

MOLECULAR MODELING ON POTENT COMPOUNDS
TOWARD ENVELOPE PROTEINS OF DENGUE AND ZIKA
VIRUSES



Mr. Kowit Hengphasatporn

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Bioinformatics and
Computational Biology
Inter-Department of Bioinformatics and Computational Biology
Graduate School
Chulalongkorn University
Academic Year 2018
Copyright of Chulalongkorn University

การจำลองโมเลกุลของสารประกอบที่มีฤทธิ์ต่อโปรตีนเปลือกหุ้มไวรัสเดงกีและซิกา



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาชีวสารสนเทศศาสตร์และชีววิทยาเชิงคอมพิวเตอร์ สหสาขาวิชาชีวสารสนเทศศาสตร์และ
ชีววิทยาทางคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

โกวิท เสงประสาทพร : การจำลองโมเลกุลของสารประกอบที่มีฤทธิ์ต่อโปรตีนเปลือกหุ้มไวรัส
 เดงกีและซิก้า. (MOLECULAR MODELING ON POTENT
 COMPOUNDS TOWARD ENVELOPE PROTEINS OF
 DENGUE AND ZIKA VIRUSES) อ.ที่ปรึกษาหลัก : ผศ. ดร.ธัญญดา
 รุ่งโรจน์มงคล

มากกว่าครึ่งของประชากรโลกมีความเสี่ยงที่จะติดเชื้อไวรัสเดงกี และซิก้า ซึ่งเป็นสาเหตุนำไปสู่การตาย และอาการผิดปกติของสมองในเด็กได้ ในขณะที่ยังไม่มีหนทางในการป้องกัน หรือยารักษาที่มีประสิทธิภาพสำหรับต่อสู้กับการติดเชื้อไวรัสดังกล่าว ในงานวิจัยก่อนหน้ามีการรายงานถึงประสิทธิภาพในการยับยั้งไวรัสเดงกี และซิก้าด้วยสารอนุพันธ์ของฟลาโวน และสารอีพิกัลโลกาเทซินกัลเลต อย่างไรก็ตามกลไกในยับยั้งของสารยังคงคลุมเครือ ปัจจุบันการใช้เทคนิคเชิงคอมพิวเตอร์มาช่วยในการออกแบบโมเลกุลยา และการคัดกรองสารจากโครงสร้างเป็นเทคนิคที่สำคัญที่ใช้อธิบายข้อมูลเชิงลึกของรูปแบบการจับระหว่างสารยับยั้งกับโครงสร้างโปรตีน และใช้สำหรับคัดกรองสร้างเนื่องจากเป็นวิธีที่ใช้เวลาน้อย และมีความแม่นยำสูง ในงานวิจัยนี้ผู้วิจัยได้แสดงรูปแบบการเข้าจับของสารออกฤทธิ์ต่อไวรัสเดงกี และซิก้าในระดับอะตอม และได้มีการค้นหาสารที่มีประสิทธิภาพในยับยั้งไวรัสเดงกีที่โปรตีนเปลือกหุ้มไวรัส โดยมีการผสมผสานหลากหลายเทคนิคเชิงคอมพิวเตอร์ ได้แก่ ระเบียบวิธีโมเลกุลาร์ค็อกกิ้ง พลวัตเชิงโมเลกุล ร่วมกับการคำนวณค่าพลังงานอิสระ รวมไปถึงการคัดกรองสารด้วยเทคนิค pharmacophore-based virtual screening โดยรูปแบบการเข้าจับของสารออกฤทธิ์กับตำแหน่งเฉพาะบนโปรตีน รวมไปถึงรูปแบบอันตรกิริยาของสารที่มีต่อโปรตีนเปลือกหุ้มของไวรัสเดงกี และซิก้า สามารถใช้เป็นต้นแบบในการออกแบบโมเลกุลยาที่มีความเฉพาะเจาะจงในอนาคต

จุฬาลงกรณ์มหาวิทยาลัย
 CHULALONGKORN UNIVERSITY

สาขาวิชา	ชีวสารสนเทศศาสตร์และชีววิทยา เชิงคอมพิวเตอร์	ลายมือชื่อนิสิต
ปีการศึกษา	2561	ลายมือชื่อ อ.ที่ปรึกษาหลัก

5887819220 : MAJOR BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

KEYWORD Dengue Zika MD simulation Molecular docking Pharmacophore
D:

Kowit Hengphasatporn : MOLECULAR MODELING ON POTENT COMPOUNDS TOWARD ENVELOPE PROTEINS OF DENGUE AND ZIKA VIRUSES. Advisor: Asst. Prof. THANYADA RUNGROTMONGKOL

More than half of the world's population has a high risk of dengue and Zika viral infection that is one of a leading cause of death and brain's abnormality in children whereas there are no effective prevention or therapeutic agents to cure this disease. Recently, the flavanone derivative (FN5Y) and epigallocatechin gallate (EGCG) have been reported the inhibiting efficacy against dengue and Zika viruses, however, the mechanism of inhibitors is still ambiguous. Nowadays, the computer-aided drug design (CADD) and structure-based virtual screening method (SBVS) become the important techniques to reveal insight into the inhibitor's binding pattern and help drug discovery and development. Herein, the binding pattern of dengue and Zika inhibitors at the atomistic level and the novel potent compounds against dengue viral envelop protein have been proposed by various computational approaches such as molecular docking, molecular dynamics simulation, free energy calculation, and pharmacophore-based virtual screening. The binding patterns and the interaction profiles of the potent molecules on the dengue and Zika E protein can be used for further drug discovery and design.



Field of Study: Bioinformatics and
Computational Biology
Academic 2018
Year:

Student's Signature
.....
Advisor's Signature
.....

ACKNOWLEDGEMENTS

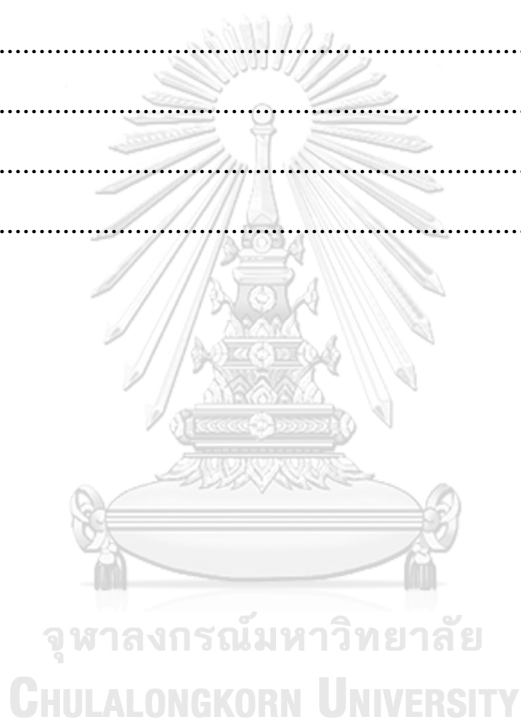
First of all, I would like to express the deepest appreciation to my main advisor, Asst. Prof. Dr. Thanyada Rungrotmongkol, for her guidance, patience, understanding and continuous encouragement throughout my study. I have had the opportunity to explore and improve my skill and knowledge that is so important for my career path in the future. I wish to thank Prof. Peter Wolschann, Prof. Thierry Langer, and lab member at the University of Vienna, Austria, for their help in providing technical guidance and kind suggestions. Assistance provided by Prof. Warinthorn Chavasiri and Prof. Siwaporn Boonyasuppayakorn were greatly appreciated. I owe a very important debt to Prof. Jin-Moon Yang, Prof. Jen-Shiang Yu, and lab member in National Chiao Tung University (NCTU), Taiwan, for their advice and good recommendations. I am also grateful to Prof. Shigeta Yasuteru and his lab members from University of Tsukuba, Japan, for providing me a great moment and the insight of quantum mechanics method that is applied in my study. My special thanks are extended to my fellows in BioSIM lab from Chulalongkorn University for their help and understanding.

Last but not least, I have received a lot of confidence and support throughout my study from my family.

TABLE OF CONTENTS

	Page
ABSTRACT (THAI)	iii
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS.....	1
Chapter I Introduction.....	3
1.1 Rationales.....	3
1.2 Background.....	3
1.3 Dengue and Zika viruses.....	6
1.4 Objectives	21
2.1 Molecular docking	22
Chapter III Material and methods.....	39
Overview of this work	39
3.1 Binding pattern evaluation.....	40
3.1.1 System preparation	40
3.1.2 Binding region identification	41
3.1.3 Binding pattern of inhibitor	43
3.2 Virtual screening.....	45
3.2.1 Pharmacophore-based virtual screening.....	45
3.2.2 Molecular docking	47
3.2.3 Binding energy calculation	48
4.1 Binding pattern evaluation.....	50
4.1.1 Binding region identification	50

4.1.2 <i>Binding pattern of inhibitor</i>	53
4.2 Virtual screening	65
4.2.1 <i>Pharmacophore-based virtual screening</i>	65
4.2.2 <i>Molecular docking</i>	71
4.2.3 <i>Binding energy calculation</i>	75
5.1 Binding pattern evaluation.....	83
5.2 Virtual screening.....	84
APPENDIX.....	85
PUBLICATION	91
CONFERENCE	91
REFERENCES	92
VITA.....	105



LIST OF TABLES

	Page
Table 1 Key steps of dengue replication and target site for pharmacological interaction. 14	
Table 2 CDocker interaction energy (kcal/mol) and hydrogen bonds of EGCG binding at the six possible binding sites on the Zika dimeric E protein.55	55
Table 3 Contributed residues of lead molecules bound to four binding sites on the dengue dimeric E protein, compare to FN5Y molecule.73	73
Table 4 Comparison of the averaged binding free energy and its components (kcal/mol) of F18 ligands bound to each binding region; K (blue), K' (red), X' (grey), and Y' (black) on dengue dimeric E protein, which is calculated from last 100 ns-MD trajectory. 77	77

LIST OF FIGURES

	Page
Figure 1	The burden of dengue virus in the world.....4
Figure 2	The transmission of the Zika virus in the world.....5
Figure 3	Transmission of dengue virus from person-to-person through its vector, <i>Aedes aegypti</i> mosquito 7
Figure 4	a) The ratio of dengue infection and its symptom illustrate as a pyramid diagram. b) The primary and c) secondary infection of dengue virus and immune response. 7
Figure 5	a) Morphology of flavivirus particle, b) envelope protein of dengue virus, and c) genome and function.9
Figure 6	The comparison of a whole genome and 9 viral protein sequences between dengue and Zika viruses using pairwise alignment method. 10
Figure 7	The major difference between dengue and Zika envelop protein. 11
Figure 8	The 3D structure of anti-parallel homodimer of dengue (PDB code: 1OKE) and Zika E protein (PDB code: 5iz7) and their amino acid sequence ... 12
Figure 9	Dengue and Zika viral life cycle..... 13
Figure 10	Schematic (A) and 3D structure (B) illustrates the conformational change of E protein when it lives in host endosome..... 14
Figure 11	Chemical structure of potent inhibitors against flavivirus at the different target proteins..... 18
Figure 12	2D structure of inhibitor against flavivirus entry process. 19
Figure 13	The schematic procedure of Molecular docking method.22
Figure 14	Ligands are docked to dengue E protein by blind docking method. 24
Figure 15	Ligand bound to specific binding region of dengue E protein using focused docking method.24
Figure 16	Generic evolution algorithm.....25
Figure 17	The procedure of <i>k</i> -mean clustering algorithm in this study.....26
Figure 18	Identification of the optimal number of clusters (<i>k</i> -value) using Elbow method for the <i>k</i> -mean clustering.26

Figure 19	Bonded and non-bonded interactions consider in MD simulation	28
Figure 20	TIP3P water model	29
Figure 21	The general procedure of MD simulation.....	30
Figure 22	Hydrogen bonding interaction between donor and acceptor atom. 31	
Figure 23	The comparison of accuracy and speed of each binding free energy calculation methods; QM, MM, SIE, and Dock.	32
Figure 24	The illustration of binding free energy calculation of ligand binds to protein receptor using MM-PBSA and MM-GBSA methods.	34
Figure 25	The schematic represents the QM and MM region identification. 35	
Figure 26	The step of paired interaction energy (PIE) calculation by fragment molecular orbital (FMO) method.	36
Figure 27	Types of pharmacophore feature are generated by Ligandscout program. 37	
Figure 28	The integration of MD trajectory and pharmacophore-based virtual screening procedure.....	38
Figure 29	The overview of this work.....	40
Figure 30	The 3D structure of dengue (a) and Zika (b) dimeric E protein with 1,000 ligand points grouped into 18 clusters as shown in different colors.....	51
Figure 31	Binding mechanism of FN5Y towards dengue envelope homodimer. 54	
Figure 32	Per-residue decomposition free energy based on MM-GBSA method for EGCG binding at the different sites	56
Figure 33	Close-up of EGCG binding orientations at all binding sites shown in grey circle, where the interacting residues are colored according to energy values from Fig. 4. The principal component analyses reveal in green circle of each binding region. The porcupine plots consist of arrow (green) and length that represent the direction of motion and amplitude of residue mobility.	57
Figure 34	The 2D projection of MD trajectories on the first two PCs (left) and the Scree plot of variances and accumulated percentage by the first ten PC modes (right) of each complex system.....	60

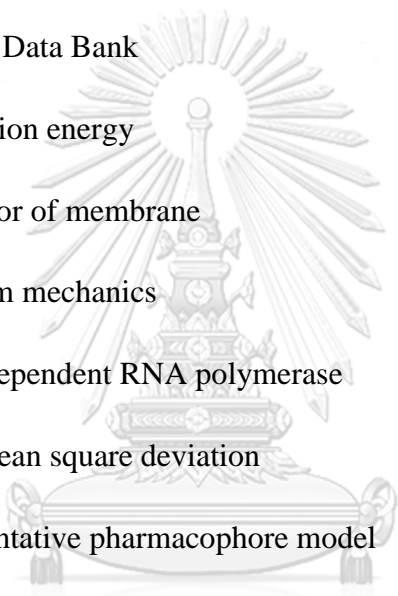
Figure 35	Hydrogen bonding interactions of EGCG at each site are presented by grid map, while hydrogen bond strength is defined by label and color in the grid cells.	62
Figure 36	Comparison of the average binding free energy (ΔG_{bind}) in kcal/mol of each EGCG binding at the different sites on Zika dimeric E protein at acidic environment and the Pearson correlation analysis (heat map) by using four different binding free energy calculation methods; SIE (triangle), MM-GBSA (square), QM/MM-GBSA with SCC-DFTB (black circle), and QM/MM-GBSA with PM3 method (gray circle).	64
Figure 37	RMSD graph of ligand/E protein (above) and F18/E protein complex based on molecular dynamics simulation trajectories during 100-ns simulations.	66
Figure 38	The 2D and 3D pharmacophore models of active ligand/E protein complex and the interacted residues in each pocket (K, K', X' and Y')	67
Figure 39	The 2D structures of 25 molecules were obtained from pharmacophore-based screening method.	69
Figure 40	Comparison between CHA score of screened molecule bound to all binding sites (right) and CDOCKER interaction energy score (left).	70
Figure 41	Tendency plot of screened molecule interaction energies compare between CDOCKER (black) and iGEMDOCK (grey) programs	72
Figure 42	Receiver operating characteristic (ROC) plot of pharmacophore model applied to dengue E protein.	74
Figure 43	Comparison of the average SIE binding free energy (ΔG_{bind}) in kcal/mol of each flavone derivative binding (FN5Y, F16, F17 and F18) at the different sites on dengue dimeric E protein (K, K', X' and Y').	76
Figure 44	RMSD plots of F18 ligand (blue) bound to 4 different sites; K, K', X', and Y', on the dengue E protein (black and grey) over 500 ns-MD simulation.	77
Figure 45	2D Pharmacophore model of potent compound (F18) interacts to residues in chain A (black box) and chain B (red box) at each binding site of dengue E protein.	78
Figure 46	Interaction profile of F18 bound to each binding sites (K, K', X', and Y') on dimeric dengue E protein along last 100 ns-md trajectory based on MM-PBSA and MM-GBSA method.	80
Figure 47	F18 binding orientations at the four binding sites	81

Figure 48 a) FMO2-MP2/6-31G(d) binding energy component between the K-site residues and F18. (b) Electrostatic attraction (red dashed line) and dispersion (blue dashed line) between K128 and flavone moieties.82



LIST OF ABBREVIATIONS

ADE	Antibody-Dependent Enhancement
C	capsid
CADD	computer-aided drug design
CHA	common hits approach
CKV	Chikungunya virus
CPE	cytopathogenic effect
DENV	dengue virus
DHF	dengue hemorrhagic fever
DSS	dengue shock syndrome
E	envelope protein
EGCG	epigallocatechin-3-gallate
FMO	fragment molecular orbital methods
FN5Y	flavanone derivative
GBS	Guillain-Barré syndrome
HF	Hartree-Fock
HTVS	high-throughput virtual screening
JEV	Japanese Encephalitis virus
MD	molecular dynamics
MM-GBSA	molecular mechanics generalized born surface area method
MM-PBSA	molecular mechanics Poisson-Boltzmann surface area method



MP2	Møller–Plesset
MTase	methyltransferase
NPT	isothermal-isobaric ensemble
NS	non-structural protein
PBC	Periodic boundary condition
PCA	principal component analysis
PDB	Protein Data Bank
PIE	interaction energy
prM	precursor of membrane
QM	quantum mechanics
RdRp	RNA-dependent RNA polymerase
RMSD	Root mean square deviation
RPM	representative pharmacophore model
SASA	Solvated accessible surface area
SBVS	structure-based virtual screening method
SCC-DFTB	Self-consistent charge density functional tight- binding
SCF	self-consistent field
SIE	solvated interaction energy
SSE	sum of squared errors
TGN	trans-Golgi network
YFV	Yellow Fever virus
ZIKV	Zika virus

Chapter I

Introduction

1.1 Rationales

More than half of the world's population lives in tropical and sub-tropical regions that shown a high risk of dengue and Zika viral infection. This infection ranges from mild fibril illness to severe symptom that is one of a leading cause of death in children. There is no effective prevention or therapeutic agents to cure this disease. However, many candidates for inhibitors are in the process of drug discovery and development. This traditional procedure is time consuming in the range of 10-15 years. Due to this limitation, the computer-aided drug design (CADD) and structure-based virtual screening method (SBVS) become important techniques to perform drug screening to a high accuracy with a reduction of time. The binding pattern of an inhibitor on its target protein is necessary for virtual screening in order to search the new potent compound. Several computational tools, such as all-atom MD simulations, principal component analysis (PCA), and binding free energy calculations are used for determining the stability and binding affinity of a ligand to the viral surface at different binding sites. Moreover, pharmacophore based-virtual screening and molecular docking techniques are also essential tools to search for the novel inhibitor.

1.2 Background

The family *Flaviviridae* is enveloped viruses. They can be transmitted between people by means of a bite from an infected insect such as a tick or mosquito. Japanese Encephalitis virus (JEV), Yellow Fever virus (YFV), Chikungunya virus (CKV), dengue virus (DENV) and Zika virus (ZIKV), belong to the Flavivirus genus [1, 2]. These viruses and their vectors are spread throughout tropical and subtropical regions (Fig. 1 and 2). Southeast Asia is the highest risk zone. In 2018, Thailand vector-borne disease center

reported 84,830 dengue infections leading to 109 deaths with a fatality rate of around 0.13%, which is the highest percentage in the last decade. This virus is not only spread in Southeast Asia, but it also emerges in East Asia, including China, Korea, and Japan. In case of Zika viral infection, there is a huge outbreak over the world, while Thailand is one of the countries in the Southeast Asia with a high impact. The World Health Organization (WHO) declares people in 84 countries and territories have infected by the Zika virus.

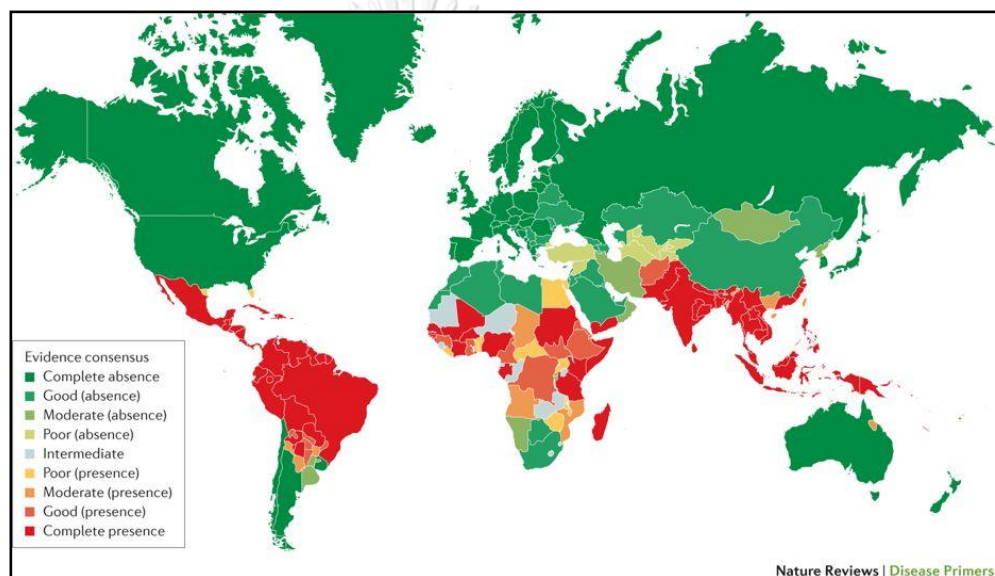


Figure 1 จุฬาลงกรณ์มหาวิทยาลัย
The burden of dengue virus in the world [3]
CHULALONGKORN UNIVERSITY

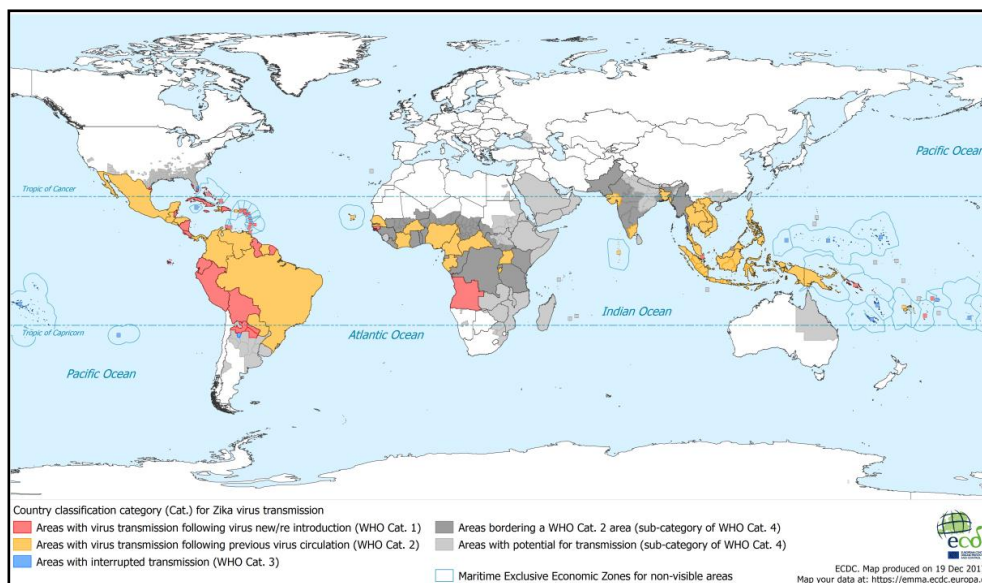
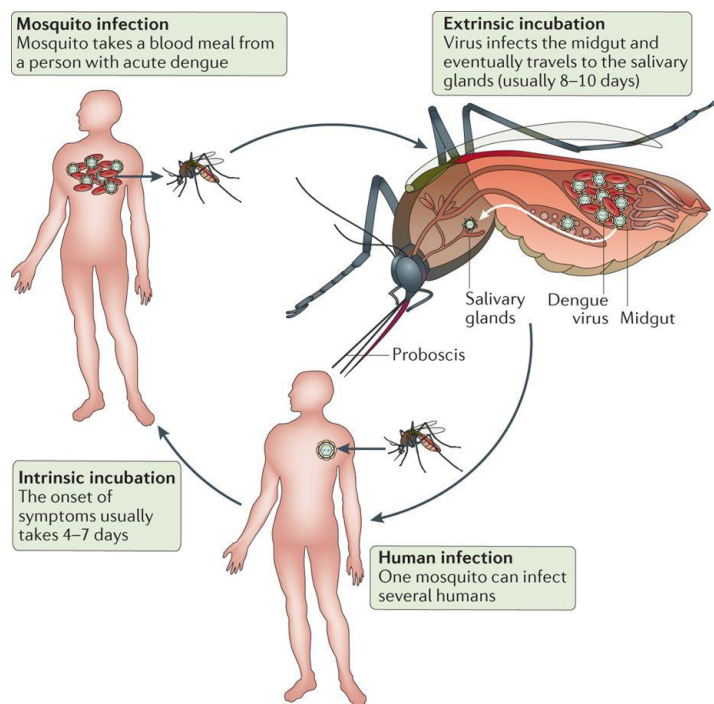


Figure 2 The transmission of the Zika virus in the world

A vaccine for dengue has very recently launched, however, its efficacy and side effects are still unclear due to the Antibody-Dependent Enhancement (ADE) to a viral infection. The commercial drug for dengue and Zika viruses is not yet available. Some inhibitors are currently in the process of drug discovery [4-6]. Many researchers have focused on a discovery of novel potent compounds from synthetic or natural products against viral targets, but it is not fast enough for fighting against these viruses. The computer-aided drug design (CADD) is very useful for high-throughput virtual screening (HTVS) of a large amount of compounds from databases. The CADD approach increases the chance of success in the drug discovery and reduces the development time and the cost of chemical reagents [7-11]. By the major key of drug discovery and design, we need to know the correct target that plays an important role in the organism's life cycle.

1.3 Dengue and Zika viruses

Dengue and Zika viruses cause a mosquito-borne disease, which is transmitted by *Aedes* mosquito species. Generally, the transmission of them cannot be directly passed from person to another. It requires the mosquitoes as the primary vectors of the virus. But in some case, the virus can be directly transmitted from mother to her fetus. These mosquitoes live in urban with the close-human environment during day time. Adult female infected mosquitoes contain viruses, in salivary glands which can exist along the mosquito's life. The infected mosquito bites persons who can get the viruses in the body and after 4-7 days, probably up to 10 days and the symptom of disease may occur (Fig. 3). Zika virus can directly spread through person-to-person by sexual transmission or through the saliva of infected animals [12]. Zika virus can be found in urine, vaginal fluid, blood, semen and saliva for at least 3 months [6, 13-15]. The infection of dengue virus related to four serotypes, of which each has slightly different antibody interaction. The symptom of dengue fever ranges from mild to severe (Fig. 4a), for example, severe headache, high fever, fatigue, severe joint and muscle pain, rash on skin, nausea, vomiting, and shortness of breath [13, 16]. The severity depends on several factors age; the child will be more severe than adult [17]. The serotype complexity and Antibody-Dependent Enhancement (ADE) of dengue can lead to dengue severity (Fig. 4b and 4c) that caused by secondary infection with the different dengue virus serotypes or Zika infection. This is able to develop from dengue fever to dengue hemorrhagic fever (DHF) or dengue shock syndrome (DSS). Immunoglobulin in blood patient is specific to each viral serotype, but all serotypes are slightly different about 35-45%, which lead to immunopathogenic response coursing the virus to rapidly replicate in multiple organs [18-21]. Many patients infected by the Zika virus may not reveal the symptom or they might show a mild fever leading to febrile illness, called Zika fever, but this virus may cause neurologic disorder in infants or microcephaly that can interrupt the fetus' brain development in every pregnancy stage. The Zika might cause Guillain-Barré syndrome (GBS) and conjunctivitis in adults [22, 23].



Nature Reviews | Disease Primers

Figure 3 Transmission of dengue virus from person-to-person through its vector, *Aedes aegypti* mosquito [19].

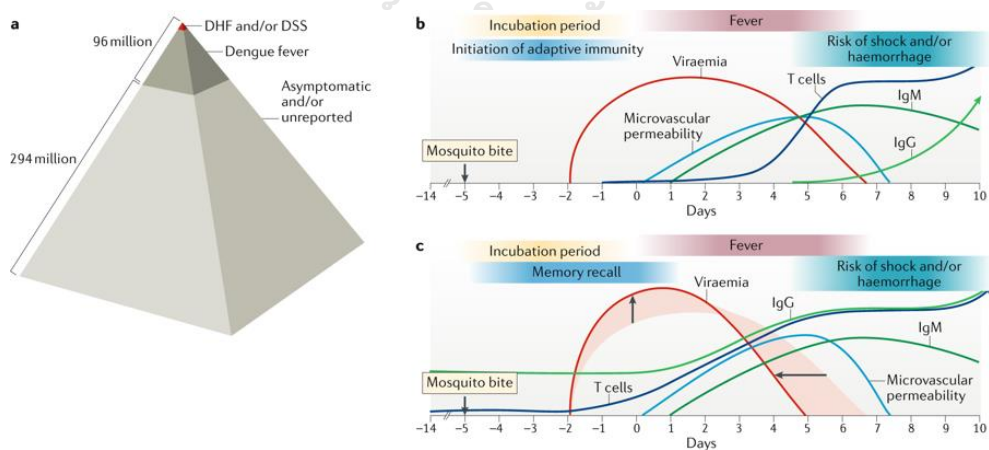


Figure 4 a) The ratio of dengue infection and its symptom illustrate as a pyramid diagram. b) The primary and c) secondary infection of dengue virus and immune response. [21].

Flaviviruses are a small spherical particles which consist of 4 main parts; envelope protein, lipid membrane, capsid and genome (Fig. 5a) [24]. The genome of flavivirus is infectious, which is approximately 11,000 nucleotides in length. RNA is positive-sense with a single strand that can be translated into a single polyprotein that can be cleaved into 10 proteins. The N-terminal region of polyprotein encodes 3 structural proteins; capsid (C), membrane (prM), and envelope (E) protein, followed by 7 non-structural (NS) proteins; NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Fig. 5c) [1, 25]. The structural proteins facilitate the viral attachment, entry, assembly, and budding processes. The NS proteins have functions in many steps of the viral life cycle, including viral replication, viral assembly and viral release. The capsid protein is involved in RNA binding and nucleocapsid packing via the opposite charged residues. It consists of about 114 amino acids in a mature stage [26, 27]. The precursor of membrane (prM) is translocated into the ER by a signal sequence provided by capsid. The immature virion shows a spike-like particle that is containing a tight complex of prM covering the fusion peptide of E protein trimer (Fig. 5a). The prM has been shown in immature virion until the virus exposed under the neutral pH. In the trans-Golgi network (TGN), the prM was cleaved and released out of E protein [28].

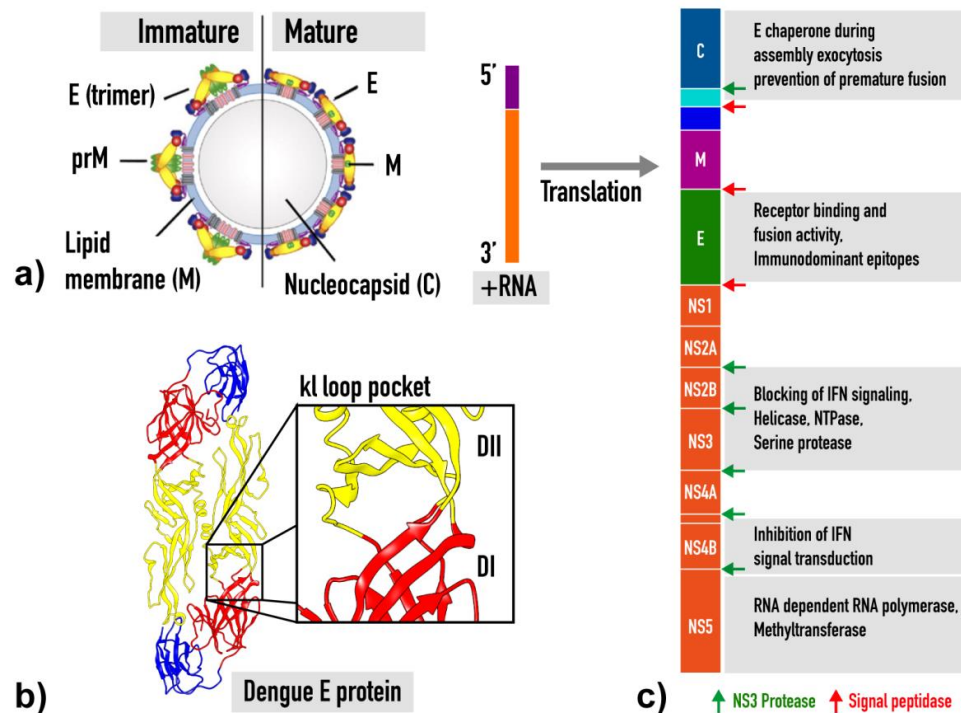


Figure 5 a) Morphology of flavivirus particle, b) envelope protein of dengue virus, and c) genome and function.

Dengue and Zika viruses are closely related species about 56.3% identity and 72.6% similarity based on their genome. They belong to a member of the genus *Flavivirus*. The comparative sequence similarity of each protein between dengue and Zika viruses based on pairwise alignment that shown in Fig. 6. The most similar protein sequence among dengue and Zika virus is NS3 (66.7 % identity and 80.8% similarity) and the lowest similarity is NS2A (27.5 % identity and 47.0% similarity). Nucleocapsid is covered by host-cell-derived lipid bilayer and the outermost is the envelope protein (Fig. 5b) [29]. Capsid protein (37.1 % identity and 62.9% similarity) is icosahedral in shape, and protects the viral genome from toxic agents. The capsid protein is essential induced to specific encapsidation of viral RNA [30]. Moreover, host ubiquitin-proteasome is the one factor for RNA release into the host cell cytoplasm [31]. The membrane is a key for viral maturation and role in forming part of the premembrane – envelope protein (prM-E) complex [32]. Envelope protein

(54.5 % identity and 74.9% similarity) is an anti-parallel homodimer, in which each monomer consists of 3 domains (I, II and III) and packed about 180 copies as raft structure on the viral surface. There is a slight difference between dengue and Zika E protein at the region near the glycosylation site and kl loop region based on pairwise alignment (Fig. 7 and 8)[33]. The hydrophobic region located at the end of E protein (DII) is particularly conserved among flavivirus. E protein is involved with viral absorption and viral entry into host cells by receptor-mediated endocytosis. Therefore, E protein is a promising target for screening, designing and developing potent compounds against dengue and Zika viral infections. The acidic condition is a major key of E protein conformational change [25, 34-36]. While these stay in the endosome, the class II fusion protein at the tip of domain II exposes and inserts to host endosomal membrane, and then nucleocapsid is released into host cytoplasm. Not only pH condition, but environmental temperature also affects to the conformational stability of E protein [37].

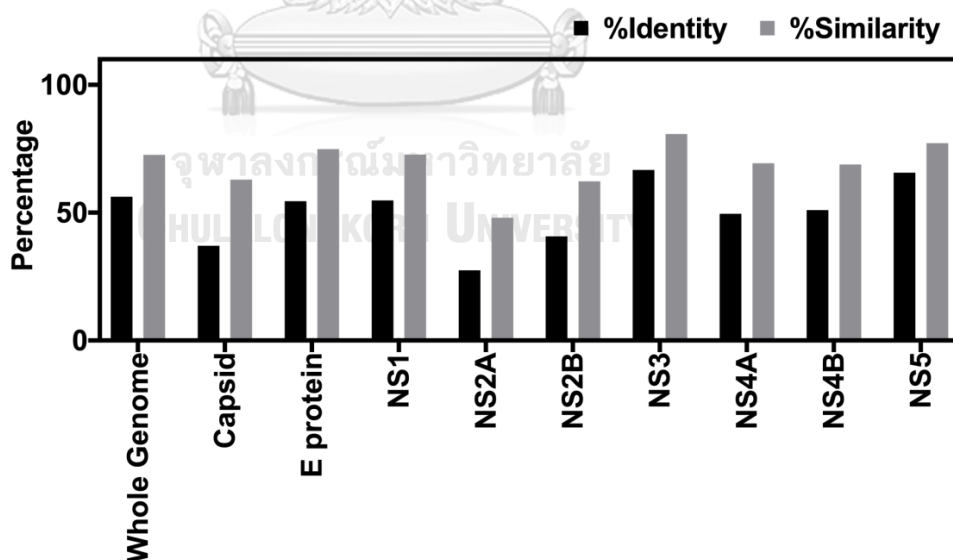


Figure 6 The comparison of a whole genome and 9 viral protein sequences between dengue and Zika viruses using pairwise alignment method.

replication. NS5 encodes to methyl-transferase (MTase) and RNA-dependent RNA polymerase (RdRP) which plays a vital role in virus replication [46, 47]. Translated viral genome and each function are shown in Fig 5c [48].

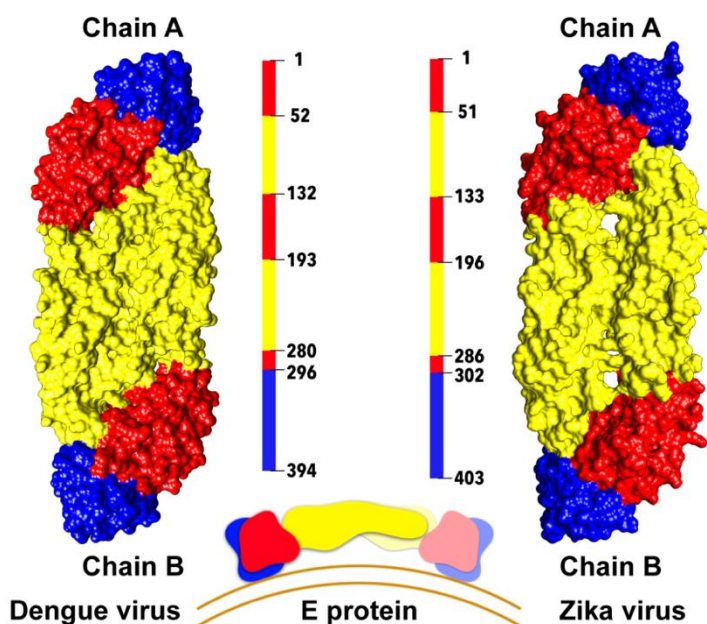
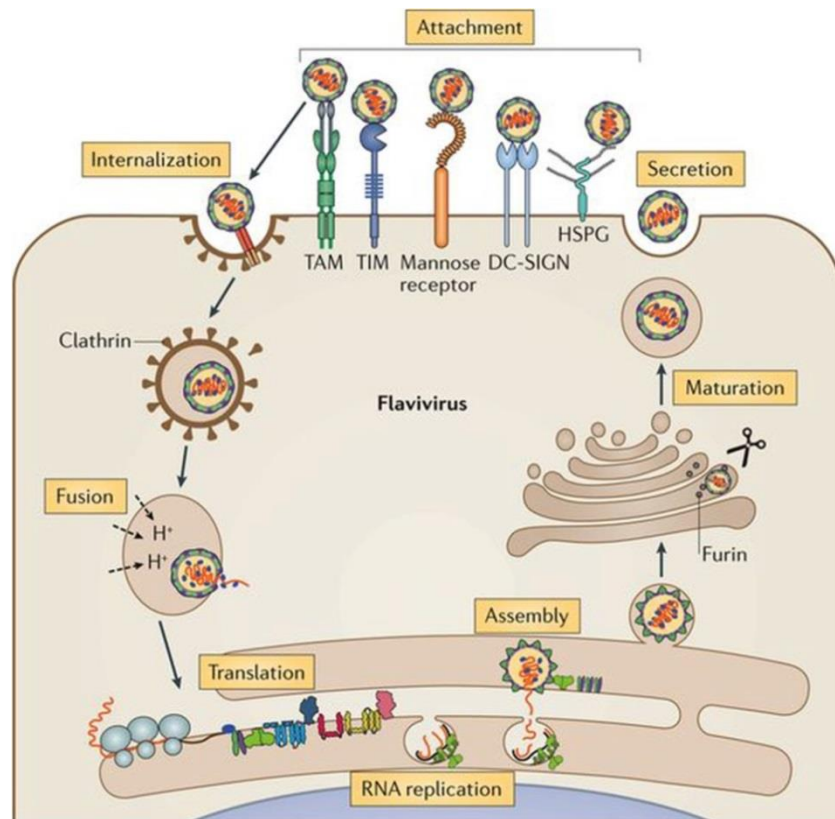


Figure 8 The 3D structure of anti-parallel homodimer of dengue (PDB code: 1OKE) and Zika E protein (PDB code: 5iz7) and their amino acid sequence, in which each monomer comprises of 3 domains; DI (red), DII (yellow) and DIII (blue).

Several natural compounds have been found to interrupt in many steps of the viral life cycle as shown in Fig. 9 [49], including viral attachment (host receptor and E protein), entry process (E protein), replication (NS protein), and assembly (C protein). Viral replication is a necessary process of the virus, when the viruses float in the blood stream. They are attached to host cell receptor such as heparin sulfate, clathrin receptor, DC-SIGN, Mannose sugar, TIM, and TAM before entering into the host cell through receptor mediated endocytosis. In the endosome, the structure of envelop protein will be changed

to an active stage due to the proton pumping from the endosomal membrane (Fig. 10). The viral RNA is releasing into host cytoplasm, called entry stage. Host genome is manipulated by viral RNA and synthesized viral non-structural proteins in the cytoplasm like functional enzyme, which is used in polymerization, transcription and translation. Then viral particles will be assembled before released out of the host cell and spread into the patient's body. Key steps of these are represented in Table 1 [50].



Nature Reviews | Microbiology

Figure 9 Dengue and Zika viral life cycle [49].

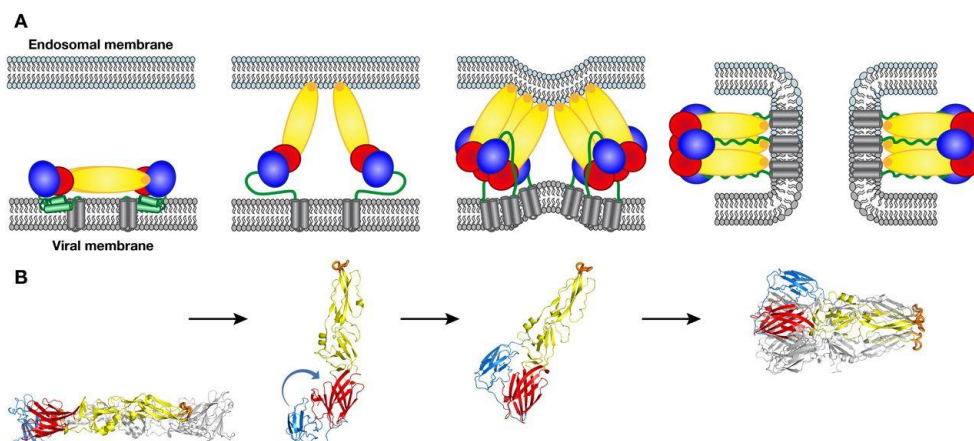


Figure 10 Schematic (A) and 3D structure (B) illustrates the conformational change of E protein when it lives in host endosome [51].

Table 1 Key steps of dengue replication and target site for pharmacological interaction.

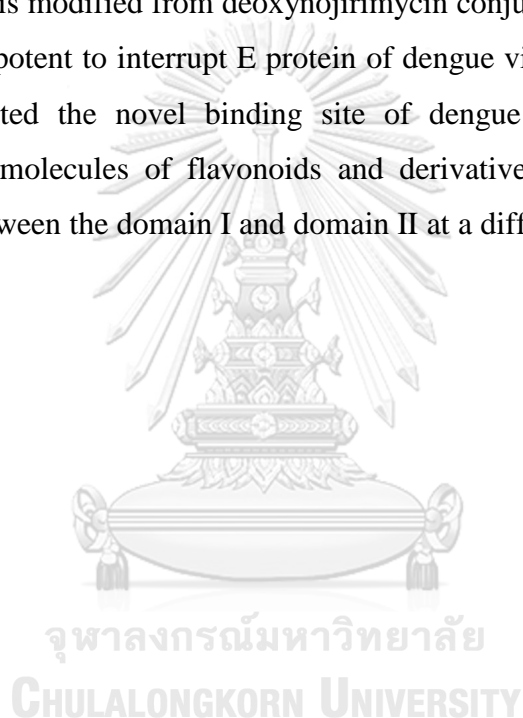
No.	Event in viral replication	Target(s)
1.	Adsorption or attachment	Dengue envelope protein; host targets
2.	Endocytosis, uncoating	Host targets
3.	Translation of viral polyprotein	Host targets
4.	Cleavage and post-translation	Viral NS3 protease; (host protease); alpha-glucosidase
5.	Replication complex	Various targets
6.	Transcription to - RNA	Viral NS5 polymerase; NS3 helicase; NS4B
7.	Transcription to + RNA	Viral NS5 polymerase and methyltransferase; NS3 helicase; NS4B; host SAHase
8.	Assembly	Viral capsid protein C; host targets
9.	Releasing	Host targets

Up to date, no effective vaccine or a specific antiviral drug is available for the treatment of dengue in human. Hence, the therapeutic application needs to be developed immediately [3, 28, 52]. Analgesic; NSAIDs and aspirin, should be abstained, therefore this drug can interfere platelet mechanism that leads to hemophagic symptoms. Many relevant antiviral drug research groups have discovered the potent compounds against dengue and Zika viruses, in which the antiviral drug is obviously in the process of drug discovery and design. Mechanism of antiviral agents against targets; virus and host-cell, has several approaches. The common timeline of drug discovery takes about 10-15 years from target selection to approved drug [53]. To overcome this limitation, the integration of experimental and computational approaches lead to time reduction and often improves quality of the process [6, 54, 55]. One of the challenges for drug discovery by computational method is how to search for the correct target. Many viral proteins have been reported about its potential to be a target protein for virtual screening.

Most researches focus on non-structural proteins which are enzymatic targets related to viral replication and translation processes. NS2B/3 protease is mostly studied to discover the potent inhibitor; small molecules, short peptide inhibitors, analogs, and allosteric ligands, for this target of dengue and Zika viruses [56-65]. NS3 helicase and NS5 MTase and RdRp are also used as a target protein to find potent inhibitors (Fig. 11) [66-71]. On the other hand, the structural proteins are able to be a potential target for a broad spectrum of inhibitors because they show highly conserved regions among flavivirus [72]. Capsid is an immensely positively charged protein consists of about 25% of arginine and lysine [73]. Two interesting inhibiting targets of the capsid are the hydrophobic core and the N-terminus region which is used for drug development. According to Byrd et al., in 2013, ST-148 (Fig. 11) can induce CPE and reduces virus titer in the low concentration for all serotypes of dengue virus [74]. The small peptide also can be used as a capsid protein inhibitor [75-77]. The related researches revealed the mechanism of this small molecule that has an antiviral effect on the entry, assembly and viral release

out of infected host cells. The E protein is a structural protein without specific binding region. Notwithstanding, one of the known inhibiting targets for the viral entry mechanism of the E protein is the kl loop or n-octyl- β -d-glucoside (β -OG) pocket [78], the hydrophobic region between domains I and II. In addition, several domains of E protein can be used as the inhibiting target sites; stem domain, the β -OG hydrophobic pocket (kl loop) and the receptor binding domain III [79]. N-octyl- β -D-glucoside (β -OG) is a detergent that is inserted into the kl loop pocket during crystallization, but it cannot be an inhibitor against dengue virus. Short peptides and small molecules play a key role as they inhibit viral agents at E protein. The designed small peptide interacts with the hinge of domain II stem region, which shows an increased hydrophobicity and enhanced inhibitory strength [79-81]. This region has been suggested as a promising ligand binding site in order to inhibit dengue virus at the earliest step of infection [78, 82-84]. According to Wang et al., a pyrimidine derivative (compound 6 in Fig. 11) and quinazoline (compound A5 in Fig. 11) can serve as molecular probes for the study of the entry of flavivirus into host cells. However, the experiment in mouse show that both compounds are agglomerated in the digestive tract [85]. Doxorubicin (Fig. 12) has shown an antiviral effect against dengue infected cell based on cell culture experiment with lower cytotoxicity. An analog of doxorubicin is not able to inhibit dengue serotype 4, but it can enhance antibody activity and can reduce the immature dengue serotype 2 [86]. According to Jadav et al., a novel hybrid inhibitor of β -OG pocket is reported that is developed from two hits identified by Zhou et al. and Li et al. [40, 87]. The receptor binding domain III of E protein shows adsorption on glycosylation site and viral replication. This binding region near glycosylation site can interact with several receptors on the host cell surface; heparan-sulfate proteoglycans or syndecans according to previous studies, which are aim at the discovery of new entry inhibitors mimicking heparan-sulfate moiety of these receptors [88]. The anticoagulant activity and low bioavailability are also the major problems of these compounds. Unlike, curdlan sulfate binds to the interface between domain II and III has shown a good level in plaque forming assay and its advantage is a

low anticoagulant activity [89]. In previous study, MLH40, a small peptide can be served as a candidate for an antiviral drug because it can reduce the viral infectivity to host cells in all dengue serotypes [90]. Teicoplanin derivative is another compound that binds to E protein and blocks replication observed by the experimental study (Fig. 12) [91]. Polyphenolic lipid compounds, cardol triene, could interrupt at kl loop of E protein of dengue virus, but not in the Zika virus according to Kanyaboon, et al. [36]. The iminosugar (UV-4) has been reported about inhibiting efficacy. This compound is modified from deoxynojirimycin conjugate with alkyl side chains that could potent to interrupt E protein of dengue virus [92]. Ismail and Jusoh have reported the novel binding site of dengue E protein based on the consensus molecules of flavonoids and derivatives. This binding region is located between the domain I and domain II at a different chain [93].



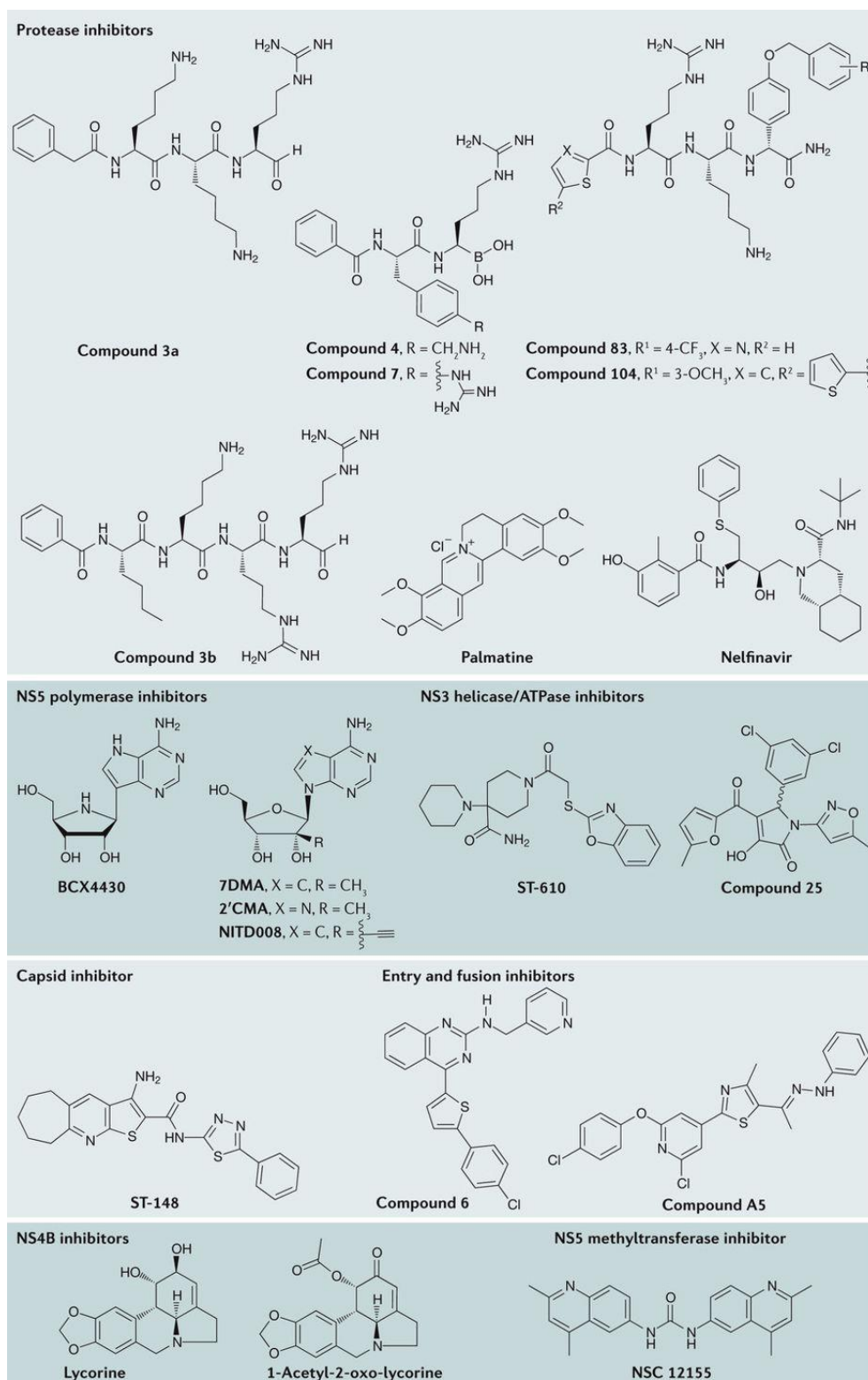


Figure 11 Chemical structure of potent inhibitors against flavivirus at the different target proteins [72].

Recently, *in vitro* and *in silico* study has shown that a flavanone derivative (FN5Y) can interfere and interrupt at envelope protein of dengue virus [4]. The FN5Y molecule (Fig. 12) attaches to 4 binding sites of dimeric dengue E protein. This compound could disturb the initial step of the viral infection by blocking the re-arrangement of E protein dimer to trimer conformation. The mechanism of inhibition has been proposed by computational studies [83, 84]. Carneiro and co-workers have reported that the Zika viruses can be inhibited by (-)-epigallocatechin-3-gallate (EGCG) as shown in Fig. 12. The EGCG is a natural polyphenol product found in green tea (*Camellia sinensis*) [6, 54, 55]. They believe that this compound may has an effect on Zika virus in the viral entry process. Moreover, the 50-ns molecular dynamics (MD) study on the Zika E protein monomer at pH 7 suggested that the binding site of EGCG is located between domains I and III [55]. However, the complete understanding on the inhibitory mechanism of EGCG on the anti-parallel homodimer of Zika E protein in microscopic level remains unclear.

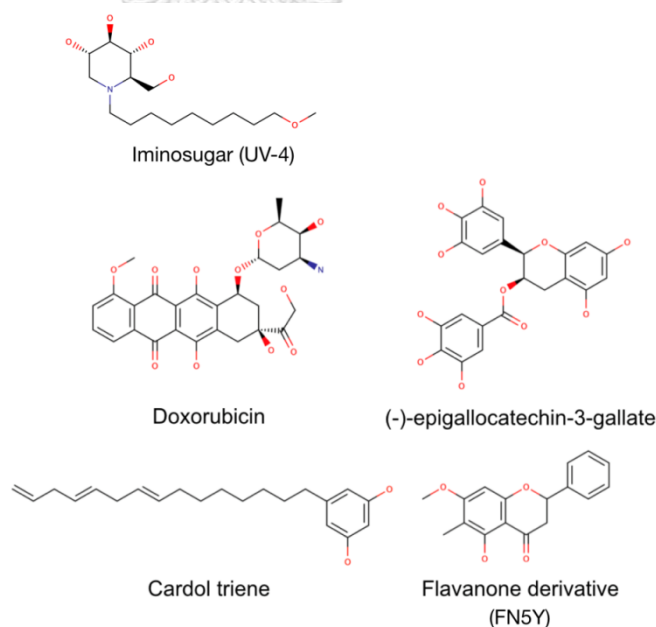


Figure 12 2D structure of inhibitor against flavivirus entry process.

This study aimed to investigate the possible binding sites of active inhibitors on the E homodimer using *in silico* methods. The possible complex structures between inhibitor and E protein dimer are predicted by molecular docking. The *k*-mean clustering algorithm is applied to group and predict the preferential binding site. Several computational tools, such as all-atom MD simulations, principal component analysis (PCA), and binding free energy calculations are used for determining the stability and binding affinity of ligands on the viral surface at different binding sites. The obtained results propose the new valuable information to assist further drug development of inhibitors against dengue and Zika viruses. It could also inspire similar studies for the other systems of the flavivirus genus, such as yellow fever virus, West Nile virus and Japanese Encephalitis virus.

Moreover, the screening novel dengue inhibitors against the dimeric E protein is carried out via *in silico* methods. Flavonoid and its derivatives are collected from online and in-house databases based on the core structural similarity of flavonoid [94, 95]. In the beginning step, the pharmacophore-based virtual screening and molecular docking methods are used for virtual screening of a numerous compounds in the data sets. Then, the hits need to be evaluated for this binding affinity by analyzing the behavior of the complexes through MD simulation. The solvated interaction energy (SIE) method is a less time-consuming and moderate accurate calculation that is commonly used to compare and evaluate binding free energy among the complex systems. The molecular mechanics generalized born surface area (MM-GBSA) and molecular mechanics Poisson–Boltzmann surface area methods (MM-PBSA) [96, 97] are used to evaluate the binding efficiency to distinguish the most preferable binding sites of the potent compound in the complex system. Insight of binding affinity of a potent compound to a binding pocket of E protein, can be evaluated by quantum mechanics (QM), fragment molecular orbital (FMO) method, that can provide the interaction energy (PIE) with the highest accuracy [98, 99]. The binding pattern of the active compound on E protein and a novel dengue viral inhibitor will be proposed.

1.4 Objectives

1.4.1 To evaluate the interaction between lead compounds and dengue and Zika viruses via *in silico* study.

1.4.2 To screen the compounds inhibit envelope protein of dengue and Zika viruses.



Chapter II

Theory and Computational details

2.1 Molecular docking

Molecular docking technique becomes a powerful method used for predicting the possible binding site on a target protein [100-102]. The common step of molecular docking is illustrated in Fig. 13. The conformations of ligand will be randomly generated by the search strategy algorithm. Then, each pose of ligand will be docked into the protein and computed the interaction energy. List of dock poses are ranked depending on binding affinity score [8, 9, 11, 103, 104].

Molecular docking can be classified into two methods; blind and focused docking, according to the search strategy. The small molecules randomly search and insert into the whole macromolecule without specific binding region, which called blind docking. Another technique is focused docking, which requires the previous knowledge in order to determine the binding sites on the target protein.

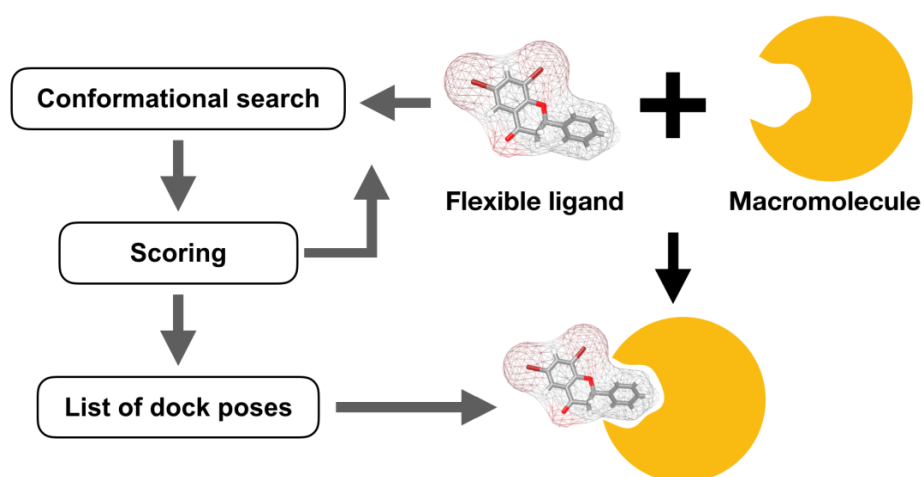


Figure 13 The schematic procedure of Molecular docking method.

The binding free energy, scoring function can be approximated by many parameters of the standard chemical potentials of the system, depended on

each algorithm. One of the most popular docking software is autodock vina, which is the flexible ligand docking method [105]. This program can be used to estimate the binding affinity of the ligand binds to macromolecules and it is also used for searching the possible binding region of structural protein by special algorithm that developed from various stochastic global optimization approaches. Ligand will be generated in various conformations including bonds, angles, and dihedrals. Then, the ligand will be randomly searched space on the surface of macromolecule (Fig. 14). All of the ligands will generate the binding poses and access the binding affinity. The most common binding region of ligand indicated the possible binding region of the target protein. In case of CDOCKER software, this program is the molecular dynamics (MD) simulated-annealing-based algorithm using CHARMM force field scoring function [106]. One of the key steps of focused docking is determining the specific binding sites of macromolecule, grid (Fig. 15). Each ligand is then subjected to high temperature MD. A simulated annealing method is a step of heating the ligand, that lead randomly generating the various ligand conformations in order to increase the performance of docking. Another focused docking program, iGEMDOCK is a flexible docking tool that uses a generic evolution method (Fig. 16). The combination of empirical and a pharmacophore-based scoring function is the key strategy of this program. The binding affinity of this program is considered as a Fitness score that can be computed by the following terms: van de Waals interaction (E_{vdW}), hydrogen-bonding potential (E_{Hbond}), and electrostatic (E_{elec}) (eq. 2.1) [107, 108].

$$Fitness\ score = E_{vdW} + E_{Hbond} + E_{elec} \quad eq. 2.1$$

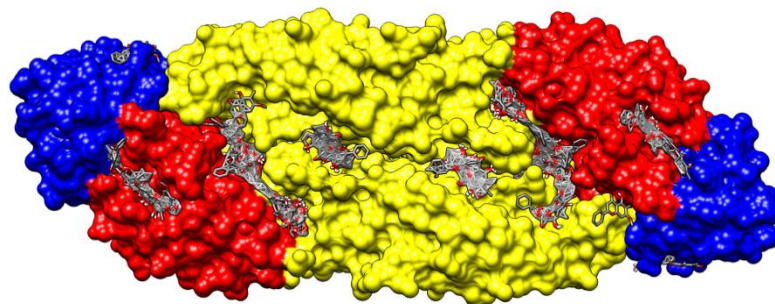


Figure 14 Ligands are docked to dengue E protein by blind docking method.

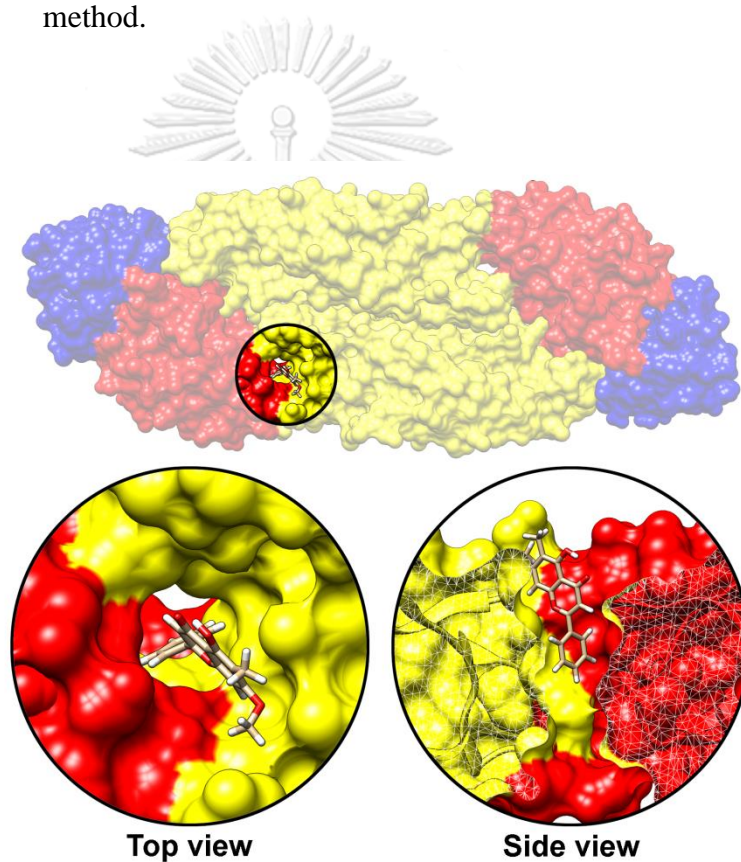


Figure 15 Ligand bound to specific binding region of dengue E protein using focused docking method.

Darwin-like evolutionary processes

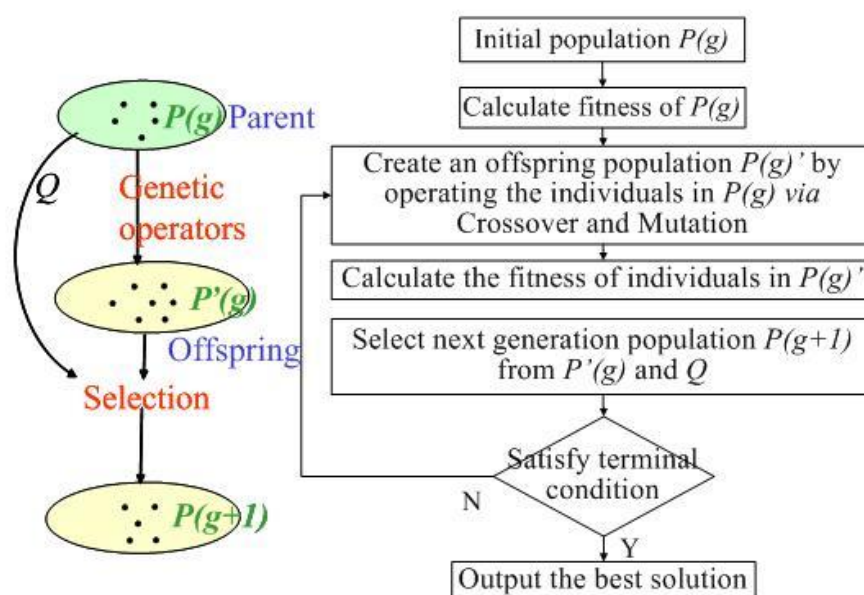


Figure 16 Generic evolution algorithm [107].

2.2 *k*-mean clustering

Among complexity of data without defining categories, the clustering is an unsupervised machine learning algorithms that used to manage and group the similar data into the same group. There are several algorithms for clustering such as hierarchical and *k*-mean clustering. According to blind docking result as shown in Fig. 14, the abundant molecules bind to dengue E protein are arduous to interpret [6]. Therefore, *k*-mean clustering algorithm can be used to cluster all ligands. The principal of this algorithm consists of several steps, beginning with determining the centroid of data and ending with the member of each cluster being proposed (Fig. 17). First, the number of the simulated center point of each group is assigned by elbow method [109, 110]. This method is used to validate the optimal number of clusters (*k*) for *k*-mean clustering by plot the sum of squared errors (SSE) that calculated from each *k* in range. The line plot will be looking like a flexor arm and the critical point or the best *k* value will be shown at the elbow. Each *k* value in the range of 1-50, is, calculated SSE and plotted as a Fig. 18. The appropriated number of

clusters is 18 (red arrow) which is the elbow of this graph. Second, 18 centroids are used to be the representative centroid of each group; they will be clustered and updated till convergence using Euclidean distance-based method [110, 111]. Each cluster can be represented as a probability value to reveal the possible binding region on a macromolecule.

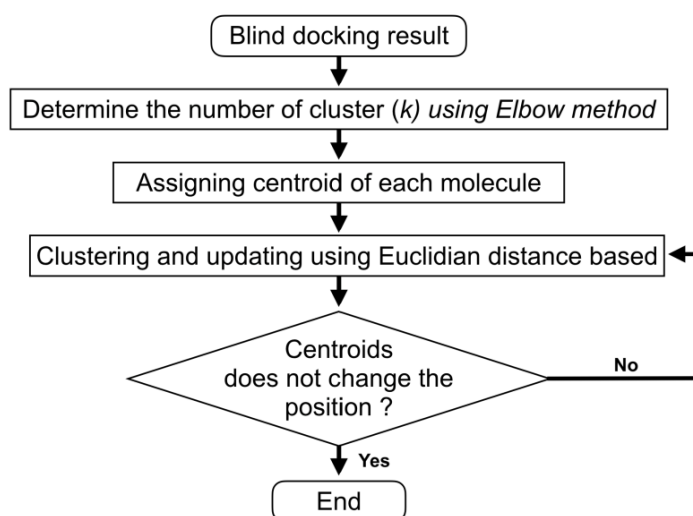


Figure 17 The procedure of k -mean clustering algorithm in this study.

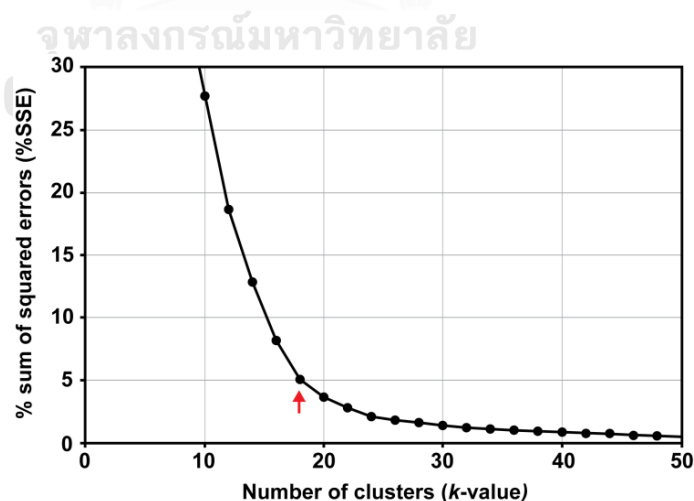


Figure 18 Identification of the optimal number of clusters (k -value) using Elbow method for the k -mean clustering.

2.3 Molecular dynamics simulation

The motion of atoms along the trajectory can be mimicked by molecular dynamics (MD) simulation in order to observe the conformational changes, molecular behaviors, the binding pattern, and ligand-receptor interaction energy. This method is very useful for improving the strategy in CADD and SBVS. For the classical MD simulation, particles (i) in the system are calculated until they converge based on the Newton's second law of motion that is applied in this method related to potential energy (V) with position r_i and time step (dt) (eq. 2.2).

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad \text{eq. 2.2}$$

The potential energy of macromolecule provides the force field parameter that can be computed from bonded and non-bonded interaction energy. The quantum harmonic oscillator and Morse potential provide the description for bonded terms; bond stretch, angle, and dihedral (Fig. 19). For non-bonded interaction energy, electrostatic and van de Waals are computed by the Coulomb's law and Lennard-Jones potential, respectively (eq. 2.3, 2.4) [85]. Much software can be used for performing the MD simulation, for example, GROMACS, NAMD, CHRMm, and AMBER. Periodic boundary condition (PBC) is applied to mimic a large system infinite by using a unit cell. The PBC can avoid the surface effect of each atom in the MD system. The particle in the simulation box can move through other box and form the interaction with their image.

$$E_{MM} = E_{bonded} + E_{non-bonded} \quad \text{eq. 2.3}$$

$$E_{MM} = E_{stretch} + E_{angle} + E_{dihedral} + E_{elec} + E_{vdw} \quad \text{eq. 2.4}$$

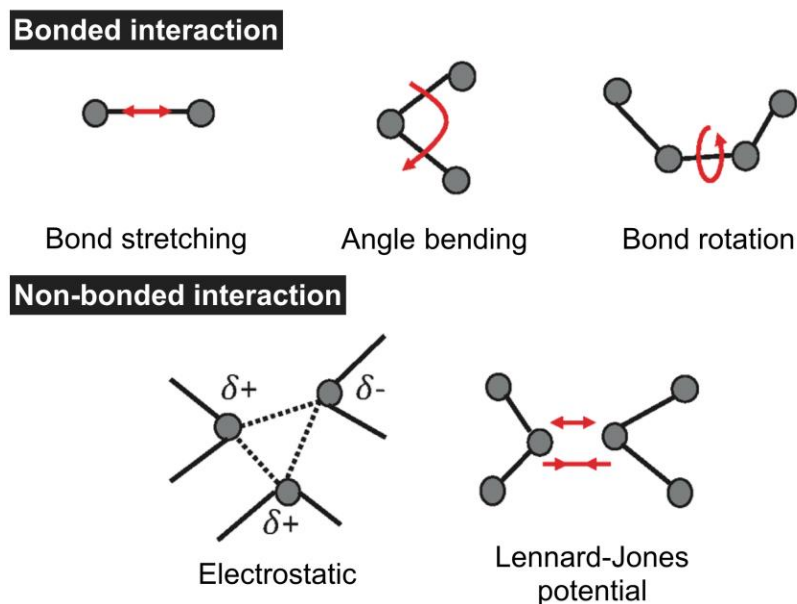


Figure 19 Bonded and non-bonded interactions consider in MD simulation [112].

In this study, trajectory and Newton's equation of motion can be generated by Verlet integrated algorithm (eq. 2.5) [113, 114]. For a time increment in the small step (Δt), a Taylor series can be implemented in the function $r(t)$ as shown in eq. 2.5. It uses the position of the atom at time and the acceleration and the position of time to compute the position of time [115]. The position (r_i) at t and $t - \Delta t$ in the previous step are used to predict the position at the time.

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + a_i(t)\Delta t^2 \quad \text{eq. 2.5}$$

Basically, the MD calculation's process begins from the macromolecule structure that can be obtained from Protein Data Bank (PDB) or homology modeling. The parameters of the system such as the number of particles in the system (N), temperature (T), pressure (V), the initial velocity of each particle, and time step, are assigned. The biomolecules are soaked in the solvent according to water model. There are essential for improving the realistic

condition in MD simulation. Actually, it has more than 46 water models that can be used in computer simulation. The 3 site explicit water models (TIP3P) is implemented in this study. The parameters of O-H distance (l), H-O-H angle (θ) are 0.9572 \AA and 104.52° , respectively (Fig. 20) [116]. In addition, Lennard-Jones parameter (σ) and charge of each hydrogen atom ($q_1=+0.417$ and $q_2=-0.834$) are also considered in this model. One of the limitations of the explicit solvated model is the time consuming and high computational cost.

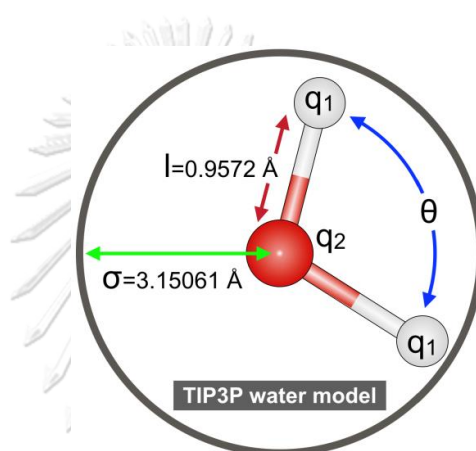


Figure 20 TIP3P water model

In the subsequent stage, the solvated structural system is heating up to the interested temperature, and then simulated it at constant temperature till the equilibrium phase. The isothermal-isobaric ensemble (NPT), temperature and pressure are constant during a production phase. After that, the equilibrium trajectory is analyzed in the term of RMSD, hydrogen bonding interaction, conformational structure, and binding free energy (Fig. 21). Nowadays, the MD trajectory can be furious simulated due to the highly computational performance.

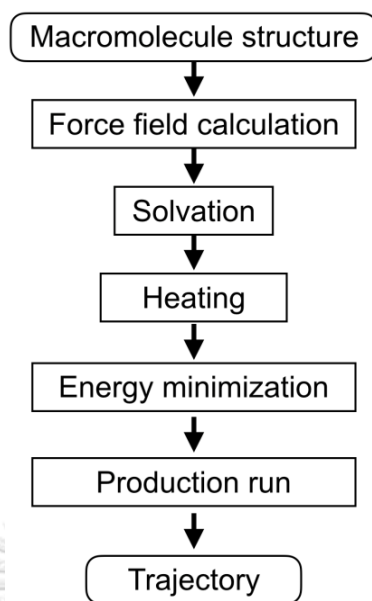


Figure 21 The general procedure of MD simulation

2.4 MD trajectory analysis

The behavior stability, structure fluctuation, and binding affinity can be analyzed from the production phases of MD trajectory. In AMBER package program, cpptraj module is an important tool that is used for analyzing the simulation data.

2.4.1 Root mean square deviation (RMSD)

This analysis is a major tool to observe the structural stability of molecules in the system along the MD trajectory [117]. The RMSD value is calculated from the average distance between current atomic coordinated data and the previous step that defined as a standard (eq. 2.6).

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2} \quad \text{eq. 2.6}$$

The RMSD is computed by the n pairs of similar atoms and distance (d) between two atoms in each time step. Then, the calculated RMSD can be plotted as a line graph between the structural fluctuations over time step. The stability of macromolecule in the system is indicated by this plot. The production phase of each system will be used to describe the interaction between ligand and protein receptor by binding free energy calculation or hydrogen bond analysis.

2.4.2 Hydrogen bond analysis

The strength of non-bonded interaction energy can be described by hydrogen bonding interaction. Hydrogen bond can be defined by the distance ($\leq 3\text{\AA}$) and angle ($\geq 120^\circ$) between H and the hydrogen atom (the positive end) and nitrogen, oxygen, or fluorine atom (strong electronegativity with a lone pair of electrons), which is the strongest of dipole-dipole interaction (Fig. 22) [118]. The strength of hydrogen bonds depended on distance between interacted atom [119].

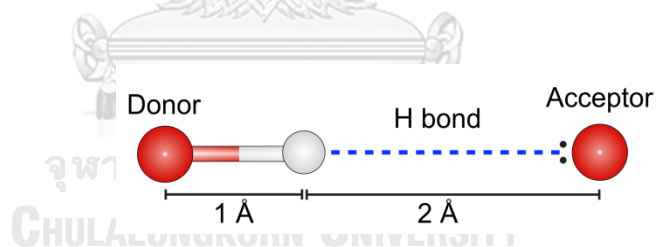


Figure 22 Hydrogen bonding interaction between donor and acceptor atom.

2.5 Interaction energy calculations

The binding affinity between ligand and protein receptor can be approximated from various methods depended on accurate and time consuming (Fig. 23). Several solvated water models can be used for binding free energy calculation. Explicit solvation, the water molecules are applied in the system that can provide the realistic condition for MD simulation.

However, it is compensated to the high computational cost and time consuming. To overcome the limitation of the explicit solvation model, the implicit model is able to figure out the solvation term in binding free energy calculation (ΔG_{bind}) by Solvated accessible surface area (SASA), Poisson–Boltzmann (PB), and Generalized Born (GB) model. According to the binding free energy equation (eq. 2.7), the different energy between bound and unbound states of macromolecule in the gas and solvent are used to computed in order to describe the interaction energy. The insight of calculation will be explained in each method.

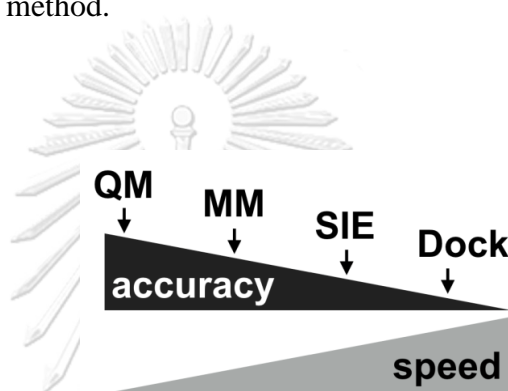


Figure 23 The comparison of accuracy and speed of each binding free energy calculation methods; QM, MM, SIE, and Dock.

$$\Delta G_{bind} = \Delta G_{complex} - (\Delta G_{receptor} + \Delta G_{ligand}) \quad \text{eq. 2.7}$$

2.5.1 Solvated interaction energy (SIE)

This method is a powerful post-processing tool that has been used to estimate the binding free energy of each snapshot from MD trajectory, and then it will compute the average SIE value. The intermolecular energy; van de Waals and electrostatic between ligand and protein receptor, and the different reaction field energy and molecular surface between bound and unbound state, are the important parameter in SIE method. In addition, the constant value such as dielectric constant, coefficient, and

entropy-enthalpy compensation, are also used to combined with this method (the details will be discussed in Chapter 3) [96]. This method is less time consuming compared to free energy simulation, however, the limitations of this method is the moderate accuracy level of calculation and some parameter in SIE equation are not really universal.

2.5.2 Molecular mechanics (MM-PBSA and MM-GBSA)

The molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) and molecular mechanics generalized Born surface area (MM-GBSA) are the end-point binding free energy calculation approaches, which are widely used in CADD and SBVS in order to approximate binding free energy between ligand and protein receptor. The macromolecule has to be removed water molecules, and then, it will be solvated by continuum model that is an implicit solvation (Fig. 24). It is applied in these method [120]. The MM-PBSA and MM-GBSA require a snapshot extracted from MD simulation of ligand-protein complex in solvation for computing the binding free energy in solvation (eq. 2.8). The binding free energy of molecule in solvent, polar contribution can be solved by Poisson-Boltzmann (PB) or the Generalized Born (GB) equation. These methods are chosen for rescoring and enhancing the accuracy of binding free energy from SIE method [121]. The dielectric constant, the ligand formal charge, the initial protein structure, and MD sampling affect the accuracy of MM-PBSA and MM-GBSA methods. One of the key strategies of MM-PBSA method is Poisson–Boltzmann equation which is the combination between Poisson equation (eq. 2.9) and Boltzmann distribution. These terms are used for understanding the electrostatic properties of biological macromolecules [122]. According to eq. 2.9, where the variation in dielectric constant is ϵ , electric potential is ϕ , charge density is ρ , and constant value is k' , and position is r .

$$\Delta G_{bind,solv} = \Delta G_{bind,gas} + \Delta G_{solv,comp} - (\Delta G_{solv,lig} + \Delta G_{solv,rec}) \text{ eq. 2.8}$$

$$\nabla \cdot \epsilon(r)\nabla\phi(r) - k'\phi(r) = -4\pi\rho(r) \text{ eq. 2.9}$$



Figure 24 The illustration of binding free energy calculation of ligand binds to protein receptor using MM-PBSA and MM-GBSA methods.

2.5.3 QM/MM-GBSA

Quantum mechanics (QM) method is the high accuracy calculation but it is very expensive computational resources and time, which is commonly used for small molecular system. Molecular mechanics (MM) approach can be applied for a large molecule [123, 124]. The MM method lacks of electron density term for the calculation, which is one of the limitations. The hybrid QM/MM-GBSA approach is able to use to calculate binding free energy based on QM with GB model.

The complex system is divided into 2 parts of the combined QM/MM calculation (Fig. 25). The ligand and protein in the binding region are computed by the QM method and the outer region is calculated by the classical MM technique that is treated by a force field parameter. Self-consistent charge density, functional tight binding (SCC-DFTB) is the highly efficient approximated DFT method that deserved to calculate the binding site of macromolecule [125]. The Kohn-Sham equation is

similar to Hartree-Fock in terms of electron wave function and orbital energy, which is accounted to charge density fluctuations.

2.5.4 Quantum mechanics: fragment molecular orbital method

The *ab initio* calculation, the highest accurate binding free energy calculation is extremely computational time and cost consuming that depended on the number of atoms in the system. To deal with this limitation, the macromolecule is separated into small fragments (monomer) and computed electronic properties in parallel calculation. Then, each paired of monomer (dimer) is computed for the quantum effect. That is fragment molecular orbital (FMO) method. The critical idea of this approach is a Fragmentation step because of the particular time reduction. The FMO calculation consists of consists of several steps as shown in Fig. 26 [126].

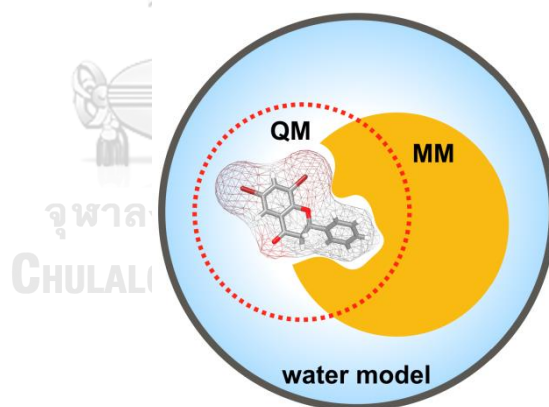


Figure 25 The schematic represents the QM and MM region identification.

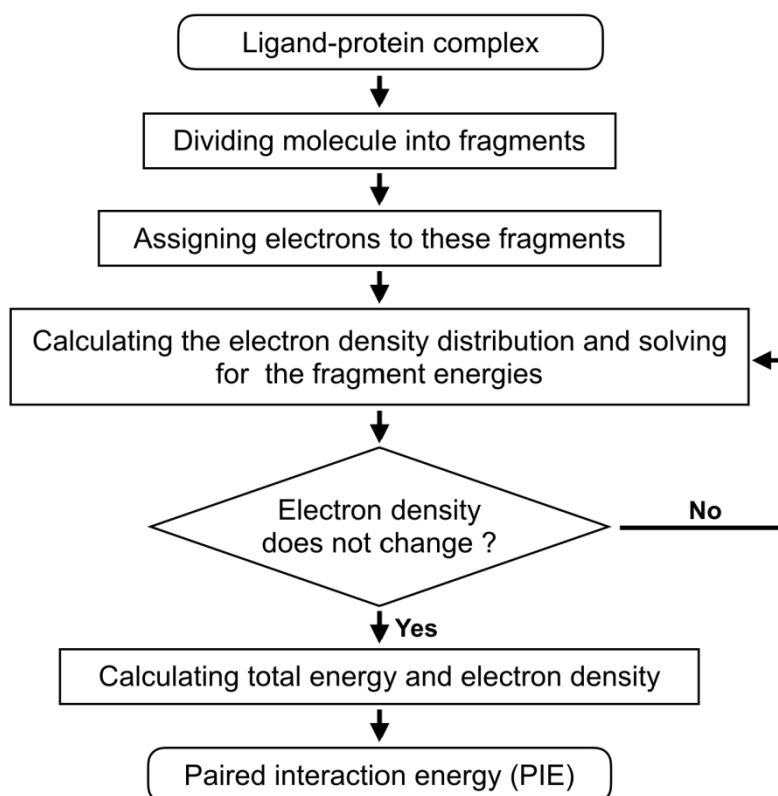


Figure 26 The step of paired interaction energy (PIE) calculation by fragment molecular orbital (FMO) method.

The total energy of each fragment was computed self-consistent field (SCF) energy (E_i) until convergent. Then, each pair of fragment was figured the paired interaction energy (PIE and PIEDA, ΔE_{ij}^{int}), which is sum of; electrostatic (E_{ij}^{ES}), charge exchange (E_{ij}^{EX}), charge transfer (E_{ij}^{CT+mix}), and dispersion (E_{ij}^{DI}), according to eq. 2.10 [99, 127]. The PIE can be used to explained both bonded and non-bonded interaction. FMO2 is applied for ligand-protein interaction energy calculation because this level is suitable for comparing the different energy between binding modes of complex.

$$\Delta E_{ij}^{int} = \Delta E_{ij}^{ES} + \Delta E_{ij}^{EX} + \Delta E_{ij}^{CT+mix} + \Delta E_{ij}^{DI} \quad \text{eq. 2.10}$$

2.6 Pharmacophore-based virtual screening

A pharmacophore-based screening approach becomes one of the most advantageous tools, which is matching the similarity of the 3D interaction pattern or pharmacophore model of drug-like compounds with known ligands in the complexes [128-135].

Pharmacophore model can be described from the interaction pattern between ligand and 3D complex structures, which is able to use it for screening the potent molecules from a huge compound library. The interaction pattern such as hydrogen bond donor and acceptor property, coordination to metal ions, charge, hydrophobic groups, aromatic ring of a compound that binding to a biological target, and the excluded volume, are included as pharmacophore properties (Fig. 27) [129, 134, 136-139]. The important strategy of this approach is the overlays-based scoring algorithm, which is able to enhance the performance of pharmacophore-based screening [140]. The pharmacophore model can be also generated from either single protein-ligand complex or molecular dynamics (MD) trajectory along the simulation in order to gather more information (Fig. 28) [137].

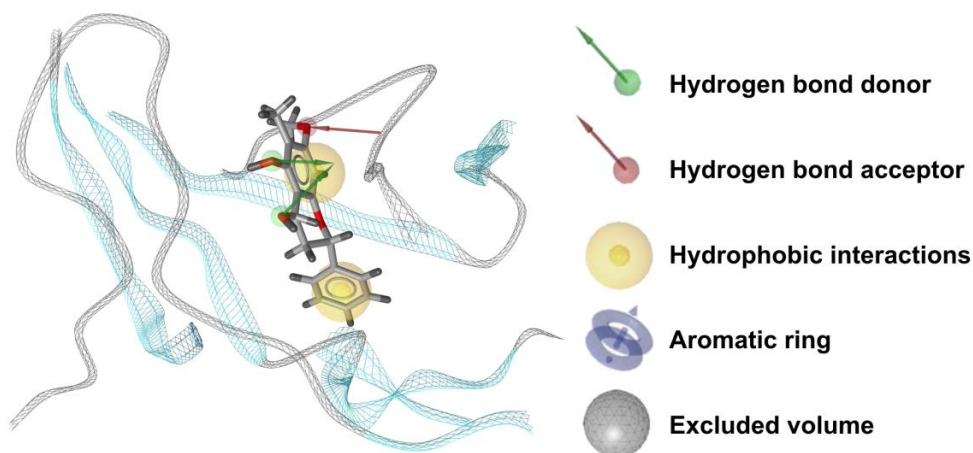


Figure 27 Types of pharmacophore feature are generated by Ligandscout program.

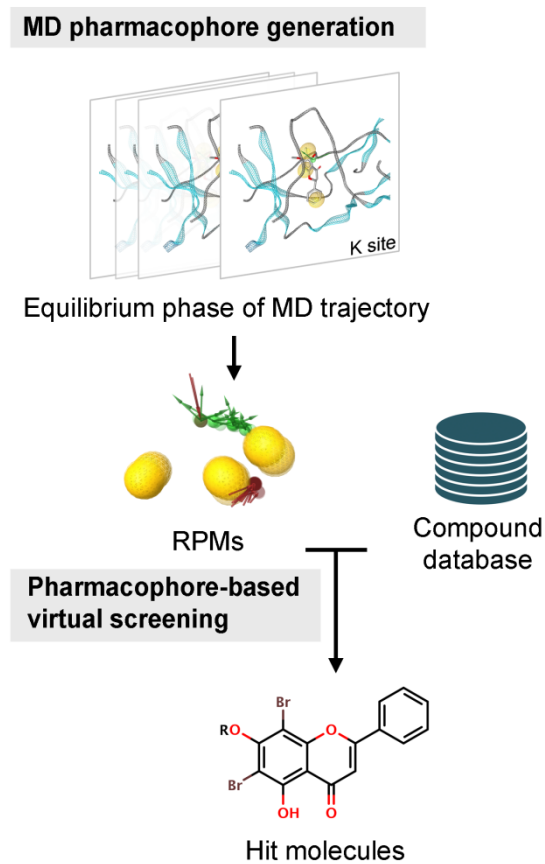


Figure 28 The integration of MD trajectory and pharmacophore-based virtual screening procedure.

Chapter III

Material and methods

Overview of this work

These studies focused on both envelope protein of dengue and Zika viruses using computation aided drug design. Many reports suggest that flavonoid derivatives and polyphenolic lipids are able to inhibit dengue virus in the early step of viral infection [4, 93, 141-146]. In addition, epigallocatechin gallate (EGCG) has shown an inhibitory effect against Zika viral infection [54, 55]. However, the viral target proteins of known inhibitors are still ambiguous because of the limitation of experiments that cannot describe the mechanism of the inhibitor at an atomistic level. In this study, several computational tools are used to identify the protein target of active inhibitors for dengue and Zika viruses. Furthermore, a novel potent compound that can inhibit dengue viral infection has been proposed using multi-virtual screening method and molecular dynamic (MD) simulation (Fig. 29). First, the possible protein target of the active compound was identified by a molecular modeling approach. Two major methods involved in this step are molecular docking and MD simulation. After that, the novel potent inhibitor was proposed by multiple virtual screening methods, and then this compound was computed the binding affinity by molecular mechanics and chemical quantum approaches. The integration of MM and QM calculations is able to provide high accuracy and precise results. Moreover, this combinatory study also leads us to overcome the limitation of each approach.

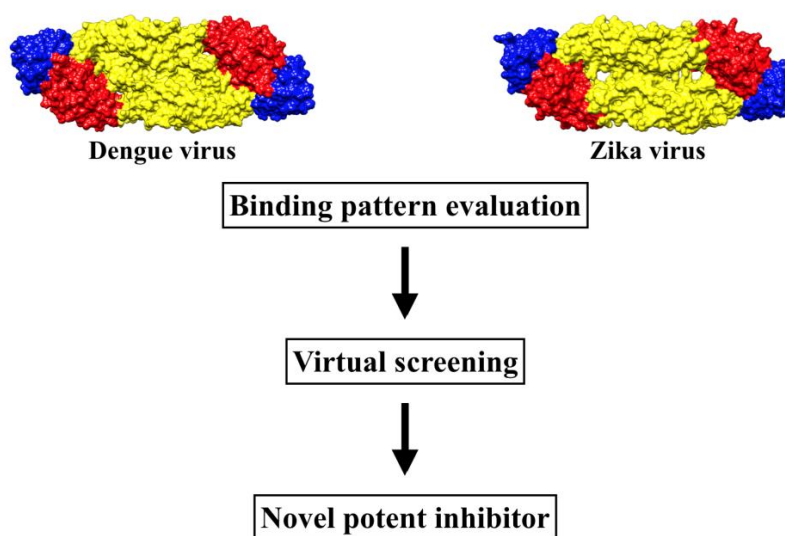


Figure 29 The overview of this work. This study is separated into 2 major works. First, this study evaluates the binding pattern of active compound on target protein of dengue and Zika viruses. Second, the novel potent inhibitor of dengue virus is screened by several steps of virtual screening techniques.

3.1 Binding pattern evaluation

3.1.1 System preparation

This study focused on the envelope protein, a viral-surface protein, which is related to the early step of viral infection [1, 2, 25, 27, 29, 34, 45, 147]. The 3D structure is obtained from Protein Data Bank (PDB). The crystal structure of dengue envelope (E) protein with n-octyl-beta-D-glucoside (β -OG) bound taken from the PDB with the entry code: 1OKE [78], and Zika virus E protein is 5IZ7 [148]. The hydrogen atoms and missing residues were added to the protein structure using Discovery Studio 2.5 package (Accelrys, Inc.). For the Zika virus, FN5Y and EGCG

can inhibit at the early step of the viral infection. Therefore, these compounds could inhibit at the E protein which is located at the viral surface [54, 55]. In order to prove the assumption, E protein of dengue and Zika viruses will be used as the target protein in this study. The structure of flavonoid derivatives and EGCG were constructed by Gaussview 05. The compound structures were optimized by Gaussian 09 with HF/6-31g (d, p) basis set [149].

The 2D structures of all active compounds were evaluated via binding affinity between ligand and target protein using autodock vina based on blind docking method [105, 150]. This method is widely used in CADD for constructing the initial complex structure or estimating the binding affinity between ligand and protein based on its preferential binding conformation and scoring function. The interaction energy was calculated from many individual energy terms such as non-bonded and bonded terms [151]. In order to increase the performance of the program, the conformations of ligand were generated by MD-based and simulated annealing algorithm. Then, the ligand is randomly moved into the binding sites and the interaction energies were calculated [150]. Molecular docking method is used to identify and evaluate the binding interaction energy between ligand and protein. The top interaction energy score of a complex structure is chosen as a possible complex, which is used in the further steps.

3.1.2 Binding region identification

Most of non-structural proteins (NS) have specific binding sites, which are active or allosteric binding regions. Unlike, a structural protein, they do not show the actual active site in the protein. However, some part of the structural proteins may play an important role in the vital mechanism of the viral infection. The outermost part of flavivirus is the envelop protein that is related in several steps to the viral life cycle. The preferential binding region of flavonoid derivatives on the dimeric E protein was predicted via blind docking method. This method is one of the

molecular docking techniques that involve docking of the ligand into the whole protein and compute the interaction energies of each pose in order to discover the possible binding region of interesting ligand to the target protein [102, 152, 153]. For the docking procedure of the dimeric E protein, the missing hydrogen atoms were added, assigned and merged with the Gasteiger charges, and saved as a receptor. Then, FN5Y and EGCG were constructed following the protein receptor's procedure. All rotatable bonds were accepted to rotate during the docking process. The grid map was computed using Autogrid module. The volume selected for the grid map was $62 \times 126 \times 126$ points with grid space of 1.0 Å covering the whole viral surface. The binding affinity energy score, hydrogen bond interactions, and ligand conformations for all possible FN5Y and EGCG binding on each dimeric E protein surface were calculated.

Due to the high number of ligands bound to each dimeric E protein, a total of 1,000 ligands was clustered and considered for the probability of each cluster on the E protein surface. The blind docking results were analyzed by *k*-mean clustering method applied in this study via Python script [6]. This algorithm process is to separate a set of molecules into *k* groups based on the Euclidean distance between each point in the cluster. The optimal number of groups (*k*) was assigned according to Elbow method. The steps of *k*-mean clustering include; giving the number of clusters (*k*) [109, 111, 154], assigning centroid of the data point, clustering and updating new cluster centroids, and iterating the procedure until centroids are converged [111]. Each cluster was calculated and ranked by probability value. Then the possible binding sites were proposed. To verify predicted binding sites and construct ligand protein complex structure, the active molecules were docked into each binding region using CDOCKER in Discovery Studio 2.5 package (Accelrys, Inc.) via focused docking approach [106].

Insight of verifying the blind docking result, the results of *k*-mean clustering were considered as the predicted binding sites. The blind

docking results suggested four possible binding sites (K-K', X-X', Y-Y' and Z-Z') of FN5Y on dengue E protein and the three possible binding sites (X-X', Y-Y' and Z-Z') of EGCG of Zika E protein dimer. To refine these results and construct the initial structure for MD simulation, the FN5Y and EGCG were then docked into each binding site (10 Å sphere radius) with 100 runs using CDOCKER in Discovery Studio 2.5 package (Accelrys, Inc.) in according to the previous studies [155-158]. The ligand conformations with the best interaction energy score at each site were selected as the initial complex structure for MD study.

3.1.3 Binding pattern of inhibitor

The FN5Y/E protein complex structure from the previous step was set as an initial structure for MD simulation. In case of dengue virus, the FN5Y binding to 8 different binding regions were used as an initial structure for all-atom molecular dynamics (MD) simulation at neutral condition. Unlike, EGCG binding to 6 possible binding region of Zika viral E protein was separate into 3 systems. These complexes of ZIKV dimeric E protein with two EGCG molecules bound at the opposite site (X-X', Y-Y', and Z-Z') were individually performed at acidic condition pH of 5, in order to simulate the system in endosome [37, 148, 159]. Both dengue and Zika E protein complexes were performed with periodic boundary condition using AMBER 16 package program, similar to previous studies. The general AMBER (GAFF)[85] and AMBER ff03.r1 force fields[160] was used to treat ligands and E protein, respectively. The partial atomic charges of ligands were prepared in according to the standard procedure by Gaussian 09 (Gaussian, Inc.). The protonation state of FN5Y and EGCG were determined at neutral and pH 5 for post-fusion conformations of E protein using pKa calculation implemented in ChemAxon software[161, 162]. In addition, all ionizable residues in protein structure were assigned by PDB2PQR version 2.0.0 [163]. All missing hydrogen atoms were added to the complex structure using the tLeaP module implemented in AMBER16. A cut-off value of 10 Å was

assigned for non-bonded interactions. The counter ions were consequently added in order to electrical neutralization. Finally, each complex was fully solvated in TIP3P water box extending 12 Å from the protein surface. The added hydrogen atoms, water molecules and the whole system were subsequently minimized with 1,500 steps of both steepest descents (SD) and conjugated gradient (CG) using the SANDER module. Each system was then heated up to 300 K for 200 ps, and consequently simulated at the same temperature till 100 ns and 500 ns for dengue and Zika system, respectively. The MD trajectories were collected at every 0.2 ps from the production phase.

Using the CPPTRAJ module [164], the convergences of energies, temperature, and global root mean-square displacement were used to verify the stability of the FN5Y/E complex. The MD snapshots extracted from the last 50 ns were used for total binding free energy analysis based on MM/GBSA method [121, 123, 157, 165] and its per-residue decomposition energy contribution.

The system stability of EGCG/E complexes was determined by root mean square displacement (RMSD). The MD snapshots extracted from the last 200 ns were used for analysis in terms of per-residue decomposition free energy based on the MM/GBSA method and intermolecular hydrogen bonding between EGCG and E protein. The total binding free energy of complex was predicted using 4 different methods of calculation, including solvated interaction energy (SIE) method by the following equation (eq. 3.1)[96, 97], molecular mechanics generalized Born surface area (MM-GBSA), and QM/MM-GBSA. For QM/MM-GBSA, the EGCG molecule was only treated with PM3 or SCC-DFTB [123, 125].

$$\Delta G_{bind}(\rho, D_{in}, \alpha, \gamma, C) = \alpha \cdot [E_{elec}(D_{in}) + \Delta G^R(\rho, D_{in}) + E_{vdW} + \gamma(\rho, D_{in}) \cdot \Delta MSA(\rho)] + C. \quad \text{eq. 3.1}$$

ρ is a van der Waals radii linear scaling coefficient of AMBER, D_{in} is the solute interior dielectric constant, γ is a molecular surface area coefficient and C is a constant. The parameter α is a global scaling factor of the total raw solvated interaction energy relating to the scaling of the binding free energy due to configurational entropy effects. Both terms of Coulomb (E_{elec}) and van der Waals (E_{vdW}) interaction were calculated using AMBER ff03 force field. The change in the reaction energy of complex between bound and unbound states (ΔG^R) was calculated from the Poisson equation with a boundary element method using BRI BEM program [96, 121]. In term of $\gamma \cdot \Delta MSA$ relates with the change in the molecular surface area upon binding. These parameters were used as the default values; $\alpha = 0.104758$, $D_{in} = 2.25$, $\gamma = 0.012894$ kcal/(mol·Å²), $\rho = 1.1$ and $C = -2.89$ kcal/mol, respectively [96, 121]. Moreover, the correlation between 4 approaches of binding free energy calculation was evaluated using Pearson correlation coefficients.

3.2 Virtual screening

Several screening techniques were used in this study, in order to search for novel potent molecules from a large number of compounds from several databases including the flavonoid database using MD trajectory-based virtual screening. The hit compound was then synthesized and tested biological activity by experimental assays.

3.2.1 Pharmacophore-based virtual screening

For the preliminary screening, the intermolecular interactions of the FN5Y/E complex were used to create the pharmacophore models based on steric and electronic features. The pharmacophore feature of each of the 4 binding site and FN5Y was automatically generated from MD trajectories by LigandScout 4.2 program [129]. A total of 10,000 frames from a trajectory over the last 20-ns simulation were used to establish pharmacophore models at individual binding site. All

pharmacophore models were clustered, filtered and selected to set as a representative pharmacophore model (RPM) for further virtual screening.

In this study, we focused on the flavonoids reported in the previous study. The first compound library was constructed by similarity search method. Six subclasses of flavonoid structure were used as templates to search for similar molecules from Zinc database by 80% Tanimoto-based similarity [166, 167]. The second compound dataset was downloaded from TimTec compound library [168]. In total, the compound library consists of 996 compounds; 450 molecules from Zinc, 507 molecules from TimTec, and 39 in-house designed molecules. For each molecule, conformers up to 400 conformations were generated with “icon best” feature. Therefore, a total of 398,400 ligand conformations was saved as the local database for further step.

Virtual screening was performed by pharmacophore-based screening technique [169]. All compounds in library were screened for each RPM by using “get best matching conformation” in retrieval mode and check exclusion volume based on pharmacophore-fit scoring function (eq.3.2 and 3.3). The hits from each screening iteration were collected and refined by common hits approach (CHA) developed by merging and rescoring each RPM run to generate the single hit-list of screened ligands. The steps between pharmacophore generation and final hit-list were implemented by KNIME workflow program [170]. The final hit-lists were created in SDF file format and visualized by LigandScout 4.2 program [129].

$$S_{RMS} = 9 - (3 * \min(\text{RMS}_{FP}, 3)) \quad \text{eq. 3.2}$$

$$S_{FCR} = c * N_{MFP} + S_{RMS} \quad \text{eq. 3.3}$$

In eq. 1 and 2, S_{RMS} is the feature count/RMS distance score, RMS_{FP} is the root mean squared (RMS) of the matched feature pair distance, S_{FCR} is the matched feature pair RMS score, c stands for the weighting factor for the number of matched feature pairs, and N_{MFP} is the

number of geometrically matched feature pairs [129].

The robustness of screening results was diagnosed and validated by receiver operating characteristic curve and the area under this curve (ROC-AUC). This curve illustrates the performance of pharmacophore screening by comparing the results between active ligands and decoy in the datasets. The area under ROC curve (AUC) was used for method validation [171]. For comparison, a set of inactive compounds (decoys) designed by mimic from the active molecule structure was obtained from Zinc database [172]. The ROC values gathered from the individual RPM screening with active and decoy datasets. Then, the values were plotted, analyzed and interpreted the term of sensitivity (true positive rate, TPR) and specificity (false positive rate, FPT) of screening results. Sensitivity (eq. 3.4) and specificity (eq. 3.5) measurements were used to validate the pharmacophore-based screening results.

$$\text{Sensitivity} = \frac{\text{Selected active ligands}}{\text{All selected ligands}} \quad \text{eq. 3.4}$$

$$\text{Specificity} = 1 - \frac{\text{Selected inactive ligands}}{\text{All selected ligands}} \quad \text{eq. 3.5}$$

3.2.2 Molecular docking

The 26 screened compounds were then separately docked into the four binding sites (K, K', X' and Y', Fig. 1c) on dimeric E protein with 100 runs based on two different algorithms, simulated annealing algorithm [106] in CDOCKER software of Discovery Studio 2.5 package and Genetic algorithm (GA) in iGEMDOCK 2.1 [108]. The protein structure for molecular docking was extracted from the last snapshot of FN5Y/E simulation from the previous study [4] without ligands, ions and waters, while the screened compounds with protonation state determined at neutral pH by ChemAxon software [161, 162] were optimized by HF/6-31G* level of theory using Gaussian 09 (Gaussian, Inc.). More details of

molecular docking can be found in the recent studies [4, 6, 36]. The docked conformation with the best interaction energy and fitness score at each specific binding site was selected as an initial complex structure for all-atom MD simulations.

3.2.3 Binding energy calculation

The top three docked complexes ranked from CDOCKER interaction energies and iGEMDOCK fitness scores were used for 100-ns MD simulation with periodic boundary condition using AMBER 16 program [85]. The partial charges of each ligand were prepared in according to the standard protocol [156-158], while the other parameters were taken from the general AMBER force field [85]. The AMBER ff14SB force field [160] was applied on the E protein. All system preparation, minimization and MD simulation at 300 K were set as in our previous studies on FN5Y and cardol triene in complex with DENV E protein [4, 36]. The system stability of each complex was accessed by root mean square displacement (RMSD). The solvated interaction energy (SIE) method [96, 121] was applied on the 100 snapshots taken from the last 40 ns in order to determine the binding free energy of the complex. Among the three simulated complexes, the simulation of the best complex with lowest SIE binding free energy at all four sites was extended to 500 ns.

The conventional approach, MM-PBSA and MM-GBSA [121, 157, 158, 165] was used to identify the most preferential binding site and to figure out the key binding residues using per-residue decomposition free energy analysis, as well as the interaction profile of the best complex system. Furthermore, chemical quantum mechanics approach was used to compute the insight of binding affinity between ligand and protein. The *ab initio* total energy values calculated at the second-order Møller–Plesset (MP2) level [99, 173, 174] with the 6-31G* basis set. This calculation was performed with General Atomic and Molecular Electronic Structure System (GAMESS) [175, 176]. The most possible binding region, only 8

Å of residues bound to each ligand were chosen and separated into small fragments at $C\alpha - C$ atom by one residues per one fragment [177]. Charge and parameter of each fragment (monomer) were assigned into an input file using Facio software [178]. Paired interaction energy (PIE) was computed in the term of electrostatic, charge exchange, charge transfer, and dispersion, in order to describe the interaction of key residue in the binding region.



Chapter IV

Results and discussion

4.1 Binding pattern evaluation

4.1.1 Binding region identification

The blind docking method was applied to identify the potential binding regions of FN5Y on the surface of the dengue dimeric E protein using AutoDock Vina version 1.1.2. [105] From a total of 1,000 docking poses, the interaction energies of FN5Y/E complexes were in a range from -6.2 to -8.4 kcal/mol, in which the ligands were found to interact either on each monomer or in the groove of E protein homodimer as shown in Fig. 30a. The docked FN5Y conformations were then grouped into 18 clusters by the *k*-means clustering algorithm based on the Cartesian coordinates of their points. It is worth noting that this method is a novel solution for clustering and counting the number of ligands in the complex structure, which is very useful for high throughput docking method. The clusters of FN5Y located under or on the side of the dimeric E protein were ignored, because the raft structure of E protein was placed on the lipid membrane and thus ligand cannot bind into these regions. Fig. 30a shows the 8 possible binding regions on dengue E protein homodimer for FN5Y. There is the kl loop or β -OG pocket (21.25% for K and K') [78], the X and X' sites located on domain I/III hinge region with 9.5% binding probability [179], the Y and Y' conserved regions among flaviviruses (25.25%) [93], and the Z and Z' interfaces between the two monomers (11.25%) [180]. The K', X', Y' and Z' sites represented the opposite positions relative to the K, X, Y and Z sites. All binding sites were involved with the conformational change from dimer to trimer [25, 29, 34, 78, 81-83, 179]. The X and X' sites were only located on the E protein monomer, while the other sites were located on the groove of protein dimer.

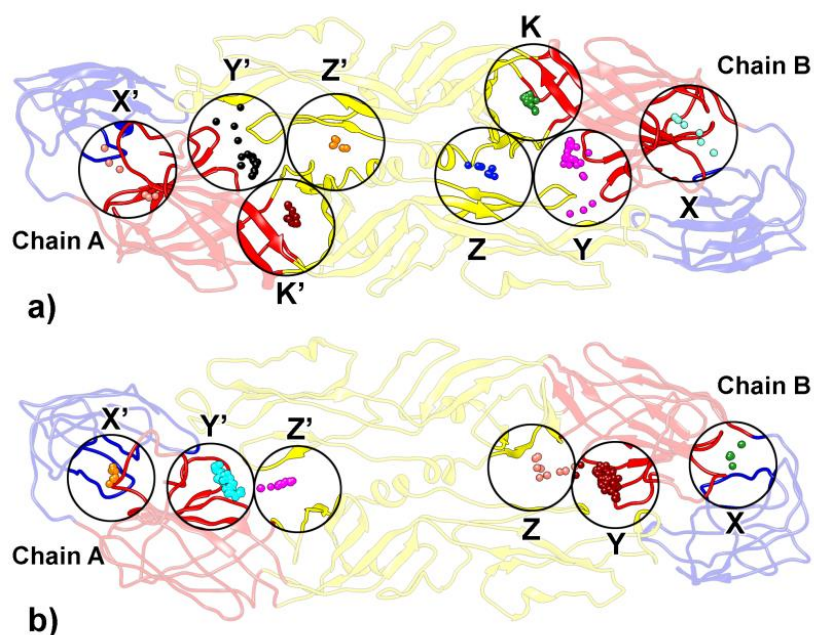


Figure 30 The 3D structure of dengue (a) and Zika (b) dimeric E protein with 1,000 ligand points grouped into 18 clusters as shown in different colors. a) Eight possible binding sites of FN5Y on molecular surface of dengue E protein: kl hairpin (K), domain I/III hinge region (X), conserved region among flaviviruses (Y) and interface between the two monomers (Z) and their opposite sites (K', X', Y' and Z'), which is evaluated by molecular docking and analyzed by k-mean clustering algorithm. Domains I, II and III are colored by red, yellow and blue, respectively. b) Six preferential binding regions of EGCG on Zika E protein.

To identify the possible binding region of EGCG on the Zika E protein and confirm the inhibiting mechanism of these inhibitors [6, 54, 55], the binding affinity of EGCG binding to E protein were in a range from -6.0 to -9.2 kcal/mol according to autodock vina program. The blind docking result was analyzed by *k*-mean clustering, demonstrated the 6 possible binding regions on the Zika E homodimer for EGCG (Fig. 30b): there are the domain I/III hinge region (2.5% for X and X') [179],

conserved region among flaviviruses (42.2% for Y and Y') [93] and interface between the two monomers (5.9% for Z and Z') [180]. Note that the rest placed in the E protein surrounding lipid membrane [29, 79], which may not be suitable for drug binding. The previous MD study described of EGCG binding at the X site on the Zika E protein monomer[55]; however, the ligand binding at the interfaces especially at the Y and Y' sites (domain I of one chain and domain II of the another) were favorably > 10-fold higher than the others. The residues 98-113 of the fusion loop at these sites were highly conserved among flaviviruses [1, 29, 34, 37, 148, 181]and were used to fuse to host endosomal membrane. In addition, the virtual screening of chemical libraries showed two preferential binding sites relevant to our results in the X-X' and Y-Y' sites [22].

Although, dengue and Zika virus belong the same genus which is flavivirus [22, 23, 27, 31, 36], there are slightly different amino acid sequence and protein structure. According to the previous study, the structures of envelop protein of dengue and Zika virus were compared by computational and experimental study [36]. The Zika E protein is lacking kl loop region in its structure or it might be closed conformation (Fig. 7).

From the all-atom MD simulation of the E protein in complex with eight FN5Y molecules (Fig. 31a), the ligands binding at the K, K', X' and Y' sites were likely stable along the 100-ns simulation, whereas the others were moved out from their pocket. MM-GBSA binding free energy values (Fig. 31b) suggested that the FN5Y preferred to bind at the kl hairpin in two different orientations. The phenyl ring (B-ring) of FN5Y favors its insertion into the kl loop (-28.4 kcal/mol at K site), while less binding was detected at the K' site (-23.5 kcal/mol) where the benzopyran ring (A-ring) is positioned in the loop instead. In addition, FN5Y is relatively less able to interact with E protein target at X' and Y' sites (-16.5 and -21.0 kcal/mol). The per-residue decomposition energy results in Fig. 31c provide the important residues associated with FN5Y binding as follows:

K (P83, L107, K128 and L198), K' (T48, E49, A50, L198, Q200 and L277), X' (Y138, V354 and I357) and Y' (V97, R99, N103 and K246). Taken altogether the binding affinity of FN5Y at either kl hairpin or protein surface could help to prevent the low-pH conformational change after the viral entry. In addition, the role of 6-methyl group of FN5Y towards the binding efficiency of dengue envelope proteins was studied. Overall, the presence of a 6-methyl group in FN5Y strengthened the binding efficiencies towards dengue E protein.

4.1.2 Binding pattern of inhibitor

Additionally, FN5Y could interact with protein at the Y' site, the conserved region among flaviviruses, in somewhat consistent with the previous MD study of the binding of flavonoids such as baicalein, quercetin and EGCG [93], and with the X' site, the domain I/III hinge region involved with the dengue vaccine development [179, 182]. This study bridged the gap between molecular simulation and biological function by showing that FN5Y bound consistently to these all three-target sites and inhibited fusion in cell-based study.

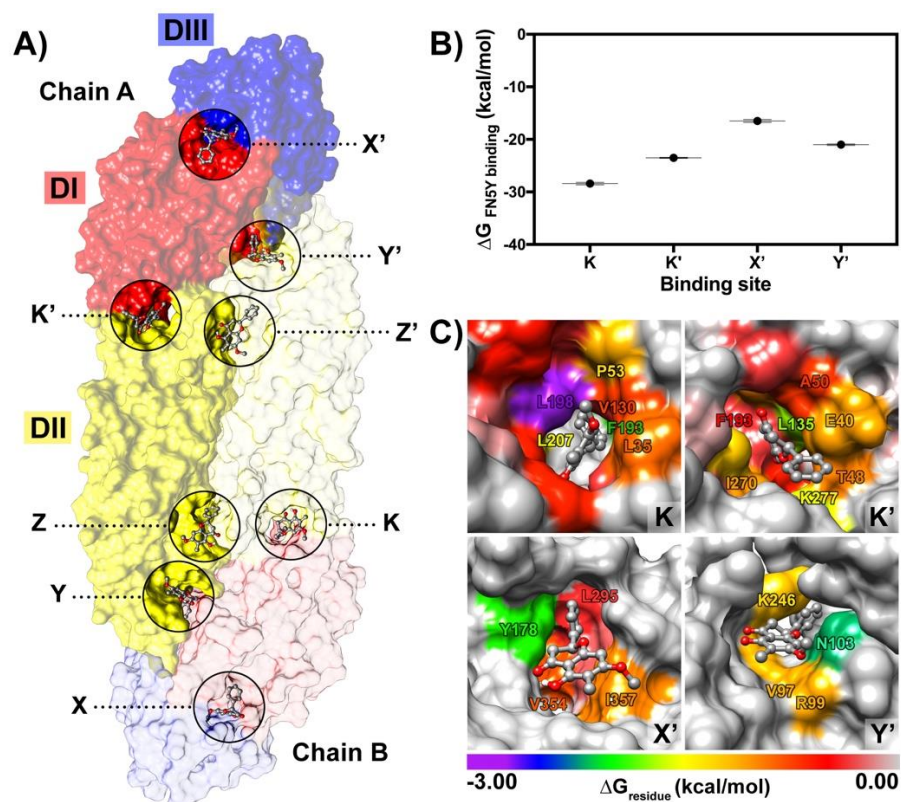


Figure 31 Binding mechanism of FN5Y towards dengue envelope homodimer. a) The initial complex structure of FN5Y/E protein generated by molecular docking method. b) Binding free energies of FN5Y/E complex ($\Delta G_{\text{FN5Y binding}}$) at the K, K', X', and Y' sites c) Closed-up of FN5Y binding orientations at the K, K', X', and Y' sites, where the interacting residues are colored according to energy values as shown in color scale. The residues with energy contribution lower than -0.5 kcal/mol are labeled.

Based on the simulated annealing algorithm [106], EGCG was separately docked into the six identified sites (X, X', Y, Y', Z and Z') of the Zika E protein homodimer using the CDOCKER module implemented in Accelrys Discovery Studio software. From 100 docking runs at each site, the lowest interaction energy of EGCG and its intermolecular hydrogen bonds with the E protein were summarized in Table 2. The

strongest EGCG binding was found at the X site with six formed hydrogen bonds with the chain-A residues Q147, H148, D161, Q162, R164 and K301. The other cases showed relatively similar binding strength (from -43.1 to -46.6 kcal/mol) and were stabilized by two to five hydrogen bonds (Table 4.1). The EGCG conformation was selected based on the best pose generated by CDOCKER interaction energy score. In order to construct the initial structure, the three different complexes are containing the dimeric E protein with two EGCG molecules bound at the opposite site (X-X', Y-Y' and Z-Z') were used for MD study.

Table 2 CDOCKER interaction energy (kcal/mol) and hydrogen bonds of EGCG binding at the six possible binding sites on the Zika dimeric E protein.

Predicted binding regions	X	X'	Y	Y'	Z	Z'
CDOCKER interaction energy score (kcal/mol)	-58.7	-46.6	-43.1	-45.1	-43.6	-46.4
Residues involved in hydrogen bond interaction	Q147	-	R2	G102	H210	T254
	H148		G5	N103	Q274	
Chain A	D161		N154	K251	A275	
	Q162				Q276	
	R164					
	K301					
Chain B	-	Q147	N103	G5	V255	K209
		H148	R99			
		R164				

The binding pattern of EGCG at different sites on surface of Zika dimeric E protein was calculated by per-residue decomposition free energy based on MM-GBSA using the 100 snapshots extracted from the

last 200 ns of 500-ns MD simulations. The resulted per-residue decomposition free energies ($\Delta G^{\text{residue}}$) for all complexes are given and compared in Fig. 32, where the EGCG orientation inside each binding site is depicted in Fig. 33. The residues with negative $\Delta G^{\text{residue}}$ value less than -1 kcal/mol are only labeled. The simulated results suggested that EGCG was able to bind at all the identified binding sites but differed in binding orientation and strength.

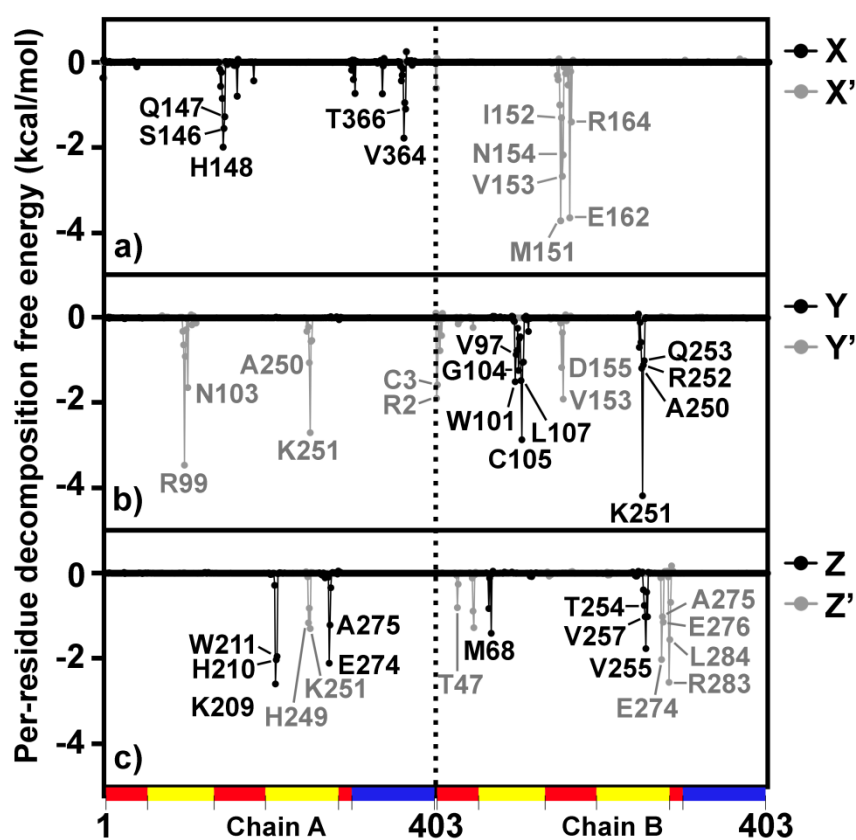


Figure 32 Per-residue decomposition free energy based on MM-GBSA method for EGCG binding at the different sites in acidic condition where the residues with energy contribution ≤ -1 kcal/mol are labeled.

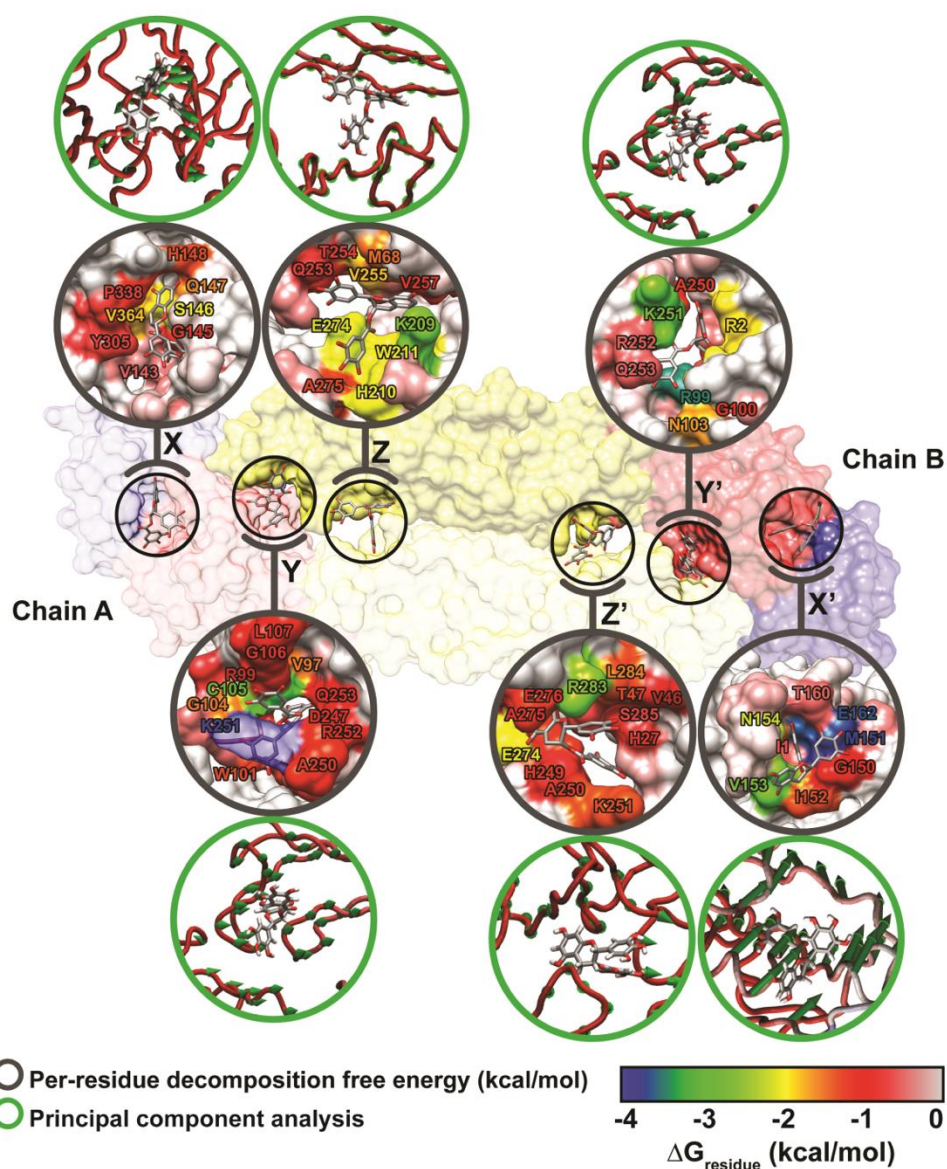


Figure 33 Close-up of EGCG binding orientations at all binding sites shown in grey circle, where the interacting residues are colored according to energy values from Fig. 4. The principal component analyses reveal in green circle of each binding region. The porcupine plots consist of arrow (green) and length that represent the direction of motion and amplitude of residue mobility.

Fig. 33 shows the possible binding orientations of EGCG at domain I/III hinge region in chain A (X) and chain B (X') were different. At X site, the gallate ester inserted into the pocket, while chromane and trihydroxyphenyl rings of EGCG were stabilized by S146, Q147, H148, V364, and T366 with energy contributions from -1.1 to -1.9 kcal/mol. The impact of electrostatic interaction from E162 was supported by the previous MD study of EGCG and small screened molecules binding at this site of Zika E protein monomer and dengue E protein dimer, respectively [6, 54, 55]. In contrast, the higher ligand-protein interactions were found at the X' site with M151, I152, V153, N154, E162 and R164 (from -1.3 to -3.7 kcal/mol) [183].

More residue contributions for EGCG binding were observed (Fig. 32b and 33) at the conserved regions among flaviviruses (Y and Y' sites). With an insertion of the chromane ring into the interface between domain I of chain A and domain II of the another, the EGCG molecule was stabilized in a range from -1.0 to -4.2 kcal/mol by the five more residues that preferentially contributed for EGCG binding at Y (V97, W101, G104, C105, G106, L107, A250, K251, R252, and Q253 from chain A) Y' (R99, N103, A250, and K251 from chain A, and R2, C3, V6, V153, and D155 from chain B) sites. These binding residues at Y site were located in the fusion loop of E protein (residues 98-109), which were highly conserved among flaviviruses [22, 179, 184]. The Zika E protein residue V153 was congruent with the V151 residues of dengue E protein based on protein sequence alignment between Zika and dengue involved with the EGCG binding at this site [184]. Moreover, the residue N154 near X-X' binding site was found to be the key residue in the glycosylation site for viral adsorption to receptor on the host cell surface vaccine besides development target region [106, 151]. For the gallate ester placed on external surface of E dimer at the interface between the domains II of the two chains, five major interacted residues of chain A (K209, H210, W211, E274, and A275) and four contact residues of chain B (M68, T254, V255,

and V257), stabilized EGCG at the Z binding region in a range of -1.0 and -2.6 kcal/mol, whilst ligand in the another site was stabilized by -1.0 to -2.5 kcal/mol from both monomers: H249, A250, and K251 of chain A and T47, E274, A275, E276, R283, and L284 of chain B.

To understand the dynamics of E protein in complex with EGCG, principal component analysis (PCA) was applied in this study [6, 185]. This method employs the coordinate covariance matrix of the protein C α atoms along the average structure in the trajectory at last 100-ns simulation. The 2D projection of each simulated system on the first two PCs acquired from diagonalizing the covariance matrix of atomic fluctuation (Fig. 34, left) showed that both PC1 and 2 of all three complexes had the uniform distributions and no significant difference. It can be clearly seen from the Scree plot of variances and accumulated percentage by the first ten PC modes (Fig. 34, right) that PC1 is dominant. The crucial motion of three complex systems was represented by the first two PC modes ratio; 2.27 (X-X'), 1.82 (Y-Y') and 2.29 (Z-Z'), and the first ten PC modes of X-X', Y-Y' and Z-Z' systems revealed the 77.00%, 68.05%, and 69.03% of accumulated variance, respectively.

In addition, the protein motion of the top PC mode for each system was visualized by porcupine plot (green arrow) and length of arrow shows the direction and amplitude of selected eigenvectors for C α atoms in protein structure (Fig. 33) [35, 186]. It can be seen that the motion of residues in E protein at domains I (red) and III (blue) around X and X' binding sites presented in a similar motion that the loops slightly shifted close to the EGCG molecule. These regions are involved to a conformational change of E protein in the acidic environment [181]. For the binding sites between the chains A and B of E protein, the dynamics of residues at Y-Y' and Z-Z' sites reveal the different patterns of motion. For Y and Y', the loop at the tip of domain II (yellow) and the loop of domain I in the opposite chain (red) partially moved towards EGCG. There was not found in the Z-Z' site.

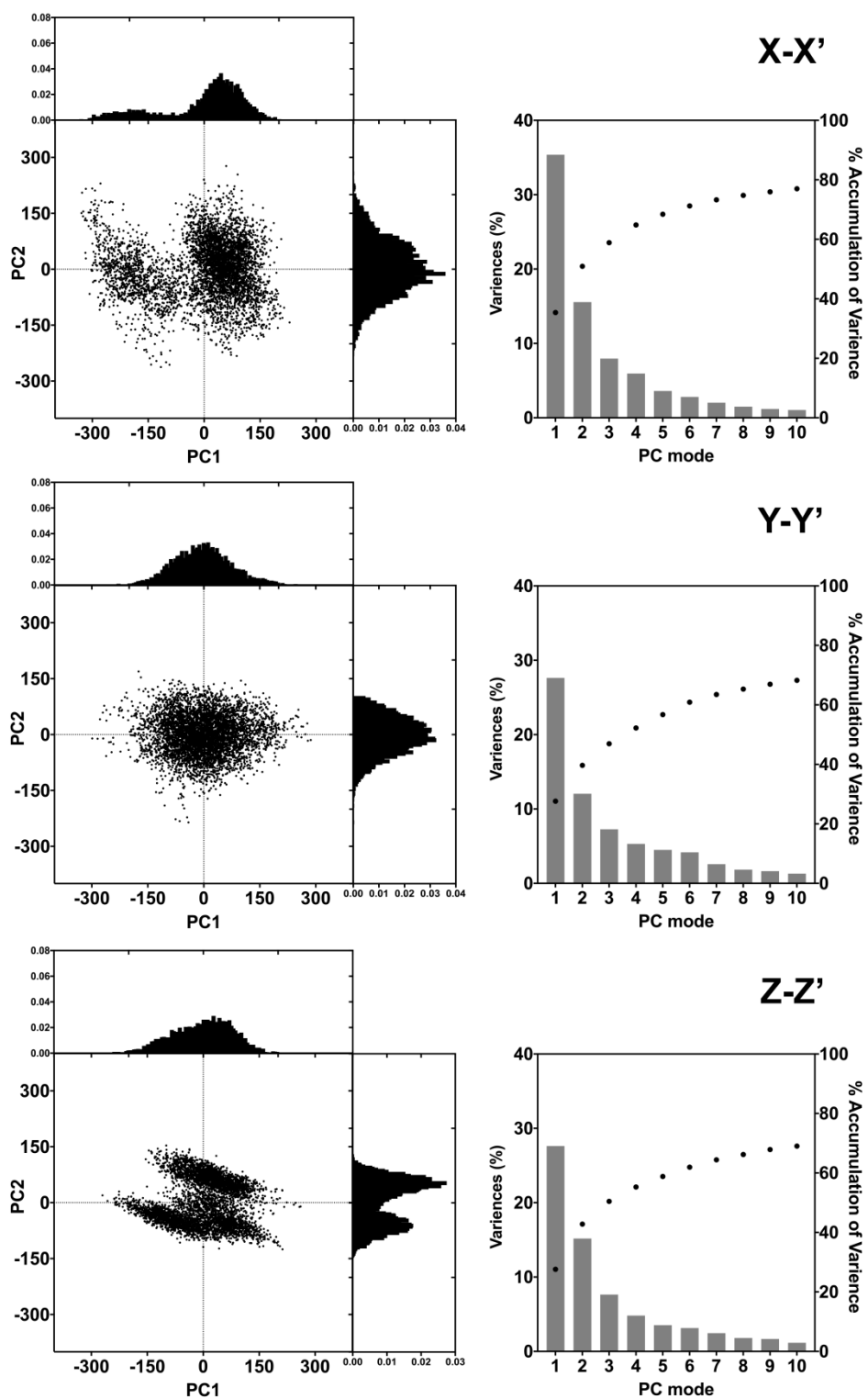


Figure 34 The 2D projection of MD trajectories on the first two PCs (left) and the Scree plot of variances and accumulated percentage by the first ten PC modes (right) of each complex system.

As known, the formation of ligand-protein complex and the ligand binding specificity is importantly involved with intermolecular hydrogen bond. Herein, hydrogen bond analysis was performed to determine the intermolecular interactions between EGCG and the viral Zika E protein residues of at each site based on criteria of: (i) distance between hydrogen bond donor (HD) and acceptor (HA) less than 3.5 Å, and (ii) hydrogen bond angle of HD_H...HA more than 120°. The hydrogen bond occupation calculated over the production phase of simulation was summarized in Fig. 35. The results were discussed in terms of hydrogen bond strength [155, 157, 158]. The probability of hydrogen bonding interactions represented by the gradient of greenish, reddish and bluish grid cells from low to high hydrogen bond strengths.

Fig. 35 shows that EGCG is able to form hydrogen bond with E protein but mostly with low occupation (<30%). EGCG was stabilized by medium hydrogen bond formation (>30%) at X' site (S146 and Y305), X' site (V153 and E162), Y site (D98 and H248), and Y' site (D98, V153 and D155). The stronger hydrogen bonding with E protein was detected at Z site (H210 and A275) and Z' site (H249 and E274). E162 was found to be a crucial residue for EGCG [55] and small molecules [93] binding at the domain I/III hinge region in E protein monomer of Zika and dengue viruses, respectively. At Y-Y' site, EGCG is able to interact with the residues D98, R99 and G104 of fusion loop, which are conserved residues among flaviviruses [22, 93, 184]. From the Fig. S3 in the Appendix, the distance between the centers of mass of domain II of Zika E protein homodimer was measured from EGCG/E protein complexes along the 500-ns simulation. The distances were ~28-31 Å, ~26-28 Å, and ~27-29 Å for X-X', Y-Y', and Z-Z' systems respectively. Such shorter distance found in the system with EGCG binding at Y-Y' sites relative to that of apo-form (28-30 Å) [35] suggested that EGCG could inhibit the Zika viral infection by inactivating the function of E protein at the fusion loop of domain II on E protein.

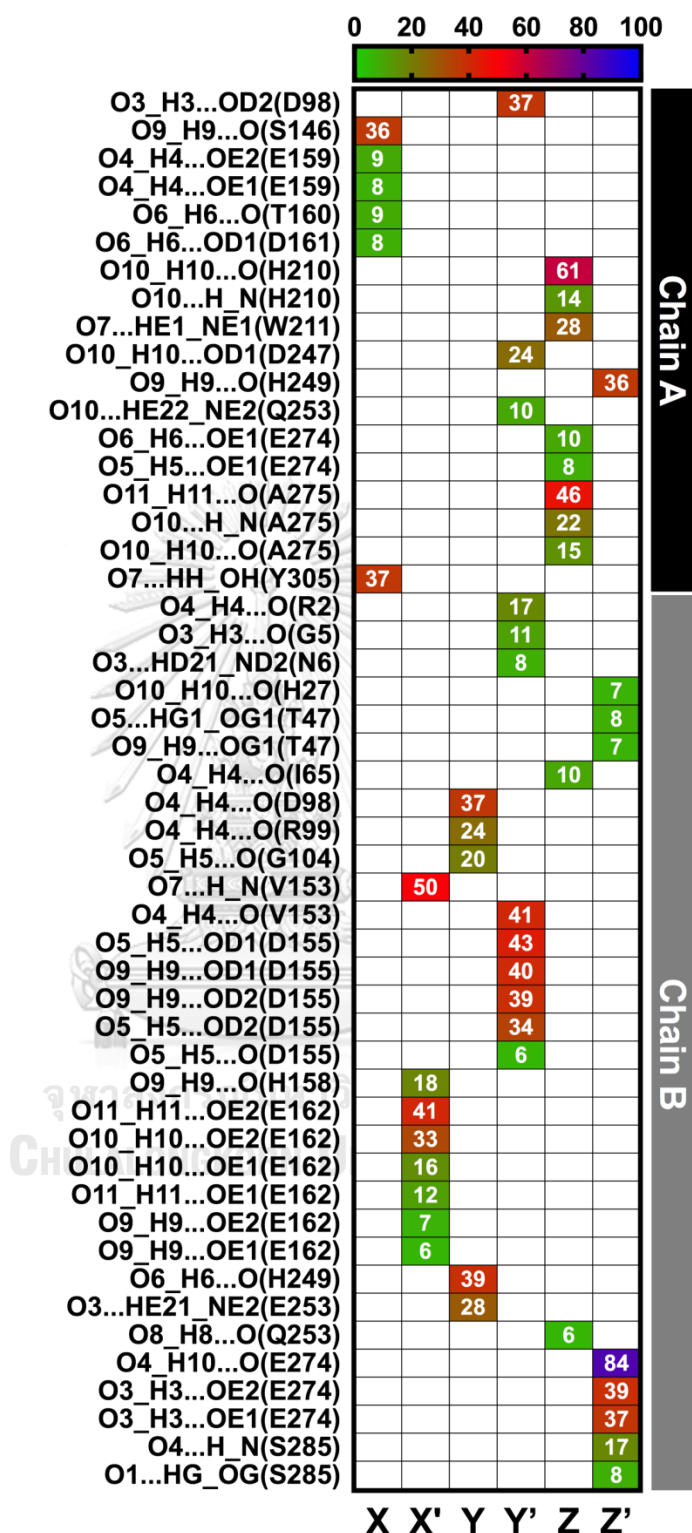


Figure 35 Hydrogen bonding interactions of EGCG at each site are presented by grid map, while hydrogen bond strength is defined by label and color in the grid cells.

The binding free energy (ΔG_{bind}) of EGCG binding on the surface of Zika E protein homodimer was calculated based on 4 different binding free energy calculation methods including SIE, MM-GBSA, QM/MM-GBSA with PM3 and SCC-DFTB methods [123, 125, 187] treated on EGCG molecule only. These approaches have different limitations and computational consumptions [97, 121]. The ΔG_{bind} based on SIE method was calculated using *sietraj* program [165]. This method estimates binding free energy in solvation (ΔG_{bind}) by a summation of Coulomb (ΔE_{elec}) and non-polar components (ΔE_{vdW}) of interactions as well as desolvation free energy contribution [96]. The MM-GBSA and QM/MM-GBSA binding free energy calculations were performed using *MMPBSA.py* module. The binding free energy calculation of complex was performed on the 100 snapshots taken from the last 200-ns of MD simulation. The ΔG_{bind} values comparison of complexes were calculated by using different techniques, which are plotted in Fig. 36, while the energy components are given in Tables S1-3 (Appendix). The correlation of binding free energy results is analyzed by Pearson correlation (Fig. 36).

From Fig. 36, the ΔG_{bind} values calculated by various methods show the similar trend for all complexes. According to these results, the favorable binding region of EGCG shows at Y and Y' sites that evaluated by SIE ($\sim -8.50 \pm 0.1$ kcal/mol), MM-GBSA ($\sim -27.50 \pm 0.7$ kcal/mol) and QM/MM-GBSA (PM3 and SCC-DFTB) calculations ($\sim -42.2 \pm 1.0$ and -35.00 ± 0.8 kcal/mol). However, EGCG can also bind on E protein surface at the other sites suggested by negative ΔG_{bind} values. Again, it is a worth stating that the preferential binding sites of EGCG are found in between the domain I of one chain and domain II of the another (Y-Y' site). Such strong binding affinities were contributed from vdW interaction, which was greater than EGCG binding at the domain I/III hinge region (X-X' site) and Z-Z' site, supported by ligand-protein interactions and PCA results as mentioned earlier.

The Pearson correlation coefficient ranged from 0.69 to 1.00 for 4 end-point free energy calculation methods indicated that both MM-GBSA and QM/MM-GBSA methods are congruent with $p=0.006$ (PM3), and $p=0.00$ (SCC-DFTB), whereas the SIE method show the lower correlated with other calculations with $p=0.129$ (MM-GBSA), $p=0.076$ (PM3), and $p=0.125$ (SCC-DFTB) as shown in Table S4 in the Appendix

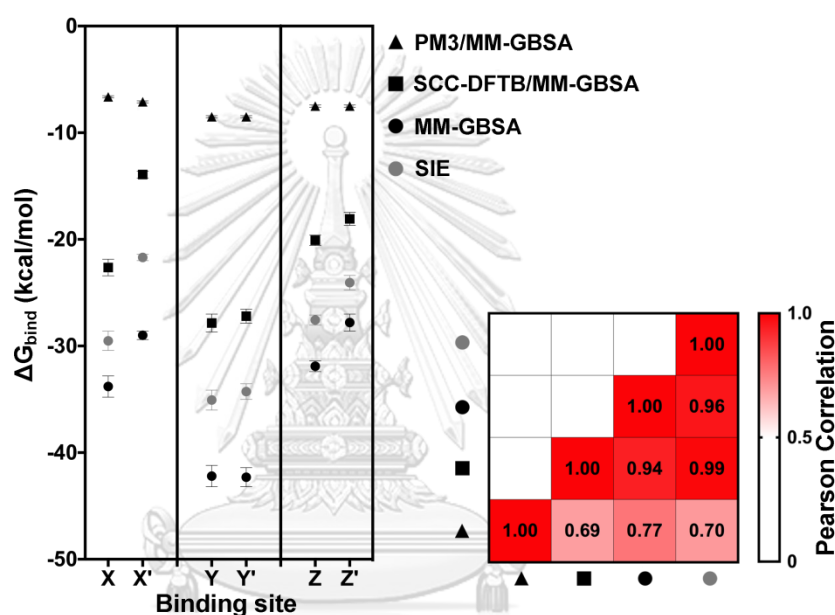


Figure 36 Comparison of the average binding free energy (ΔG_{bind}) in kcal/mol of each EGCG binding at the different sites on Zika dimeric E protein at acidic environment and the Pearson correlation analysis (heat map) by using four different binding free energy calculation methods; SIE (triangle), MM-GBSA (square), QM/MM-GBSA with SCC-DFTB (black circle), and QM/MM-GBSA with PM3 method (gray circle).

4.2 Virtual screening

According to the previous study, a flavanone derivative (FN5Y) has been reported for the inhibiting effect on dengue infected cell by cell-based assay and time of addition method. This inhibitor might be possible to interrupt at the early step of dengue viral infection, which is related to E protein [4]. In this step, we would like to screen the novel potent flavonoid molecule that could inhibit dimeric dengue E protein and we also illustrate the binding pattern and interaction profile of potent molecules at an atomic level by computational study.

4.2.1 Pharmacophore-based virtual screening

To increase the virtual screening performance, every snapshot during equilibrium phase of FN5Y/E complex trajectory was computed by means of representative pharmacophore model (RPM). The system stability of all ligand-protein complex structure is determined by the root mean square displacement (RMSD) as plotted in Fig. 37. A total of 10,000 MD-snapshots extracted from equilibrium phase at the last 20 ns were used to generate pharmacophore features of FN5Y bound to each binding site of dengue E protein based on RMSD of the complex system using LigandScout 4.2 program [129]. A total of 10,000 MD snapshots extracted from production phase of the DENV2 FN5Y/E complex [4] were used to generate pharmacophore features using LigandScout 4.2 program [129] (Fig. 38). All RPMs were aligned and merged, then, they were used as the template for pharmacophore-based virtual screening [169]. According to Fig. 4, the major pharmacophore features among the 4 binding regions are mainly described by hydrophobic properties of the aromatic ring system of the FN5Y molecule (yellow sphere). Additionally, the 3D interaction pattern reveals hydrogen bond donor (green arrow), hydrogen bond acceptor (red arrow), and an excluded volume (grey sphere). These properties are collected and aligned into 40,000 RPMs and filtered into 136 RPMs in order to reduce the

computational afford. All of these RPMs are obtained from each binding site: K (31), K' (26), X' (42) and Y' (3), with 1,536 pharmacophore features. The pharmacophore-based virtual screening method is performed on each RPM using Hungarian matching algorithm implemented in LigandScout 4.2 program [129].

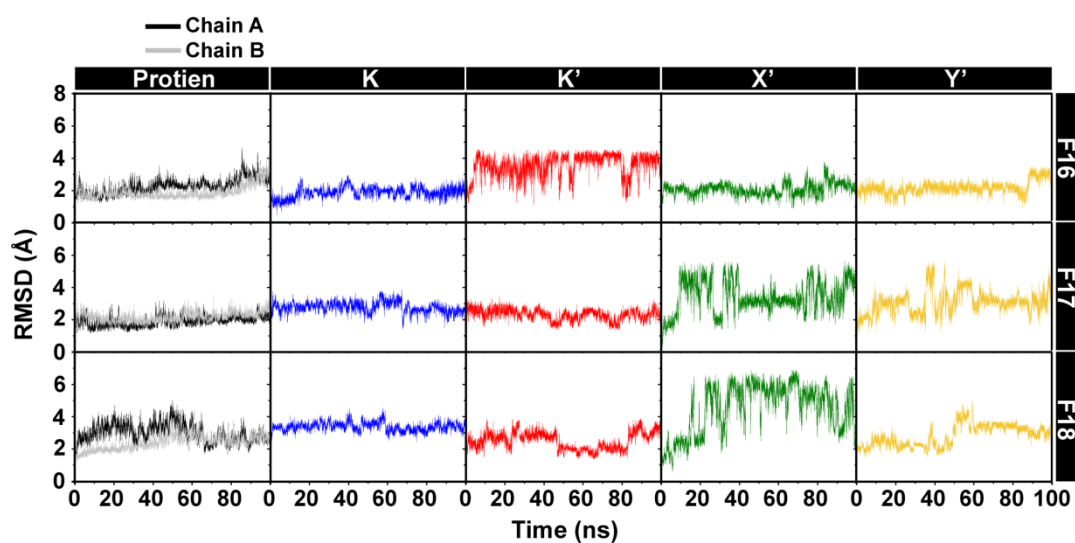


Figure 37 RMSD graph of ligand/E protein (above) and F18/E protein complex based on molecular dynamics simulation trajectories during 100-ns simulations.

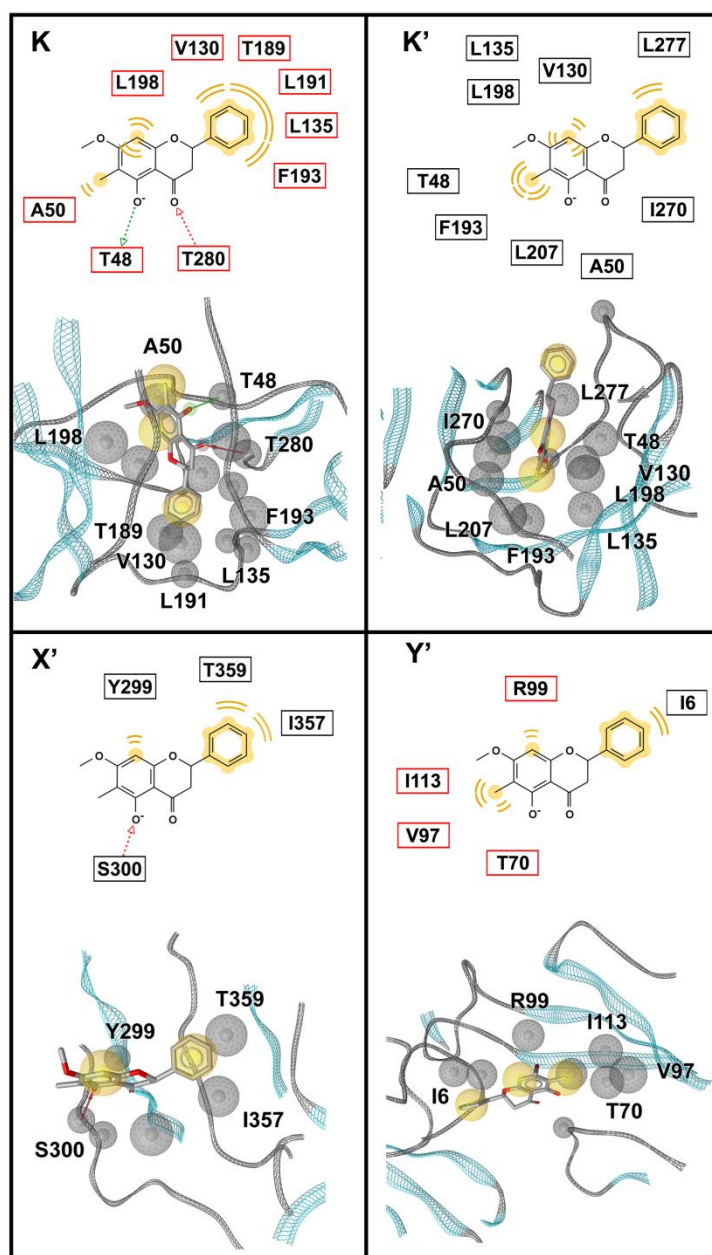


Figure 38 The 2D and 3D pharmacophore models of active ligand/E protein complex and the interacted residues in each pocket (K, K', X' and Y') were extracted from the first snapshot of a last 20 ns-md trajectory. The pharmacophore features are represented as green arrow (HBD), red arrow (HBA), yellow sphere (hydrophobic property) and grey sphere (excluded volume). The interacted residues from chain A and B were labeled in black and red box, respectively.

The screening results from 136 different RPMs are merged and rescored. The consensus molecules are ranked by common-hits approach (CHA) [169]. From the 996 compounds in the focused library, the result shows that only 26 flavonoids have a higher CHA score than 50 (Fig. 40, right): from in-house (F) 17 molecules including FN5Y, TimTec (ST) 7 molecules, and Zinc (ZINC) 1 molecule (Fig. 39). F15 reveals the highest CHA score of 77, whereas ZINC000236568961 shows the lowest CHA score (50). F15 is obtained from 82 from 136 pharmacophore-based virtual screening run (60%) with 57 different conformations.



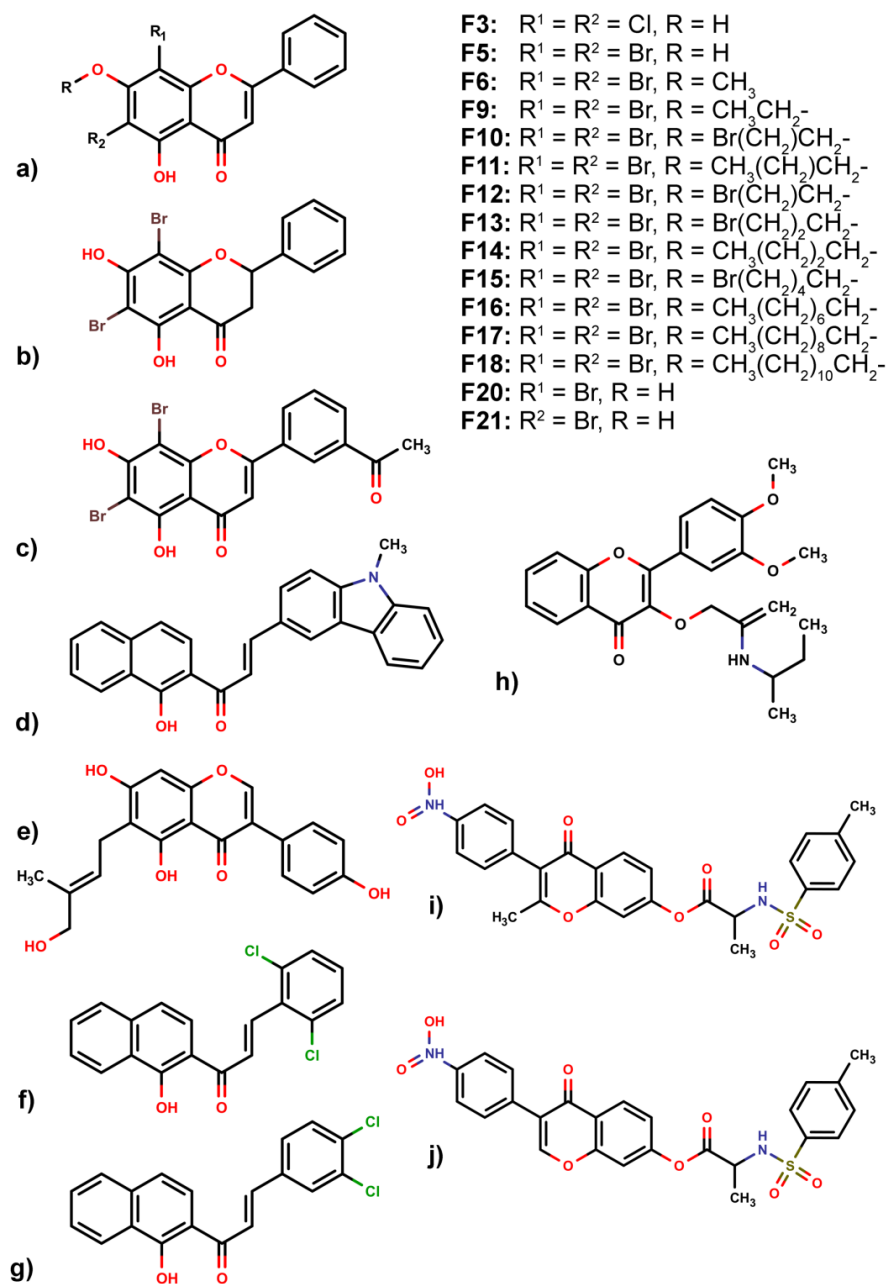


Figure 39 The 2D structures of 25 molecules were obtained from pharmacophore-based screening method. Twenty-five flavone derivatives were derived from 3 compound databases; in-house (F), TimTec (ST) and ZINC database. a): F3-6, 9-18 and 20-21, b): F4, c) F30, d): ST080183, e): ST077115, f): ST070287, g): ST070182, h): ST090053, i): ST080306 and j): ST080278 and k): ZINC236568961

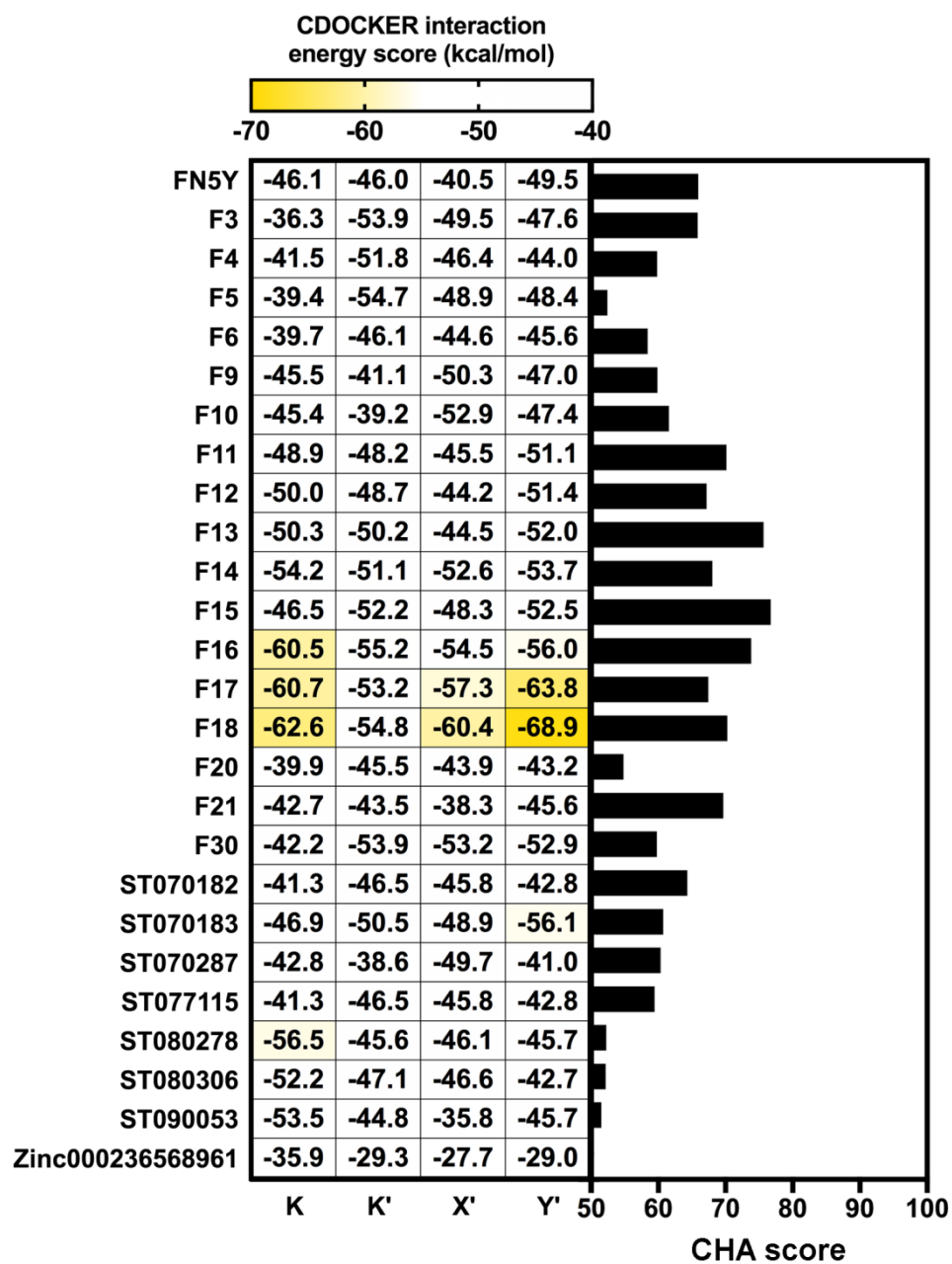


Figure 40 Comparison between CHA score of screened molecule bound to all binding sites (right) and CDOCKER interaction energy score (left). The screened molecules which have CHA score over 50, are chosen and plot as a bar graph. All of these molecules were docked into 4 binding sites of dengue E protein using CDOCKER software.

4.2.2 Molecular docking

A total of 26 screened molecules including FN5Y were docked into the 4 specific binding sites (K, K', X', and Y') of DENV2 E protein. The results show all of the flavonoid derivatives from 3 databases can fit well into every binding pocket based on CDOCKER interaction energy score (Fig. 40, left), which is similar to the pattern of iGEMDOCK's fitness score (Fig. 41). Particularly, in-house flavone derivatives F16, F17 and F18 are ranked in the top three with the lowest CDOCKER interaction energy score (from -53.24 to -68.93 kcal/mol). These molecules can be potent ligands for designing broad-spectrum inhibitors against DENV at E protein target. The common structure of the top three candidate molecules contains long hydrophobic tails $-(\text{CH}_2)_8$, $-(\text{CH}_2)_{10}$ and $-(\text{CH}_2)_{12}$ (Fig. 39) highly interacting with the hydrophobic region in the binding site of E protein (Fig. 40). Moreover, the relevant contact residues of hit/E protein complex at each site are described in Table 3. The important residues of all binding sites are non-polar (I6, V130, and F193), polar (T155 and T359), and charged (K247 and K295) amino acids. In order to characterize the behavior of ligand in the complexes, the conformations of F16, F17 and F18 at individual binding sites (K, K', X' and Y') were studied by MD simulations.

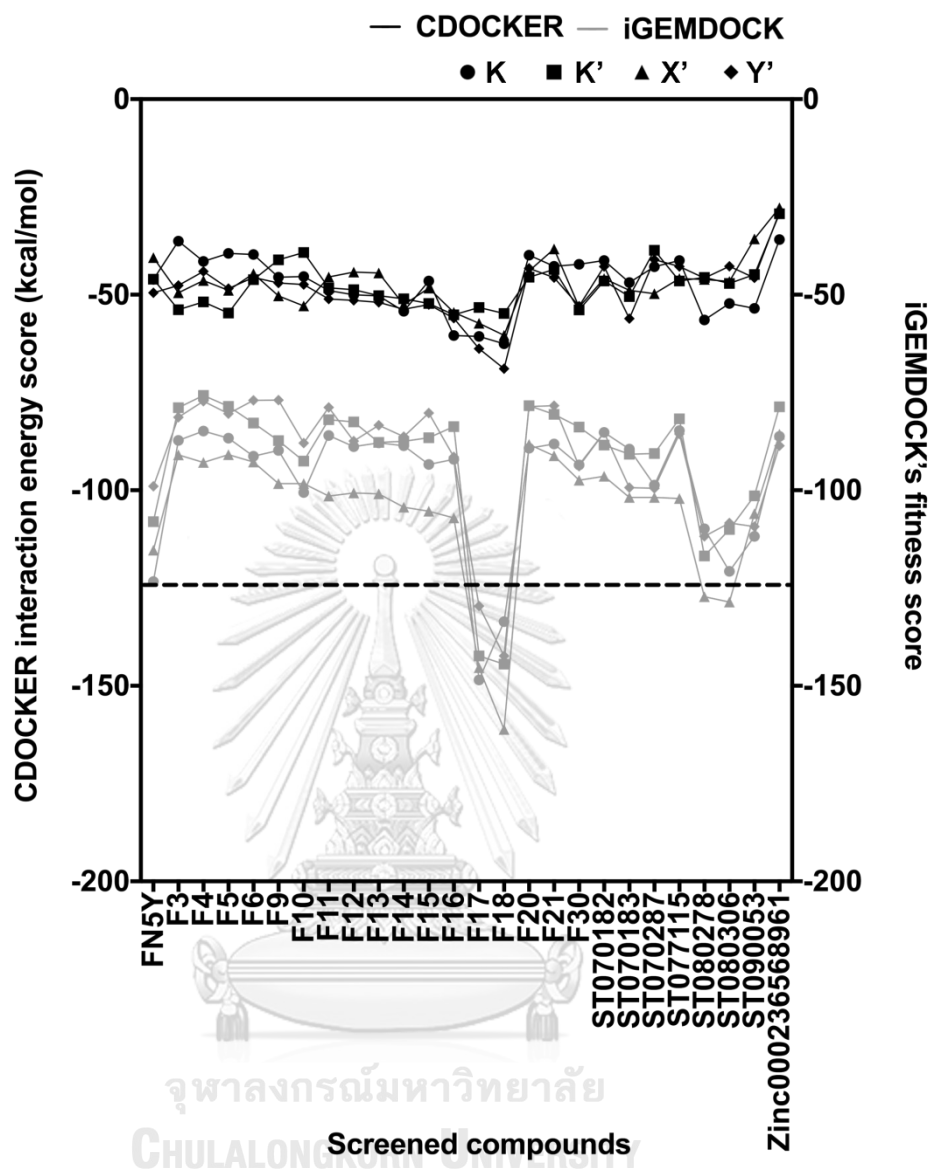


Figure 41 Tendency plot of screened molecule interaction energies compare between CDOCKER (black) and iGEMDOCK (grey) programs which is computed using different algorithms. The dash line represents the minimum iGEMDOCK's fitness score as a standard of this study.

Table 3 Contributed residues of lead molecules bound to four binding sites on the dengue dimeric E protein, compare to FN5Y molecule. The CDOCKER interaction energy is represented follow by Figure 4. The conserved residues of each binding regions revealed as bold letter.

	K	K'	X'	Y'
FN5Y	T48	A50	K295	I6
	A50	K128	Y299	T155
	V130	L135	I357	V97
	L135	F193	T359	K247
	L191	L198		
	F193	I270		
	L198			
	L207			
	L277			
F16	T48	T48	K295	I6
	E49	E126	Y299	T155
	A50	K128	I357	T70
	V130	L135	T359	K247
	F193	F193		
	L198	I270		
	L207	L277		
	I270			
	L277			
F17	T48	K47	K295	I6
	E49	A50	Y299	T155
	A50	V130	I357	T69
	V130	L135	T359	T70
	L135	F193		I113
	T189	L198		K247
	L198			
F18	V130	K47	K295	I6
	L191	A50	T303	T155
	F193	V130	T359	T69
	Q200	F193		T70
	L207	L198		K247
	L277			

The performance of pharmacophore-based screening results and the ability of identification between 6,190 decoys and 26 active compounds were achieved by the Receiver operating characteristic (ROC) plot. Note, that the area under the curve (AUC) represents the quality of ROC plot. The AUC value > 0.50 suggests that the results from this method are reliable in which the active molecules are likely screened [169]. In Fig. 42, for the present work the AUC values are 0.82 (1%), 0.96 (5%), 0.97 (10%), and 0.88 (100%), therefore, the hit compounds from this pharmacophore-based screening are acceptable for further antiviral drug development.

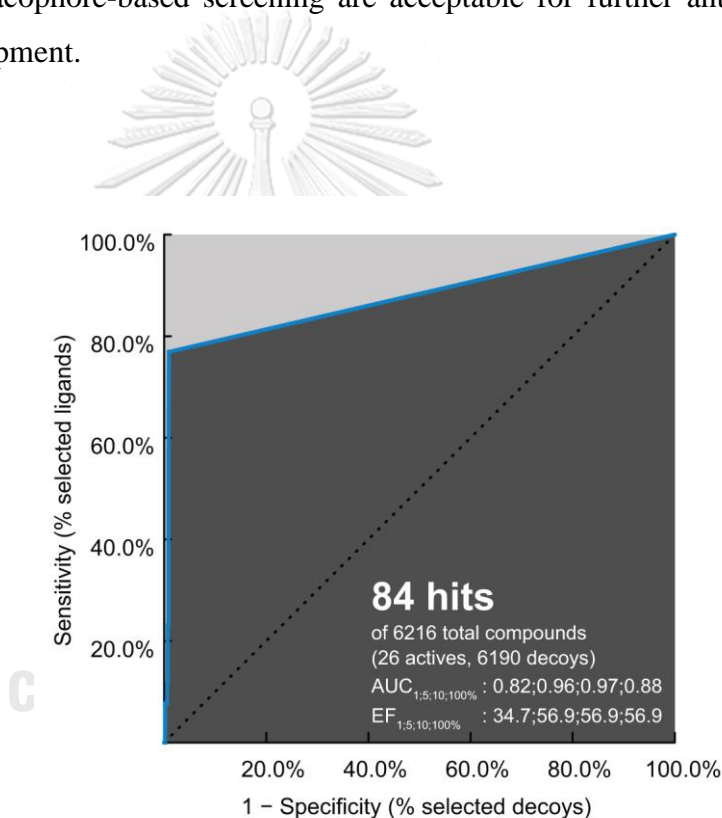


Figure 42 Receiver operating characteristic (ROC) plot of pharmacophore model applied to dengue E protein. The enrichment factor (EF) and area under the curve (AUC) are represent at 1, 5, 10, and 100% of database.

4.2.3 Binding energy calculation

To identify the most potent flavone derivative binding to DENV E protein, the binding free energy calculation based on solvated interaction energy (SIE) method was applied on the three complexes (F16, F17 and F18) using the last 40-ns trajectories, in comparison to the known compound, FN5Y. The SIE binding free energy in solvation (ΔG_{bind}) is estimated by a summation of non-polar components (ΔE_{vdW}) and Coulomb interactions (ΔE_{elec}) as well as desolvation free energy contributions [96, 121]. The ΔG_{bind} values of the considered complexes are plotted in Fig. 7, while the energy components are given Table S5 (Appendix).

In Fig. 43, the calculated ΔG_{bind} values of F16, F17 and F18 complexes are in a range of -10.48 to -13.06, -8.56 to -10.81, and -11.11 to -14.42 kcal/mol, respectively. These predictions suggest a binding of these halogenated flavones with the E protein significantly stronger than that of FN5Y (-3.43 to -4.35 kcal/mol). In addition, vdW interaction is the major contribution for ligand/E protein complexes (ΔE_{vdW} , Table S4). F18 has the highest efficacy to bind on surface of the E protein at all sites than other ligands, and its most preferential binding site is located at the kl loop region, K and K'. The bromine substitutions in F18 could strengthen the interactions at all pockets [188-191], which may lead to an improved drugability of this potent molecule [192].

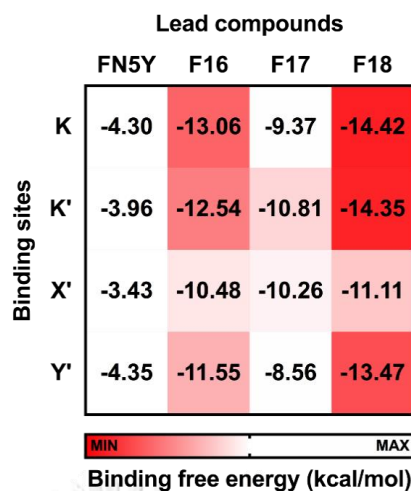


Figure 43 Comparison of the average SIE binding free energy (ΔG_{bind}) in kcal/mol of each flavone derivative binding (FN5Y, F16, F17 and F18) at the different sites on dengue dimeric E protein (K, K', X' and Y').

The insight of each binding region in the F18/E complex is figured out by end-point binding free energy calculation; MM-GBSA and MM-PBSA method, and QM calculation (FMO). The MD simulation of this complex was extended into 500 ns, then only last 100 ns-MD trajectory was analyzed the binding free energy of each binding sites (Table 4.3). The stability of system can be described by RMSD plotted (Fig. 44). In the kl loop pocket, both vdW and electrostatic interaction based on MM-GBSA/PBSA calculation are mainly contributed energy with contact residues in these sites (Fig. 45). The halogen, bromine atoms of F18 interact to bound residue that might increase the interaction at all pockets [188-191]. Moreover, this interaction is able to improve a drug ability of potent molecule, as well [192]. The average binding free energy suggests that F18 molecules preferentially bound to K and K' sites at -45.81 ± 0.3 and -44.57 ± 0.2 kcal/mol according to MM-GBSA calculation (Table 4). However, X' (domain I/III hinge region) and Y' (conserved region among flaviviruses) also serve as a binding region for F18.

Table 4 Comparison of the averaged binding free energy and its components (kcal/mol) of F18 ligands bound to each binding region; K (blue), K' (red), X' (grey), and Y' (black) on dengue dimeric E protein, which is calculated from last 100 ns-MD trajectory.

	K	K'	X'	Y'
ΔE_{vdW}	-50.25 ± 0.3	-47.72 ± 0.2	-36.86 ± 0.4	-47.8 ± 0.3
ΔE_{elec}	-40.69 ± 0.7	-30.42 ± 0.8	6.83 ± 1.0	-9.83 ± 1.2
ΔE_{MM}	-90.94 ± 0.8	-78.15 ± 0.7	-30.03 ± 1.1	-57.69 ± 1.4
PB	$\Delta G_{\text{polar solvation}}$	-6.79 ± 0.0	-6.37 ± 0.0	-5.81 ± 0.1
	$\Delta G_{\text{non-polar solvation}}$	51.66 ± 0.7	42.41 ± 0.7	9.72 ± 1.0
	$\Delta G_{\text{solvation}}$	44.87 ± 0.6	36.04 ± 0.7	3.91 ± 1.0
	$\Delta G_{\text{MM-PBSA}}$	-46.06 ± 0.2	-42.11 ± 0.2	-26.12 ± 0.3
GB	$\Delta G_{\text{polar solvation}}$	-4.72 ± 0.0	-4.49 ± 0.0	-3.36 ± 0.0
	$\Delta G_{\text{non-polar solvation}}$	49.85 ± 0.7	33.58 ± 0.2	6.82 ± 1.0
	$\Delta G_{\text{solvation}}$	45.13 ± 0.7	38.06 ± 0.7	3.46 ± 1.0
	$\Delta G_{\text{MM-GBSA}}$	-45.81 ± 0.3	-44.57 ± 0.2	-26.57 ± 0.3

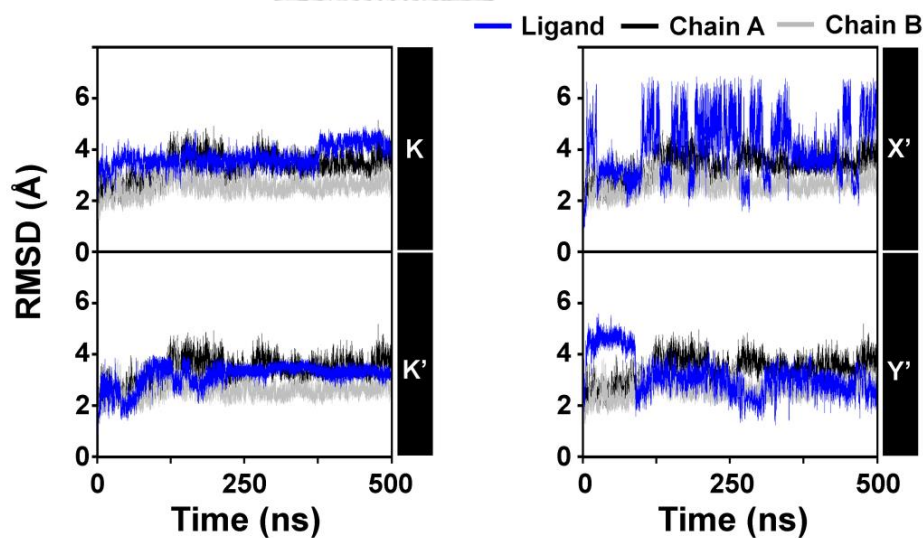


Figure 44 RMSD plots of F18 ligand (blue) bound to 4 different sites; K, K', X', and Y', on the dengue E protein (black and grey) over 500 ns-MD simulation.

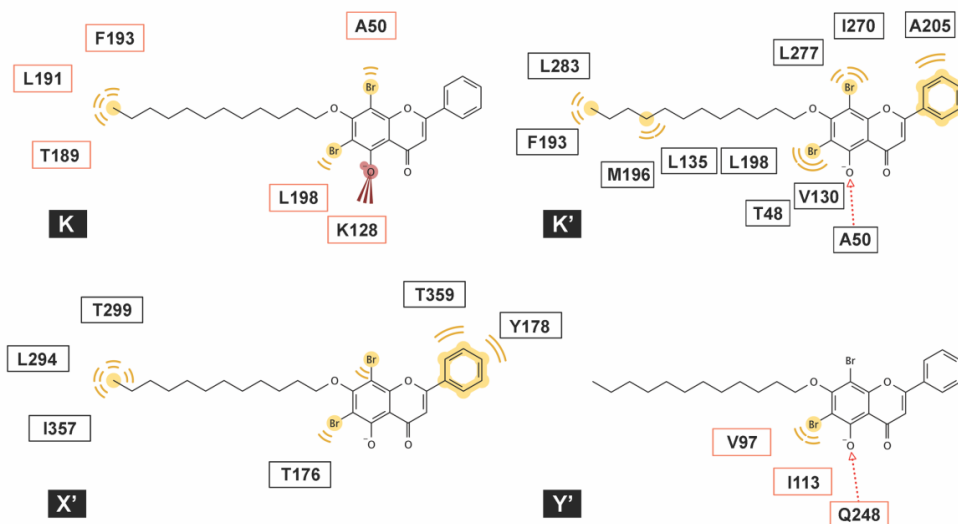


Figure 45 2D Pharmacophore model of potent compound (F18) interacts to residues in chain A (black box) and chain B (red box) at each binding site of dengue E protein. Pharmacophore properties are represented in each moiety of molecule structure that consist of hydrophobicity (yellow), hydrogen bond interaction (red arrow), and electrostatic interaction (rayed red).

Simulation on the F18/E complex was then extended to 500 ns to investigate the binding pattern and interaction profile of F18. Per-residue decomposition free energies ($\Delta G_{\text{residue}}$) for F18 at the four different binding sites over and along the last 100-ns simulation are given in Figs. 8a and 8b, respectively. The F18 orientations inside each binding site are depicted in Fig. 47, where the surrounding residues with $\Delta G_{\text{residue}} \leq -1$ kcal/mol are labeled. There are several residues important for F18 binding at K (E49, A50, P53, K128, L135, L198 and Q200), K' (T48, E49, A50, L135, L198, I270, Q271 and T280), X' (H149, K157, H158 and Y299), and Y' (V97, I113, P243, K247 and Q248). By considering the contributed residues and their stabilization energies, again F18 likely

prefers to interact at the K and K' binding regions. Then, the FMO calculation was used to rigorously reveal the strong paired interaction energy (E_{ij}) [127] between F18 and the residues at K site.

At the K site, the residue K128 provides the highest stabilization ($\Delta G_{\text{residue}}$ of -3.24 kcal/mol in Fig. 46a) mainly through electrostatic attraction with the negatively charged oxygen and partially dispersion to bromine [193] on F18 flavone (-85.24 and -3.99 kcal/mol from FMO MP2/6-31G(d) calculation, Fig. 48). Although the two residues Q200, previously reported as the key amino acid for cyanohydrazone (3-110-22) binding [5], and E49 favorably stabilized F18 (Fig. 46), they showed electrostatic repulsion (34.03 and 31.37 kcal/mol, Fig. 48) with the two bromine atoms (Fig. 47). The other residues A50, P53, L135, and L198 contributed for stabilizing the non-polar tail of F18.

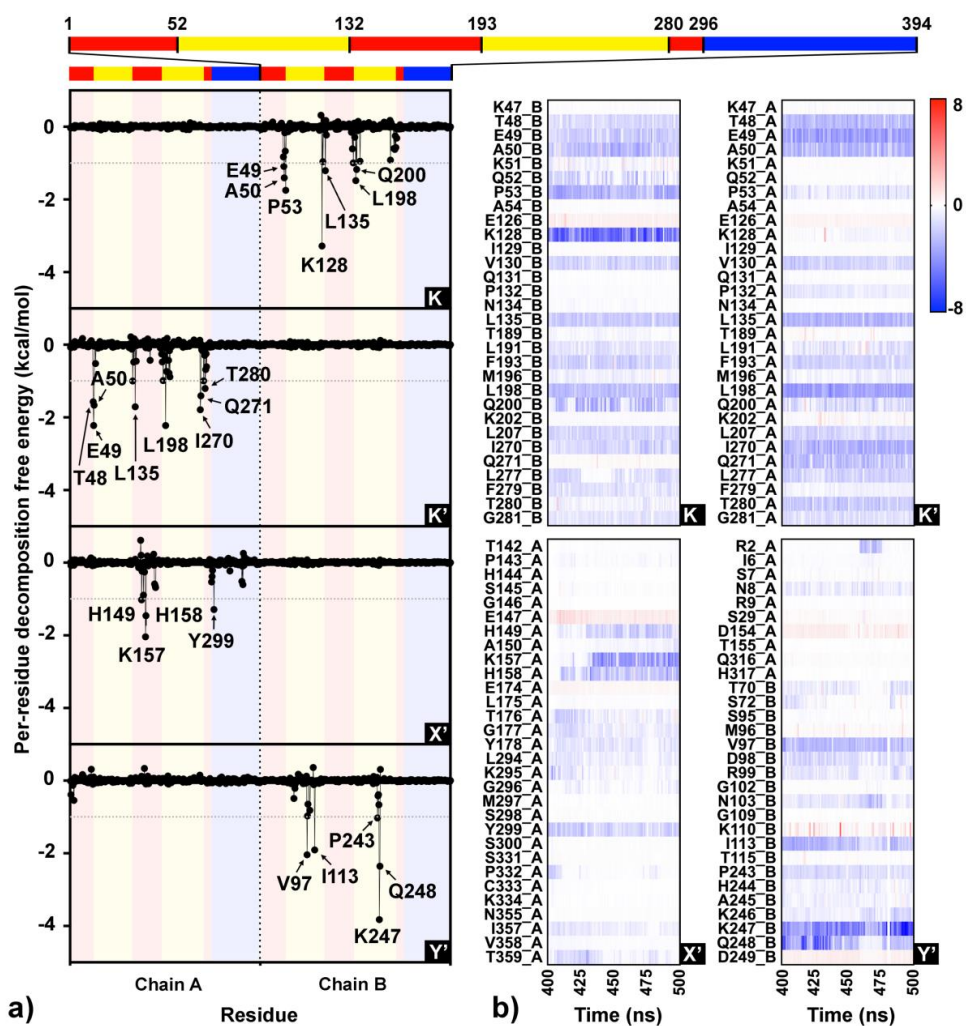


Figure 46 Interaction profile of F18 bound to each binding sites (K, K', X', and Y') on dimeric dengue E protein along last 100 ns-md trajectory based on MM-PBSA and MM-GBSA method. The monomeric E protein shows in 3 different domains; DI (red), DII (yellow) and DIII (blue). a) Per-residue decomposition free energy where the residues with energy contribution ≤ -1 kcal/mol are labeled with average within trajectory. b) The energetic component of each residue within 8 Å around F18 was plotted per time labeled with red-white-blue spectrum color represented as a heat map.

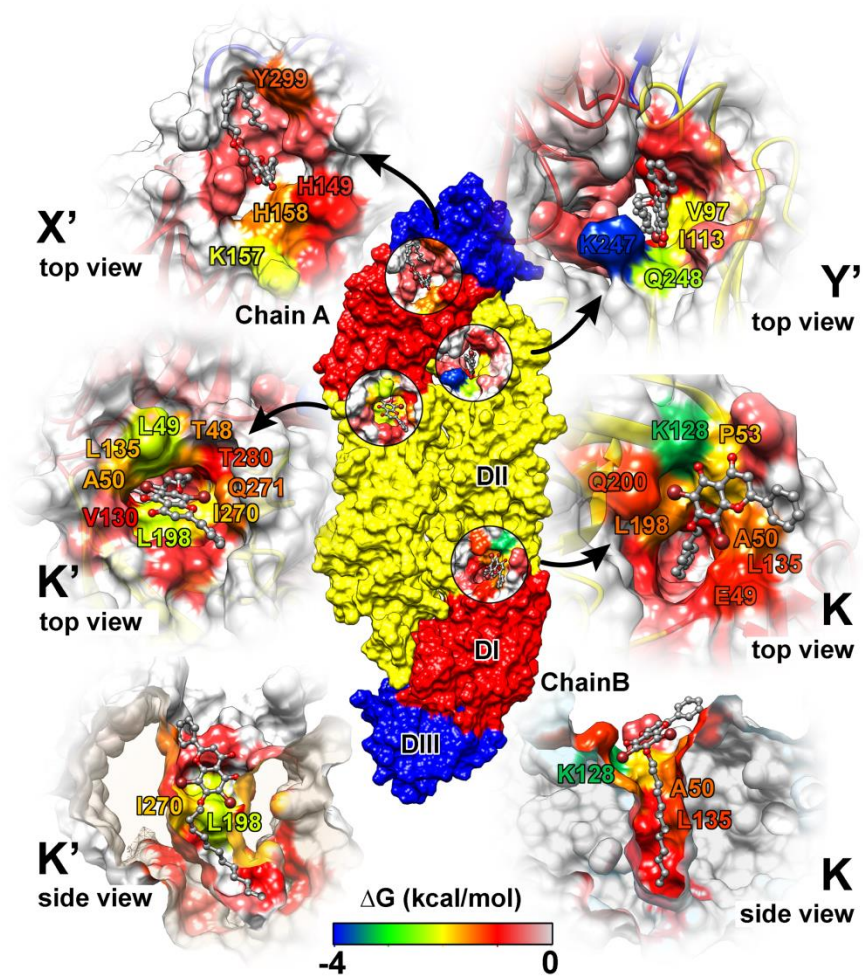


Figure 47 F18 binding orientations at the four binding sites, where the interacting residues are colored according to the energy values from Fig. 46a. The residues with decomposition free energy contribution lower than -1 kcal/mol are labeled.

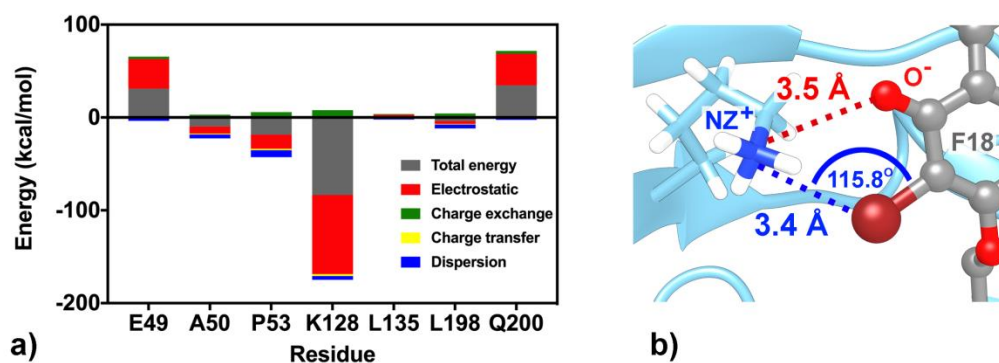


Figure 48 a) FMO2-MP2/6-31G(d) binding energy component between the K-site residues and F18. (b) Electrostatic attraction (red dashed line) and dispersion (blue dashed line) between K128 and flavone moieties.

Chapter V

Conclusion

5.1 Binding pattern evaluation

The binding pattern of ligand on its binding region is able to describe the mode of the inhibitory mechanism of active molecules. Moreover, the information of ligand and protein is also used as a template to search for the novel potent inhibitor by the virtual screening method. The binding pattern of FN5Y on dimeric dengue envelope protein was predicted by molecular modeling approach. The molecular docking study revealed 8 possible binding regions of FN5Y on dengue viral surface, which are the K and K' sites located on the domain I/II or kl loop region, the X and X' sites located on the domain I/III hinge region, the Y and Y' conserved regions among flaviviruses, and the Z and Z' interfaces between the two monomers. After 100 ns-MD simulation, the ligand at only 4 binding sites can bind well along the MD trajectory that are K, K', X', and Y' binding regions according to the binding free energy calculation; SIE. Therefore, the drug elicited strong bindings to K, K', X', and Y' sites of dengue envelope protein and prevented the pH-dependent fusion.

In order to elucidate the binding region of EGCG on Zika viral surface at E protein homodimer by computational simulations, the docking results suggested that there are 6 possible binding regions for EGCG that the same as FN5Y binding to dengue E protein except kl loop binding region. The ligand-protein binding interaction and total binding free energy of the complexes were then studied by all-atom MD simulation for 500 ns at acidic pH environment. The results suggested that EGCG bind at all sites on the surface of E protein. However, the orientations, contacted residues, and binding affinities of EGCG were different. The more favorable binding region for EGCG was found at the Y and Y' sites than others compare to X-X' and Z-Z' sites. The binding of EGCG towards E protein homodimer may prevent its conformational change induced by acidic condition, and thus the activation of fusion process of Zika virus is interrupted. These finding are useful for the

further designing better inhibitor against flavivirus envelope protein, which can block the entry process of the Zika virus and the other viruses in the flavivirus genus.

5.2 Virtual screening

About 40,000 pharmacophore models were filtered into non-redundant 136 RPMs. The common-hits approach was used to refine the result from pharmacophore-based virtual screening. Molecular docking results from both algorithms are mostly congruent to the results from previous step. The binding free energy calculations; SIE, MM-PBSA, and MM-GBSA, are used to carry out the strength and stability of potent compound bound to each binding site on dengue E protein. The binding pocket with the lowest binding energy was obtained via FMO approach. The potent compound was synthesized and tested the inhibiting efficiency by *in vitro* study. This potent compound will be tested in the animal for the further step.

Moreover, the hydrophobic tail modification of halogenated flavone derivative (F16-18) increases the ability of binding on the viral surface. Furthermore, the halogen interaction of bromine atom is vital for stabilizing the potent molecule in its binding region. The binding pose of potent compound at the best binding region of dengue E protein, kl loop, has been proposed. The binding pattern and interaction profile will be used for designing the inhibitor against dengue viral infection. Moreover, the potent compound might need to optimize for finding the novel inhibitor using combination between the key residue interaction profiling and potent scaffold.

APPENDIX

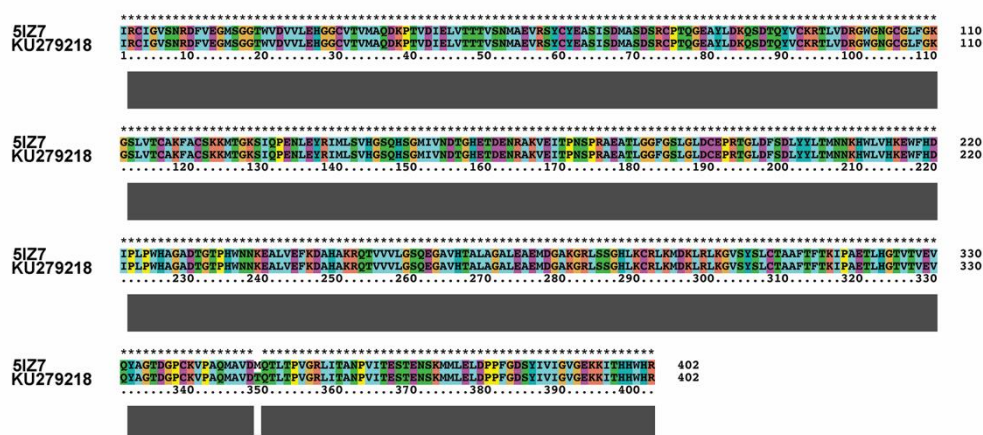


Figure S1. Progressive pairwise sequence alignment of ZIKV E protein sequences from French Polynesia strain H/F/2013, PDB entry code: 5IZ7 and Brazilian (BeH 828305, GenBank accession number: KU729218 [12] using ClustalX 2.1.

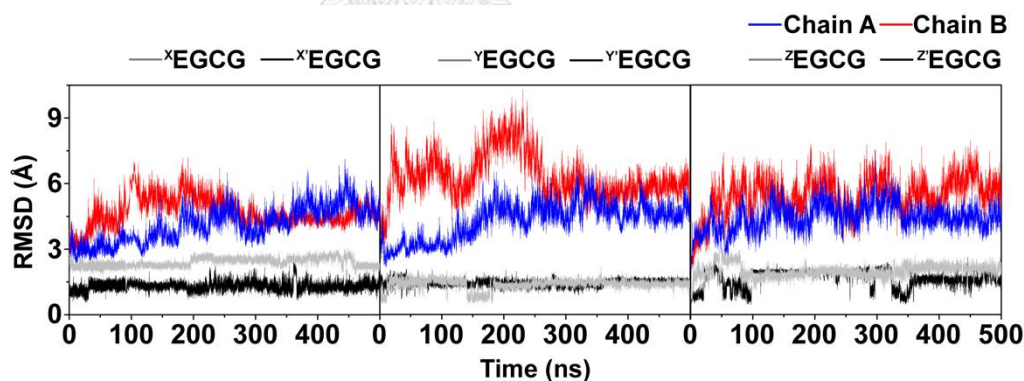


Figure S2. RMSD plots for all atoms of EGCG and E protein chains A and B for all three complexes with two EGCGs binding to E protein homodimer at X-X', Y-Y' and Z'-Z' sites versus simulation time under pH 5.

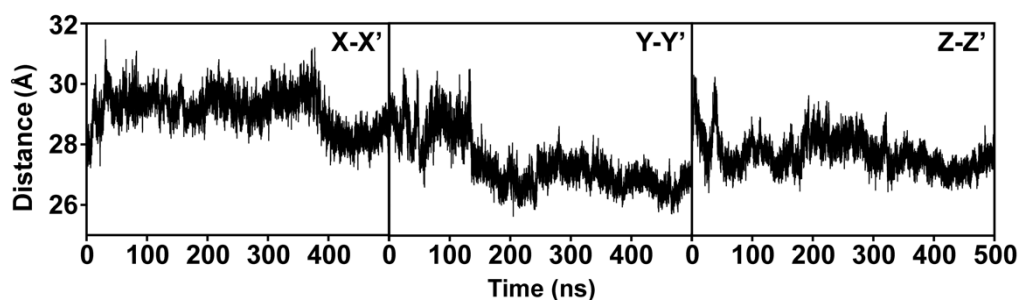


Figure S3. Distance plot between the centers of mass of domain II on each monomer along the simulation time.

Table S1. Comparison of the average binding free energy (ΔG_{bind}) and its components (kcal/mol) of each EGCG binding to ZIKV E protein at acidic pH condition based on SIE method. The calculation occurs with several terms; van de Waals interaction (ΔE_{vdW}), electrostatic interaction (ΔE_{elec}), the change of reaction energy (ΔG^R), the change of molecular surface upon binding ($\gamma \cdot \Delta MSA$), conformational upon binding (α), constant (C) value and molecular surface area coefficient (γ) [96, 97, 121].

	X	X'	Y	Y'	Z	Z'
ΔE_{vdW}	-35.2±0.5	-28.1±0.8	-45.6±0.6	-48.4±1.2	-33.1±0.7	-40.2±0.5
ΔE_{elec}	-15.2±0.7	-27.3±0.9	-28.3±0.9	-15.5±0.6	-26.6±0.7	-15.9±0.5
ΔG^R	21.6±0.6	20.3±0.6	28.9±0.6	18.1±0.4	21.0±0.5	19.8±0.3
$\gamma \cdot \Delta MSA$	-6.9±0.9	-5.0±0.1	-8.2±0.1	-7.3±0.2	-5.4±0.1	-7.4±0.1
C				-2.89		
α				0.104758		
γ				0.012894		
ΔG_{bind}	-6.63±0.1	-7.1±0.1	-8.5±0.1	-8.5±0.2	-7.5±0.1	-7.5±0.1

Table S2. Binding free energy and its components (kcal/mol) of EGCG ligand binding to each binding region based on MM-GBSA method. Note that entropy contribution was neglected due to inadequate computing resource and therefore the relative binding free energy values are discussed.

	X	X'	Y	Y'	Z	Z'
ΔE_{vdW}	-28.3±0.8	-35.5±0.4	-49.0±0.9	-45.0±0.9	-40.2±0.5	-33.1±0.7
ΔE_{elec}	-62.3±1.9	-34.0±1.5	-35.9±1.0	-64.7±1.8	-35.7±1.1	-60.0±1.5
ΔE_{MM}	-90.7±2.3	-69.5±1.4	-84.9±1.8	-109.7±2.2	-76.0±1.1	-93.0±2.0
$\Delta G_{polar\ sol.}$	-2.8±0.0	-3.7±0.0	-3.9±0.0	-4.3±0.1	-3.8±1.1	-3.3±0.1
$\Delta G_{non-polar\ sol.}$	70.8±1.9	59.2±1.4	60.9±1.2	86.8±1.9	59.6±0.8	78.2±1.6
$\Delta G_{solvation}$	68.0±1.8	55.6±1.4	57.1±1.1	82.5±1.8	55.8±0.8	74.9±1.5
$\Delta G_{MM-GBSA}$	-22.6±0.8	-13.9±0.3	-27.8±0.8	-27.2±0.7	-20.1±0.5	-18.1±0.6

Table S3. QM/MM-GBSA binding free energy and its components (kcal/mol) of EGCG ligand binding to each binding region while only EGCE was treated with PM3 or SCC-DFTB method. Note that entropy contribution was neglected due to inadequate computing resource and therefore the relative binding free energy values are discussed.

	X	X'	Y	Y'	Z	Z'
ΔE_{ndW}	-28.3±0.8	-35.5±0.4	-49.0±0.9	-45.0±0.9	-40.2±0.5	-33.1±0.7
ΔE_{GAS}	-28.3±0.8	-35.5±0.4	-49.0±0.9	-45.0±0.9	-40.2±0.5	-33.1±0.7
QM method = PM3						
$\Delta G_{polar\ sol.}$	-4.1±0.1	-5.3±0.0	-5.6±0.1	-6.2±0.1	-5.5±0.0	-4.7±0.1
$\Delta G_{non-polar\ sol.}$	54.7±1.5	39.7±1.1	45.3±1.0	61.4±1.3	47.6±0.6	63.1±1.3
$\Delta G_{Solvation}$	50.6±1.4	34.5±1.2	39.8±0.9	55.2±1.2	42.0±0.6	58.4±1.2
$\Delta G_{PM3/MM-GBSA}$	-33.8±1.0	-29.0±0.4	-42.2±1.0	-42.3±0.9	-31.9±0.5	-27.8±0.8
QM method = SCC-DFTB						
$\Delta G_{polar\ sol.}$	-4.1±0.1	-5.3±0.0	-5.6±0.1	-6.2±0.1	-5.5±0.0	-4.7±0.1
$\Delta G_{non-polar\ sol.}$	71.2±1.9	51.5±1.3	56.3±1.1	76.2±1.7	54.5±0.7	73.5±1.6
$\Delta G_{Solvation}$	67.1±1.8	46.2±1.2	50.7±1.1	70.1±0.7	48.9±0.7	68.8±1.5
$\Delta G_{SCC-DFTB/MMGBSA}$	-29.5±0.9	-21.7±0.3	-35.1±0.9	-34.3±0.7	-27.6±0.5	-24.1±0.7

Table S4 *p* values of Pearson correlation analysis

SCC-DFTB /MM-GBSA	0.125	0.000	0.002	-
PM3/MM- GBSA	0.076	0.006	-	0.002
MM-GBSA	0.129	-	0.006	0.000
SIE	-	0.129	0.076	0.125
	SIE	MM-GBSA	QM/MM- GBSA (PM3)	QM/MM- GBSA (SCC- DFTB)

Table S5. Comparison of the average binding free energy (ΔG_{bind}) and its component (kcal/mol) of each flavone derivative bound to DENV E protein based on SIE method. The calculation occurs with several terms; van de Waals interaction (ΔE_{vdW}), electrostatic interaction (ΔE_{elec}), the change of reaction energy (ΔG^R), the charge of molecular surface upon binding ($\gamma \cdot \Delta MSA$), conformational upon binding (α), constant (C) value and molecular surface area coefficient (γ).

	ΔE_{vdW}	ΔE_{elec}	ΔG^R	$\gamma \cdot \Delta MSA$	C	α	γ	ΔG_{bind} (kcal/mol)
FNSY	K	-27.33±0.1	6.14±0.1	6.3±0.4	1.48±0.2			-4.30±0.0
	K'	-23.36±0.2	5.95±0.2	6.07±0.4	1.14±0.2	-2.89	0.104758	-3.96±0.0
	X'	-5.09±0.4	-7.19±0.1	5.88±0.4	1.20±0.2		0.012894	-3.43±0.1
	Y'	-13.06±1.0	-8.75±0.3	5.93±0.4	1.98±0.2			-4.35±0.1
F16	K	-46.28±0.3	-13.71±0.7	-27.46±1.1	-9.61±0.1			-13.06±0.9
	K'	-33.72±0.3	-21.37±0.4	-27.53±1.1	-9.52±0.1			-12.54±0.1
	X'	-33.98±0.4	-1.31±0.7	-27.52±1.1	-9.64±0.1	-2.89	0.104758	-10.48±0.1
	Y'	-41.98±0.3	-12.20±0.5	-20.56±0.6	-7.91±0.1			-11.55±0.1
F17	K	-48.82±0.2	-12.95±0.4	9.80±0.1	-9.89±0.1			-9.37±0.1
	K'	-53.24±0.2	-22.25±0.3	9.72±0.6	-9.86±0.1			-6.28±0.1
	X'	-32.53±0.2	-37.01±0.3	9.59±0.7	-10.35±0.1	-2.89	0.104758	-7.34±0.1
	Y'	-37.8±0.5	-16.16±0.5	9.72±0.6	-9.83±0.1			-8.56±0.1
F18	K	-51.91±0.2	-7.73±0.4	-39.92±0.7	-10.53±0.1			-14.42±0.1
	K'	-47.90±0.2	-15.00±0.4	-35.74±0.7	-10.73±0.1			-14.35±0.1
	X'	-28.72±0.3	-3.41±0.7	-35.74±0.7	-10.62±0.1	-2.89	0.104758	-11.11±0.1
	Y'	-46.94±0.2	-7.53±0.4	-35.78±0.7	-10.73±0.1			-13.47±0.1

PUBLICATION

K. Hengphasatporn, A. Garon, P. Wolschann, T. Langer, W. Chavasiri, T. Saelee, S. Yasuteru, S. Boonyasuppayakorn, T. Rungrotmongkol*, Multiple Virtual Screening Strategies for the Optimization of the Novel Compound Against Dengue Virus: A Drug Discovery Study. *JCIM*. **2019** (*submitted*)

K. Hengphasatporn, N. Kungwan, T. Rungrotmongkol*, Binding Pattern and Susceptibility of Epigallocatechin Gallate Against Envelope Protein Homodimer of Zika Virus: A Molecular Dynamics Study. *J Mol Liq*. **2019**; 274: 140-147.

P. Kanyaboon, T. Saelee, A. Suroengrit, **K. Hengphasatporn**, T. Rungrotmongkol, W. Chavasiri, and S Boonyasuppayakorn*, Phenolic lipids as potential dengue virus inhibitors. *Sci Rep*. **2019**;

P. Srivarangkul, W. Yuttithamnon, A. Suroengrit, S. Pankaew, **K. Hengphasatporn**, T. Rungrotmongkol, P. Phuwapraisirisane, K. Ruxrungham, S. Boonyasuppayakorn*, A novel flavanone derivative inhibits dengue virus fusion and infectivity. *Antiviral Res*. **2018**; 151: 27-38.

CONFERENCE

The Computational Sciences Workshop 2019 (CSW2019) on January 16-18, **2019**

The 22nd International Annual Symposium on Computational Science and Engineering (ANSCSE22) on August 2-4, **2018**

The 1st Taiwan-Thailand-Vietnam Workshop on Theoretical and Computational Chemistry (TTV), National Taiwan University, Taipei, Taiwan on March 22-23, **2018**

Consortium of Biological Science 2017 (CONBIO), Kobe, Japan on December 6-9, **2017**

Thai-Japan Symposium in Chemistry and The 13th Thai Summer School of Computational Chemistry Workshop, **2016** (2nd poster presentation awarded)

REFERENCES



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1. Lindenbach, B.D. and C.M. Rice, *Molecular biology of flaviviruses*, in *Advances in Virus Research*. 2003, Academic Press. p. 23-61.
2. Gubler, D.J., G. Kuno, and L. Markoff, *Flaviviruses: Past, Present And Future*, in *Fields Virology*, D. Knipe and P. Howley, Editors. 2007, Walters Kluwer/Lippincott Williams and Wilkins: Philadelphia, PA. p. 1153–1252.
3. Guzman, M.G., et al., *Dengue infection*. Nature Reviews Disease Primers, 2016. **2**: p. 16055.
4. Srivarangkul, P., et al., *A novel flavanone derivative inhibits dengue virus fusion and infectivity*. Antiviral Research, 2018. **151**: p. 27-38.
5. de Wispelaere, M., et al., *Inhibition of Flaviviruses by Targeting a Conserved Pocket on the Viral Envelope Protein*. Cell Chemical Biology, 2018. **25**(8): p. 1006-1016.e8.
6. Hengphasatporn, K., N. Kungwan, and T. Rungrotmongkol, *Binding pattern and susceptibility of epigallocatechin gallate against envelope protein homodimer of Zika virus: A molecular dynamics study*. Journal of Molecular Liquids, 2019. **274**: p. 140-147.
7. Tambunan, U.S.F., et al., *In silico identification of 2-oxo-1,3-thiazolidine derivatives as novel inhibitor candidate of class II histone deacetylase (HDAC) in cervical cancer treatment*. Arabian Journal of Chemistry.
8. Bajorath, J., *Integration of virtual and high-throughput screening*. Nature Reviews Drug Discovery, 2002. **1**: p. 882.
9. Song, C.M., S.J. Lim, and J.C. Tong, *Recent advances in computer-aided drug design*. Briefings in Bioinformatics, 2009. **10**(5): p. 579-591.
10. Langer, T., *Pharmacophores in Drug Research*. Molecular Informatics, 2010. **29**(6-7): p. 470-475.
11. Hughes, J.P., et al., *Principles of early drug discovery*. British Journal of Pharmacology, 2011. **162**(6): p. 1239-1249.
12. Faria, N.R., et al., *Zika virus in the Americas: Early epidemiological and genetic findings*. Science (New York, N.Y.), 2016. **352**(6283): p. 345-349.
13. Agumadu, V.C. and K. Ramphul, *Zika Virus: A Review of Literature*. Cureus, 2018. **10**(7): p. e3025-e3025.
14. Barrera, R., et al., *A comparison of mosquito densities, weather and infection rates of Aedes aegypti during the first epidemics of Chikungunya (2014) and Zika (2016) in areas with and without vector control in Puerto Rico*. Medical and Veterinary Entomology, 2019. **33**(1): p. 68-77.
15. Gao, H., et al., *Role of heparan sulfate in the Zika virus entry, replication, and cell death*. Virology, 2019. **529**: p. 91-100.
16. Cobo, F., *17 - Viral haemorrhagic fevers*, in *Imported Infectious Diseases*, F. Cobo, Editor. 2014, Woodhead Publishing. p. 243-255.
17. Rothman, A.L., *Immunology and Immunopathogenesis of Dengue Disease*, in *Advances in Virus Research*. 2003, Academic Press. p. 397-419.
18. Halstead, S.B. and E.J. O'Rourke, *Dengue viruses and mononuclear phagocytes. I. Infection enhancement by non-neutralizing antibody*. The Journal of Experimental Medicine, 1977. **146**(1): p. 201-217.
19. Guzman, M.G., M. Alvarez, and S.B. Halstead, *Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: an historical*

- perspective and role of antibody-dependent enhancement of infection.* Archives of Virology, 2013. **158**(7): p. 1445-1459.
20. Flipse, J., et al., *Antibody-Dependent Enhancement of Dengue Virus Infection in Primary Human Macrophages; Balancing Higher Fusion against Antiviral Responses.* Scientific Reports, 2016. **6**: p. 29201.
 21. St. John, A.L. and A.P.S. Rathore, *Adaptive immune responses to primary and secondary dengue virus infections.* Nature Reviews Immunology, 2019.
 22. Barba-Spaeth, G., et al., *Structural basis of potent Zika–dengue virus antibody cross-neutralization.* Nature, 2016. **536**(7614): p. 48-53.
 23. Weaver, S.C., et al., *Zika virus: History, emergence, biology, and prospects for control.* Antiviral Research, 2016. **130**: p. 69-80.
 24. Hoang, L.T., et al., *The Early Whole-Blood Transcriptional Signature of Dengue Virus and Features Associated with Progression to Dengue Shock Syndrome in Vietnamese Children and Young Adults.* Journal of Virology, 2010. **84**(24): p. 12982-12994.
 25. Stiasny, K. and F.X. Heinz, *Flavivirus membrane fusion.* Journal of General Virology 2006. **87**(10): p. 2755-2766.
 26. Yamshchikov, V.F. and R.W. Compans, *Formation of the flavivirus envelope: role of the viral NS2B-NS3 protease.* Journal of Virology, 1995. **69**(4): p. 1995-2003.
 27. Oliveira, E.R.A., et al., *The flavivirus capsid protein: Structure, function and perspectives towards drug design.* Virus Research, 2017. **227**: p. 115-123.
 28. Yacoub, S. and J. Farrar, *15 - Dengue*, in *Manson's Tropical Infectious Diseases (Twenty-third Edition)*, J. Farrar, et al., Editors. 2014, W.B. Saunders: London. p. 162-170.e2.
 29. Kuhn, R.J., et al., *Structure of Dengue Virus: Implications for Flavivirus Organization, Maturation, and Fusion.* Cell, 2002. **108**(5): p. 717-725.
 30. Cao-Lormeau, V.-M., et al., *Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study.* The Lancet. **387**(10027): p. 1531-1539.
 31. Byk, L.A., et al., *Dengue Virus Genome Uncoating Requires Ubiquitination.* mBio, 2016. **7**(3): p. e00804-16.
 32. Wong, K.L., et al., *Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets.* Blood, 2011. **118**(5): p. e16.
 33. Chakraborty, S., *Computational analysis of perturbations in the post-fusion Dengue virus envelope protein highlights known epitopes and conserved residues in the Zika virus.* F1000Research, 2016. **5**: p. 1150-1150.
 34. Heinz, F.X. and S.L. Allison, *Structures and mechanisms in flavivirus fusion.* Advances in Virus Research, 2000. **55**: p. 231-269.
 35. Sun, J., et al., *Study of the mechanism of protonated histidine-induced conformational changes in the Zika virus dimeric envelope protein using accelerated molecular dynamic simulations.* Journal of Molecular Graphics and Modelling, 2017. **74**: p. 203-214.
 36. Kanyaboon, P., et al., *Cardol triene inhibits dengue infectivity by targeting kl loops and preventing envelope fusion.* Scientific Reports, 2018. **8**(1): p. 16643.

37. Slon Campos, J.L., et al., *Temperature-dependent folding allows stable dimerization of secretory and virus-associated E proteins of Dengue and Zika viruses in mammalian cells*. Scientific Reports, 2017. **7**: p. 966.
38. Rodriguez-Roche, R. and E.A. Gould, *Understanding the Dengue Viruses and Progress towards Their Control*. BioMed Research International, 2013. **2013**: p. 690835.
39. Alcon, S., et al., *Enzyme-Linked Immunosorbent Assay Specific to Dengue Virus Type 1 Nonstructural Protein NS1 Reveals Circulation of the Antigen in the Blood during the Acute Phase of Disease in Patients Experiencing Primary or Secondary Infections*. Journal of Clinical Microbiology, 2002. **40**(2): p. 376-381.
40. Zhao, B.T., et al., *Expression of dengue virus structural proteins and nonstructural protein NS1 by a recombinant vaccinia virus*. Journal of Virology, 1987. **61**(12): p. 4019-4022.
41. Aguirre, S., et al., *Dengue virus NS2B protein targets cGAS for degradation and prevents mitochondrial DNA sensing during infection*. Nature Microbiology, 2017. **2**: p. 17037.
42. Xie, X., et al., *Inhibition of Dengue Virus by Targeting Viral NS4B Protein*. Journal of Virology, 2011. **85**(21): p. 11183-11195.
43. Miller, S., et al., *The Non-structural Protein 4A of Dengue Virus Is an Integral Membrane Protein Inducing Membrane Alterations in a 2K-regulated Manner*. Journal of Biological Chemistry, 2007. **282**(12): p. 8873-8882.
44. Yang, C.-C., et al., *Novel Dengue Virus-Specific NS2B/NS3 Protease Inhibitor, BP2109, Discovered by a High-Throughput Screening Assay*. Antimicrobial Agents and Chemotherapy, 2011. **55**(1): p. 229-238.
45. Boonyasuppayakorn, S., et al., *Amodiaquine, an antimalarial drug, inhibits dengue virus type 2 replication and infectivity*. Antiviral research, 2014. **106**: p. 125-134.
46. Dong, H., et al., *Biochemical and genetic characterization of dengue virus methyltransferase*. Virology, 2010. **405**(2): p. 568-578.
47. Lim, S.P., et al., *Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling*. PLOS Pathogens, 2016. **12**(8): p. e1005737.
48. Rodenhuis-Zybert, I.A., et al., *Immature Dengue Virus: A Veiled Pathogen?* PLOS Pathogens, 2010. **6**(1): p. e1000718.
49. Neufeldt, C.J., et al., *Rewiring cellular networks by members of the Flaviviridae family*. Nature Reviews Microbiology, 2018. **16**: p. 125.
50. Behnam, M.A.M., et al., *The Medicinal Chemistry of Dengue Virus*. Journal of Medicinal Chemistry, 2016. **59**(12): p. 5622-5649.
51. Rey, F.A., et al., *The bright and the dark side of human antibody responses to flaviviruses: lessons for vaccine design*. EMBO reports, 2017: p. e45302.
52. Ooi, E.E. and D.J. Gubler, *CHAPTER 75 - Dengue and Dengue Hemorrhagic Fever*, in *Tropical Infectious Diseases: Principles, Pathogens and Practice (Third Edition)*, R.L. Guerrant, D.H. Walker, and P.F. Weller, Editors. 2011, W.B. Saunders: Edinburgh. p. 504-510.

53. Rocha-Roa, C., D. Molina, and N. Cardona, *A Perspective on Thiazolidinone Scaffold Development as a New Therapeutic Strategy for Toxoplasmosis*. *Frontiers in cellular and infection microbiology*, 2018. **8**: p. 360-360.
54. Carneiro, B.M., et al., *The green tea molecule EGCG inhibits Zika virus entry*. *Virology*, 2016. **496**: p. 215-218.
55. Sharma, N., et al., *Epigallocatechin gallate, an active green tea compound inhibits the Zika virus entry into host cells via binding the envelope protein*. *International Journal of Biological Macromolecules*, 2017. **104**: p. 1046-1054.
56. Constant, D.A., et al., *Targeting intramolecular proteinase NS2B/3 cleavages for trans-dominant inhibition of dengue virus*. *Proceedings of the National Academy of Sciences*, 2018. **115**(40): p. 10136.
57. Lin, K.-H., et al., *Dengue Virus NS2B/NS3 Protease Inhibitors Exploiting the Prime Side*. *Journal of Virology*, 2017. **91**(10): p. e00045-17.
58. Cabarcas-Montalvo, M., et al., *Discovery of antiviral molecules for dengue: In silico search and biological evaluation*. *European Journal of Medicinal Chemistry*, 2016. **110**: p. 87-97.
59. Pambudi, S., et al., *A small compound targeting the interaction between nonstructural proteins 2B and 3 inhibits dengue virus replication*. *Biochemical and Biophysical Research Communications*, 2013. **440**(3): p. 393-398.
60. Lu, D., et al., *Discovery and optimization of phthalazinone derivatives as a new class of potent dengue virus inhibitors*. *European Journal of Medicinal Chemistry*, 2018. **145**: p. 328-337.
61. Liu, A.L., et al., *Anti-influenza virus activities of flavonoids from the medicinal plant *Elsholtzia rugulosa**. *Planta Med*, 2008. **8**.
62. Tomlinson, S.M. and S.J. Watowich, *Use of parallel validation high-throughput screens to reduce false positives and identify novel dengue NS2B-NS3 protease inhibitors*. *Antiviral Research*, 2012. **93**(2): p. 245-252.
63. Weigel, L.F., et al., *Phenylalanine and Phenylglycine Analogues as Arginine Mimetics in Dengue Protease Inhibitors*. *Journal of Medicinal Chemistry*, 2015. **58**(19): p. 7719-7733.
64. Xu, G., et al., *Inhibitory effects of baicalein on the influenza virus in vivo is determined by baicalin in the serum*. *Biol Pharm Bull*, 2010. **33**.
65. Brecher, M., et al., *A conformational switch high-throughput screening assay and allosteric inhibition of the flavivirus NS2B-NS3 protease*. *PLOS Pathogens*, 2017. **13**(5): p. e1006411.
66. Stahla-Beek, H.J., et al., *Identification of a Novel Antiviral Inhibitor of the Flavivirus Guanylyltransferase Enzyme*. *Journal of Virology*, 2012. **86**(16): p. 8730.
67. Pelliccia, S., et al., *Inhibition of dengue virus replication by novel inhibitors of RNA-dependent RNA polymerase and protease activities*. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 2017. **32**(1): p. 1091-1101.
68. Smith, T.M., et al., *Identifying Initiation and Elongation Inhibitors of Dengue Virus RNA Polymerase in a High-Throughput Lead-Finding Campaign*. *Journal of Biomolecular Screening*, 2014. **20**(1): p. 153-163.

69. Tarantino, D., et al., *Targeting flavivirus RNA dependent RNA polymerase through a pyridobenzothiazole inhibitor*. Antiviral Research, 2016. **134**: p. 226-235.
70. Cannalire, R., et al., *Functionalized 2,1-benzothiazine 2,2-dioxides as new inhibitors of Dengue NS5 RNA-dependent RNA polymerase*. European Journal of Medicinal Chemistry, 2018. **143**: p. 1667-1676.
71. Benmansour, F., et al., *Novel 2-phenyl-5-[(E)-2-(thiophen-2-yl)ethenyl]-1,3,4-oxadiazole and 3-phenyl-5-[(E)-2-(thiophen-2-yl)ethenyl]-1,2,4-oxadiazole derivatives as dengue virus inhibitors targeting NS5 polymerase*. European Journal of Medicinal Chemistry, 2016. **109**: p. 146-156.
72. Boldescu, V., et al., *Broad-spectrum agents for flaviviral infections: dengue, Zika and beyond*. Nature Reviews Drug Discovery, 2017. **16**: p. 565.
73. Ma, L., et al., *Solution structure of dengue virus capsid protein reveals another fold*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(10): p. 3414-3419.
74. Byrd, C.M., et al., *A Novel Inhibitor of Dengue Virus Replication That Targets the Capsid Protein*. Antimicrobial Agents and Chemotherapy, 2013. **57**(1): p. 15-25.
75. Faustino, A.F., et al., *Dengue virus capsid protein interacts specifically with very low-density lipoproteins*. Nanomedicine: Nanotechnology, Biology and Medicine, 2014. **10**(1): p. 247-255.
76. Saudi, M., et al., *Synthesis and evaluation of imidazole-4,5- and pyrazine-2,3-dicarboxamides targeting dengue and yellow fever virus*. European Journal of Medicinal Chemistry, 2014. **87**: p. 529-539.
77. Martins, Ivo C., et al., *The disordered N-terminal region of dengue virus capsid protein contains a lipid-droplet-binding motif*. Biochemical Journal, 2012. **444**(3): p. 405.
78. Modis, Y., et al., *A ligand-binding pocket in the dengue virus envelope glycoprotein*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(12): p. 6986-6991.
79. Cruz-Oliveira, C., et al., *Receptors and routes of dengue virus entry into the host cells*. FEMS Microbiology Reviews, 2015. **39**(2): p. 155-170.
80. Zhou, Z., et al., *Antiviral Compounds Discovered by Virtual Screening of Small-Molecule Libraries against Dengue Virus E Protein*. ACS chemical biology, 2008. **3**(12): p. 765-775.
81. Schmidt, A.G., P.L. Yang, and S.C. Harrison, *Peptide Inhibitors of Dengue-Virus Entry Target a Late-Stage Fusion Intermediate*. PLoS Pathogens, 2010. **6**(4): p. e1000851.
82. Modis, Y., et al., *Structure of the dengue virus envelope protein after membrane fusion*. Nature, 2004. **427**(6972): p. 313-319.
83. Wang, Q.-Y., et al., *A Small-Molecule Dengue Virus Entry Inhibitor*. Antimicrobial Agents and Chemotherapy, 2009. **53**(5): p. 1823-1831.
84. Tambunan, U.S.F., et al., *Screening Analogs of β -OG Pocket Binder as Fusion Inhibitor of Dengue Virus 2*. Drug Target Insights, 2015. **9**: p. 33-49.
85. Wang, J., et al., *Development and testing of a general amber force field*. Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.

86. Kaptein, S.J.F., et al., *A Derivate of the Antibiotic Doxorubicin Is a Selective Inhibitor of Dengue and Yellow Fever Virus Replication In Vitro*. *Antimicrobial Agents and Chemotherapy*, 2010. **54**(12): p. 5269-5280.
87. Jadav, S.S., et al., *Design, synthesis, optimization and antiviral activity of a class of hybrid dengue virus E protein inhibitors*. *Bioorganic & Medicinal Chemistry Letters*, 2015. **25**(8): p. 1747-1752.
88. Zhang, X., et al., *Dengue structure differs at the temperatures of its human and mosquito hosts*. *Proceedings of the National Academy of Sciences of the United States of America*, 2013. **110**(17): p. 6795-6799.
89. Ichiyama, K., et al., *Sulfated Polysaccharide, Curdlan Sulfate, Efficiently Prevents Entry/Fusion and Restricts Antibody-Dependent Enhancement of Dengue Virus Infection In Vitro: A Possible Candidate for Clinical Application*. *PLoS Neglected Tropical Diseases*, 2013. **7**(4): p. e2188.
90. Panya, A., et al., *Peptide inhibitors against dengue virus infection*. *Chem Biol Drug Des*, 2014. **84**: p. 148-157.
91. Crill, W.D. and J.T. Roehrig, *Monoclonal Antibodies That Bind to Domain III of Dengue Virus E Glycoprotein Are the Most Efficient Blockers of Virus Adsorption to Vero Cells*. *Journal of Virology*, 2001. **75**(16): p. 7769-7773.
92. Perry, S.T., et al., *An iminosugar with potent inhibition of dengue virus infection in vivo*. *Antiviral Research*, 2013. **98**(1): p. 35-43.
93. Ismail, N.A. and S.A. Jusoh, *Molecular Docking and Molecular Dynamics Simulation Studies to Predict Flavonoid Binding on the Surface of DENV2 E Protein*. *Interdisciplinary Sciences: Computational Life Sciences*, 2016: p. 1-13.
94. Seyedi, S.S., et al., *Computational Approach Towards Exploring Potential Anti-Chikungunya Activity of Selected Flavonoids*. *Scientific Reports*, 2016. **6**: p. 24027.
95. Akimoto, N., et al., *FlavonoidSearch: A system for comprehensive flavonoid annotation by mass spectrometry*. *Scientific Reports*, 2017. **7**(1): p. 1243.
96. Naim, M., et al., *Solvated Interaction Energy (SIE) for Scoring Protein-Ligand Binding Affinities. 1. Exploring the Parameter Space*. *Journal of Chemical Information and Modeling*, 2007. **47**(1): p. 122-133.
97. Sulea, T. and E.O. Purisima, *The Solvated Interaction Energy Method for Scoring Binding Affinities*, in *Computational Drug Discovery and Design*, R. Baron, Editor. 2012, Springer New York: New York, NY. p. 295-303.
98. Kurauchi, R., et al., *Novel type of virtual ligand screening on the basis of quantum-chemical calculations for protein-ligand complexes and extended clustering techniques*. *Computational and Theoretical Chemistry*, 2015. **1061**: p. 12-22.
99. Kitaura, K., et al., *Fragment molecular orbital method: analytical energy gradients*. *Chemical Physics Letters*, 2001. **336**(1): p. 163-170.
100. Lee, M. and D. Kim, *Large-scale reverse docking profiles and their applications*. *BMC Bioinformatics*, 2012. **13**(Suppl 17): p. S6-S6.
101. Kumar, A. and K.Y.J. Zhang, *Hierarchical virtual screening approaches in small molecule drug discovery*. *Methods*, 2015. **71**: p. 26-37.

102. Rognan, D., *Docking Methods for Virtual Screening: Principles and Recent Advances*, in *Virtual Screening*. 2011, Wiley-VCH Verlag GmbH & Co. KGaA. p. 153-176.
103. Taft, C.A., V.B. da Silva, and C.H.T.d.P. da Silva, *Current topics in computer-aided drug design*. *Journal of Pharmaceutical Sciences*, 2008. **97**(3): p. 1089-1098.
104. Settivari, R.S., et al., *Predicting the Future: Opportunities and Challenges for the Chemical Industry to Apply 21st-Century Toxicity Testing*. *Journal of the American Association for Laboratory Animal Science : JAALAS*, 2015. **54**(2): p. 214-223.
105. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading*. *Journal of computational chemistry*, 2010. **31**(2): p. 455-461.
106. Wu, G., et al., *Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMM-based MD docking algorithm*. *Journal of Computational Chemistry*, 2003. **24**(13): p. 1549-1562.
107. Yang, J.-M. and C.-C. Chen, *GEMDOCK: A generic evolutionary method for molecular docking*. *Proteins: Structure, Function, and Bioinformatics*, 2004. **55**(2): p. 288-304.
108. Hsu, K.-C., et al., *iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis*. *BMC Bioinformatics*, 2011. **12**(Suppl 1): p. S33-S33.
109. Azar, A.T., S.A. El-Said, and A.E. Hassanien, *Fuzzy and hard clustering analysis for thyroid disease*. *Computer Methods and Programs in Biomedicine*, 2013. **111**(1): p. 1-16.
110. Syakur, M.A., et al., *Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster*. *IOP Conference Series: Materials Science and Engineering*, 2018. **336**: p. 012017.
111. Hartigan, J.A. and M.A. Wong, *Algorithm AS 136: A K-Means Clustering Algorithm*. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979. **28**(1): p. 100-108.
112. Waidyasooriya, H., M. Hariyama, and K. Kasahara, *An FPGA accelerator for molecular dynamics simulation using OpenCL*. Vol. 5. 2017. 52.
113. Mazur, A.K., *Common Molecular Dynamics Algorithms Revisited: Accuracy and Optimal Time Steps of Störmer–Leapfrog Integrators*. *Journal of Computational Physics*, 1997. **136**(2): p. 354-365.
114. Spreiter, Q. and M. Walter, *Classical Molecular Dynamics Simulation with the Velocity Verlet Algorithm at Strong External Magnetic Fields*. *Journal of Computational Physics*, 1999. **152**(1): p. 102-119.
115. Chen, J., *The Development and Comparison of Molecular Dynamics Simulation and Monte Carlo Simulation*. *IOP Conference Series: Earth and Environmental Science*, 2018. **128**: p. 012110.
116. Mahoney, M. and W. L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid*. Vol. 112. 2000. 8910-8922.
117. Carugo, O., *Statistical validation of the root-mean-square-distance, a measure of protein structural proximity*. *Protein Engineering, Design and Selection*, 2007. **20**(1): p. 33-37.

118. Motiejunas, D. and R.C. Wade, 4.09 - *Structural, Energetic, and Dynamic Aspects of Ligand–Receptor Interactions*, in *Comprehensive Medicinal Chemistry II*, J.B. Taylor and D.J. Triggle, Editors. 2007, Elsevier: Oxford. p. 193-213.
119. Aldred, E.M., C. Buck, and K. Vall, *Chapter 3 - Bonds found in biological chemistry*, in *Pharmacology*, E.M. Aldred, C. Buck, and K. Vall, Editors. 2009, Churchill Livingstone: Edinburgh. p. 11-19.
120. Tsui, V. and D.A. Case, *Theory and applications of the generalized born solvation model in macromolecular simulations*. *Biopolymers*, 2000. **56**(4): p. 275-291.
121. Virtanen, S.I., S.P. Niinivehmas, and O.T. Pentikäinen, *Case-specific performance of MM-PBSA, MM-GBSA, and SIE in virtual screening*. *Journal of Molecular Graphics and Modelling*, 2015. **62**: p. 303-318.
122. Leach, A.R., 4.05 - *Ligand-Based Approaches: Core Molecular Modeling*, in *Comprehensive Medicinal Chemistry II*, J.B. Taylor and D.J. Triggle, Editors. 2007, Elsevier: Oxford. p. 87-118.
123. Wichapong, K., et al., *Application of Docking and QM/MM-GBSA Rescoring to Screen for Novel Myt1 Kinase Inhibitors*. *Journal of Chemical Information and Modeling*, 2014. **54**(3): p. 881-893.
124. Chen, J., et al., *A Comparative Study of Trypsin Specificity based on QM/MM Molecular Dynamics Simulation and QM/MM GB SA Calculation*. 2015. 1-38.
125. Cui, Q., et al., *A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method*. *The Journal of Physical Chemistry B*, 2001. **105**(2): p. 569-585.
126. Heifetz, A., et al., *Using the fragment molecular orbital method to investigate agonist-orexin-2 receptor interactions*. *Biochemical Society transactions*, 2016. **44**(2): p. 574-581.
127. Maruyama, K., et al., *Application of singular value decomposition to the inter-fragment interaction energy analysis for ligand screening*. *Computational and Theoretical Chemistry*, 2018. **1132**: p. 23-34.
128. Osman, G., C. Omoshile, and K. Yasuhisa, *Pharmacophore Modeling and Three Dimensional Database Searching for Drug Design Using Catalyst: Recent Advances*. *Current Medicinal Chemistry*, 2004. **11**(22): p. 2991-3005.
129. Wolber, G. and T. Langer, *LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters*. *Journal of Chemical Information and Modeling*, 2005. **45**(1): p. 160-169.
130. Yang, S.-Y., *Pharmacophore modeling and applications in drug discovery: challenges and recent advances*. *Drug Discovery Today*, 2010. **15**(11): p. 444-450.
131. Rodolpho, C.B. and H.A. Carolina, *Assessing the Performance of 3D Pharmacophore Models in Virtual Screening: How Good are They?* *Current Topics in Medicinal Chemistry*, 2013. **13**(9): p. 1127-1138.
132. Shin, W.-J. and B.L. Seong, *Recent advances in pharmacophore modeling and its application to anti-influenza drug discovery*. *Expert Opinion on Drug Discovery*, 2013. **8**(4): p. 411-426.

133. Sanders, M.P.A., et al., *From the protein's perspective: the benefits and challenges of protein structure-based pharmacophore modeling*. MedChemComm, 2012. **3**(1): p. 28-38.
134. Langer, T. and G. Wolber, *Pharmacophore definition and 3D searches*. Drug Discovery Today: Technologies, 2004. **1**(3): p. 203-207.
135. Horvath, D., *Pharmacophore-Based Virtual Screening*, in *Chemoinformatics and Computational Chemical Biology*, J. Bajorath, Editor. 2011, Humana Press: Totowa, NJ. p. 261-298.
136. Daniela, S., et al., *Development and Validation of an In Silico P450 Profiler Based on Pharmacophore Models*. Current Drug Discovery Technologies, 2006. **3**(1): p. 1-48.
137. Rakers, C., et al., *In Silico Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations*. Journal of Biological Chemistry, 2016. **291**(1): p. 58-71.
138. Murgueitio, M.S., et al., *In silico virtual screening approaches for anti-viral drug discovery*. Drug Discovery Today: Technologies, 2012. **9**(3): p. e219-e225.
139. Sakkiah, S., et al., *Pharmacophore based virtual screening, molecular docking studies to design potent heat shock protein 90 inhibitors*. European Journal of Medicinal Chemistry, 2011. **46**(7): p. 2937-2947.
140. Sanders, M.P.A., et al., *Comparative Analysis of Pharmacophore Screening Tools*. Journal of Chemical Information and Modeling, 2012. **52**(6): p. 1607-1620.
141. Shinozuka, K., et al., *Inhibitory effect of flavonoids on DNA-dependent DNA and RNA polymerases*. Experientia, 1988. **44**.
142. Ono, K., et al., *Differential inhibitory effects of various flavonoids on the activities of reverse transcriptase and cellular DNA and RNA polymerases*. Eur J Biochem, 1990. **190**.
143. Sánchez, I., et al., *Antiviral effect of flavonoids on the dengue virus*. Phytother Res, 2000. **14**.
144. Chiang, L.C., et al., *In vitro antiviral activities of Caesalpinia pulcherrima and its related flavonoids*. J Antimicrob Chemother, 2003. **52**.
145. Zandi, K., et al., *Antiviral activity of four types of bioflavonoid against dengue virus type-2*. Virol J, 2011. **8**.
146. Zandi, K., et al., *Novel antiviral activity of baicalein against dengue virus*. BMC Complementary and Alternative Medicine, 2012. **12**(1): p. 214.
147. Mackenzie, J.S., D.J. Gubler, and L.R. Petersen, *Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses*. Nat Med, 2004.
148. Kostyuchenko, V.A., et al., *Structure of the thermally stable Zika virus*. Nature, 2016. **533**: p. 425.
149. Frisch, M.J., et al., *Gaussian 09, Revision B.01*. 2009: Wallingford CT.
150. Hetényi, C. and D. van der Spoel, *Efficient docking of peptides to proteins without prior knowledge of the binding site*. Protein science : a publication of the Protein Society, 2002. **11**(7): p. 1729-1737.

151. Brooks, B.R., et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
152. Lengauer, T. and M. Rarey, *Computational methods for biomolecular docking*. Current Opinion in Structural Biology, 1996. **6**(3): p. 402-406.
153. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications*. Nature Reviews Drug Discovery, 2004. **3**: p. 935.
154. Tripathi, S., A. Bhardwaj, and P. E., *Approaches to Clustering in Customer Segmentation*. Vol. 7. 2018. 802.
155. Kongkaew, S., et al., *Molecular Dynamics Simulation Reveals the Selective Binding of Human Leukocyte Antigen Alleles Associated with Behçet's Disease*. PLOS ONE, 2015. **10**(9): p. e0135575.
156. Meeprasert, A., S. Hannongbua, and T. Rungrotmongkol, *Key Binding and Susceptibility of NS3/4A Serine Protease Inhibitors against Hepatitis C Virus*. Journal of Chemical Information and Modeling, 2014. **54**(4): p. 1208-1217.
157. Nutho, B., et al., *Binding mode and free energy prediction of fisetin/ β -cyclodextrin inclusion complexes*. Beilstein Journal of Organic Chemistry, 2014. **10**: p. 2789-2799.
158. Sangpheak, W., et al., *Enhanced stability of a naringenin/2,6-dimethyl β -cyclodextrin inclusion complex: Molecular dynamics and free energy calculations based on MM- and QM-PBSA/GBSA*. Journal of Molecular Graphics and Modelling, 2014. **50**: p. 10-15.
159. Dai, L., et al., *Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody*. Cell Host & Microbe, 2016. **19**(5): p. 696-704.
160. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*. Journal of Computational Chemistry, 2003. **24**(16): p. 1999-2012.
161. Csizmadia, F., et al., *Prediction of Distribution Coefficient from Structure. I. Estimation Method*. Journal of Pharmaceutical Sciences. **86**(7): p. 865-871.
162. Dixon, S.L. and P.C. Jurs, *Estimation of pKa for organic oxyacids using calculated atomic charges*. Journal of Computational Chemistry, 1993. **14**(12): p. 1460-1467.
163. McCammon, J.A., et al., *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations*. Nucleic Acids Research, 2004. **32**(suppl_2): p. W665-W667.
164. Roe, D.R. and T.E. Cheatham, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data*. Journal of Chemical Theory and Computation, 2013. **9**(7): p. 3084-3095.
165. Mishra, S.K. and J. Koca, *Assessing Performance of MM/GBSA, MM/PBSA and QM-MM/GBSA Approaches on Protein/Carbohydrate Complexes: Effect of Implicit Solvent Models, QM Methods, and Entropic Contribution*. The Journal of Physical Chemistry B, 2018.
166. Chen, X. and C.H. Reynolds, *Performance of Similarity Measures in 2D Fragment-Based Similarity Searching: Comparison of Structural Descriptors*

- and Similarity Coefficients*. Journal of Chemical Information and Computer Sciences, 2002. **42**(6): p. 1407-1414.
167. Irwin, J.J. and B.K. Shoichet, *ZINC – A Free Database of Commercially Available Compounds for Virtual Screening*. Journal of chemical information and modeling, 2005. **45**(1): p. 177-182.
168. Manner, S., et al., *Systematic Exploration of Natural and Synthetic Flavonoids for the Inhibition of Staphylococcus aureus Biofilms*. International Journal of Molecular Sciences, 2013. **14**(10): p. 19434-19451.
169. Wieder, M., et al., *Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations*. Journal of Chemical Information and Modeling, 2017. **57**(2): p. 365-385.
170. Berthold, M.R., et al. *KNIME: The Konstanz Information Miner*. in *Data Analysis, Machine Learning and Applications*. 2008. Berlin, Heidelberg: Springer Berlin Heidelberg.
171. Fawcett, T., *An introduction to ROC analysis*. Pattern Recognition Letters, 2006. **27**(8): p. 861-874.
172. Mysinger, M.M., et al., *Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking*. Journal of Medicinal Chemistry, 2012. **55**(14): p. 6582-6594.
173. Tsukamoto, T., et al., *Partial geometry optimization with FMO-MP2 gradient: Application to TrpCage*. Chemical Physics Letters, 2012. **535**: p. 157-162.
174. Pakiari, A.H. and A. Mohajeri, *Møller–Plesset perturbational self-consistent field theory*. Journal of Molecular Structure: THEOCHEM, 2004. **684**(1): p. 15-20.
175. Schmidt, M.W., et al., *General atomic and molecular electronic structure system*. Journal of Computational Chemistry, 1993. **14**(11): p. 1347-1363.
176. Fedorov, D.G., *The fragment molecular orbital method: theoretical development, implementation in GAMESS, and applications*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2017. **7**(6): p. e1322.
177. Choi, J., et al., *Application of the fragment molecular orbital method to discover novel natural products for prion disease*. Scientific Reports, 2018. **8**(1): p. 13063.
178. Suenaga, M., *Facio: New computational chemistry environment for PC GAMESS*. Vol. 4. 2005. 25-32.
179. Yennamalli, R., et al., *Identification of novel target sites and an inhibitor of the dengue virus E protein*. Journal of Computer-Aided Molecular Design, 2009. **23**(6): p. 333.
180. Hernández-Santoyo, A., et al., *Protein-Protein and Protein-Ligand Docking*, in *Protein Engineering - Technology and Application*, T. Ogawa, Editor. 2013, InTech: Rijeka. p. Ch. 03.
181. Bressanelli, S., et al., *Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation*. The EMBO Journal, 2004. **23**(4): p. 728-738.
182. Dubey, K.D., G. Tiwari, and R.P. Ojha, *Targeting domain-III hinging of dengue envelope (DENV-2) protein by MD simulations, docking and free energy calculations*. Journal of Molecular Modeling, 2017. **23**(4): p. 102.

183. Fuzo, C.A. and L. Degrève, *The pH dependence of flavivirus envelope protein structure: insights from molecular dynamics simulations*. Journal of Biomolecular Structure and Dynamics, 2014. **32**(10): p. 1563-1574.
184. Fuzo, C.A. and L. Degrève, *New pockets in dengue virus 2 surface identified by molecular dynamics simulation*. Journal of Molecular Modeling, 2013. **19**(3): p. 1369-1377.
185. David, C.C. and D.J. Jacobs, *Principal component analysis: a method for determining the essential dynamics of proteins*. Methods in molecular biology (Clifton, N.J.), 2014. **1084**: p. 193-226.
186. Kaiyawet, N., T. Rungrotmongkol, and S. Hannongbua, *Effect of Halogen Substitutions on dUMP to Stability of Thymidylate Synthase/dUMP/mTHF Ternary Complex Using Molecular Dynamics Simulation*. Journal of Chemical Information and Modeling, 2013. **53**(6): p. 1315-1323.
187. Hayes, J.M., *Chapter 2 - Computer-Aided Discovery of Glycogen Phosphorylase Inhibitors Exploiting Natural Products A2 - Brahmachari, Goutam*, in *Discovery and Development of Antidiabetic Agents from Natural Products*. 2017, Elsevier. p. 29-62.
188. Brammer, L., *Halogen bonding, chalcogen bonding, pnictogen bonding, tetrel bonding: origins, current status and discussion*. Faraday Discussions, 2017. **203**(0): p. 485-507.
189. Cavallo, G., et al., *The Halogen Bond*. Chemical Reviews, 2016. **116**(4): p. 2478-2601.
190. Lisac, K., et al., *Halogen-bonded cocrystallization with phosphorus, arsenic and antimony acceptors*. Nature Communications, 2019. **10**(1): p. 61.
191. Suroengrit, A., et al., *Halogenated Chrysin Inhibit Dengue and Zika Virus Infectivity*. Scientific Reports, 2017. **7**(1): p. 13696.
192. Mendez, L., et al., *Looking Back, Looking Forward at Halogen Bonding in Drug Discovery*. Molecules, 2017. **22**(9).
193. Lu, Y.-X., et al., *Substituent effects on noncovalent halogen/ π interactions: Theoretical study*. International Journal of Quantum Chemistry, 2007. **107**(6): p. 1479-1486.

VITA

NAME Kowit Hengphasatporn
DATE OF BIRTH 25 July 1988
PLACE OF BIRTH Bangkok, Thailand
INSTITUTIONS ATTENDED Chulalongkorn University
HOME ADDRESS 222/56, Premium place II, Ramindra 14 Rd., Ladprao, Bangkok, Thailand, 10230



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY