino boronic acid has also been shown by Dunsdon and co-workers (Dunsdon et al., 2000) to inhibit the activities of HCV’s NS3 protease. Such successful cases of inhibiting viral replication by employing inhibitor design based on their related serine proteases has attracted more studies of serine proteases related diseases in order to find the best serine protease inhibitor to inhibit viral infection. It is also important to have the inhibitor that is only selective towards the targeted protease in order to minimize the risk of adverse effects.

1.4.3 Dengue Virus NS2B/NS3 Serine Protease inhibitor

Serine proteases were recognised as a useful target for their inhibitor design and discovery in the recent trend of drug discovery development work. While there have not been many reports on bioactive small molecules against dengue viruses, there have been some work in progress in this area. Leung and co-workers synthesised several small peptide substrates (Figure 1.4) with potent inhibitory activity against CF40.gyl.NS3 protease (Leung et al., 2001). Chanprapaph and co-workers designed synthetic tripeptides such as KKR that were found to act as competitive inhibitors for
NS3 serine protease with the substrate GRR coupled with aminomethyl coumarin (or AMC) (Chanprapaph et al., 2005) while Ganesh and his co-workers have identified five small molecules with inhibitory activity against the NS2B(H)-NS3 protease (Figure 1.5) through molecular docking experiments (Ganesh et al., 2005). Small molecules from natural product extracts have also been reported to exhibit inhibitory activities against NS2B/NS3 dengue serine protease. For example, 4-hydroxpanduratin A (1) and panduratin A (2) extracted from Boesenbergia Rotunda, have been reported to competitively inhibit the activity of the DEN 2 serine protease (Tan et al., 2006).

Figure 1.4: Small peptide substrate. A: AcGRR-α-keto-SL-COMH₂, B: AcGRR-CHO (Leung et al., 2001)
Many researchers have taken the advantage of the advancement in computational techniques in drug design and development work. There are many successful examples from the computer-aided drug design of the HIV protease which provides many drug candidates for further phase of processes. Amongst them, Oscarsson and co-workers utilised the crystal structure of HIV protease and the substrates information to design a tetrahyrofuran P2 analogues that inhibit HIV protease in nanomolar scale (Oscarsson et al., 2003). More recently, Durdagi and co-workers developed a series of fullerene derivatives based on an in silico virtual screening study on these compounds. The compounds with good binding scores were found to be active on HIV protease when subjected to biological studies (Durdagi et al., 2009).
1.5 **Aims and Objectives**

Through the understanding of the structure and conformation of the DEN2 serine protease and its binding interactions to the inhibitors, the new drug candidate could be designed. Amongst the aim of this work is to use computational technique to study the molecular binding interactions between the DEN2 NS2B/NS3 serine protease with competitive inhibitors observed in vitro (Tan et al., 2006). Subsequently, new ligands with better inhibitory activities towards the NS2B/NS3 DEN2 serine protease will be designed and synthesised. The designed molecule will then be screened to validate the template used for the design of novel active molecules.
CHAPTER 2

HOMOLOGY, DOCKING AND NEW LIGAND DESIGN OF DEN2 NS2B/NS3 SERINE PROTEASE INHIBITION

2.1 Molecular Modelling in Drug Design

In the past few decades, from new compounds to new drug discovery, methods employed by scientist were mostly on trial-and-error basis. Million of compounds, from natural products to chemically-synthesized, have been screened against targeted systems to obtain a lead compound for further development. Rationalisation for screening of compounds in searching for bioactivity is usually based on the experience of the researchers and/or by chemical intuitions. However, this routine work for drug discovery and development is very expensive, laborious, time consuming and perhaps in the context of modern drug design and development research, somehow inelegant. In spite of this, this “classical” approach has provided several successful drugs, from minor infections to the life-threatening diseases. For instance, Taxol® (Figure 2.1), a well-known compound to date that is commonly used to treat cancer, was firstly isolated by Wall and co-worker and reported their findings in 1971 (Wani et al., 1971)

![Figure 2.1: Structure of Taxol®](image-url)
Today, classical drug discovery approach is often coupled with more rational approaches, whereby structural information is channelled to the processes involved in the underlying illness. For this, one begins with identifying a related molecular target (enzyme, receptor, etc) that causes the problem or disease, understanding of their mechanism, followed by selecting a suitable drug candidate or a lead compound that interacts in the biological activity of the disease or the target. In the process of approaching rational drug design work against diseases, molecular modelling has become a powerful tool.

Molecular modelling can simply be defined as utilisation of computational resources to study, model and or, to mimic the molecules behaviour and molecular system. Molecular modelling involves computational approaches combined with multi-disciplinary knowledge, incorporating the field of physics, chemistry, biology and mathematics. Such techniques used to be restricted to a small number of scientists with access to the computer hardware and software, where the programs, systems and maintenance were all done by themselves. Today, however, with the fast developing computing technology, computing facilities cost has become relatively low, yet still powerful enough to handle complicated calculations. Computational methods and molecular modelling are now very popular techniques in many academic institutions as well as world leading pharmaceutical companies. There are now many molecular modelling softwares available as open source for academic institutions which have benefited many scientists since they do not need to write their own programs but just to understand the working operations with some backgrounds on the software development. Molecular modelling is now blossoming with many successful approaches on drug discovery and development research. This can be seen by the exponential rise in the number recent scientific publications incorporating molecular
modelling techniques. This field of science is now more matured. However, there is still room for improvement in which more robust and more complicated molecular level calculations are required in drug discovery research. In order to make the drug design approach more rational with help of molecular modelling, the homology modelling and docking were used in this work.

2.2 Homology Modelling

In the absence of a crystal structure of a protein of interest, homology modelling is one of the approaches used to predict the protein structure. Homology modelling, or comparative modelling, is a structural prediction method that is commonly used for protein structure prediction and building. Here, the amino acid sequence of the protein of interest is aligned with one or more known protein structures (known as "templates") (Blundell et al., 1987; Sali, and Blundell, 1993, Fiser et al., 2002;). The protein of interest and the templates used usually contain structurally conserved region when they are aligned with proteins from the same family that have nearly identical structures. The observed sequence similarities usually imply the significant structural similarity since the three dimensional structures of proteins from the same family is more conserved than their primary sequences (Lesk, Chothia, 1980). The aligned sequences and the template structure are then used to build a structural model of the targeted protein (protein of interest). Homology modelling is the only method remaining technique that can reliably predict a protein structure with an accuracy that is comparable to a low-resolution experimentally determined structure (Marti-Renom et al., 2002).
Basically, homology modelling procedure consists of four sequential steps: template selection, target-template alignment, model construction, and model evaluation. Template selection is usually initiated by PDB searching (Westbrook et al., 2002) of known protein structures, using the target sequence as a query of the search. This search is done by comparing the targeted protein sequence with the sequence of each of the structures of proteins in the database (Fiser and Sali, 2003).

### 2.2.1 Target-template selection

A list of potential templates is obtained from the search earlier which would contain one or more templates that should be appropriate for the particular modelling problem. Since the quality of the model generated increases with the overall sequence similarity of the selected template to the target and decreases with the number and length of gaps in the alignment, the best template selected would be the structure with the highest sequence similarity to the modelled sequence. Occasionally, one should also consider the similarity between the “environment” (e.g., solvent, pH, ligands, and quaternary interactions) of the template and the environment in which the targeted protein needs to be modelled.

In addition, a template bound to the same or similar ligands as the modelled sequence is the best choice of template used for the modelling. Besides, the resolution and R-factor of a crystallographic structure and the number of restraints per residue for an NMR structure is the key to the accuracy of the structure. Thus, the highest resolution should generally be selected. The purpose of a comparative model generation could sometime alter the template. On the other hand, the template that contains a similar ligand to the targeted protein is probably more important than the
resolution of the template itself. For the generation of a model to be used for the analysis of the geometry of an active site in an enzyme, it may be preferable to use a high-resolution template structure (Srinivasan and Blundell, 1993; Sanchez and Sali, 1997).

2.2.2 Target-template alignment

Following a suitable template selection, an alignment method is used to align the target sequence with the template structures (Briffeuil et al., 1998; Baxevanis, 1998; Smith, 1999). The alignment is easier and more reliable when the target and template protein have sequence identity higher than 40%. For sequence identity below 40%, regions that have low local sequence similarity become frequent (Saqi et al., 1998). The sequence alignment is said to be difficult or in the “twilight zone” when their sequence identities are less than 30% (Rost, 1999). In such cases, alignments may contain increasingly large number of gaps and alignment errors, regardless of whether they are prepared automatically or manually. Therefore, it is worth the effort to get the most accurate alignment possible because there is no current comparative modelling method available to recover from an incorrect or bad sequence alignment. Multiple sequence and structure alignment may help in the more difficult target-template sequence alignment. There are various web-based protein sequence alignment, including CLUSTAL (Thompson et al., 1994; Higgins et al., 1996), FASTA3 (Pearson et al., 1990), BCM (Smith et al, 1996), BLAST2 (Altschul et al., 1990), BLOCK MAKER (Henikoff et al., 1995) and MULTALIN (Corpet, 1988).
2.2.3 Model construction

After the sequence alignment between targeted protein and template were determined, a three-dimensional (3D) protein model is built. There are various ways to construct a target protein. One of the very early time and still widely used method is the rigid body assembly (Browne et al., 1969; Greer, 1990; Blundell et al., 1987). Modelling by segment matching is another method that depends on the approximate positions of conserved atoms in the templates (Jones and Thirup, 1986; Claessens et al., 1989; Levitt, 1992). Yet another method involves modelling by satisfaction of spatial restraints, where the distance geometry or optimization techniques were used to fulfil spatial restraints obtained from the alignment (Sali and Blundell, 1993 and references therein). All model building methods are said to be accurate and relatively similar when used optimally (Marti-Renom et al., 2002). As mentioned earlier, other factors such as template selection and target-template sequence alignment will give more impact to the model accuracy, especially when the models are based on less than 40% sequence identity to the templates (Marti-Renom et al., 2000 and references therein). MODELLER 6V2, the comparative modelling software based on satisfaction of spatial restraints was used in this study due to its popularity on various homology modelling in many recent works (Sali and Blundell, 1993).

2.2.4 Model evaluations

The constructed 3D protein model of interest has to be evaluated to check for its accuracy. The evaluation can be performed on either individual regions or the whole protein itself. The folding and stereochemistry of the model will first be checked. The reliability of the generated protein model is generally increased depending on the
following factors; i.e. when the sequence similarity is increased between the target and template, the pseudo-energy Z-score (Sippl, 1993; Sanchez and Sali, 1998) is increased, and conservation of the key functional or structural residues in the target sequence is increased.

Stereochemistry of the model can be verified with the help of some commonly used programs such as PROCHECK (Laskowski et al., 1998), PROCHECK-NMR (Laskowski et al., 1996), AQUA (Laskowski et al., 1996), SQUID (Oldfield, 1992), and or, WHATCHECK (Hooft et al., 1996). These programs are available to check the bond lengths, bond angles, peptide bond and side chain ring planarities, chirality, main chain and side chain torsion angles, and clashes between non-bonded pairs of atoms in a built protein model. Program such as VERIFY3D (Luthy et al., 1992), PROSAII (Sippl, 1993), HARMONY (Topham et al., 1994), and ANOLEA (Melo and Feytmans, 1998) are amongst the methods available for inspecting spatial features of built model based on 3D profiles and statistical potentials of mean force (Sippl, 1990; Luthy et al., 1992). Errat (Colovos and Yeates, 1993) is used to check the pairwise non-covalently bonded interactions of carbon (C), oxygen (O) and nitrogen (N) atom (CC, CN, CO, NN, NO, and OO). The environment of each residue in a built model will be evaluated with respect to the expected environment found in the high-resolution X-ray structures. The theoretical validity of the energy profiles will then enable regional error detection in the models (Fiser et al., 2002).
2.3 Molecular Docking

2.3.1 Introduction

Molecular docking is one of the molecular modelling techniques that is used to predict binding interactions and molecular orientation between macromolecules (mainly are proteins, enzymes, DNA or RNA) and other molecules (either proteins, nucleic acids or small drug-like molecules), where the bindings are later evaluated geometrically and energetically.

It is known that the ability of macromolecules to interact with small molecules affects their biological function. It has also been observed that the binding between ligands and nucleic acids to form supra-molecular complexes helps in the control of many biological pathways. Due to these observations, molecular docking has become very popular and has significantly grown in its applications in computational biology such as in rational drug design research.

Molecular docking was inspired by the “lock-and-key” model that was first proposed by Emil Fisher in 1890 to represent protein and ligand interactions. The suitable “key” that is able to open up a “lock” from a given a set of keys mimics the protein that behave as the “lock” and the ligand as the “key”. Current docking methods treat protein structures as rigid entities, leaving the ligand to be flexible during the binding process to find the best spatial and energetic fit to the protein’s binding site. It is therefore possible to use molecular docking with different “keys” (ligands) that can bind to the same protein and optimise it in order to discover the “best-fit” ligand that binds to a protein of interest.
Two main matters need to be considered while approaching the molecular docking protocols; namely searching algorithm and scoring function (Taylor et al., 2002). For searching algorithm, there are two basic approaches that are commonly employed. The first approach uses the matching techniques that describe the protein and the ligand as complementary surfaces. Matching methods resembles the active site of a protein model that is usually rigid and its binding surface was described by including hydrogen bonding sites and sites that are sterically accessible. Attempts to dock various ligands of interest were then performed into the protein as a rigid body based on its geometric matching to the active site. This approach is typically fast and robust and allows a quick scan through thousands of ligands in matter of seconds and determines whether they can bind to the active site, regardless of the ligand size. One of the most successful examples of this approach is DOCK which has been used efficiently to screen an entire chemical database for lead compounds rapidly (Kuntz et al, 1982; Shoichet and Kuntz, 1993). Unfortunately, DOCK is unable to accurately estimate the dynamic changes in the protein-ligand conformations. However, recent developments have allowed molecular docking methods to investigate ligand flexibility.

The other approach involves modelling of ligands by positioning it randomly outside the protein and exploring their translations, orientations, and conformations until an ideal site is found. Compared to the matching technique earlier, this technique is relatively more time consuming. However, they allow flexibility within the ligand to be modelled and a more detailed molecular mechanics could be utilised to calculate the energy of the ligand when it interacts with the putative active site. This approach mimics the actual protein-ligand interaction better because the total energy of the system is calculated following every move of the ligand in the protein’s active site. One
of more popular software that is based on this approach is AUTODOCK, which is
developed by Olson and his co-workers at the Scripp’s Institute, San Diego.

Search algorithm could be performed to produce an optimum number of
configurations that contained experimentally determined binding modes. These
configurations are evaluated using scoring functions to search all possible binding
modes between the ligand and protein.

It is impractical to search through all degrees of freedom (translational and
rotational) for the protein-ligand molecules interaction due to the gigantic size of search
space that require long computing duration with the recent computing resources and
technology (Taylor et al., 2002). As a compromise, the amount of search space
examined with the computational expenses, constraints, restraints and approximations
were applied while sampling the search space. This helps to reduce the dimensionality
of the problem while locating the global minimum efficiently. Some common search
algorithms include molecular dynamics, Monte Carlo methods, genetic algorithms,
fragment-based methods, point complementary methods, distance geometry methods,
Tabu searches and systematic searches. It is also possible to use a combination of search
algorithms.

After all possible bound conformations of ligands have been explored with the
appointed search algorithm, a scoring function is required to rank all ligands to
determine the plausible binding mode. Usually, scoring function includes
approximation of the free energy of binding between the ligand and the protein (Leach,
2001) by adding entropic terms to the molecular mechanics equations as shown below:
\[ \Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{hbond}} + \Delta G_{\text{elec}} + \Delta G_{\text{conform}} + \Delta G_{\text{tor}} + \Delta G_{\text{solv}} \]

where \( \Delta G_{\text{vdw}} \) is dispersion/repulsion energy, \( \Delta G_{\text{hbond}} \) is hydrogen bonding energy, \( \Delta G_{\text{elec}} \) is electrostatic energy, \( \Delta G_{\text{conform}} \) is the energy deviations arises from conformational change, \( \Delta G_{\text{tor}} \) corresponds to the energy changes due to the restriction of internal rotors and global rotation and translation; and \( \Delta G_{\text{solv}} \) is attributed by desolvation upon binding and the hydrophobic effect (solvent entropy changes at solute-solvent interfaces). The first four terms are derived from molecular mechanic, and the latter term is the most challenging (Morris et al., 1998).

The complexity of the scoring function is usually reduced in order to adapt the computational expenses. This often resulted in distorting its accuracy. There are various force fields used in scoring functions, ranging from molecular mechanics force fields such as AMBER (Cornell et al., 1995), OPLS (Jorgensen and Tirado-Rives, 1988) or CHARMM (Brooks et al., 1983), to empirical free energy scoring functions (Eldridge et al., 1997) or knowledge based functions (Muegge and Martin, 1999). Usually, there are two ways to define the scoring functions in most docking methods. One uses the scoring function to rank a particular ligand conformation, followed by the modification of the ligand conformation by a search algorithm, and the scoring function is again used to rank the newly generated conformation. Another is by applying the scoring function in a two-stage scoring function: first, the search strategy is directed by a reduced scoring function. This is followed by a more rigorous scoring function to rank the various conformer generated from the studied ligand which is directed to the putative binding site as determined by the reduced scoring functions.
The second method is modified to adapt to the computational expenses by omitting the terms such as electrostatics and only consider some binding interactions (e.g., hydrogen bond), as well as making assumptions on the energy hypersurface. Other term such as the solvation effect, is either neglected or defined in a snap-shot fashion, where it involves the generation of structures in vacuo, followed by ranking with a scoring function that includes a solvent model (Taylor et al., 2002).

2.3.2 AUTODOCK

AUTODOCK is one of the widely used molecular docking software developed by Olson and his co-workers at the Scripp Institute. AUTODOCK is a flexible ligand-oriented docking technique by random positioning of the ligand outside the protein and exploring its translations, orientations, and conformations to get the ideal binding site. The original search algorithm employed was the Metropolis method, or more commonly known as the Monte Carlo simulated annealing (SA). This algorithm directs the ligand to perform a random walk in the spaces around the protein while the protein remained static throughout the simulation. A small and random displacement (translation of its centre of gravity or root atom; orientation; and dihedral angles around each of its flexible bond) is applied to each of the degrees of freedom of the ligand while each step in the simulation is performed. As a result, a new conformer is generated and its energy is evaluated using the grid interpolation procedure. Different searching methods which have been claimed to have a better accuracy than SA have been developed. These searching methods are called Genetic Algorithm and Lamarckian Genetic Algorithm, which are outlined below.
2.3.3 Searching methods for AUTODOCK

The version of AUTODOCK (AUTODOCK 3.0) (Morris et al., 1998) used in these studies employed a few options of search algorithm. While maintaining its initial Monte Carlo simulated annealing (SA) searching method, genetic algorithm (GA), local search (LS) were also used to perform energy minimization. In addition, the hybrid methods of GA and LS based on the work of Hart’s and Belew’s co-workers (Hart et al., 1994; Belew and Mitchell, 1996) was used. This hybrid method is also termed as “Lamarckian genetic algorithm” (LGA).

Lamarckian was initiated by Jean Batiste de Lamarck whose postulated that phenotypic characteristics acquired during and individual’s lifetime can become heritable traits (discredited) (Lamarck, 1914).

GA (Holland, 1975) is a mathematical language that used the idea of Darwin’s theory of evolution, which was initially used to explain the natural genetics and biological evolution. In AUTODOCK, the translation, orientation, and conformation of the ligand with respect to the protein are defined by a set of values called ligand’s “state variable”. In the context of GA, each state variable corresponded to a “gene”, the ligand’s state corresponded to the “genotype”, whereas its coordinates corresponded to the “phenotype”. The total interaction energy of the ligand with the protein which corresponded to the “fitness” is calculated using the energy function. The “crossover” processes then occur to generate new individuals that inherit genes a random pair of individuals. “Mutation” may happen to some offspring to alter their “genes” for variation. The “elitist” strategy is applied when “selection” is made from the offspring of the current generation based on the individual’s “fitness” calculated from the implemented scoring function. Offspring that is better suited to their environment
(lowest energy) will proceed to reproduce new generations, whereas poorer ones will die or stop from reproducing. (Morris et al., 1998)

The crossover or binary mutation for new individual generation being inefficient due to the generation of value that is outside of the domain of interest. Thus, the GA search performance is improved by implementing a local search method. The local search method is based on Solis and Wets’ protocol (Solis and Wets, 1981). This protocol facilitates the torsional space search which does not require gradient information about the local energy landscape. The local search method is more adaptive because the step size can be adjusted according to the recent history of the calculated energies: a user-defined number of consecutive failures or increases in energy doubled the step size; whereas the success will reduce the step size into halves. Putting the GA and LS methods together resulted in the hybrid method called Lamarckian Genetic Algorithm. This searching method is claimed to enhance AUTODOCK performance and allows more degrees of freedom. In addition, the force field used in docking could also be used for ligands energy minimization. For each new population, in which GA uses two point crossover and mutation operators, a user-determined fractions will undergo a local search procedure with a random mutation operator where the step size is adjusted to give an appropriate acceptance ratio (Morris et al., 1998).

In summary, a generation of new conformers would have undergone five stages consecutively: mapping and fitness evaluation, selection, crossover, mutation, and elitist selection. These processes were repeated until a user-defined total number of final conformers are achieved. Three different search algorithms (SA, GA and LGA) were tested on seven crystal structure of protein-ligand complexes. GA and LGA showed better results than SA, with their lowest energy structures are within 1.14 Å RMSD of
the crystal structure (Morris et al., 1998). Figure 2.2 showed the workflow of LGA search method.

![Diagram](image)

**Figure 2.2:** The protocol of Lamarckian Genetic Algorithm (LGA) search method. The lower horizontal line represents the space of the phenotypes, whereas the upper one represents the space of the genotypes. The mapping function maps the genotypes to phenotypes. \( F(x) \) represents fitness function. The genotypic mutation operator from the parent’s genotype with the corresponding phenotype is shown on right-hand side of the diagram, whereas the local search operator is shown on the left-hand side. Searching is usually performed in phenotypic space to gain information about the fitness value. With sufficient iterations of the local search to arrive at a local minimum, an inverse mapping function is then used to convert phenotype to its corresponding genotype. AUTODOCK perform local search by continuously converting the genotype to the phenotype, hence inverse mapping is not required, where the genotype of the parent is replaced by the resulting genotype, in accordance with Lamarckian principles (Source: Morris et al., 1998)

### 2.3.4 Scoring function of AUTODOCK

Scoring function in AUTODOCK is implemented to evaluate the “fitness” or how good the docked energy between ligand and protein is. Five terms were implemented in AUTODOCK based on the thermodynamic cycle of Wesson and Eisenberg (Wesson and Eisenberg, 1992): a Lennard-Jones 12-6 dispersion/repulsion
term for Van der Waals potential energy calculation; a directional 12-10 hydrogen bond term for hydrogen bonds modelling; a coulombic electrostatic potential; a term proportional to the number of sp\(^3\) bonds in the ligand to represent unfavourable entropy of ligand binding due to the restriction of conformational degrees of freedom; and a desolvation term that is derived from inter-molecular pair wise summation combining an empirical desolvation weight for ligand carbon atoms and a pre-calculated volume term for the protein grid (Taylor et al., 2002). The empirical free energy coefficients of these five terms are derived using linear regression analysis from a set of 30 protein-ligand complexes with known binding constants. AMBER force field is implemented into AUTODOCK for the protein and ligands parameters (Morris et al., 1998).

### 2.3.5 Programs in AUTODOCK

To run molecular docking using AUTODOCK, there are three main programs involved: Autotors, Autogrid, and Autodock. “Autotors” is used to define the torsion in the ligands by determining their bonds, either by making all bonds rotatable, selective rotatable or rigid; and defining the root atom (fixed portion of the ligand, from which rotatable ‘branches’ sprout).

“Autogrid” is used on protein (or termed as “macromolecule” in AUTODOCK) to build a three dimensional grid of interaction energies map based on the atom type of the protein target. This grid map is a three dimensional lattice of uniformly spaced points that positioned surrounding or is centered in the site-of-interest of the protein. Each point contains a probe atom that has the pre-calculated affinity potential energy for each atom type of interest that it is assigned to (Morris et al., 2001). By using a distant-dependent dielectric function (Mehler and Solmajer, 1991), the grid map is able to
include the electrostatic interactions by interpolating the electrostatic potential and by multiplying the atom charges. The pre-calculated energy functions stored in the grid map makes the protein-ligand binding interaction energy calculation solely dependent on the number of atom in the ligand, hence accelerating the molecular docking in AUTODOCK.

Finally, the “Autodock” is the program that execute the docking simulation based on user-defined parameters (ligands, searching methods, number of docking runs, etc.) which gave the output of the “elitist” (best conformer) in terms of its docked energy, estimated free energy of binding, estimated inhibition constant, internal energy of ligand, together with some user defined analyses, such as clustering histogram, ranking of found conformers and rmsd.
2.4 Materials and Methods

2.4.1 Homology model of DEN2 NS2B/NS3 Serine Protease

Homology model of NS2B/NS3 of dengue virus type 2 was built using the HCV serine protease NS3/NS4A (pdb ID: 1jxp) as the template. The Modeller (mod6v2) software package was used to perform model building. The sequence alignment was done based on the published results of Brinkworth et al., 1999. The quality of the backbone of rough model generated from Modeller was then evaluated using PROCHECK (Laskowski et al., 1993), VERIFY3D (Bowie et al. 1991) and ERRAT (Colovos and Yeates, 1993) on the UCLA bioinformatics server (http://nihserver.mbi.ucla.edu/SAVES/, 16 April 2005). Energy minimization (100 steps of steepest decent plus 50 steps of conjugate gradient) was performed onto the model, using Hyperchem software package (Hypercube, Inc.) to reduce the bumps and bad contacts while keeping the backbone of the protein restrained. The model evaluation was then repeated. Figure 2.3 illustrated the workflow of this work performed.
Figure 2.3: Work flow of homology model construction for 3D structure of DEN2NS2B/NS3 serine protease

2.4.2 Comparison of the homology model with crystal structures of and DEN2 NS3 and HCV NS3/4A

The similarities and differences of the structure and conformation around the catalytic triad in of the constructed homology model of DEN2 NS2B/NS3 serine protease were evaluated using the crystal structures of DEN2 NS3 (pdb id: 1bef) and the HCV NS3/NS4A (pdb id: 1jxp).

2.4.3 Docking experiment using homology model

The docking of three competitive bioactive molecules, 4-hydroxypanduratin A (1), panduratin A (2) and ethyl 3-(4-(hydroxymethyl)-2-methoxy-5-nitrophenoxy) propanoate (3) (termed as “ester (3)” in later discussion), onto the catalytic triad of the serine protease were performed using AUTODOCK 3.05 software package (Morris et
al., 1998). The homology model of DEN2 NS2B/NS3 protease molecule was added polar hydrogen atoms and its non-polar hydrogen atoms were merged to the heteroatom connected to them. Kollman charges were assigned and solvation parameters were added to this enzyme molecule. For the ligands, non-polar hydrogen atoms were merged with Gasteiger charges assigned. All rotatable bonds of ligands were set to be rotatable. Docking was performed using genetic algorithm and local search methods (or termed as Lamarkian Genetic Algorithm). A population size of 150 and 10 millions energy evaluations were used for 100 times searches, with a 60 x 60 x 60 dimension of grid box size and 0.375 Å grid spacing around the catalytic triad. Clustering histogram analyses were performed after the docking searches. The best conformations were chosen from the lowest docked energy that populated in the highest number of molecules in a particular cluster with not more than 1.5 Å root-mean-square deviation (rmsd). The H-bond, van der Waals and other binding interactions were analysed using Viewerlite 4.2 (Accelrys Software Inc.). Figure 2.4 illustrated the workflow of the docking experiment was performed.
Figure 2.4: Workflow of performing docking experiment using AUTODOCK 3.05
2.4.4 Design of the new ligand from the docked bioactive molecules

The conformer of the studied molecule that has the lowest docked energy was extracted its coordinates and the binding interactions between molecules and protease were studied with the help of molecule viewer software, Viewer Lite, to locate the important interactions between protease and molecules. Superimpositions between the different docked molecules to the protease were performed to find the common and redundant functional groups among the docked molecules. The important fragments and functionalities were then implemented in the new designed ligand. The designed ligand was then docked into the protease and the docked energy was reevaluated.
2.5 Homology Model of DEN2 NS2B/NS3 Serine Protease

2.5.1 Results

2.5.1.1 Homology model building and model evaluation

In order to enable the \textit{in silico} binding interaction study to be carried out, a homology model of DEN2 NS2B/NS3 serine protease was built based on the crystal structure of HCV NS3/NS4A serine protease. The model was built by spatial restrain that is applied in MODELLER 6v2 software. The built model was then refined by several minimisations and was sent to a web-based structural verification to gain details about the quality of the generated model. In this study, PROCHECK, VERIFY 3D and ERRAT was performed.

In the Ramachandran plot obtained from PROCHECK (Figure 2.5), an overall 100 \% non-glycine residue was shown to be in the allowed region. This implies a good protein backbone structure and folding, where the distribution of the $\phi/\psi$ angle of the model were within the allowed region.

In addition, analysis of the homology model from VERIFY 3D (Figure 2.6) showed 90.4\% of the residues having a 3d-1d score of greater than 0.2. This suggests a reasonable conformation of the residues in the model. However, the region with 3d-1d scores of lower than 0.2 was found in the range of Glu-91-Gln-110. This indicates a lower confidence in its conformations and folding, implying a lower homology between DEN2 serine protease and HCV serine protease in this particular region.

Besides PROCHECK and VERIFY 3D, ERRAT was also used to examine the non-bonded structures of the protein model and to compare with a reliably high-
resolution structures from the database of protein crystals. The DEN2 N2SB/NS3 homology model showed about 77% overall quality factor of the sequence to be below 95% rejection limit for each chain in the input structure (Figure 2.7). This indicated an improved three-dimensional profile of the protein after several minimisations, as compared to the pre-generated homology model (data not shown). All these verification procedures performed on the NS2B/NS3 protease model indicated this model to have reached a satisfactory fold quality. Thus, no further loop modelling was carried out on the model.

Figure 2.5: Ramachandran plot of built homology model of DEN2 NS2B/NS3 complex
Figure 2.6: VERIFY 3D plot of DEN2 NS2B/NS3 homology model
2.5.2 Discussions

2.5.2.1 Comparison of the homology model with crystal structures of DEN2 NS3 and HCV NS3/NS4A

Overall, the homology model showed almost the same folding pattern as that observed in the DEN2 NS3 crystal structure. One alpha-helix and 6 beta sheets are observed in the first domain of both homology model as well as DEN2 NS3 crystal structure (Figure 2.8). The differences between two models, however, were observed in the second NS3 domain where more loop regions were observed in the crystal structure of NS3 compared to those observed in the homology model. In addition, only one alpha helix and 7 beta sheets in C terminal region was observed in the crystal structure for
NS3, whilst one extra beta sheets in the same domain was observed in the homology model.

In the reported crystal structure of HCV serine protease, the NS3 protein is incorporated with the NS4A residues as co-factor into the N-terminal domain β-sheet, thus led to a more rigid and precise framework for “prime-side” substrate binding channel residues which provided a better catalytic cavity making the NS3 enzyme more active in proteolytic process (Kim *et al.*, 1996). Superimposition (Figure 2.8d) of the crystallographic structure for NS3 with that of the homology model revealed a difference in the folding in the region between Gly-114-Val-126 which would explain the importance of NS2B as the co-factor of the NS3 protein. In the homology model, the protein has repacked into a more rigid and stabilised conformation, particularly at the C-terminal domain, where more secondary structure was observed. This is contrary to the NS3 crystal structure of the protein when less secondary structure was observed in absence of the NS2B co-factor (Murthy *et al.*, 1999).

The catalytic site of a protease is crucial for the initiation of the proteolytic process. It is therefore the catalytic triad for HCV NS3/NS4A and DEN2 NS2B/NS3 serine proteases were observed and found to be structurally conserved with the identical conformations among these catalytic triad residues. The RMSD value found between the catalytic triad residues of the HCV NS3/NS4A crystal (His-57, Asp-81 and Ser-139) and the homology model of DEN2 NS2B/NS3 (His-51, Asp-75 and Ser-135) is 0.6, whilst the RMSD on the catalytic triad of the homology model of DEN2 NS2B/NS3 and the DEN2 NS3 crystal is 1.1. The hydrogen bonding between the hydroxyl group of Ser-135 and cycloimine of the His51 side chain was observed in the catalytic triad of the reported DEN2 NS3 crystal structure (Figure 2.9a). The side chain carboxyl oxygen
of Asp-75, however, is pointed away from His-51 (Figure 2.9a), thus caused the inability to form a hydrogen bond between carboxyl group of Asp-75 and cycloamine of His-51 that disrupt the proton transfer from Asp-75 to Ser-135 which is required to activate the proteolytic process. However, in the homology model, the carboxyl oxygen of Asp-75 and His-51, as well as that of His-75 and the hydroxyl of Ser-135, was found to be at 1.6 Å which is within the hydrogen bonding distance (Figure 2.9b), suggesting a better arrangement of catalytic residues that enables the catalytic process of the protease.

The structural verifications via PROCHECK, VERIFY 3D and ERRAT for the generated homology model of DEN2 NS2B/NS3, HCV NS3/NS4A crystal structure and DEN2 NS3 crystal structure are tabulated in Table 2.1. Generally, structural verification from PROCHECK revealed that all the structures gave a reasonable reading of the Ramachandran plot, where more than 90% of non-glycine residues were located in the allowed region and no residues were located in disallowed. This indicated that the backbone of the serine protease of HCV and DEN2 have a reasonably high degree of homology, in spite of their low sequence identity (Brinkworth et al., 1999).

Comparison of the VERIFY 3D and ERRAT analyses of the three proteins (crystal structure of HCV NS3/NS4A, homology model of NS2B/NS3 and crystal structure of DEN2 NS3), displayed a better reading in VERIFY3D and ERRAT for the homology model compared to that crystal structure of NS3. This seems to suggest a better side chain packing in the computer-generated model. In addition, the absence of the NS2B co-factor in the crystal structure of NS3 is attributed to a lower quality 3d structure of the crystals. This information confirmed the role of the protease co-complexed with NS2B co-factor, which seems to re-orientate the active pocket of the DEN2 NS3, exhibiting a better side chain packing for a more efficient proteolytic
cleavage (Yusof et al., 2000). The structural verification studies of various methods performed on DEN2 NS3 crystal structure showed a low confidence in the structural information. Thus, it is presumably not a viable template. On the other hand, structural verifications performed on the crystal structure of HCV NS3/NS4A showed a remarkably high level of confidence. Hence, the homology model generated using HCV crystal structure as a template should provide a better and more accurate picture of the DEN2 serine protease structure.

Figure 2.8: Structures of flavivirus serine proteases; a: DEN2 NS2B/NS3 complex homology model, b: DEN2 NS3 crystal structure, c: HCV NS3/NS4A complex crystal structure, d: superimposition of DEN2 NS3 crystal (green) and homology model (blue). Fragment that exhibiting the difference in the protein folding is shown in red.
Figure 2.9: Spatial arrangement of catalytic triad. In a: DEN2 NS3 crystal structure (pdb id: 1bef) and b: DEN2 NS2B/NS3 complex homology model. Distance between the carboxyl oxygen of Asp-75 and His-51, as well as that of His75 and the hydroxyl of Ser-135 are indicated.

Table 2.1: Structural verification (PROCHECK, VERIFY3D, ERRAT) and comparison between structure of HCV NS3/NS4A crystal, homology model of DEN2 NS2B/NS3 and DEN2 NS3 crystal

<table>
<thead>
<tr>
<th>Structural Verification</th>
<th>HCV crystal (1jxp)</th>
<th>NS2B/NS3 model</th>
<th>homology</th>
<th>NS3 crystal (1bef)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramachandran Plot</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>80.7</td>
<td>86.3</td>
<td>82.7</td>
<td></td>
</tr>
<tr>
<td>allowed</td>
<td>18.0</td>
<td>11.0</td>
<td>15.1</td>
<td></td>
</tr>
<tr>
<td>Generously allowed</td>
<td>1.3</td>
<td>2.7</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>Disallowed</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>VERIFY3D</td>
<td>97.4</td>
<td>90.4</td>
<td>52.2</td>
<td></td>
</tr>
<tr>
<td>ERRAT</td>
<td>92.0</td>
<td>77.1</td>
<td>49.4</td>
<td></td>
</tr>
</tbody>
</table>

In summary, although the overall sequence identity between DEN2 NS2B/NS3 and HCV NS3/NS4A was only 14.8% (Brinkworth et al., 1999), the region surrounding the catalytic triad of the protease and the residues involved in the substrate binding showed a high level of sequence identity and is conserved. In addition, the information gathered from the results obtained from PROCHECK (all the non-glycine residue are in allowed region), VERIFY3D (90.4) and ERRAT (77.1) revealed the reasonable structure of this generated homology model that could make some plausible prediction.
In summary, the homology model of DEN2 NS2B/NS3 complex serine protease has provided information on the similarities and differences between the structures of HCV NS3/NS4A crystal and uncomplexed DEN2 NS3 crystal. The crystal structure of HCV NS3/NS4A appears to be a better choice to be used as a template in order to generate a model for DEN2 serine protease since results indicated better structure verification values for HCV NS3/NS4A than that of DEN2 NS3 crystal. This model was then used as the target protein (macromolecule) for the molecular docking studies of the binding interactions between this protein and found competitive inhibitor.

2.6 Molecular Docking Studies

2.6.1 Results

2.6.1.1 Inhibition of bioactive compounds towards DEN2 NS2B/NS3

Several competitive inhibitors towards DEN2 serine protease NS2B/NS3 activity have been discovered through a substrate-based approach by mimicking the polyprotein cleavage junctions (Chanprapaph et al., 2005). In addition, α-ketopeptidomimetic (Leung et al., 2001) compounds and guanidine derivatives (Ganesh et al., 2005) have also been targeted as competitive inhibitors for the DEN2 serine protease. In our laboratory, the natural products, 4-hydroxypanduratin A (1) and panduratin A (2) as well as a synthesised compound, ethyl-3-(4-(hydroxymethyl)-2-methoxy-5-nitrophenoxy)propanoate, (3) were found to competitively inhibit the activity of the DEN2 serine protease. The structures of these compounds are shown in Figure 2.10. The \( K_i \) values for these compounds were obtained from inhibition studies using recombinant DEN2 NS2B/NS3 enzyme. These inhibition studies showed 4-hydroxypanduratin A (1) to be the most potent, followed by panduratin A (2) and the ester (3) (Table 2.2).
Table 2.2: $K_i$ values of the competitive inhibitors (Tan et al., 2006)

<table>
<thead>
<tr>
<th>Compound</th>
<th>$K_i$, $\mu M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-hydroxy panduratin A (1)</td>
<td>21</td>
</tr>
<tr>
<td>Panduratin A (2)</td>
<td>25</td>
</tr>
<tr>
<td>Ester (3)</td>
<td>59</td>
</tr>
</tbody>
</table>

Figure 2.10: Structure of the selected competitive inhibitors; (1): 4-hydroxy panduratin A, (2): panduratin A and (3): ethyl 3-(4-(hydroxymethyl)-2-methoxy-5-nitrophenoxy)propanoate

2.6.1.2 Active site docking

These three compounds (Figure 2.10) that demonstrated competitive inhibition of the DEN2 serine protease were used as ligands in the binding interaction studies on the active site of the DEN2 serine protease. All these ligands showed reasonably low internal energy, indicating that the docked conformers were in their most favourable conformations. The spatial arrangement of the three ligands bound to the active site of DEN2 NS2B/NS3 serine protease is shown in Figure 2.11.

Docking of these compounds to the active sites showed 4-hydroxy panduratin A (1) to have the lowest docked energy, followed by panduratin A (2) and the ester (3) (Table 2.3). These results are parallel to the $K_i$ value observed experimentally for these compounds.
Figure 2.11: Connolly surface representations of the active site of DEN2 NS2B/NS3 protease with the bound ligands. (a): 4-hydroxypanduratin A (1), (b): ester (3) and (c): panduratin A (2), which is shown in stick model. Connolly surface of the active site of the protease is coloured according to a charge spectrum: acidic groups are red, basic groups are blue and neutral groups are white. Residues labelled in black are those that may involve in H-bond/salt bridge interaction with ligand. Residues in labeled blue are involved in van der Waals interactions, while those in green are involved in both the van der Waals and H-bond interactions.

Although panduratin A (2) showed the best free energy of binding, estimated inhibition constant and intermolecular energy, it has a higher ligand torsional free energy and internal energy values than that of 4-hydroxypanduratin A (Table 2.3). This may results in a weaker binding of this ligand to the enzyme when compared to the 4-hydroxypanduratin A (1). The ester (3), with the most number of rotatable torsion points, suffer the highest torsional free energy among these three compounds and hence resulted in a lower binding affinity as compared to the other ligands.
Table 2.3: Energies (in kcal/mol) calculated using AUTODOCK 3.05

<table>
<thead>
<tr>
<th>Entry</th>
<th>Ligand\textsuperscript{a}</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Estimated Free Energy of Binding\textsuperscript{b}</td>
<td>-7.4</td>
<td>-7.7</td>
<td>-6.1</td>
</tr>
<tr>
<td>2</td>
<td>Estimated Inhibition Constant, (\mu)M (K\textsubscript{i})</td>
<td>+3.9</td>
<td>+2.3</td>
<td>+33.6</td>
</tr>
<tr>
<td>3</td>
<td>Final Docked Energy\textsuperscript{c}</td>
<td>-10.2</td>
<td>-10.1</td>
<td>-9.2</td>
</tr>
<tr>
<td>4</td>
<td>Final Intermolecular Energy</td>
<td>-8.9</td>
<td>-9.6</td>
<td>-8.9</td>
</tr>
<tr>
<td>5</td>
<td>Final Internal Energy of Ligand</td>
<td>-1.3</td>
<td>-0.6</td>
<td>-0.3</td>
</tr>
<tr>
<td>6</td>
<td>Torsional Free Energy</td>
<td>+1.6</td>
<td>+1.9</td>
<td>+2.8</td>
</tr>
</tbody>
</table>

\textsuperscript{a} three different ligands were numbered as follows:
(1): 4-hydroxy panduratin A
(2): Panduratin A
(3): Ester
\textsuperscript{b} Estimated Free Energy of Binding is derived from the Final Internal Energy (entry 4) and Torsional Free Energy (entry 6)
\textsuperscript{c} Final Docked Energy is the combination of Final Internal Energy (entry 4) and Final Internal Energy of Ligand (entry 5)

2.6.2 Discussions

2.6.2.1 Interactions between inhibitors and residues in DEN2 NS2B/NS3

Analysis was performed on the docked DEN2 NS2B/NS3 protein complex to establish the hydrogen bonding interactions between the ligands and the active site of NS2B/NS3 protease. The criteria for hydrogen bond interaction used are when the distance between the hydrogen and the heteroatom is within the range of 2.5-3.5 Å and the bond angle is between 109°-180°. In general, these ligands exhibited hydrogen bonding within the active site and the residues as reported by Bazan and Fletterick (1989).

The formation of tetrahedral intermediates of the substrates in the oxyanion hole is an essential part of the catalysis during proteolytic process. This tetrahedral formation enables the stabilisation of the substrate which is to be cleaved by the protease in addition to lowering the activation energy to ease the proteolytic process. Oxyanion holes have been observed to interact with ligands within the active site of the
serine protease at the residues Gly-133 and Ser-135 by Murthy and co-workers (Murthy et al., 1999). Interestingly, the carbonyl group of Gly-151 and the hydroxyl in Ser-135 side chain were found to be involved where the plausible hydrogen bond were observed in all ligands studied (Table 2.4). This seemed to suggest a different binding mode of the ligands to the enzyme than those observed by Murthy and his co-workers where they reported the residues backbone of Gly-133 and Ser-135 to be involved in forming the oxyanion hole. In our study, the docked structures indicated the involvement of Gly-133 and Ser-135 solely in the binding interactions with the ester (3) but not with the 4-hydroxypanduratin A (1) and panduratin A (2) (Table 2.4).

Figure 2.12: Hydrogen bond analysis of the docked ligands. (a): 4-Hydroxypanduratin A; (b): Panduratin A; (c): Ester 3. The catalytic triad residues were illustrated in ball and stick model, whereas other residues were depicted in stick model. The docked ligands were shown in thicker stick model, where the suggested hydrogen bonds were displayed in green dotted lines.
Table 2.4: Residues in the active site of DEN2 NS2B/NS3 that are involved in hydrogen bonding with the various ligands

<table>
<thead>
<tr>
<th>Residues</th>
<th>4-Hydroxypanduratin A (1)</th>
<th>Panduratin A (2)</th>
<th>Ester (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>His-51</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp-75</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser-131</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly-133</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr-134</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser-135</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Gly-151</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Asn-152</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val-155</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although the involvement of the Ser-135 and Gly-151 in the formation of the oxyanion hole with all ligands was observed in the docked models, other hydrogen bonding interactions were found to vary with these three ligands (Figure 2.12). Firstly, with the two similar ligands, 4-hydroxypanduratin A (1) and panduratin A (2), replacement of the hydroxyl group in the 4-hydroxypanduratin A (1) with a methoxy group seemed to result in a higher $K_i$ value observed for 4-hydroxypanduratin A (1). Presumably, substituting the hydroxyl group with the methoxy group in 4-hydroxypanduratin A (1) resulted in the loss of hydrogen bonding with the carboxyl group in Asp-75 side chain of the DEN2 NS2B/NS3 protease. With the ester (3), the high $K_i$ value observed could be attributed to its structural flexibility. However, this flexibility was somewhat compensated by additional hydrogen bonds formed between the ester and Ser-131, Thr-134, Asn-152 and Val-155 (Table 2.4).

All three ligands were found to exhibit van der Waals interactions with Ser-131, Pro-132 and Try-150, suggesting the importance of these three residues as part of the van der Waals specificity pocket. Likewise, His-51, Gly-133, Ser-135, Asn-152 and
Gly-153 may also play important roles in such binding interactions (Table 2.5). The small hydrophobic specificity pocket for P1 which is formed by the residues Leu-135, Phe-154 and Ala-157 in HCV NS3/NS4A protease (Love et al., 1996) is said to be equivalent to the residues Ser-131, Tyr-150 and Gly-153 in the DEN2 NS2B/NS3 protease (Brinkworth et al., 1999). These latter residues were found to be involved in the van der Waals interactions with the ligands (1), (2) and (3). This suggested more hydrophilic interactions of the enzyme to the ligands rather than the more hydrophobic interactions observed in HCV NS3/NS4A protease. In addition, Ser-131 which has been shown not to be an important binding residue in previous study of DEN2 NS2B/NS3 protease (Brinkworth et al., 1999), is now observed to be involved in the van der Waals interaction with the ligands in our docking models (Figure 2.13).

**Figure 2.13:** Van der Waals interactions and hydrophobic interactions between the docked ligands (1, 2 and 3) and the DEN2 NS2B/NS3 serine protease protein model. a: 4-hydroxypanduratin A; b: panduratin A; and c: ester(3) Solid surfaces in green colour represent the amino acid residues that were involved in Van der Waals interactions, whereas the solid surfaces in grey colour are the amino acid residues that were involved in hydrophobic interactions. Pro132 that was in yellow surface has both Van der Waals and hydrophobic-type of interactions.
Table 2.5: Residues in the active site of DEN2 NS2B/NS3 that are involved in Van der Waals interaction

<table>
<thead>
<tr>
<th>Residues</th>
<th>4-Hydroxypanduratin A (1)</th>
<th>Panduratin A (2) Ester (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>His-51</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ser-131</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Pro-132</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Gly-133</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Ser-135</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Tyr-150</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Asn-152</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Gly-153</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Additional interactions were observed between the ligands (1) and (2) with the residues Val-52, Leu-128, Pro-132 and Val-155 in the hydrophobic site in the binding pocket of DEN2 serine protease whereby the orientation of these ligands (1) and (2) seemed to fit into these hydrophobic residues. Here, the phenyl group of the hexacyclic portion in ligands (1) and (2) fitted into the hydrophobic pocket consisting of the residue Val-52, whilst the 2-methyl-2-butylene, another hydrophobic substituent of the ligands seemed to fit into a second hydrophobic pocket formed by the residues Leu-128, Pro-132 and Val-155 (Figure 2.13, a and b). The ester group in ligand (3) also fitted into the Leu-128, Pro-132 and Val-155 hydrophobic pocket but did not interact with the residue Val-52 (Figure 2.13, c).

Another form of interaction observed with 4-hydroxypanduratin A (1) and panduratin A (2) which was not observed in previous studies was the $\pi-\pi$ aromatic interaction (Figure 2.14). In this case, the position of the trisubstituted phenyl moiety in both the ligands (1) and (2) is located at about 3 - 4.2 Å and is parallel to the pentacyclodiazo side-chain of the His-51. This enabled a $\pi-\pi$ type aromatic interaction.
to occur. However, such interaction was not been observed with the ester (3). The lack of both the hydrophobic group that could interact with Val-52 and aromatic interaction, presumably, has made the ester (3) less active as observed from both its experimental and calculated $K_i$ values when compared to the two ligands.

![Figure 2.14](image)

**Figure 2.14**: Molecular orientation of the docked ligand at the catalytic triad of DEN2 NS2B/NS3. (a): 4-hydroxypanduratin A; and (b): panduratin A. Docked ligand are shown in stick model, whereas catalytic triad residues are shown in ball and stick model. The double-ended arrow showed the suggested $\pi-\pi$ type aromatic interaction between pentacyclodiazao side-chain of His-51 and the phenolic moiety of the ligands.

In summary, several modes of interactions such as hydrogen bonding, Van der Waals, hydrophobic interaction as well as $\pi-\pi$ interaction were observed between these ligands and the DEN2 NS2B/NS3 active sites. These findings have provided further understanding on the binding interactions of the catalytic triad of the DEN2 NS2B/NS3 serine protease, thus giving input into the mode of action of the catalytic triad.
2.6.2.2 New ligand design strategy

The goal in new ligand design was to generate a new class of compounds that could exhibit better binding affinity for the active site of the DEN2 NS2B/NS3 serine protease. For this, piperidinyl group was impregnated into the design to give a different class of compound from that of the active compounds found (cyclohexenyl chalcones). A new ligand was designed based on the model of the interaction studies between the active and competitively inhibiting compounds with the DEN2 NS2B/NS3 serine protease recombinant. The conformers with the best docked energy from their most populated cluster were superimposed to each other in order to construct a suitable pharmacophore model for a new ligand design (see Figure 2.15).

Both in vitro and molecular docking studies had showed that 4-hydroxy panduratin A has the best binding affinity. Therefore, the new ligand should possess similar skeleton with this cyclohexynyl chalcone.

Superimposition of the three ligands revealed the importance of the hydrophobic chain as exhibited by 2-methyl-2-butylene of the cyclohexynyl chalcones as well as the ethyl-3-ethoxypropanoate of ester (3). Another hydrophobic interaction was attributed by the phenyl moiety of the both cyclohexynyl chalcones (ligands 1 and 2). This phenylring of the phenolic group on the cyclohexynyl chalcones not only contributed to the hydrophobic interactions but also seemed to be important in providing the $\pi-\pi$ stacking with the His-51. In addition, hydrogen bonding interaction was also observed between the phenolic group in the cyclohexynyl chalcones and the hydroxyl side chain of Ser-135 in the protease. Keeping this information in mind, introducing a hydrophobic long chain as well as keeping the aromatic moiety in the new ligand should
be strongly considered. In addition, the position of the hydroxyl group on the aromatic ring should also be taken into account during the design of the new ligand.

Finally, previous reports revealed the importance of generating peptide substrates of DEN2 NS2B/NS3 serine protease that contain dibasic amino acids due to the dibasic cleavage junction recognition of the binding pocket. Hence, we intuited the idea of an amine functionality placement into the newly designed ligand. In this case, the branched methyl group from the cyclohexynyl ring of the cyclohexynyl chalcones was replaced by the secondary amine. In addition, introduction of different functional group allow us to explore the better drug candidates from different chemical spaces.

![Figure 2.15: Superimposition of the best docked conformer of the three competitive inhibitor. All hydrogens were removed for clarity. Red: 4-Hydroxypanduratin A; Yellow: Panduratin A; Blue: Ester (3).](image)
2.6.2.3 Virtual screening of newly designed ligand

To fulfil all the required characteristics on the newly generated ligand, a piperidinyl class of compound as shown in Figure 2.16 was constructed. This newly constructed compound was screened in silico through ligand-protein docking to the model of DEN2 NS2B/NS3 serine protease using AUTODOCK 3.05.

![Molecule structure of the newly designed ligand](image)

*Figure 2.16: Molecule structure of the newly designed ligand*

Interestingly, the docking results revealed this designed ligand to give better binding affinity than the three ligands (ligands 1, 2 and 3) used in constructing the pharmacophore model. Its docked energy was observed to be about -11.4 kcal/mol. The modes of interactions such as hydrogen bonding, van der Waals and hydrophobic interaction were analysed and illustrated in Figure 2.17. This piperidenyl molecule (4) seemed to be able to disrupt the electron transfer from the carboxyl group of Asp-75 to Ser-135 via His-51, presumably by forming a hydrogen bond with these residues. In addition, hydrogen bonding interactions were also observed between this molecule (4) and Gly-151 as postulated from the previous study, as well as with the Gly-133 and Thr-134. These interactions possibly helped to stabilise the protein-ligand complex further. Van der Waals interactions were found with Pro-132 and Asn-152. In addition, two hydrophobic binding interactions consisting of Val-52, as well as Pro-132, Leu-128,
Val-155 and Asn-152 form hydrophobic binding interactions with the molecule (4). Pi-pi (π–π) stacking with His-51 was also observed with this molecule. Thus, the newly designed ligand (4) offers a new candidate that may open up a new chemical space of compounds with the binding interactions to the protease of interest.

**Figure 2.17:** Binding interactions illustration between the newly designed ligand and the homology model of DEN2 NS2B/NS3 serine protease. a: Hydrogen bond were shown in green dotted lines; b: residues that involved in Van der Waals were shown as solid surfaces in green and hydrophobic interactions were solid surfaces in grey. Pro-132 that was in yellow surface has both Van der Waals and hydrophobic-type of interactions.
CHAPTER 3
SYNTHESIS OF THE DESIGNED MOLECULE

3.1 Retrosynthetic Analysis

The newly designed ligand (4) predicted a good binding interaction with the DEN2 NS2B/NS3 serine protease model. Thus, as part of the project, the molecule (4) was synthesised for in vitro screening in order to validate the pharmacophoric model that was constructed. The molecule (4) shares a similar skeleton to Panduratin A. However, it falls into piperidinyl class of compound instead of the cyclohexynyl chalcones of 4-hydroxypanduratin A (1) and panduratin A (2). The molecule was disconnected into 4 different moieties (Scheme 3.1) in the retrosynthetic analysis.

From our retrosynthesis, nicotinic acid was chosen to be the starting material for the synthesis since it is cheap and available in abundant. Making an ethyl ester to mask the carboxylic acid group in the nicotinic acid, incorporation of the moiety C into the nicotinic acid (D) through 1,4-nucleophilic addition should be possible following the pyridinyl ring activation by chloroformate. The α,β-unsaturated ethyl ester (E) generated from the dihydropyridine synthesis could then be subjected to Michael addition with the moiety B to give the ester (F). Finally, hydrogenation of the unsaturated bond would form the piperidine (G) which was then converted to the Weinreb amide (H). Grignard reaction between the moiety B and the Weinreb amide (H) should give the carbamate (I). The deprotection of the carbamate (I) followed by methylenedioxy ring opening should furnish the targeted product (4).
Scheme 3.1: Retrosynthesis analysis of the targeted compound
In summary, there are 3 key steps involved in this strategy to connect all 4 moieties, starting from a tandem piperidinyl ring activation and 1,4-nucleophilic addition, followed by Michael 1,4-addition, and finally ketone synthesis from Weinreb amide. The background of these key steps syntheses will be discussed.

3.2 Chemistry of Pyridinyl, Dihydropyridinyl and Piperidinyl Ring Synthesis: Synthesis of Dihydropyridines

The chemistry of dihydropyridines began with Hantzsch in 1882 when he performed the reaction to synthesise dihydropyridines (the Hantzsch reaction) as antihypertensive drugs (Hantzsch, 1882). The dihydropyridines was formed through the condensation reaction between 2 moles of ethyl acetoacetate and 1 mole of aldehyde-ammonia under warm ethanol to give a good yield (Eisner and Kuthan, 1972) (Scheme 3.2).

Later, many syntheses of dihydropyridine were carried out based on this reaction by substituting different acetoesters, aldehydes and the ammonia species to make various kind of dihydropyridine of interests (Stout and Meyers, 1982). However, only the symmetrical dihydropyridines can be obtained based on this method.

Various approaches to synthesise dihydropyridine, particularly the asymmetrical and or, functionalised 1, 4-dihydropyridine, apart from Hantzsch reaction, have been
reported. This included many reactions such as rearrangement, fragmentation cycloaddition and insertion reactions, which were reviewed by Stout and Meyers, in 1982 (Stout and Meyers, 1982). More recently, Lavilla updated the development of dihydropyridine synthesis in his review. One of the reviewed methods that attracted the attention is the nucleophilic addition to the pyridinium salt to make multifunctional dihydropyridine. Here, two kinds of pyridinium salt that were highlighted, the N-alkylpyridinium salt and N-acylpyridinium salt. The nuleophilic addition of various nucleophile to the N-alkylpyridinium salt that were generated \textit{in situ} afforded dihydropyridine in good yield, making this method a good choice for natural product synthesis, particularly in total syntheses of alkaloids. Nevertheless, the generation N-acylpyridinium salt from more reactive electrophilic species are possible. The more reactive N-acylpyridinium is generated \textit{in situ} and reacts readily with wide range of nucleophile species to afford various dihydropyridine with high degree of regio- and stereoselectivity (Lavilla, 2002). Due to the well-developed chemistry, this method of dihydropyridine generation is adopted for this work.

Hilgeroth and co-workers, in their synthesis of 6,12-diazatetrakishomocubanes and 3,9-diazatetraasteranes used various multi-functioned 1,4-dihyropyridines (Hilgeroth and Baumeister, 2000; Hilgeroth et al., 2002) as the starting material. Hilgeroth syntheses of cubanes began with the formation of 1,4-dihydropyridine as shown below (Scheme 3.3).

![Scheme 3.3: Hilgeroth’s synthesis of dihydropyridine as the precursor of cubanes](image-url)

64
3.3 1, 4-Michael Addition

Michael addition typically refers to the base-promoted nucleophilic addition of the nucleophile (or termed as Michael donor) to an $\alpha,\beta$-unsaturated substrates (or Michael acceptor) that contains carbonyl functionality, such as esters, aldehydes, ketones and amides (Bergman et al., 1959). The development of Michael addition has grown tremendously with various forms of Michael donors and acceptors and has widened the scope of the organic reaction that can be classified as Michael addition.

However, to date, there are few publications on Michael addition reaction on $\alpha,\beta$-unsaturated ester with dihydropyridine as the Michael acceptor. The closest related work is on the conjugate addition is on the Michael acceptor, 3-alkoxycarbonyl-dihydropyridones which is in the synthesis of inhibitors of 2,3-oxidosqualene-lanosterol cyclase by Dodd and his co-workers (Scheme 3.4). Here, the conjugate addition was the key step in incorporating a medium-sized alkyl chain to the 3-alkoxycarbonyl-dihydropyridones, followed by several functional groups conversions to furnish some 2-alkylpiperindin-4-ol as the final product (Dodd and Oehlschlager, 1992). Starting from the model study reaction by using higher order cuprates generated from 2 equivalents of n-BuLi and 1 equivalent of CuCN at -78°C as the Michael donor that gave a reasonable good yield (87-91%), they used the squalene-type of alkyl chain and thienyl cyanocuprate to generate the Michael donor species to give the Michael adduct was in excellent yield (90%) (Dodd and Oehlschlager, 1992).
Scheme 3.4: Synthesis of inhibitors of 2,3-oxidosqualene-lanosterol cyclase using 1,4-Michael addition as a key step

The Michael donors that are of interest are the organometallic reagent, due to the ease of generation. There are varieties of organometallic reagents such as organozinc, organomagnesium, organolithium and organocupper that could be used as good Michael donors. Typically, cuprates reagent, or Gilman reagent that is generated from the organolithium and copper (I) salt are among the most popular method employed in many 1,4-additions due to its good regioselectivity.

Since the extensive research and development were done on the Michael addition on \(\alpha,\beta\)-unsaturated carbonyl species, which resulted many publications and reviews, utilisation of the 1,4-Michael addition may be possible to connect another moiety into the molecule, after the \(\alpha,\beta\)-unsaturated ester was formed form the 1,4-dihydropyridine generated from nucleophilic addition to acylpyridinium salt (Scheme 3.5).
Scheme 3.5: Proposed reaction using 1,4-Michael addition as the key step to incorporate the butyl moiety to the dihydropyridine

3.4 Weinreb Amide

Since the work was pioneered by Nahm and Weinreb, the use of N-methoxy-N-methylamides in the synthesis of ketone derivatives, this Weinreb amide has rapidly popularised itself in organic synthesis. Nahm and Weinreb found N-methoxy-N-methylamides to react cleanly with Grignard and organolithium reagents to form ketones in a one-pot reaction (Scheme 3.6) (Nahm and Weinreb, 1981).

Scheme 3.6: General example of ketone synthesis from Weinreb amide using Grignard reagent

Unlike other nucleophilic attack on carbonyl functionalities that usually give rise to the secondary and tertiary alcohol as the adduct, the Weinreb amide is always utilised to make ketones in the organic synthesis, because of the ease of preparation with only a small amount of side-reactions during the nucleophilic addition and mild, selective acidic workup. These advantages can be attributed to the stability of the tetrahedral metal-chelated intermediate which is formed by addition of nucleophiles to N-methoxy-
N-methylamides which reduced the propensity of producing a secondary or tertiary alcohol by an over reactive Grignard and/or organolithium reagents to the substrate. In addition, reactions between organometallic reagents and the Weinreb amides generally give good-to-excellent yields under fairly mild conditions (at -78°C or 0°C, and solvents such as THF, ether or DME). (Mentzel and Hoffmann, 1997)
3.5 Materials and Methods

3.5.1 Materials and instruments used

The chemicals used in this synthesis were purchased from Merck and Sigma Aldrich. NMR spectra were recorded from the following instruments: JEOL ECA 400, JEOL Lambda 400. Thin layer chromatography (TLC) was conducted on silica gel F\textsubscript{254} TLC plates purchased from Merck. Flash column chromatography was carried out using silica gel (40 to 63 µm) purchased from Merck and Mallinckrodt. Diethyl ether (Et\textsubscript{2}O), tetrahydrofuran (THF) were distilled from sodium-benzophenone prior to use, Dichloromethane (CH\textsubscript{2}Cl\textsubscript{2}) were distilled from CaH\textsubscript{2} prior to use. All the air-sensitive reactions were performed under nitrogen, with the related glasswares were heat-dried under high vacuum using hot air gun as a heating source. Table 3.1 illustrates the molecules involved in the synthesis work.

### Table 3.1: List of molecules that were used and synthesised in this work

<table>
<thead>
<tr>
<th>No.</th>
<th>Molecular structure</th>
<th>Name of the molecule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="" alt="Nicotinic acid" /></td>
<td>Nicotinic acid</td>
</tr>
<tr>
<td>2</td>
<td><img src="" alt="Ethyl Nicotinate" /></td>
<td>Ethyl Nicotinate</td>
</tr>
<tr>
<td>3</td>
<td><img src="" alt="Diethyl-4-phenylpyridine-1,3(4H)-dicarboxylate" /></td>
<td>Diethyl-4-phenylpyridine-1,3(4H)-dicarboxylate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ethyl (1-phenoxycarbonyl-4-phenylpyridinyl)-3(4H)-carboxylate</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Ethyl (1-tert-butoxycarbonyl-4-phenylpyridinyl)-3(4H)-carboxylate</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Ethyl (1-tert-butoxycarbonyl-2-butyl-4-phenyl)-3,4-dihydropyridinyl-3-carboxylate</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Ethyl (1-tert-butoxycarbonyl-2-butyl-4-phenyl)-piperidinyl-3-carboxylate</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>(1-tert-butoxycarbonyl-2-butyl-4-phenyl)-piperidinyl-3-carboxylic acid</td>
</tr>
<tr>
<td></td>
<td>Structure</td>
<td>Chemical Formula</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>------------------</td>
</tr>
<tr>
<td>9</td>
<td><img src="image9.png" alt="Structure 9" /></td>
<td>1-tert-butoxycarbonyl-2-butyl-3-(methoxy(methyl)carbamoyl)-4-phenylpiperidine</td>
</tr>
<tr>
<td>10</td>
<td><img src="image10.png" alt="Structure 10" /></td>
<td>1-tert-butoxycarbonyl-2-butyl-3-(benzo-1,3-dioxol-4-carbonyl)-4-phenylpiperidine</td>
</tr>
<tr>
<td>11</td>
<td><img src="image11.png" alt="Structure 11" /></td>
<td>1-tert-butoxycarbonyl-2-butyl-3-hydroxymethyl-4-phenylpiperidine</td>
</tr>
<tr>
<td>12</td>
<td><img src="image12.png" alt="Structure 12" /></td>
<td>1-tert-butoxycarbonyl-2-butyl-3-formyl-4-phenylpiperidine</td>
</tr>
<tr>
<td>13</td>
<td><img src="image13.png" alt="Structure 13" /></td>
<td>1-tert-butoxycarbonyl-3-(benzo-1,3-dioxol-4-carbonyl)-2-butyl-4-phenylpiperidine</td>
</tr>
</tbody>
</table>
3.5.2 Synthesis of Ethyl Nicotinate (2)

37.02 g of nicotinic acid 1(300.7 mmol) was refluxed together with 115 cm$^3$ of EtOH absolute and 50 cm$^3$ of H$_2$SO$_4$ for 4 hours in oil bath at 135 °C. Reaction mixture was then cooled in ice bath, followed by addition 25% NH$_3$ aqueous until slightly alkaline, and evaporated to remove excessive EtOH. The pale yellow oil formed was then extracted with Et$_2$O followed by washing with brine. 82% yield was obtained as pale yellow oil. $^1$H NMR (400MHz, CDCl$_3$): 1.41 (t, 3H, $J$=7.2Hz, H1’), 4.41 (q, 2H, $J$=7.2Hz, H2’), 7.39 (q, 1H, $J$=4.8Hz, H5), 8.29 (dt, 1H, $J$=4.8Hz, 2.1Hz, H4), 8.76 (dd, 1H, $J$=2.1Hz, 4.8Hz, H6), 9.22 (s, 1H, H2). (Sambrook et al., 2005)
3.5.3 Synthesis of diethyl 4-phenylpyridine-1,3(4H)-dicarboxylate (3)

Ethyl nicotinate 2 (3.32 g, 22.0 mmol) was dissolved in anhydrous THF (200 cm$^3$) and cooled to -10ºC. Ethyl chloroformate (2.38 g, 22.0 mmol) was then added dropwise and stirred for 30 minutes. The reaction mixture was then added catalytic amount of copper (I) iodide (0.22 g, 1.29 mmol) followed by 12.0 cm$^3$ PhMgCl (2 M in THF). Reaction mixture was then stirred at room temperature for additional 2 hours. The reaction mixture was then quenched by ammonium chloride, vacuumed dry to remove THF. The remaining aqueous layer was then extracted with Et$_2$O, followed by washing of organic layer with NH$_3$ (20%):NH$_4$Cl = 1:1, H$_2$O, 10% HCl, H$_2$O, brine. The solvent was then evaporated to dryness and recrystallized with MeOH to give 69% yield as white powder. $^1$H NMR (400MHz, CDCl$_3$): 1.16 (t, 3H, $J$=7.1Hz, H1’), 1.38 (t, 3H, $J$=6.8, H1”), 4.08 (m, 2H, H2’), 4.36 (q, 2H, $J$=7.2Hz, H2”), 4.47 (d, 1H, $J$=4.8Hz, H4), 5.17 (brs, 1H, H5), 6.88 (brd, 1H, H6), 7.26 (m, 5H, H Phenyl), 8.07 (brs, 1H, H2). (Hilgeroth and Baumeister, 2000)
3.5.4 Synthesis of ethyl (1-phenoxy carbonyl-4-phenylpyridinyl)-3(4H)-carboxylate (4)

Ethyl nicotinate 2 (3.32g, 22.0 mmol) was dissolved in anhydrous THF (200 cm$^3$) and cooled to -10°C. Phenyl chloroformate (3.44g, 22.0 mmol) was then added dropwise and stirred for 30 minutes. The reaction mixture was then added catalytic amount of copper(I) iodide (0.246 g, 1.29 mmol) followed by 13.01 cm$^3$ PhMgCl (1.69 M in THF). Reaction mixture was then stirred at room temperature for additional 2 hours. The reaction mixture was then quenched by ammonium chloride, vacuumed dry to remove THF. The remaining aqueous layer was then extracted with Et$_2$O, followed by washing of organic layer with NH$_3$ (20%):NH$_4$Cl=1:1, H$_2$O, 10% HCl, H$_2$O, brine. The solvent was then evaporated to dryness and recrystallized with MeOH to give 62% yield as white powder. $^1$H NMR (400MHz, CDCl$_3$): 1.19 (t, 3H, $J$=7.1Hz, H1’), 4.13 (m, 2H, H2’), 4.50 (d, 1H, $J$=4.4Hz, H4), 5.30 (brs, 1H, H5), 7.03 (brd, 1H, $J$=5.4Hz, H6), 7.31 (m, 10H, H Phenyl), 8.22 (brs, 1H, H2). $^{13}$C NMR (100MHz, CDCl$_3$): 14.05, 38.08, 60.39, 113.22, 120.46, 121.32, 126.31, 126.82, 128.08, 128.49, 129.56, 131.51, 144.69, 149.68, 150.58, 166.27.
3.5.5 Synthesis of ethyl (1-tert-butoxycarbonyl-4-phenylpyridinyl)-3(4H)-carboxylate (5)

Product 4 (2.37 g, 6.78 mmol) was dissolved in anhydrous THF and cooled to -42 °C. In another flask, t-BuOK (1.53g, 13.6 mmol) was dissolved in 30 cm³ of anhydrous THF (5 cm³ for rinsing purposes), then cannulated to substrate dropwise. Reaction mixture was then stirred for 30 minutes, quenched with H₂O, followed by extraction with 3 times of Et₂O. The extracted product was washed with NH₃ (20%):NH₄Cl=1:1, H₂O, then H₂O, brine, followed by Na₂SO₄ treatment then rotadry. The crude wash purified from flash column chromatography using petroleum ether and EtOAc as eluting solvent.

¹H NMR (400MHz, CDCl₃): 1.12 (t, 3H, J=7.0Hz, H1’), 1.53 (s, 9H, H1”), 4.03 (m, 2H, H2’), 4.43 (d, 1H, J=4.0Hz, H4), 5.08 (brs, 1H, H5), 6.80 (brs, 1H, H6), 7.20 (m, 5H, H Phenyl), 8.02 (brd, 1H, H2). ¹³C NMR (100MHz, CDCl₃): 14.03, 28.04, 38.94, 60.07, 83.60, 110.95, 111.75, 120.65, 126.52, 126.56, 128.00, 128.33, 132.16, 145.39, 149.64, 166.68.
In anhydrous condition, 2 equivalents of n-BuLi (1.6 M in hexanes, 12.2 ml) was added dropwise into a cooled (-40 °C to -50 °C) CuCN (0.875 g, 9.77 mmol) in 100 ml THF and stirred for 30 minutes under cold bath. The resulting 2 immiscible layers, where cuprate was formed at the lower layer as pale yellow coloured solution was then cooled to -78 °C and the resulting solution (86.7 ml) was cannulated into the substrate 5 (2.79g, 8.47 mmol) at this temperature. The reaction was quenched with NH₄Cl saturated solution, extracted with Et₂O, followed by washing with more NH₄Cl, then water and brine, finally dried with Na₂SO₄ anhydrous. Further purification was performed using flash column chromatography using petroleum ether and EtOAc as eluting solvent. The product was purified as a mixture of enantiomer with 31% yield. ¹H NMR (400MHz, CDCl₃): 0.81 (m, 3H, H1'”), 1.05 (t, 3H, J=7.2Hz, H1’), 1.28 (m, 6H, H2””, H3””, H4””), 1.46 (s, 9H, H1”), 2.81 (dd, 1H, J=62.7Hz, 3.8Hz, H3), 3.73 (brd, 1H, J=11.6Hz, H4), 3.94 (m, 2H, H2’), 4.55 (brd, 1H, J=10.8Hz, H2), 4.72 (brd, 1H, J=8.4Hz, H5), 6.76 (d, 1H, J=7.6Hz, H6), 7.14 (m, 5H, H Phenyl). ¹³C NMR (100MHz, CDCl₃): 13.95, 22.55, 27.96, 28.21, 37.98, 50.65, 51.14, 51.58, 52.94, 60.51, 81.04, 109.41, 122.92, 123.16, 126.63, 128.30, 142.96, 152.20, 171.72. LRMS (EI): m/z 387 (M⁺, 1), 214 (65), 156 (100), 69 (77), 41 (91). HRMS (EI) calculated for C₂₃H₃₃O₄N (M⁺): 387.2410, found 387.2403.
3.5.7 Synthesis of ethyl (1-tert-butoxycarbonyl-2-butyl-4-phenyl)-piperidinyl-3-carboxylate (7)

The reaction flask was charged with substrate 6 (0.107g, 0.277 mmol) and Pd/C catalyst (0.029g, 0.0277 mmol Pd). 10 ml of MeOH was added and then the mixture was purged with nitrogen gas to remove oxygen. The reaction began when it was then purged with hydrogen and stirred under hydrogen at room temperature. The completion of reaction was monitored by TLC. After completion of reaction, the reaction mixture was again purged with nitrogen and the catalyst was then filtered off. The product was obtained for 99% yield after the solvent was evaporated off. $^1$H NMR (400MHz, CDCl$_3$): 0.82 (m, 3H, H1”), 0.96 (m, 3H, H1’), 1.17 (m, 6H, H2’’, H3’’,H4’’), 1.44 (s, 9H, H1”), 1.71 (m, 1H, H5), 1.81 (m, 1H, H5), 2.81 (m, 1H, H6), 2.94 (dt, 1H, $J$=12.4Hz, 4.8Hz, H3), 3.11 (m, 1H, H4), 3.88 (m, 2H, H2’), 3.95 (brd, 1H, $J$=13.2Hz, H6), 4.62 (dd, 1H, $J$=79.1Hz, 10.0Hz, H2), 7.15 (m, 5H, H Phenyl). $^{13}$C NMR (100MHz, CDCl$_3$): 13.84, 22.45, 25.6, 27.70, 28.40, 33.63, 37.19, 38.36, 42.76, 50.62, 51.99, 57.65, 79.73, 126.68, 127.31, 128.32, 144.13, 154.82, 209.88. LRMS (EI): m/z 389 (M$^+$, 5), 276 (80), 232 (100), 128 (21), 84 (17), 57 (57). HRMS (EI) calculated for C$_{23}$H$_{35}$O$_4$N (M$^+$): 389.2566, found 389.2559.
3.5.8 Synthesis of (1-tert-butoxycarbonyl-2-butyl-4-phenyl)-piperidinyl-3-carboxylic acid (8)

The substrate 7 was dissolved in little amount of THF, then 2M KOH solution in EtOH 95% was added. The mixture was then refluxed. The mixture was acidified with 2M HCl and extracted with EtOAc. The combined extracted product was then washed with water brine then dried with Na$_2$SO$_4$ anhydrous, followed by evaporation of the solvent to get the product in quantitative yield. $^1$H NMR (400MHz, CDCl$_3$): 0.82 (m, 3H, H1'), 1.20 (m, 6H, H2', H3', H4'), 1.41 (s, 9H, H1''), 1.66 (brd, 1H, J=13.2Hz, H5), 1.77 (m, 1H, H5), 2.72 (m, 1H, H6), 2.95 (dt, 1H, J=12.2Hz, 4.8Hz, H3), 3.04 (m, 1H, H4), 3.95 (dd, 1H, J=64.3Hz, 11.2Hz, H6), 4.60 (dd, 1H, J=71.9Hz, 13.2Hz, H2), 7.09 (m, 5H, H Phenyl). $^{13}$C NMR (100MHz, CDCl$_3$): 14.00, 22.33, 26.11, 27.80, 28.40, 29.67, 33.73, 37.91, 50.73, 52.35, 80.15, 126.51, 127.23, 128.40, 143.79, 154.83, 176.51. LRMS (EI): m/z 361 (M$^+$, 5), 304 (22), 260 (29), 204 (38), 57 (100). HRMS (EI) calculated for C$_{21}$H$_{35}$O$_4$N (M$^+$): 361.2253, found: 361.2242.
3.5.9 Synthesis of 1-tert-butoxycarbonyl-2-butyl-3-(methoxy(methyl)carbamoyl)-4-phenylpiperidine (9)

The substrate 8 was dissolved in CH₂Cl₂. DMAP and PyBrOP were then added and stirred for 1 hour. NHMe(OMe).HCl was then added and stirred for 19 hours at room temperature. The reaction mixture was quenched by adding water, extracted by CH₂Cl₂ and the organic layer was washed with brine. Purification of the crude through column chromatography (hex:EtOAc= 3:17) afforded 49% of the product. ¹H NMR (400MHz, CDCl₃): 0.81 (m, 3H, H1’”), 1.22 (m, 4H, H2’”, H3’”), 1.42 (s, 9H, H1”), 1.58 (m, 2H, H4’”), 1.80 (m, 2H, H5), 2.84 (m, 1H, H6), 2.92 (s, 3H, H1’), 3.30 (m, 2H, H3, H4), 3.62 (s, 3H, H2’), 3.99 (dd, 2H, J=72.3Hz, 13.6Hz, H6), 4.60 (brd, 1H, J=43.16Hz, H2), 7.13 (m, 5H, H Phenyl). ¹³C NMR (100MHz, CDCl₃): 13.84, 22.10, 25.61, 27.64, 28.25, 31.84, 33.76, 36.93, 38.40, 47.22, 50.91, 61.49, 79.23, 125.99, 127.09, 128.03, 144.08, 154.51, 172.18. LRMS (EI): m/z 404 (M⁺, 1), 303 (24), 273 (40), 247 (43), 57 (100). HRMS (EI) calculated for C₂₃H₃₆O₄N₂ (M⁺): 404.2675, found 404.2682.
3.5.10 Synthesis of 1-tert-butoxycarbonyl-2-butyl-3-(benzo-1,3-dioxol-4-carbonyl)-4-phenylpiperidine (10)

![Chemical structure](image)

**a. from 9**

The Grignard of 4-Bromo-1,3-benzodioxole was prepared in prior to use by gently reflux it with magnesium turnings in THF for about 1 hour. 1.1 eqv. of TMEDA was then added to the Grignard reagent and stirred for 1 hour at room temperature. In the other pot, substrate 9 was dissolved in THF and transferred to the Grignard reagent. The mixture was stirred at room temperature for 1 1/2 hour, followed by quenching with saturated NH₄Cl, extraction with Et₂O, and dried with anhydrous MgSO₄. Purification of the crude through column chromatography (Hex: EtOAc = 19:1 to 9:1) afforded 7.8 % of product.

**b. from 13**

Substrate 13 (0.042g, 0.090 mmol) was dissolved in CH₂Cl₂, and 1.1 eqv. of DMP (0.03ml, 0.3 M in CH₂Cl₂) was added and stirred at room temperature for 1 hour. 20% Na₂S₂O₃ in saturated Na₂CO₃ was then added into the reaction mixture and the reaction mixture was diluted with Et₂O. Extraction was performed with Et₂O, followed by treatment with anhydrous MgSO₄. Evaporation of the solvent afforded the pure product in quantitative yield without further purifications. ¹H NMR (400MHz, CDCl₃): 0.74 (m, 3H, H1”), 1.04 (m, 4H, H2”, H3”), 1.42(s, 9H, H1”), 1.58 (m, 2H, H4”), 1.80 (m,
2H, H5), 2.84 (m, 1H, H6), 3.40 (dt, 1H, \(J=11.9, 3.6\text{Hz}, H4\)), 4.01 (dd, 1H, \(J=79.7\text{Hz},13.4\text{Hz}, H6\)), 4.05 (m, 1H, H3), 4.77 (dd, 1H, \(J=72.4, 12.7\text{Hz}, H2\)), 6.05 (d, 2H, \(J=3.2\text{Hz}, H4'\)), 6.75 (dd, 1H, \(J=16.8\text{Hz}, 8.1\text{Hz}, H2'\)), 6.88 (dd, 1H, \(J=15.9\text{Hz}, 7.6\text{Hz}, H1'\)), 7.03 (m, 1H, H3’), 7.16 (m, 5H). \(^{13}\text{C NMR (100MHz, CDCl}_3\))): 13.98, 22.29, 25.82, 27.55, 28.60, 29.68, 34.12, 38.25, 50.73, 51.71, 55.63, 79.43, 101.46, 112.42, 119.42, 121.31, 126.14, 127.31, 128.35, 144.56, 147.27, 148.47, 154.61, 196.54. LRMS (EI): m/z 465 (M\(^+\), 5), 408 (22), 364 (55), 244 (68), 57 (100). HRMS (EI) calculated for C\(_{28}\)H\(_{35}\)O\(_3\)N\(_1\) (M\(^+\)): 465.2515, found 465.2527.

3.5.11 Synthesis of 1-tert-butoxycarbonyl-2-butyl-3-hydroxymethyl-4-phenyl piperidine (11)

![Chemical structure of 1-tert-butoxycarbonyl-2-butyl-3-hydroxymethyl-4-phenyl piperidine (11)]

Under anhydrous and nitrogen condition, substrate 7 was dissolved in THF and cooled to -78°C. 3 eqv. of LiAlH\(_4\) (2M solution in THF) was added dropwise and reaction mixture was stirred for 1 hour by slowly heated up to room temperature. Reaction flask was then placed into cold bath, quenched by slow addition of MeOH into the reaction mixture, diluted with water, followed by 10% NaOH. Extraction was done with Et\(_2\)O and the extract was dried with anhydrous MgSO\(_4\). The extract was then evaporated to afford quantitative pure product without further purification. \(^1\text{H NMR (400MHz, CDCl}_3\))): 0.84 (m, 3H, H1’’’), 1.19 (m, 6H, H2’’’, H3’’’, H4’’’), 1.42 (s, 9H, H1’”), 1.65 (m, 2H, H5), 2.06 (m, 1H, H3), 2.60 (m, 1H, H4), 2.80 (m, 1H, H6), 3.25 (m, 2H, H1’”), 81
4.02 (dd, 1H, J=75.9Hz, 13.2Hz, H6), 4.48 (dd, 1H, J=59.9Hz, 12.0Hz, H2), 7.15 (m, 5H). $^{13}$C NMR (100MHz, CDCl$_3$): 14.16, 22.68, 24.23, 28.47, 34.46, 37.38, 38.78, 40.72, 46.76, 51.62, 62.89, 79.37, 126.67, 127.20, 128.70, 143.80, 155.10. LRMS (EI): m/z 347 (M$^+$, 1), 290 (20), 234 (100), 204 (18), 190 (42), 57 (55). HRMS (EI) calculated for C$_{21}$H$_{33}$O$_3$N$_1$ (M$^+$): 347.2460, found 347.2466.

### 3.5.12 Synthesis of 1-tert-butoxycarbonyl-2-butyl-3-formyl-4-phenylpiperidine

![Structure of 1-tert-butoxycarbonyl-2-butyl-3-formyl-4-phenylpiperidine](structure.png)

**a. from 9**

Under anhydrous and nitrogen condition, substrate 9 was dissolved in THF and cooled to 0°C. 3 equiv. of LiAlH$_4$ (2M solution in THF) was added drop wise and reaction mixture was stirred at 0°C for 1 hour. Reaction was then quenched by slow addition of MeOH into the reaction mixture, diluted with water, followed by 10% NaOH. Extraction was done with Et$_2$O and the extract was dried with anhydrous magnesium sulphate. The extract was then evaporated to afford 91.1% of pure product without further purification.

**b. from 11**

Substrate 11 was dissolved in CH$_2$Cl$_2$, and 1.1 eqv. of DMP (0.3 M in CH$_2$Cl$_2$) was added and stirred at room temperature for 1 hour. 20% Na$_2$S$_2$O$_3$ in saturated Na$_2$CO$_3$
was then added into the reaction mixture and the reaction mixture was diluted with 
Et₂O. Extraction was performed with Et₂O, followed by treatment with anhydrous 
MgSO₄. Evaporation of the solvent afforded the pure product in 87.7% yield without 
进一步提纯。¹H NMR (400MHz, CDCl₃): 0.82 (m, 3H, H1‴), 1.18 (m, 6H, 
H2‴, H3‴, H4‴), 1.43 (s, 9H, H1″″), 1.86 (m, 2H, H5), 2.82 (m, 1H, H3), 2.95 (m, 1H, 
H6), 3.19 (m, 1H, H4), 4.05 (dd, 1H, J=84.7Hz, 12.0Hz, H6), 4.71 (dd, 1H, J=92.3Hz, 
9.6Hz, H2), 7.22 (m, 5H, H Phenyl), 9.39 (d, 1H, J=15.6Hz, H1′). ¹³C NMR (100MHz, 
CDCl₃): 13.97, 22.35, 26.64, 27.63, 28.39, 29.65, 34.01, 37.22, 51.70, 56.84, 80.12, 
126.98, 127.46, 128.84, 142.51, 154.67, 203.10. LRMS (EI): m/z 345 (M⁺, 1), 288 (26), 
244 (45), 217(43), 57 (100). HRMS (EI) calculated for C₂₁H₃₁O₃N₁ (M⁺): 345.2304, 
found 345.2314.

3.5.13 Synthesis of 1-tert-butoxycarbonyl-3-(benzo-1,3-dioxol-4-carbonyl)-2-butyl-
4-phenylpiperidine (13)

The Grignard of 4-bromo-1,3-benzodioxole was prepared in prior to use by gently 
reflux it with magnesium turnings in THF for about 1 hour. In another flask, was 
charged with substrate 12 and dissolved in THF, then added into the Grignard reagent 
and stirred at room temperature for 2 hours. Reaction mixture was then quenched with 
saturated NH₄Cl, extracted with Et₂O, dried with anhydrous MgSO₄. Purification of the 
crude product from column chromatography (hex:EtOAc = 19:1 to 4:1) afforded 58.4 %
of product. $^1$H NMR (400MHz, CDCl$_3$): 0.78 (t, 3H, $J$=7.0Hz, H1’’’), 1.19 (m, 4H, H2’’’ , H3’’’), 1.24 (s, 9H, H1’”), 1.56 (m, 2H, H4’’’), 1.69 (m, 2H, H5), 2.36 (m, 1H, H3), 2.82 (m, 1H, H6), 3.12 (m, 1H, H4), 3.86 (m, 1H, H2), 4.02 (dd, 1H, $J$=14.4Hz, 4.4Hz, H2), 4.61 (s, 1H, H5’’), 5.85 (d, 2H, $J$=34.0Hz, H4’’), 6.64 (d, 1H, $J$=7.6Hz, H1’’), 6.73 (t, 1H, $J$=7.8Hz, H2’’), 6.83 (d, 1H, $J$=8.0Hz, H3’’), 7.19 (m, 5H, H Phenyl). $^{13}$C NMR (100MHz, CDCl$_3$): 14.16, 22.25, 25.98, 27.94, 228.25, 35.13, 37.35, 39.64, 48.34, 51.99, 68.15, 78.91, 100.63, 107.19, 118.74, 121.46, 124.88, 126.54, 127.72, 128.63, 142.69, 143.90, 146.81, 154.92. LRMS (EI): m/z 467 (M$^+$, 1), 410 (24), 366 (28), 247 (56), 57 (100). HRMS (EI) calculated for C$_{28}$H$_{37}$O$_5$N$_1$ (M$^+$): 467.2672, found 467.2671.

3.5.14 Synthesis of (2-butyl-4-phenylpiperidin-3-yl)(2,3-dihydroxyphenyl) methanone (14)

Substrate 13 was dissolved in anhydrous CH$_2$Cl$_2$ and cooled to -78 °C. 5 eqv. of BCl$_3$ (1M in hexanes) was then added dropwise and the reaction mixture was stirred at -78°C for 2 hours, then heated up slowly and stirred at room temperature for additional 24 hours under nitrogen. MeOH was then added and stirred overnight. Reaction mixture was then evaporated and purified by column chromatography (100% CH$_2$Cl$_2$ to CH$_2$Cl$_2$: MeOH = 9:1) to afford 81.7 % of product. $^1$H NMR (400MHz, CDCl$_3$): 0.71 (t, 3H, $J$=7.2Hz, H1’’’), 1.19 (m, 4H, H2’’’ , H3’’’), 1.56 (m, 1H, H4’’’), 1.82 (m, 1H, H4’’’), 1.95 (m, 1H, H5), 2.29 (m, 1H, H5), 3.23 (m, 1H, H6), 3.44 (m, 1H, H4), 3.72 (m, 1H, H6), 4.15 (m, 1H, H2), 4.79 (m, 1H, H3), 6.85 (t, 1H, $J$=8.0Hz, H5’’), 7.01 (d, 1H,
$J=8.0\text{Hz, }H^4$), 7.08 (t, 1H, $J=7.2\text{Hz, }H$ Phenyl), 7.15 (d, 2H, $J=7.4\text{Hz, }H$ Phenyl), 7.22 (t, 2H, $J=9.0\text{Hz, }H$ Phenyl), 7.69 (d, 1H, $J=8.0\text{Hz, }H^6$), 8.05 (brs, 1H, 3’OH), 8.95 (brs, 1H, H1”), 11.94 (s, 1H, 2”OH). $^{13}\text{C NMR (100MHz, CDCl}_3$): 13.57, 22.18, 25.81, 27.96, 29.45, 29.69, 36.81, 39.80, 48.12, 54.75, 118.52, 120.50, 121.21, 121.56, 127.22, 128.87, 141.32, 145.34, 149.81, 203.78. LRMS (EI): m/z 353 (M$^+$, 1), 303 (24), 273 (40), 247 (43), 57 (100). HRMS (EI) calculated for C$_{22}$H$_{27}$O$_3$N$_1$ (M$^+$): 353.1991, found 353.1992.
3.6 Results and Discussions

3.6.1 Synthesis setup for the designed ligand: (2-butyl-4-phenylpiperidin-3-yl)(2,3-dihydroxyphenyl)methanone

![Structure of the target compound](image)

Scheme 3.7: Structure of the target compound

The synthesis of the target compound (shown in Scheme 3.7) began with esterification of the starting material, nicotinic acid, in the presence of sulfuric acid in ethanol. The esterification of the nicotinic acid was performed in order to mask the carboxylic acid from interfering in a later step involving a 1,4-nucleophilic addition on the pyridinyl ring using organomagnesium reagent. Similar reaction conditions and workup protocols were used by Sambrook and co-worker in their work on heterocyclic molecule construction (Sambrook et al., 2005). Refluxing the reaction mixture for 4 hour, followed by treatment with ammonia aqueous, gave a reasonably good yield of 82% (Scheme 3.8).

![Scheme 3.8: Esterification of nicotinic acid](image)

Following the procedure of Hilgeroth and Baumeister (Hilgeroth and Baumeister, 2000), the ester was first reacted with ethyl chloroformate to form the
carbamate intermediate which was subsequently reacted with phenylimagnesium bromide through a 1,4-nucleophilic addition on the pyridinyl ring in the presence of a catalytic amount of copper (I) iodide (5%) to give the carbamate 3 in 69% yield (Scheme 3.9. The rationale for first treating ethyl nicotinate with ethyl chloroformate is to active the pyridinyl ring in order to make the nucleophilic addition easier. In addition, the presence of 5% copper iodide as a catalyst is required to exclusively produce 4-phenylpiperindinyl as a product. Earlier attempts in the absence of this copper catalyst gave lower regioselectivity and 3 regioisomers, which are inseparable by chromatography, were produced.

Scheme 3.9: 1,4-nucleophilic addition of the phenyl moiety to ethyl nicotinate activated by ethyl chloroformate

The next step in the synthesis involved putting a n-butyl group at the 2 position of the dihydropyridine 3 through a 1,4-addition reaction with butylcuprate. Initially, due to the difficulties in the purification steps, since the by-product has the same retention as the desired product, only about 10 % of product was isolated from this reaction. Attempts to produce more yield of the product by varying the reaction conditions, reagents and catalysts were unsuccessful. Varying reaction conditions such as the reaction temperature from -80°C to room temperature resulted in more by-product instead. Increasing the amount of cuprate (prepared either from copper (I) iodide or copper cyanide) used also did not help to increase the yield, but instead caused the deprotection of the ethyl carbamate protecting group to occur. Addition of Lewis acid led to immediate decomposition of the dihydropyridine 3. Use of other organometallic
reagent such as organomagnesium, organozinc and also silanes, did not give any isolable yield of the desired product. Organomagnesium gave a mixture of the 1,2- and the 1,4- adducts as well as several by-product such as a deprotected compound. Both organozinc and organosilane reagents did not give any reaction and only the starting material was recovered. Lengthening the reaction time also did not help in producing the desired compound. Ethyl carbamate compounds are sensitive towards a series of nucleophile and can be cleaved readily by mild to strong nucleophilic agent to form an unstable unprotected 3,4-substituted dihydropyridine which may undergo air-oxidation to rearomatised and formed 3,4-disubstituted pyridine (Scheme 3.10).

Scheme 3.10: Proposed reaction mechanism related to the deprotection of the dihydropyridine 4 and the rearomatisation. Nu = nucleophile, mainly from organometallic reagent

Since the substrate 3 is sensitive to organometallic reagent, the protecting group was changed from ethyl carbamate to t-butyl carbamate (t-BOC) in order to stabilise the compound under basic condition. Following the method of Hilgeroth and Baumeister (Hilgeroth and Baumeister, 2002) phenyl chloroformate was reacted with the nicotinate
ester 2 and subsequently subjected to treatment phenyl magnesium chloride in presence of CuI as catalyst (Scheme 3.11) to give the dihydropyridine 4 as the product.

Scheme 3.11: 1,4-nuceophilic addition of the phenyl moiety to ethyl nicotinate activated by phenyl chloroformate

The substrate 2 was then transformed from phenyl carbamate to a BOC protected adduct by stirring with t-BuOK in THF at -42 °C. BOC protecting group is acid labile but stable under basic condition as well as against nucleophilic attack when compared to the phenyl carbamate. Thus, the dihydropyridine 5 readily underwent 1,4-addition reaction with cuprate prepared from copper (I) cyanide to give 31% yield (Scheme 3.12).

Scheme 3.12: Functional group interconversion from phenyl carbamate to t-butyl carbamate followed by 1,4-Michael addition of butyl moiety insertion

In the subsequent step, the double bond from the piperidinyl ring was subjected to a reduction reaction. Three different reagents were attempted, i.e., hydrogen gas with 10 % palladium on activated carbon, HOAc with zinc dust, and sodium
cyanoborohydride with HCl. Hydrogen gas with 10% palladium on activated carbon was found to be the most efficient of the three reducing agents where conversions using the hydrogen gas on palladium/carbon gave 99% yield. Reductions of 6 with sodium cyanoborohydride or zinc dust were rather inefficient and conversions were incomplete (Scheme 3.13).

![Scheme 3.13: Reduction of 6 using 10% palladium on activated carbon](image)

In summary, the first half of the synthesis pathway beginning from the nicotinic acid to adduct 7 hinges upon the workability of the second key step (i.e. Michael 1, 4-addition reaction of the butyl chain) which required many attempts of various nucleophilic reagent, ranging from organocopper, organozinc, organomagnesium to silanes as well as the choice of Lewis acids that were all incompatibility with the ethyl carbamate group in the molecule. The low tolerance of the ethyl carbamate with low reactivity of the α, β-unsaturated ester to the Michael 1,4-addition reaction resulted in the revision of the choice of protecting group used in the step from ethyl carbamate to t-butyl carbamate (Scheme 3.14).
Scheme 3.14: Partial synthesis of designed ligand, with 3 moieties attached (1 to 7). Reagents and conditions: a: EtOH, H₂SO₄, reflux, 4h, then NH₃(aq), 82%; b: EtOCOCl, -10°C, 30min, then PhMgCl, 5% Cul, THF, r.t., 2h, 69%; c: PhOCOCl, -10°C, 30min, then PhMgCl, THF, r.t., 2h, 62%; d: t-BuOK, THF, -42°C, 30 min, 72%; e: Bu₂(CuCN)Li₂, THF, -78°C, 31%, 6h; f: 10% Pd/C, H₂, MeOH, r.t, 24h, 99%.

The next step in the synthesis involved a conversion of the ethyl ester to Weinreb amide. Initially, a method reported by Williams and co-workers was employed, where the ester was reacted with NHMe(OMe) and an organomagnesium reagent, preferably i-PrMgCl, to form Weinreb amide in one pot with low by-product and easy purification. However, several attempts with the similar reaction conditions to make a Weinreb amide from ester 7 were unsuccessful. Varying the reaction temperatures did not make any changes and only the starting material was recovered. Presumably, the ester 7 is too hindered to be attacked by the nucleophile. Therefore, strategy was modified to make a carboxylic acid first, by hydrolysing the ester 7 with KOH, followed by general amide formation reaction. The ester 7 was easily hydrolysed with KOH under reflux to form carboxylic acid 8 in quantitative yield. The acid 8 was then subjected to a reaction with NHMe(OMe) with DMAP and PyBrOP as the carboxylic acid activator (Scheme 3.15) to form Weinreb amide 9 in 49% yield.
The Weinreb amide 9 was then subjected to insertion of the phenolic moiety, 4-bromo-l,3-benzodioxole, via nucleophilic substitution to form a ketone. The 4-bromo-l,3-benzodioxole was synthesised in 3 steps, starting from guaiacol following the method of Klix and co-workers (Klix et al., 1995). Nahm and Weinreb reported that Weinreb amide would react cleanly with organometallic reagent to form ketone as a final product, where organomagnesium and organolithium are amongst the popular choices of organometallic used (Nahm and Weinreb, 1981). However, in our hands, the Grignard reagent of the 4-bromo-l,3-benzodioxole did not work as expected. Varying the reaction temperature, reactant load (more equivalent of organometallic reagent) or replacing the organomagnesium with organolithium reagent was also unable to convert the amide 9 to the desired ketone 10 (scheme 3.16).
The addition was only successful, albeit a very low yield of 10 at 8%, when the chelating agent TMEDA was added to the organomagnesium generated from 4-bromo-1,3-benzodioxole with. Due to the low efficacy of the conversion the Weinreb amide to the desired phenolic moiety, the synthesis route was revised to enable a better yield of product to be obtained. The Weinreb amide 9 was converted to aldehyde 12 by LiAlH₄ in about 91% yield. The aldehyde 12 proved to be a better candidate for the conversion when it underwent the nucleophilic attack by the Grignard reagent generated from 4-bromo-1,3-benzodioxole to make the secondary alcohol 13 in a more reasonable yield of 58%. The racemic mixture of the alcohol 13 was then oxidised by Dess Martin periodinane to produce the ketone 10 in quantitative yield. With this alternative route, the same objective was achieved but with a better yield of 58% from the aldehyde 12 to the ketone 10 although additional 2 steps were introduced. The final step in the synthesis involved the removal of two protecting group with BCl₃, followed by the hydrolysis with methanol to furnish the targeted product 14 in 87% yield (Scheme 3.17).
Scheme 3.17: Revised route from Weinreb amide 9 to furnish the targeted product

Since it has been demonstrated that the targeted product can be synthesised from aldehyde 12 in a relatively good yield, an alternative strategy is taken in order to further reduce the number of synthesis steps using the aldehyde 12 as one of the intermediate in the synthesis. Aldehyde 12 was produced by first reducing the ester 7 to the alcohol 11 with LiAlH₄ in quantitative yield. This alcohol was then treated to DMP oxidation to produce the aldehyde 12 in 88 % yield (Scheme 3.18).

Scheme 3.18: Revised route to synthesise aldehyde 12 from ester 7
The originally planned route that utilised Weinreb amide as a key intermediate gave only 3% yield from ester 7 while the alternative route involving the aldehyde 12 formed from Weinreb amide 11 as a key intermediate gave 27% yield from ester 7. Using the aldehyde 12 formed directly by reduction of the ester 7 as the key intermediate (without going through Weinreb amide 11) gave an even better yield of 42% (from ester 7). The three different routes of synthesis employed are illustrated in Scheme 3.19 and their related yields are summarised in Table 3.1. Considering the best route among the three, the overall yield starting from nicotinic acid to the targeted molecule was found to be about 4.7%.

**Scheme 3.19:** Different routes to synthesise target molecule 14. Reagents and conditions: a: KOH, EtOH, reflux, 1h, 99%; b: NH Me(OMe).HCl, DMAP, PyBroP, CH₂Cl₂, r.t., 19h, 49%; c: 4-bromo-l,3-benzodioxole, Mg, THF, TMEDA, r.t., 8%; d: LiAlH₄, THF, -78 °C, 3h, quantitative of 11 and 91% of 12; e: DMP, CH₂Cl₂, r.t., 1h, 88% yield from 11 to 12 and quantitative yield from 13 to 10; f: 4-bromo-l,3-benzodioxole, Mg, THF, r.t., 3h, 58%; g: BCl₃, CH₂Cl₂, -78 °C to r.t., 24h then MeOH, rt., overnight, 87%.
Table 3.2: Percent yield of the targeted product from 3 different route of synthesis

<table>
<thead>
<tr>
<th>Route</th>
<th>No. of steps</th>
<th>% yield (from 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: 7, 8, 9, 10, 14</td>
<td>4</td>
<td>3%</td>
</tr>
<tr>
<td>B: 7, 8, 9, 12, 13, 10, 14</td>
<td>6</td>
<td>27%</td>
</tr>
<tr>
<td>C: 7, 11, 12, 13, 10, 14</td>
<td>5</td>
<td>42%</td>
</tr>
</tbody>
</table>

3.6.2 Stereochemical control of the proposed synthesis

Since no stereocontrol was exercised in the above synthesis, the products obtained are of a mixture of stereoisomers (enantiomer, diastereomer, etc.). However, only a pair of enantiomers of intermediates was observed throughout the synthesis, the exception being the secondary alcohol 13, where 2 pairs of enantiomers were obtained (Scheme 3.17). This observation is corroborated by further examination of the Michael adduct 6 with the polarometric as well as the NOE NMR experiment.

The polarograph indicated the compound 6 to show no significant reading in its optical rotation, indicating the compound 6 to be in the form of a racemic mixture. The relationship for the NOE NMR experiment performed on the compound 6 is as shown in Scheme 3.20(a). Irradiation of H-2 enhanced the signal for H-3 without enhancing H-4. This affirmed the assignment of the H-2 and H-3 to a cis-relationship while H-3 and H-4 are in the trans-configuration. Irradiation of H-3 enhanced the signals for H-2, H-4 and phenyl proton. Irradiation of H-4 enhanced the signals for H-3, H-5 and phenyl proton. The observed enhancement of the phenyl proton from irradiation of H-3, H-4 and H-5 signals were further their position to be near the phenyl group. Irradiation of H-5 enhanced the signals for H-4, H-6 and phenyl proton while irradiation of the H-6 signal enhanced the signal for H-5 only. Through this series of NOE experiment, the relative configuration of the compound 6 was assigned as depicted in Scheme 3.20.
Scheme 3.20: (a) The NOE correlation (showed in solid arrow) of the corresponding proton in compound 6. (b) Relative configuration of compound 6 assigned in two-dimensional representation.
CHAPTER 4
INHIBITION STUDY OF THE DESIGNED AND SYNTHESISED
COMPOUND AGAINST DEN2 NS2B/NS3 SERINE PROTEASE

4.1 Introduction

The implementation of a therapeutic strategy by inhibiting proteases of HIV-1 protease activity (Seife, 1997) has successfully generated a numbers of compounds (West and Fairlie, 1995) that inhibits HIV replication. By preventing the activity of the viral proteases that are responsible for producing structural and functional HIV proteins in host cells, HIV replication can be inhibited which may eventually halt the illness caused by HIV infection. In a similar manner, for the development of antiviral agent for DEN2 in this project, NS2B/NS3 serine protease was chosen to be the target enzyme since it is found to be important for proteolytic cleavage activity in the host cells.

Earlier work has shown some natural products inhibit the activities of the NS2B/NS3 DEN2 protease in a competitive manner with the peptide substrate Boc-Gly-Arg-Arg-4-methylcoumaryl-7-amide (Tan et al., 2006). These natural products were then used as templates to design a novel compound which is anticipated to possess inhibitory activities against DV. This chapter discusses the evaluation of the
synthesised compound (named CP14) in its ability to inhibit the DEN2 NS2B/NS3 serine protease activity and the validation of the model used in designing the inhibitor. For this purpose, CP14 was initially evaluated on the DENV-infected HepG2 cells. Inhibition or reduction the cell cytopathic effect (CPE) exhibited by the DENV-infected HepG2 cells indicated the efficacy of the compound to inhibit the viral activity. Kinetic assay of CP14 was then performed on the DEN2 NS2B/NS3 serine protease recombinant in accordance to the work reported by Yusof and co-workers (Yusof et al., 2000). Finally, RT-PCR experiment was then carried out to investigate the effect of CP14 against the replication process of DEN2 NS2B/NS3 serine protease.

4.2 Cell Cytopathic Effect

A viral infection especially dengue virus infection can be identified by examining the cytopathic effect (CPE) exhibited by the virus-treated cell culture. Cell cytopathic effect refers to morphological changes in cells, especially in tissue culture. These morphological changes are usually associated with the viral replication that invade and destroy the cell lines. When the infection is performed in tissue cultures, the virus spread is constrained by an overlay of the nutrient medium and the cytopathic effect is visualised as viral plaques. A study has to be carried out first to determine the possibility of a virus infection onto the tissue culture of interest, since not all types of
viruses are able to present the CPE on the tissue culture.

In the case of DENV infection, CPE on various cell lines has been carried out as early as 1961 when Buckley reported the observed CPE on the DENV-treated HeLa cell line (Buckley, 1961). Other cell lines, such as larval tissue (C6/36) (Corner and Ng, 1987), kidney cells (BHK) (Cleaves, 1985) and human liver cells were reported elsewhere. In this work, CP14 was evaluated on the human hepatoma cells (HepG2) for its ability to inhibit CPE. HepG2 cells were used in this study since it has been reported that liver injury is found to be persistent in DF and DHF patients (Phoolcharoen and Smith, 2004).

4.3 Analysis of Enzyme Kinetics Data

Enzyme kinetics is the study of the rate of chemical reaction catalysed by enzymes. For a simple, single substrate and irreversible reaction, Michaelis-Menten equation (see equation 1) is commonly used to describe the relationship between the rates of substrate conversion by an enzyme and the concentration of the substrate as shown in equation (2).

\[
V = \frac{V_{\text{max}} [S]}{K_m + [S]}
\]  

(1)
where [E], [S], [ES] and [P] represent the concentration of enzyme, substrate, enzyme-substrate complex and product, respectively. The Lineweaver–Burk plot is widely used to determine important terms in enzyme kinetics, such as $K_m$ and $V_{max}$. A graphical representation of the Michaelis-Menten equation gives a quick, visual impression of the different forms of enzyme inhibition; i.e., either as competitive, non-competitive or uncompetitive inhibition. These inhibitions are reversible inhibitions which involve non-covalent interactions such as hydrogen bonds, hydrophobic interactions and ionic bonds between the inhibitor and the enzyme.

The Lineweaver-Burk plot, also termed double reciprocal plot, uses the modified Michaelis-Menten equation. Rearrangement of equation (1) through inversion and factorisation produced a straight line equation used in the Lineweaver-Burk plot as follows:

$$\frac{1}{v} = \left( \frac{K_m}{V_{max}} \right) \frac{1}{[S]} + \frac{1}{V_{max}}$$  \hspace{1cm} (3)

A plot of $1/v$ versus $1/[S]$ gives a straight line whose y-intercept is $1/V_{max}$ and the slope is $K_m/V_{max}$. Substitution of $1/v$ with zero gives the x-intercept (i.e., $1/[S]$) as $-1/K_m$ (Figure 4.1). The $K_m$ and $V_{max}$ values can then be obtained through the y- and
x-intercept by plotting the reaction velocity at different substrate concentration. The $K_m$ and $V_{max}$ values are used to evaluate the type of inhibition exhibited by the inhibitors against the enzyme of interest, when the assays are performed in the set of different inhibitor concentration.

**Figure 4.1**: Lineweaver-Burk plot of $1/v$ versus $1/[S]$ to evaluate $K_m$ and $V_{max}$ (Murray *et al.*, 2003)

### 4.3 Competitive, non-Competitive and Uncompetitive inhibition: A Different Reversible Enzyme Inhibition Overview

Generally, the reversible enzyme inhibition can be classified into three different inhibition modes: competitive, non-competitive and uncompetitive inhibition. In terms of their mode of action, a competitive inhibitor competes with the enzyme’s substrate to enter the active site or the binding site of the enzyme. A non-competitive inhibitor may bind at a site other than the active site of the enzyme (or termed as allosteric site) without affecting the substrate’s binding, where the formation of enzyme-inhibitor, or
enzyme-substrate, or enzyme-substrate-inhibitor complexes are all possible. Conversions of the substrate to the product by the enzyme inhibited by non-competitive inhibitors are still possible but usually at a reduced efficacy. Uncompetitive inhibition is exhibited by the inhibitor that binds to an enzyme-substrate complex to form enzyme-substrate-inhibitor complex and prevent the enzyme from converting the substrate to a product. A truly uncompetitive inhibitor is said to bind exclusively to the enzyme-substrate complex and have no affinity for its related free enzyme.

While the maximum velocity of the enzyme ($V_{\text{max}}$) and the Michaelis constant, ($K_m$) are compared among the studies of competitive, non-competitive and uncompetitive inhibitor that present in the enzyme respectively, with the different concentration of inhibitor, [I] screened over a series of substrate concentration, their difference are readily be distinguished by a serial of reciprocal plot of 1/v versus 1/[S] (or termed as Lineweaver-Burk plot) (Figure 4.2). The Table 4.1 below summarises the different properties exhibited by the inhibitors.

<table>
<thead>
<tr>
<th>Type of Inhibitor</th>
<th>Competitive</th>
<th>Non-Competitive</th>
<th>Uncompetitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding site</td>
<td>Active sites</td>
<td>Allosteric sites</td>
<td>Allosteric sites</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>Unchanged</td>
<td>Reduced</td>
<td>Reduced</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Increased</td>
<td>Unchanged</td>
<td>Increased</td>
</tr>
</tbody>
</table>
Figure 4.2: Lineweaver-Burk plot of different inhibitor. A: competitive; B: non-competitive; C: uncompetitive

Thus, analysis of enzyme kinetics of a key enzyme would provide insight into the viral activities such as replication and give important information which can help in drug design research.
4.4 Materials and methods

4.4.1 Materials

Ni\(^{2+}\)-nitritotriacetic acid (NTA)-agarose resin was from Qiagen (Chadowrth, CA). Fluorogenic peptide substrate Boc-Gly-Arg-Arg-MCA was obtained from Peptide Institute, Inc (Osaka, Japan). Amersham Pharmacia supplied the Sephadex G-75 for column separation. AMC (7-amino-4-methylcoumarin) was purchased from Sigma-Aldrich.

4.4.2 Instrument used for Analysis and Bioassay

The intensity of the fluorogenic moiety (7-amino-4-methylcoumarin, AMC) from the cleaved fluorogenic peptide substrate was measured at excitation 385 nm and emission 465 nm using Cary Eclipse Varian fluorescent spectrophotometer. Shimadzu UV-Visible Recorder Spectrophotometer (UV-160) was used for OD and quantitative protein assay.

4.4.3 Expression and Purification of DEN2 NS2B/NS3 serine protease complex

The protein precursor consist of N-terminal six histidine tag that fused sequentially to 40-residue NS2B cofactor, a linker of 10 residues and the first 185 amino acids of NS3 was expressed using transformed competent Escherichia coli strain XL1-Blue MRF; then harvested, purified and refolded by employing the established procedures (Murthy et al., 1999; Yusof et al., 2000) to synthesise proteolytically active protease complex, DEN-2 NS2B/NS3pro. Twelve litres of competent Escherichia coli strain XL1-Blue MRF was cultured in LB medium in the presence of ampicillin
(100µg/ml) at 37°C until O.D. 600 nm reached approximately 0.6. Isopropyl-β-D-thiogalactoside (IPTG) was used to induce the overproduce of protease for 2 hours, collected by centrifugation, and stored at -70 °C until used. After resuspension in buffer A (100 mM Tris-HCl, pH 7.5, 300 mM NaCl), the cell lysis was performed using lysozyme (1mg/ml) on ice for 30 min, followed by centrifugation and pellet resuspension for 1 h at 4 °C in buffer B (100 mM Tris-HCl, pH 8.0, 300 mM NaCl, 6 M urea). The suspension was later undergoes sonication on ice using a cell disruptor. The denatured lysate was kept on ice for 1 h and clarified by centrifugation for 1 h at 4 °C.

To purify the protein of interest, the NS2B/NS3pro was isolated using a Ni²⁺-NTA affinity column and subsequently purified using Sephadex G-75 gel filtration column. The ammonium sulfate precipitated proteins were then refolded by successive dialysis to yield the active NS2B/NS3pro that was then stored at -70°C until used. 12 % SDS-polyacrylamide gel electrophoresis (PAGE) was used to trace the protease-containing fractions both after Ni²⁺-NTA and Sephadex G-75 gel filtration, and determine the purity of the enzyme after re-naturing by dialysis. Protein concentration was determined by UV-visible spectrometer using bovine albumin standards and Bradford reagent. Figure 4.3 surmised the scheme of harvesting and purification of DEN2 NS2B/NS3 serine protease complex.
Figure 4.3: Workflow of harvesting and purification of DEN2 NS2B/NS3 serine protease complex. Buffer A: 100 mM Tris-HCl, pH 8.0/300 mM NaCl; Buffer B: 100 mM Tris-HCl, pH 8.0/300 mM NaCl, 6M urea; Buffer C: 100 mM Tris-HCl, pH 7.5/300 mM NaCl, 6M urea; Buffer D: 100 mM Tris-HCl, pH 7.5/300 mM NaCl
4.4.4 DEN2 NS2B/NS3 Inhibition assay using fluorogenic peptides

Methods employed for DEN2 NS2B/NS3 inhibition assay was similar to the one that established by Yusof et al., (2000). The protease assay was performed in 96-well microtiter plates with a Cary Eclipse Varian Fluorescent Spectrophotometer at an excitation wavelength of 355 nm and emission wavelength of 460 nm at 37 °C. The standard reaction mixtures (200 µl) contained 200 mM Tris-HCl, pH 8.5, and fluorogenic peptide substrate. Preliminary studies were done to establish and the rate of product (7-amino-4-methyl-coumarin or AMC) released into the supernatant solution was determined fluorometrically [l (excitation) =385 nm, l (emission) =465 nm]. The substrate used was Boc-Gly-Arg-Arg-4-methylcoumaryl-7-amide (MCA). Stock solutions of substrate (100 mM) were prepared by dissolving the peptides in dimethyl sulfoxide, followed by dilution in water to 1 mM working stock before use. Substrate concentration was then varied between 25 and 250 µM, and signals were converted to concentrations by comparison with standard amounts of free AMC (Sigma). After enzyme addition, the reaction mixtures were incubated for 30 min at 37°C and the reaction was terminated with 1.9 ml of 125 mM ZnSO₄ (Brenner et al., 1992). The supernatant was then subjected to micro-centrifugation for 1 min at 15 000 rpm to remove the precipitate. The rate of the fluorogenic fragment (AMC) released from the substrate used was fluorometrically determined [l (excitation) = 385 nm, l (emission) = 465 nm] and monitored continuously over a period of 30 min. The kinetic parameters were calculated from non-linear regression.
4.4.5 Determination of $K_i$ for the synthesised compound

The inhibition constant, $K_i$, of the synthesised compounds were determined essentially the same way as in DEN-2 protease inhibition kinetic assays by varying the concentrations of substrate and inhibitors. The standard 100 µL reaction mixtures contained 2 µM DEN2 protease complex and Tris-HCl buffer of concentration 200 mM at pH 8.5. Four different concentrations of the synthesised compound, varying from 0 – 400 µM were used for the assays. For each concentration of inhibitor tested, the fluorogenic peptide substrate Boc-Gly-Arg-Arg-MCA concentrations were varied from 0 - 100 µM. The readings were taken in triplicates.
4.5 Results

4.5.1 Cytopathic effect study of the compound 14

In vitro cell-based study on the cytopathic effect of the compound 14 (termed “CP14”) was carried out at the Department of Molecular Medicine, Faculty of Medicine, University of Malaya. Biological activity studies were initially carried out on cell culture system to determine the response of CP14 to virus-host interaction. Human hepatoma cells (HepG2) were used as host cells in this study for the reason mentioned previously (see Chapter 4, section 4.2). In addition, participation of the dengue virus was also exhibited in the HepG2 cells since the intracellular DEN virus particles were found from the intracytoplasmic vesicles of the infected HepG2 cells (Marianneau et al., 1998)

The liver infection was identified by the cytopathic effect (CPE) on the infected HepG2 cell, where changes in the liver cell morphology were observed. An increment of the cell size (hyperthrophy) or necrosis that is presented as clear and opaque circles was observed under electron micrograph of the infected HepG2 cell as shown in Figure 4.4.

![Figure 4.4: HepG2 cell morphology. A: normal cell; B: cell infected with DEN2 virus and shown CPE.](image-url)
Prior to the biological screening, a cytotoxicity profile of the compound towards the HepG2 cells was established to prevent overdose effect that would kill the cells. The maximum non-toxic dose (MNTD) of CP14 was determined and found to be as high as 25 μg/ml. Thus, the screening for CPE on the cells was performed with CP14 at the concentration of 25 μg/ml or lower. Different virus titres were used (10, 100 and 1000 TCID₅₀/ml) for the screening. At the highest virus titre of 1000 TCID₅₀/ml, no CPE was observed on the infected HepG2 cells when 10μg/ml of CP14 was used. This indicates the ability of CP14 to protect the cells from DENV2 infection. In addition, cell-based screening assay (Figure 4.5) showed the highest inhibitory activity at this concentration to be 80% inhibition for the virus titre of 1000 TCID₅₀/ml. This observation suggested effective inhibition by CP14 at a concentration of lower than that of the MNTD value.

![Graph showing percent inhibition of CP14 on various DEN2 virus titres in HepG2 cells](image)

**Figure 4.5:** Percent inhibition of CP14 on various DEN2 virus titres in HepG2 cells
4.5.2 *In vitro* kinetic assay of CP14

The kinetic assay of the DEN2 NS2B/NS3 was carried out in the Department of Molecular Medicine, Faculty of Medicine, University of Malaya. Prior to enzyme inhibition assay on CP14, the enzyme concentrations as well as substrate concentrations were optimised. A standard plot of intensity as a response to the concentration of fluorogenic moiety of the peptide substrate, AMC (7-amino-4-methylcoumarin), was obtained (Figure 4.6) for the conversion of intensity to the concentration to be carried out and the enzyme velocity to be calculated. The compound was diluted into three different concentrations in a range of substrate concentrations.

![Intenstity vs. AMC concentration](image)

**Figure 4.6**: Plot of intensity versus concentration of fluorogenic moiety of the peptide substrate, AMC

At the optimum protease concentration, the inhibition assays were performed in several series of substrate concentration, [S], from 25 mM to 150 mM. As a control, a series of [S] with no compound was added to the enzyme. CP14 with the concentration of 70.7 μM to 176.8 μM to 353.7 μM were added into these series of [S], resulting 4 different curves of enzyme velocity versus [S] plotted. The respective regressions (R²)
of the curve fitting were exhibited to be more than 0.96, indicating the reasonable curve fittings were presented (Figure 4.7).

Generally, reduction of the enzyme velocity as well as the $V_{\text{max}}$ with CP14 present in the system resembled the inhibition effect exhibited by the CP14. However, while the curve of no inhibitor and the concentration of CP14, $[I] = 70.7 \, \mu$M, 176.8 $\mu$M gave a hyperbolic curve, the plot for $[I] = 353.7 \, \mu$M gave a sigmoidal curve. The deviation of the plot from a normal hyperbolic curve to a sigmoidal curve could be attributed by the elevated concentration of the inhibitor, [I] that could cause an early saturation of the inhibited enzymatic activity. However, the incremental concentration of the substrate, [S] under the saturated condition caused the prevalence of the alternative enzymatic mechanism. The data sets were then analysed further by plotting a Lineweaver-Burk plot.

![Velocity vs. Substrate Concentration](image)

**Figure 4.7**: Curves with different concentration of CP14, [I], of enzyme velocity versus substrate concentrations
The study of the mode of inhibition for CP14 was carried out by fitting the kinetic data to the Lineweaver-Burk plot (Figure 4.8). The control data (no inhibitor added) gave a gradient of the slope at 347.73 min⁻¹. Increasing the concentration of CP14 from 70.7 μM to 176.8 μM to 353.7 μM resulted in the increase in the gradient of the slope (Figure 4.8). This reflects the increase in inhibitory activity with increasing concentration of CP14. The series of [I] = 70.7 μM and [I] = 176.8 μM resembles the non-competitive behaviour, when these lines were intersected close to the x-axis.

However, the plot obtained at [I] = 353.7 μM did not show the same trend as that of the two with the lower inhibitor concentration. Presumably, the mode of inhibition of the inhibitor may have changed from a non-competitive one to another mode of inhibition. As mentioned earlier, the high concentration of [I] may caused early saturation and alteration in the mode of inhibition observed when the [S] was increased.
4.5.4 Effect of CP14 against DEN2 Viral Replication

To investigate further the inhibition mechanism of CP14 against the viral replication in the HepG2 cell lines, Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) experiments were carried out. RT-PCR is used to detect and quantify a specific sequence in a DNA sample as absolute number of copies or relative amount when normalized to DNA input or additional normalizing genes. The RT-PCR experiments were carried out in the Department of Molecular Medicine, University of Malaya (Mudiana, 2010) to ascertain the numbers of gene copies of the DENV2 when exposed to CP14.

![RT-PCR of DENV2 serine protease from HepG2 cell in the presence of CP14](source: Mudiana Thesis, 2010, taken with permission)

The profile generated for CP14 (Figure 4.9) indicated no apparent changes in the DENV2 protease gene copies. This is seen from 820 bp band which neither decreased nor increased in its intensity as compared to the positive control of DENV2. It is therefore concluded that CP14 has no influence on the DENV2 protease gene expression and was not affecting the viral replication process.
4.6 Discussions

The compound, CP14 was designed based on a template generated from two natural products and a synthesised small ester molecule which have shown competitive inhibition towards the NS2B/NS3 serine protease activities. To validate the ligand design, CP14 was initially bioassayed with the DENV2-infected HepG2 cell lines. The results exhibited a good CPE inhibition by CP14. At 10 μg/ml of CP14, which was way below its MNTD (25 μg/ml), high percent inhibition was exhibited (80%) at the highest virus titre (1000 TCID) used. This result warranted a further investigation on the inhibition mechanism exhibited by CP14. The kinetic assays were therefore carried out to examine the effect of CP14 against the DEN2 NS2B/NS3 serine protease.

When the CP14 was screened on the NS2B/NS3 serine protease recombinant, the Lineweaver-Burk plot analysis exhibited a non-competitive-like inhibition behaviour at two lower concentration series ([I] = 70.7 μM and 176.8 μM). However, while the plot of velocity versus [S] gave hyperbolic curves for [I] = 70.7 μM and 176.8 μM, doubling the [I] from 176.8 μM to 353.7 μM seemed to make the enzyme behave differently, i.e., with [I] =353.7 μM, a hyperbolic curve and the sigmoidal curve was observed. Thus, the mode of inhibition exhibited by CP14 could not be confirmed and further investigation is required.

For the RT-PCR experiments, non-apparent changes of the intensity of the NS2B/NS3 gene copies observed when CP14 was introduced to the DENV2-infected HepG2 cells (Figure 4.9). This observation was indicated that CP14 has no significant
effect on the viral replication process. However, since inhibition of CPE of the infected HepG2 cells was observed, it can be concluded that CP14 is still inhibiting the virulence of DENV2.

The design of CP14 was performed based on the site where the natural products, (panduratin A and 4-hydroxy panduratin A), were observed to bind in silico, to the binding site, presumably the active site of the enzyme. These natural products exhibited competitive inhibition in the protease assay (Tan et al. 2006). However, the bioassay results of CP14 did not clearly reflect the expected inhibition mode. Although in theory, CP14 showed good binding to the active site in silico (docked energy = -11.4 kcal/mol), it may not even reach the binding site of the protein experimentally to participate in competitive inhibition. To address this issue, a blind docking (Hetényi and van der Spoel, 2002) was performed between the whole DEN2 NS2B/NS3 serine protease and CP14 to locate other plausible ligand binding sites using AUTODOCK3.05 software package (Morris et al., 2001).

The blind docking results revealed 4 binding pockets including the active sites (Figure 4.10). From a hundred runs of blind docking on the whole NS2B/NS3 serine protease, about 65% of the molecule CP14 populated at the active site. However, 35% of CP14 docked at sites other than the active sites which could be alternative binding sites for CP14. The difference of the mean docked energy for these two top docked clusters was only 0.46 kcal/mol. This suggested that there could be other possible binding sites that may bind CP14 readily, and consequently the possibility of other mode of action by CP14.
Figure 4.10: Plausible bindings suggested by AUTODOCK3.05. A, B, C, D represents 4 different sites of binding, where A is the enzyme active sites. The binding ligands of CP14 were superimposed and displayed as CPK.

CP14 used for the assay were racemic since no chiral resolution was carried out on the mixture prior to the screening. Although not very likely, this could cause some variation in the inhibitory pattern observed since the pair of enantiomers may have different effect from each other on the enzyme and could alter the profile of action of the enzyme.
CHAPTER 5

GENERAL DISCUSSION

Dengue is now an emerging disease, particularly in tropical countries. While the effective medications or vaccines still remain unavailable, many efforts are made to find the best therapeutics for this disease.

Competitive inhibition against DEN2 NS2B/NS3 serine protease was observed by 4-hydroxypanduratin A and panduratin A that were purified from the rhizome extract of Boesenbergia rotunda (Tan et al., 2006). Using this information, the SBDD work was employed to design a new class of drug candidate. This work began with the construction of the protein model of dengue virus serine protease, NS2B/NS3 using homology modelling techniques. The sequence alignment and protein folding of the DEN2 NS2B/NS3 serine protease were built based on the protein crystal template of HCV NS3/NS4A (Kim et al., 1996), where this technique (Brinkworth et al., 1999) was revisited with more analyses and comparisons. Although the NS3 protein were successfully crystallised (Murthy et al. 1999), the absence of the NS2B as the protease co-factor within the crystal has caused some uncertainty for it to be the appropriate model to work on, since this serine protease requires NS2B to bind in as the co-factor for a more effective proteolytic activity. NS3 without NS2B exhibited comparatively lower proteolytic activity by more than 3000 fold (Yusof et al., 2000). In addition, the better reading of VERIFY3D (90.4) and ERRAT (77.1) gave better score on the structural verification of the homology model NS2B/NS3 serine protease which may implicate a better protein structure compared to DEN2 NS3 crystal structure.
The homology model was then used to study the binding interactions between the protein and the competitive inhibitors. To do this, protein-ligand docking was performed using AUTODOCK 3.05 as the molecular docking tools. 4-hydroxypanduratin A, panduratin A and ethyl 3-(4-(hydroxymethyl)-2-methoxy-5-nitrophenoxy)propanoate were used as the ligands. The docking studies were focused on the catalytic triad of the protein, and these bound ligands were observed to exhibit hydrogen bonding with the residues Ser-135 and Gly-151; Van der Waals interactions with Ser-131, Pro-132, Tyr-150 and Asn-152; and hydrophobic interactions in the hydrophobic pocket comprising Leu-128, Pro-132 and Val-155. The hydrogen bonding observation may implicate the involvement of these residues to the oxanion hole formation, which was found to be different from previously reported observations (Murthy et al., 1999). The lack of the additional $\pi$--$\pi$ interaction as well as the additional hydrophobic interactions with Val-52 could result in the lower inhibitory activity exhibited by ester 3, as compared to panduratin A and 4-hydroxypanduratin A.

A new ligand was designed based on the template for interactions obtained between the natural products and the protease. This newly designed ligand is hoped to exhibit the similar binding interaction or maybe even better binding affinity for the DEN2 NS2B/NS3 serine protease. For the design, the phenol moiety found in the panduratin A was kept constant but the position of the hydroxyl group was readjusted to give better binding possibility. In addition, it was observed in previous study (Yusof et al., 2000) that the DEN2 NS2B/NS3 serine protease recognises a dibasic site for cleavage. Based on this observation and to explore new chemical space, a piperidine functionality was included into the design of this new ligand. Docking of this newly designed ligand showed a better docked energy at -11.4 kcal/mol compared to 4-
hydroxypanduratin A (-10.2 kcal/mol) and panduratin A (-10.1 kcal/mol). This opened an opportunity of new drug candidate for antiviral activity for dengue virus serine protease.

A convergent synthesis plan that connects four different moieties was then adopted in the system of the newly designed ligand. This involved the synthetic strategy of dihydropyridine synthesis, 1,4-Michael addition and Weinreb ketone synthesis. The synthesis was carried out using nicotinic acid as the starting material. The newly designed compound, 1(benzo[1,3]dioxol-4-yl)(2-butyl-4-phenylpiperidin-3-yl) methanone (or CP14 in the chapter 4), was obtained in 10 step in about 4.7% overall yield. Although the synthesis attempted above was not a asymmetric strategy, it has exhibited some kind of stereocontrol, particularly during the Michael addition reaction. This is then give some insights into the strategy which could be used for a stereocontrolled synthesis in future work.

Preliminary biological studies to evaluate the effectiveness of this newly designed ligand on DEN2 virus activities were carried out in the Department of Molecular Medicine, University of Malaya. On the DEN2 virus infected HepG2 cells, about 80% inhibition of the CPE was exhibited by this ligand at 10 μg/ml, where it was below the MNTD of 25 μg/ml. Since the design of the new ligand was focused on the active site of the protease, it is expected to exhibit competitive inhibitory activities similar to its mould (panduratin A and 4-hydroxypanduratin A). However, this compound did not show competitive inhibition when tested against the protease assay. Although the compound showed, in theory, good binding to the DEN2 NS2B/NS3 serine protease, experimentally, it may not bind to the active site as expected. Thus, although inhibitory activity was observed, it may not be competitive in nature. RT-PCR experiment indicated that the ligand has little or no influence on the DENV2 serine
protease gene expression, and thus it was not affecting the viral replication. Blind docking of the newly designed ligand showed about 65% probability of the ligand docked to the active site. About 35% probability indicated ligand to be docked at other site of the protease. This result seems to support possibility of other binding site for the newly designed ligand which could explain the lack of competitive nature of the ligand.
CHAPTER 6

CONCLUSIONS

In an effort to generate new drug candidates as therapeutic agents against dengue disease, structural-based drug design, synthesis and biological evaluation approaches were employed in this work. In the absence of a protein crystal structure, a homology model of DEN2 NS2B/NS3 serine protease was built. The protein structure was verified under PROCHECK, VERIFY 3D and ERRAT.

Docking of ligands were performed using the existing competitive inhibitors (panduratin A, 4-hydroxypanduratin A and ethyl 3-(4-(hydroxymethyl)-2-methoxy-5-nitrophenox)propanoate to study the ligand binding interactions that contribute to enzyme inhibition activity.

Based on the binding interaction study, this ligand was then synthesised starting from nicotinic acid in 10 steps. Bioassay study a protease assay of the synthesised ligand, however, did not show competitive inhibition for the catalytic site of the NS2B/NS3 serine protease.