

**STUDIES ON THE BINDING AND INTERACTION OF
NEOLACTOTETRAOSYLCERAMIDE AND PEPTIDES WITH
DENGUE VIRUS TYPE 2 ENVELOPE PROTEIN**

ASFARINA BINTI AMIR HASSAN

**FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2019

**STUDIES ON THE BINDING AND INTERACTION OF
NEOLACTOTETRAOSYLCERAMIDE AND PEPTIDES
WITH DENGUE VIRUS TYPE 2 ENVELOPE PROTEIN**

ASFARINA BINTI AMIR HASSAN

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

**FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2019

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Asfarina Binti Amir Hassan

Registration/Matric No: MHA120002

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

Studies on the Binding and Interaction of Neolactotetraosylceramide and Peptides with Dengue Virus Type 2 Envelope Protein

Field of Study: Computer Aided Drug Design

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature

Date:

Subscribed and solemnly declared before,

Witness’s Signature

Date:

Name:

Designation:

**STUDIES ON THE BINDING AND INTERACTION OF
NEOLACTOTETRAOSYLCERAMIDE AND PEPTIDES WITH DENGUE
VIRUS TYPE 2 ENVELOPE PROTEIN**

ABSTRACT

Dengue fever is a common tropical infection and this acute febrile illness can be a deadly infection in cases of severe manifestation, causing dengue haemorrhagic shock syndrome. Due to the nature of the mosquito-borne infection, dengue has become a significant public health threat in many developing tropical countries. In this study, envelope (E) protein is selected as a target for drug design because it is believed to be responsible for the initial viral attachment to target cells and for mediating cellular entry of the virus. Domain III of the E-protein is found to be critical for virus adsorption to the receptors expressed on the host cell surface. In the present work, the ability of potential inhibitors to inhibit DENV was assessed *in silico* through docking of ligands into the 1OKE pdb structure, followed by binding modes analysis and calculations of the free energy of binding. The ligands involved were neolactotetraosylceramide (nLc₄Cer) and peptides (19-28 amino acids) that were believed to inhibit the trimeric conformation formation of DENV E-protein. The E-protein-nLc₄Cer and peptide E-protein complex models were generated using AutoDock 4.2 and Swarmdock docking programs. Then, the interaction of complexes were analysed using molecular dynamics simulation, followed by evaluation of the binding free energy and per-residue free energy decomposition analysis using the Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) and Molecular Mechanics Generalized Born Surface Area (MMGBSA). The amino acid residues in the DENV E-protein important for the interactions with inhibitors had also been highlighted. These computational studies suggest that nLc₄Cer and the peptides involved are proposed as potential inhibitors of DENV E-protein and are feasible to be developed as antiviral drugs for the treatment of

dengue fever. The information gathered may pave the way for the design of new anti-dengue drugs thereby aiding the discovery of a therapeutic cure for this infectious disease.

Keywords: dengue, peptide, envelope protein, molecular dynamics simulation, free energy of binding

Universiti Malaya

**KAJIAN DALAM MEMAHAMI PENGIKATAN DAN INTERAKSI
“NEOLACTOTETRAOSYLCERAMIDE” DAN PEPTIDA-PEPTIDA DENGAN
PROTEIN ENVELOPE VIRUS DENGGI JENIS 2**

ABSTRAK

Demam denggi adalah penyakit berjangkit yang biasanya ditemui di kawasan tropika dan penyakit febril akut ini boleh menjadi jangkitan yang membawa maut dalam kes-kes dengan manifestasi teruk yang menyebabkan sindrom kejutan denggi berdarah. Penyakit yang disebabkan oleh jangkitan nyamuk ini telah menjadi satu ancaman kesihatan di kebanyakan negara tropika yang sedang membangun. Dalam kajian ini, protein envelop (E) dipilih sebagai sasaran untuk merekabentuk ubatan kerana ia dipercayai bertanggungjawab sebagai langkah awal pelekatan virus kepada sel-sel sasaran sebelum kemasukan virus tersebut ke dalam sel. Domain III protein-E didapati penting bagi pelekatan virus pada permukaan sel perumah. Keupayaan perencat yang berpotensi untuk menghalang DENV telah dinilai secara “*in silico*” dengan cara memasukkan ligan ke dalam struktur pdb 1OKE diikuti dengan analisis mod ikatan dan pengiraan tenaga bebas pengikatan. Dalam penyelidikan ini, ligan-ligan yang terlibat adalah neolactotetraosylceramide (nLc₄Cer) dan peptida-peptida (asid amino 19-28) yang dipercayai boleh merencat pembentukan konformasi “trimeric” bagi protein-E. Model-model E-protein-nLc₄Cer dan peptida E-protein kompleks telah dihasilkan dengan menggunakan program AutoDock 4.2 dan Swarmdock. Kemudian, interaksi kompleks-kompleks tersebut dianalisa dengan menggunakan simulasi dinamik molekul, diikuti oleh penilaian tenaga bebas pengikatan dan analisis penguraian tenaga bagi setiap asid amino dengan menggunakan algoritma Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) dan Molecular Mechanics Generalized Born Surface Area (MMGBSA). Residu asid amino dalam protein-E DENV yang penting untuk interaksi dengan perencat juga ditekankan kepentingannya. Keputusan kajian

mencadangkan bahawa nLc₄Cer dan peptida yang terlibat berpotensi sebagai perencat kepada protein-E virus denggi dan berdayamaju untuk dibangunkan sebagai ubat-ubatan antiviral untuk merawat demam denggi. Kajian ini boleh membuka jalan untuk merekabentuk ubat-ubatan antidenggi baru, sekali gus membantu penemuan penawar terapeutik untuk penyakit berjangkit ini.

Kata kunci: denggi, peptida, protein envelop, simulasi dinamik molekul, tenaga bebas pengikatan

Universiti Malaya

ACKNOWLEDGEMENTS

First and foremost, gratitude and praises goes to Allah, in whom I have put my faith and trust in. During the entire course of this study, my faith has been tested countless times and with help of the Almighty, I have been able to go pass the obstacle that stood in my way.

I also would like to take this opportunity to express profound gratitude to my research supervisors Dr. Rozana Othman, Dr. Vannajan and Prof. Dr. Noorsaadah Abd. Rahman for the noble guidance and valuable advance throughout the period of study. Their patience, time, and understanding are highly appreciated. A word of thanks also goes to Dr. Thomas Anthony, Dr. Sasi Kumar and Mr Amarnadh during my attachments in Aurigene Discovery Technologies, Bangalore, India. Special thanks also go to Dr. Bimo Ario Tejo, Mr. Izuddin and Mrs. Azren from Universiti Putra Malaysia which providing assistance in many ways. They have made working in the laboratory during my attachment to these institutes a very enjoyable and often times entertaining experience.

I would also acknowledge MyBrain15 from Malaysia Ministry of Education, Lembaga Zakat Selangor and grants from University of Malaya (UMRG RP002/2012C and UM-QUB collaboration) which support this study.

Last but not least I would like to thank my parents and fellow friends whom I owe a debt of attitude for their prayers, encouragement and moral support throughout the whole duration of studies.

TABLE OF CONTENTS

Title Page	
Original Literary Work Declaration Form.....	ii
Abstract.....	iii
Abstrak.....	v
Acknowledgements.....	vii
Table of Contents.....	viii
List of Figures.....	xiv
List of Tables.....	xvii
List of Symbol and Abbreviations.....	xviii
List of Appendices.....	xxi
CHAPTER 1: INTRODUCTION.....	1
1.1 Background of Study.....	1
1.2 Problem Statement.....	3
1.3 Aim and Objectives.....	4
1.4 Scope of this Study.....	4
1.5 Output to be Predicted.....	5
1.6 Benefit of Research.....	5
1.7 Thesis Organization.....	6
CHAPTER 2: LITERATURE REVIEW.....	8
2.1 Dengue.....	8
2.1.1 Global Burden of Dengue Fever.....	8
2.1.2 Dengue in South East Asia Region.....	10
2.1.3 Dengue in Malaysia.....	12

2.1.4	Dengue Virus Structure and Life Cycle.....	13
2.1.5	Dengue Fever & Dengue Treatment.....	15
2.1.6	Dengue Virus Envelope (E) Protein as Target for Viral Attachment.....	16
2.2	<i>In Silico</i> Study.....	18
2.2.1	Automated Docking.....	18
2.2.2	Molecular Dynamic Simulation Studies on DENV E-Protein Systems..	20
2.2.3	Free Energy of Binding.....	22
2.2.4	Protein-Ligand Interactions.....	24
2.2.4.1	Hydrogen Bonding Interactions.....	24
2.2.4.2	Van der Waals Interactions (vdW).....	26
2.3	Inhibitors Targeting Envelope Proteins.....	27
2.3.1	Carbohydrate as a Small Molecule Inhibitor.....	27
2.3.2	Peptide.....	30
2.3.2.1	Peptide as an Inhibitor.....	30
2.3.2.2	Peptide Folding in Peptide Inhibitor Design.....	31
2.3.3	Small Molecule Inhibitors.....	33

CHAPTER 3: STUDIES INVOLVING NEOLACTOTETRAOSYLCERAMIDE

	(nLc₄Cer).....	35
3.1	Introduction.....	35
3.2	Methods.....	36
3.2.1	Characterizing Ligand Binding Site In Dengue Virus Type 2 (DENV2) Envelope Protein.....	36
3.2.1.1	Materials.....	36
3.2.1.2	Molecular Dynamics (MD) Simulation of DENV Envelope (E) Protein.....	37

3.2.1.3	Clustering.....	38
3.2.1.4	Binding Site Prediction.....	38
3.2.1.5	Docking of nLc ₄ Cer to DENV E-protein.....	39
3.2.1.6	Calculation of Free Energy of Binding.....	40
3.2.1.7	Analysis of the Decomposition of Free Energy.....	42
3.2.1.8	Analyses of Results.....	43
3.2.1.9	Flow Chart.....	44
3.2.2	Binding Pathway of Neolactotetraosylceramide towards Dengue Virus Type 2 (DENV2) Envelope Protein.....	45
3.2.2.1	Materials.....	45
3.2.2.2	Docking Pathway of nLc ₄ Cer towards DENV E-Protein.....	45
3.3	Results and Discussion.....	48
3.3.1	Characterizing Ligand Binding Site In Dengue Virus Type 2 (DENV2) Envelope Protein.....	48
3.3.1.1	Simulation of DENV E-Protein for 100 ns.....	48
3.3.1.2	Prediction of Binding Site on DENV E-Protein.....	48
3.3.1.3	Refinement of Docked E-Protein-nLc ₄ Cer Complex using MD simulation.....	53
3.3.1.4	Cluster Analysis.....	62
3.3.1.5	Binding Free Energy of E-Protein-nLc ₄ Cer Complex...66	
3.3.1.6	Per-Residue Free Energy Decomposition (DC) Analysis.....	70
3.3.1.7	Analysis of DENV E-Protein-nLc ₄ Cer Complex Interactions.....	74
3.3.1.8	Secondary Structure Analysis.....	77

3.3.2	Binding Pathway of Neolactotetraosylceramide towards Dengue Virus Type 2 (DENV2) Envelope Protein.....	80
3.3.2.1	Analysis of E-Protein-nLc ₄ Cer Interactions.....	80
3.4	Conclusion.....	92
CHAPTER 4: STUDIES INVOLVING PEPTIDES.....		94
4.1	Introduction.....	94
4.2	Methods.....	96
4.2.1	Binding Free Energies of Peptides Inhibiting Dengue Virus (DENV) Envelope Protein by using Docking & Molecular Dynamics Simulation.....	96
4.2.1.1	Materials.....	96
4.2.1.2	<i>In Silico</i> Peptide Folding.....	97
4.2.1.3	Docking of Peptides to DENV E-Protein.....	98
4.2.1.4	Molecular Dynamics Simulations of Peptide-E-Protein Complexes.....	99
4.2.1.5	Analyses of Results.....	100
4.2.1.6	Calculation of Free Energy of Binding.....	101
4.2.1.7	Analysis of the Decomposition of Free Energy.....	102
4.2.1.8	Calculation of Hydrogen Bond (H-bond) Occupancies.....	103
4.2.1.9	Assessment of the Secondary Structure of the Folded Peptides.....	103
4.2.1.10	Flow Chart.....	104
4.2.2	Peptide Synthesis.....	105
4.2.2.1	Solid Phase Peptide Synthesis (SPPS) Method.....	106
4.2.2.2	Purification of Peptides.....	109
4.2.2.3	Lyophilisation.....	109

4.3	Results and Discussion.....	110
4.3.1	Binding Free Energies of Peptides Inhibiting Dengue Virus (DENV) Envelope Protein by using Docking & Molecular Dynamics Simulation.....	110
4.3.1.1	<i>In Silico</i> Peptide Folding.....	110
4.3.1.2	Docking of Peptides to DENV E-Protein.....	111
4.3.1.3	Refinement of Docked Peptide-E-Protein Complexes using MD Simulation.....	118
4.3.1.4	Root Mean Square Deviation (RMSD).....	118
4.3.1.5	Radius of Gyration (R_g).....	122
4.3.1.6	Root Mean Square Fluctuation (RMSF).....	124
4.3.1.7	Validation of 3D Model Structure.....	125
4.3.1.8	Binding Free Energy of Peptide-E-Protein Complexes.....	129
4.3.1.9	Per-Residue Free Energy Decomposition (DC) Analysis.....	134
4.3.1.10	Analysis of Peptide-E-Protein Interaction.....	140
4.3.1.11	Hydrogen Bond Occupancy in Molecular Dynamics Simulation.....	145
4.3.1.12	Secondary Structure Analysis.....	147
4.3.2	Peptide Synthesis.....	152
4.3.2.1	Solid Phase Peptide Synthesis (SPPS).....	152
4.3.2.2	Lyophilisation.....	155
4.3.2.3	Purification of Peptides.....	156
4.4	Conclusion.....	161

CHAPTER 5: GENERAL DISCUSSION & CONCLUSION.....	162
5.1 General Discussion.....	162
5.2 Conclusion.....	164
5.3 Significant of the Study.....	164
5.4 Suggestion for Future Studies.....	165
References.....	166
List of Publications and Papers Presented.....	192
Appendix.....	197

Universiti Malaya

LIST OF FIGURES

Figure 1.1 : Workflow of this research project.....	7
Figure 2.1 : The <i>Aedes aegypti</i> mosquito in action.....	10
Figure 2.2 : The dengue virus genome encodes three structural (capsid [C], membrane [M] and envelope [E] and seven nonstructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) proteins.....	13
Figure 3.1 : The sliding grid box location on the E-protein where docking has performed.....	47
Figure 3.2 : Flow chart showing summarized method involved in the simulation of binding pathway of nLc ₄ Cer towards DENV E-protein.....	47
Figure 3.3 : Positions of the three binding sites, (a) BS1, (b) BS2 and (c) BS3, as predicted by Discovery Studio 4.0. Ligand binding site is shown in CPK presentation. DI, DII and DIII are illustrated in red, yellow and blue colours, respectively.....	51
Figure 3.4 : The energy of the system: red line (positive) shows the kinetic energy, and black line (negative) shows the potential energy. The green line shows the total energy of E-protein-nLc ₄ Cer complex during 50 ns simulation. The heating process is not shown.....	54
Figure 3.5 : Volume of E-protein-nLc ₄ Cer complex during 50 ns MD simulation.....	55
Figure 3.6 : Density of E-protein-nLc ₄ Cer complex during 50 ns MD simulation.....	56
Figure 3.7 : Pressure of E-protein-nLc ₄ Cer complex during 50 ns MD simulation.....	56
Figure 3.8 : Temperature of E-protein-nLc ₄ Cer complex during 50 ns MD simulation.....	57
Figure 3.9 : RMSD of E-protein-nLc ₄ Cer complex during 50 ns MD simulation.....	58
Figure 3.10: RMSF of E-protein-nLc ₄ Cer complex during 50 ns MD simulation.....	59
Figure 3.11: Radius of gyration (R_g) of E-protein-nLc ₄ Cer complex for backbone heavy atoms during 50 ns MD simulation.....	62
Figure 3.12: The conformational structures of DENV E-protein identified by clustering technique.....	64
Figure 3.13: Histogram showing the calculated per-residue free energy decomposition using MMPBSA and MMGBSA approaches for E-protein-nLc ₄ Cer complex at 50 ns simulation.....	73
Figure 3.14: Histogram showing the calculated per-residue free energy decomposition using MMGBSA approaches for E-protein-nLc ₄ Cer complex at different time simulation.....	73
Figure 3.15: Connolly surface representation of nLc ₄ Cer at the binding site.....	75

Figure 3.16: Schematic diagram (2D) of residues in the binding site which exhibit interactions with nLc ₄ Cer at 50 ns MD simulation, obtained using the Ligplot program.....	76
Figure 3.17: View of nLc ₄ Cer (stick) at the binding site of DENV E-protein (line). Residues interacting with ligand are shown as sticks.....	76
Figure 3.18: Secondary structure of E-protein-nLc ₄ Cer complex at 50 ns.....	79
Figure 3.19: Chemical structure of neolactotetraosylceramide (nLc ₄ Cer).....	81
Figure 3.20: Ligand pathway along x-axis shown in stick (a) and CPK (b).....	82
Figure 3.21: Ligand pathway along y-axis shown in shown in stick (a) and CPK (b)...	83
Figure 3.22: Ligand pathway along z-axis shown in stick (a) and CPK (b).....	84
Figure 3.23: Optimized ligand pathway shown in stick (a) and CPK (b).....	85
Figure 4.1 : Flow chart showing the methods involved in the study of peptide-E-protein complex interactions.....	104
Figure 4.2 : The flow chart shows summarized method involved in the peptide synthesis.....	105
Figure 4.3 : Flow chart of solid support peptide synthesis.....	108
Figure 4.4 : The top five docked conformation of peptides (a-j) generated from SwarmDock docking server.....	114
Figure 4.5 : Three possible sites for peptide binding towards DENV E-protein.....	117
Figure 4.6 : Pressure of peptide-E-protein complexes during 50 ns MD simulation...	119
Figure 4.7 : Temperature of peptide-E-protein complexes during 50 ns MD simulation.....	119
Figure 4.8 : Density of peptide-E-protein complexes during 50 ns MD simulation....	120
Figure 4.9 : Potential energy of top five peptide-E-protein complexes during 50 ns simulation.....	120
Figure 4.10: RMSD of top five peptide-E-protein complexes during 50 ns MD simulation.....	121
Figure 4.11: Radius of gyration (R_g) of top five peptide-E-protein complexes for backbone heavy atoms during 50 ns MD simulation.....	123
Figure 4.12: RMSF profile of top five peptide-E-protein complexes during 50 ns MD simulation.....	125
Figure 4.13: Verify3D plot for DS36wt at 50 ns simulation. 89.86% of the residues had	

an average 3D-1D score of more than 0.2.....	128
Figure 4.14: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DS36wt-E-protein complex.....	137
Figure 4.15: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DN81opt-E-protein complex.....	138
Figure 4.16: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DN57opt-E-protein complex.....	138
Figure 4.17: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DS03opt-E-protein complex.....	139
Figure 4.18: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DS04opt-E-protein complex.....	139
Figure 4.19: Connolly surface representation of residue Ile129 (brown) at the binding site.....	140
Figure 4.20: Secondary structure plots of peptide-E-protein complexes using Visual Molecular Dynamics (VMD).....	149
Figure 4.21: Automated peptide synthesizer used in this study.....	153
Figure 4.22: Freeze dryer used in this study for lyophilisation process.....	156
Figure 4.23: LCMS spectrum of peptide with code DS36opt.....	159
Figure 4.24: LCMS spectrum of peptide with code DS36wt.....	160
Figure 4.25: LCMS spectrum of peptide with code DN81opt.....	160

LIST OF TABLES

Table 3.1 : Binding energy of E-protein-nLc ₄ Cer complex subjected to various binding sites defined by Discovery Studio 4.0 program.....	52
Table 3.2 : Binding energy and RMSD of E-protein-nLc ₄ Cer complex subjected to different conformations generated from MD trajectories.....	64
Table 3.3 : Binding free energies predicted using the MMPBSA and MMGBSA methods at different simulation times (All data are given in kcal/mol).....	67
Table 3.4 : Binding energy (kcal/mol) of nLc ₄ Cer towards E-protein from different pathways.....	89
Table 3.5 : Binding energy (kcal/mol) of optimized pathway for nLc ₄ Cer-E-protein complex selected from x, y and z axes.....	89
Table 4.1 : Peptides used in this study.....	96
Table 4.2 : List of peptides used in this study.....	107
Table 4.3 : List of synthesized peptides chosen for LCMS analysis.....	109
Table 4.4 : Potential energy for peptide folding at different temperatures.....	111
Table 4.5 : Binding energy for both monomer and dimer E-protein docked with peptides generated from Swarmdock docking server.....	112
Table 4.6 : Relationship between the binding activities from fluorescence quenching assay and the binding energies calculated using the Swarmdock docking algorithm.....	113
Table 4.7 : Validation of peptide-E-protein complex structures using different verification programs.....	126
Table 4.8 : Binding free energies predicted using the MMPBSA and MMGBSA methods at different time simulation (All data are given in kcal/mol).....	131
Table 4.9 : H-bond and hydrophobic interactions analyses using Ligplot, for the top five peptide-E-protein complexes.....	143
Table 4.10: Hydrogen bond analysis between peptides and E-protein at 50 ns.....	146

LIST OF SYMBOLS AND ABBREVIATIONS

2D	:	Two dimensional
3D	:	Three dimensional
Å	:	Amstrong
ACN	:	Acetonitrile
AMBER	:	Assisted Model Building & Energy Refinement
π	:	pie
ΔG_{bind}	:	Estimated free energy of binding
Boc	:	Tert-Butyloxycarbonyl
BE	:	Binding energy
C	:	Capsid
CTC	:	Chlorotrietyl chloride
DI	:	Domain I
DII	:	Domain II
DIII	:	Domain III
DCM	:	Dichloromethane
DENV1	:	Dengue virus type 1
DENV2	:	Dengue virus type 2
DENV3	:	Dengue virus type 3
DENV4	:	Dengue virus type 4
DENV	:	Dengue virus
DF	:	Dengue fever
DHF	:	Dengue hemorrhagic fever
DIC	:	Diisopropylcarbodiimide
DIEA	:	N,N-Diisopropylethylamine
DMF	:	Dimethylformamide
DSS	:	Dengue shock syndrome
E	:	Envelope protein

Fmoc	:	Flourenylmethyloxycarbonyl
GPF	:	Grid parameter file
H ₂ O	:	Water
H-bond	:	Hydrogen bond
HATU	:	1-[Bis(dimethylamino) methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBPLUS	:	Hydrogen bonding calculation program
HOBt	:	1-hydroxybenzotriazole
kcal	:	Kilocalories
K _i	:	Inhibition constant
LCMS	:	Liquid chromatography mass spectrometry
LIGPLOT	:	A program to plot schematic diagrams of protein-ligand interactions.
MD	:	Molecular dynamics
MMPBSA	:	Molecular Mechanics Poisson Boltzmann Surface Area
MMGBSA	:	Molecular Mechanics Generalized Born Surface Area
m/z	:	mass/charge
ns	:	Nanoseconds
PDB	:	Protein Data Bank
PME	:	Particle Mesh Ewald
ps	:	Picoseconds
RMSD	:	Root mean square deviation
RMSF	:	Root mean square fluctuation
TBTU	:	<i>O</i> -(Benzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium tetrafluoroborate
TFA	:	Trifloureacetic acid
TIS	:	Triisopropylsilane
WHO	:	World Health Organization

Amino acids

Ala (A)	:	alanine
Arg (R)	:	arginine
Asn (N)	:	asparagines
Asp (D)	:	aspartic acid
Cys (C)	:	cysteine
Gln (Q)	:	glutamine
Glu (E)	:	glutamic acid
Gly (G)	:	glycine
His (H)	:	histidine
Ile (I)	:	isoleucine
Leu (L)	:	leucine
Lys (K)	:	lysine
Met (M)	:	methionine
Phe (F)	:	phenylalanine
Pro (P)	:	proline
Ser (S)	:	serine
Thr (T)	:	threonine
Trp (W)	:	tryptophan
Tyr (Y)	:	tyrosine
Val (V)	:	valine

LIST OF APPENDICES

Appendix A: Sample of Swarmdock docking output for DN58opt-E-protein complex.....	197
Appendix B: An example of clustering histogram from docking result of nLc ₄ Cer docked towards DENV E-protein by AutoDock 4.2 software.....	198
Appendix C: An example of the calculation of estimated free energy of binding from docking result (towards DENV E-protein) for the chosen nLc ₄ Cer conformation.....	199
Appendix D: Sample of AMBER MD input file.....	201
Appendix E: Sample AMBER MMPBSA input file.....	202
Appendix F: Verify3D plot for peptide-E-protein complexes at 50 ns simulation.....	204

Universiti Malaysia

CHAPTER 1: INTRODUCTION

1.1 Background of Study

Dengue is a viral infection caused by four types of viruses (DENV1, DENV2, DENV3 and DENV4) belonging to the *Flaviviridae* family. The viruses are transmitted through the bite of infected *Aedes aegypti* mosquitoes that feed both indoors and outdoors during daytime. These mosquitoes thrive in areas with standing water, including puddles, water tanks, containers and old tyres. Lack of reliable sanitation and regular garbage collection can also contribute to the spread of the mosquitoes.

Dengue occurs nationally in Malaysia, with increased risk in urban and periurban areas. Recent cases have been confirmed in Johor, Malacca, Negeri Sembilan, Kuala Lumpur, Selangor, Putrajaya, Perak, Penang, Pahang, Perlis and Sarawak (Zaki et al., 2019). Peak transmission occurs in the late monsoon season (October until February in East Peninsular Malaysia, Sabah and Sarawak, while in West Peninsular Malaysia is in July until August). A vaccine is available for people living in some dengue endemic countries, but not commercially available for travellers.

In mammalian cells, there are many candidate molecules that may act as receptors for dengue recognition, which include neolactotetraosylceramide (nLc₄Cer). nLc₄Cer had been reported as a putative receptor and the most promising target for DENV entry inhibitor design (Hidari, Abe, & Suzuki, 2013). The non-reducing terminal disaccharide residue, Gal β 1-4GlcNAc, of nLc₄Cer is a critical determinant for the binding of DENV2. The virus entry process mediated through host carbohydrate molecule is crucially involved in virus propagation and the pathological progression of dengue disease. Therefore, by understanding the molecular mechanism of virus entry from this study, the future development of effective new therapies to treat dengue patients can be done.

On the other hand, peptides were also used in this study as they have been reported to be highly selective against their target, possess low toxicity and low accumulation in tissues (Chew, Poh, & Poh, 2017). They readily exist in the human body and exert diverse biological roles, especially as signalling and regulatory molecules in a variety of physiological processes. Peptides that either interact with the virus particle or target at viral replication steps of the life cycle have a probability to be used as a cure for dengue disease. By targeting the attachment factor, peptides may prevent the attachment and binding of viral proteins with the host cell, thus inhibit the DENV entry.

In this study, computational approaches such as molecular dynamics (MD) simulation and docking methods were used to study the interactions of potential inhibitors, which were neolactotetraosylceramide (nLc₄Cer) and peptides, with dengue virus (DENV) envelope (E) protein. This was followed by evaluation of the binding free energy and per-residue free energy decomposition analysis using the Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) and Molecular Mechanics Generalized Born Surface Area (MMGBSA) methods. In addition, these methods were used to identify possible binding sites in the domain (DIII) region of the E-protein. These approaches allowed us to explore the various conformational states and protein flexibility, and eventually obtained the binding energies of the ligand binding towards these multiple structures. Currently, these methods are becoming common in computational tools for drug discovery. The advantage of this MD simulation method is it can explicitly treat structural flexibility and give entropic effects (Vivo, Masetti, Bottegoni, & Cavalli, 2016). This would allow more accurate estimation of the thermodynamics and kinetics associated with drug-target recognition and binding.

1.2 Problem Statement

The rise in dengue cases reported can be due to the increase in people mobility in Malaysia, which facilitate the spread of the virus all around the country (Pang & Loh, 2016). Moreover, Malaysia has a long dry season, and together with occasional heavy rains produce a perfect mosquito-breeding condition.

Currently, there is no available drug that is effective against all four dengue virus serotypes (DENV1 to 4). Once a person has immunity against one dengue serotype, the individual will never be infected with that same serotype again. If the person is subsequently infected with a different dengue serotype, however, he or she will have an increased risk of developing a more severe dengue illness. Hence, a safe drug must provide immunity against all four DENVs and further studies are compulsory in order to protect people from severe dengue infections.

Outcomes from phase III clinical trials showed that the first dengue vaccine, Dengvaxia (CYD-TDV), successfully reduced dengue hospitalizations by 80%. However, its average efficacy against DENV was low, especially against DENV1 at approximately 50% and against DENV2 at 39%. Furthermore, previous clinical trials revealed that CYD-TDV vaccination caused elevated risks of hospitalization for children less than nine years of age. The World Health Organization has therefore recommended the use of CYD-TDV vaccine only in countries where epidemiological data indicated a high burden of dengue (Chew et al., 2017).

Presently, the structural study towards understanding the binding of nLc₄Cer and designed peptides to DENV2 E-protein is rarely done especially in Malaysia. Thus, the current study will provide important call to deliver a durable and effective drug in order to fight this fatal global threat.

1.3 Aim and Objectives

This study focuses mainly towards inhibiting the attachment of DENV E-protein to the host cell. The aim of this research is to understand the binding modes of neolactotetraosylceramide (nLc₄Cer) and peptides to DENV E-protein, and screen for potential inhibitors of DENV E-protein through docking and molecular dynamics simulations.

Objectives of this study are:

- i. To evaluate the binding stability of nLc₄Cer to DENV E-protein.
- ii. To investigate the binding pathway of nLc₄Cer towards the binding site.
- iii. To identify potential peptide inhibitors that can inhibit the trimeric conformation formation of DENV E-protein.
- iv. To synthesize some of the peptides identified as potential inhibitors.

1.4 Scope of this Study

The research is focused on exploring for potential inhibitors of DENV E-protein. The scope of this research is limited to the computational studies and can be classified into two major parts which involved binding mechanisms of neolactotetraosylceramide (nLc₄Cer) towards DENV E-protein and the study of peptides as potential inhibitors. The movement of nLc₄Cer towards the binding site was also observed through docking study. In this section, a cost effective technique known as “pathway docking” was used to illustrate ligand-receptor binding free energy in order to find potential pathways of nLc₄Cer approaching E-protein binding site. Meanwhile, for the peptide study, only the best and potent peptides were chosen for syntheses for future inhibition assay and their binding free energy were evaluated through the molecular dynamics simulation method.

1.5 Output to be Predicted

The output of this study shall include inhibitors that may inhibit DENV entry into host cell by interacting directly with the virus E-protein and interfere with the viral attachment to host cell surface. Inhibition occurs as a result of inhibitor binding to E-protein, induction of structural changes in the DENV surface, and interference with virus-cell binding. Inhibitor targeting the cell entry of DENV may help to overcome dengue infection.

The binding energies between peptides and nLc₄Cer towards E-protein obtained for this study should also show negative values indicating favourable binding. In this study, important amino acids that are involved in the binding interaction are identified. It is hoped that the designed peptides and nLc₄Cer may act as inhibitors of DENV E-protein and may lead to the development of drugs against DENV infections.

1.6 Benefit of Research

This study is intended to highlight the importance of finding potential inhibitors against dengue activities that can have impact on the quality of life of the society. This research is very important as it can become a significant step in the design of drugs against dengue diseases which eventually leads to the reduction of the society burden by introducing potential inhibitors.

The peptides and nLc₄Cer have the potentials to block DENV entry into target cells, which result in the inhibition of viral infection. Hopefully, the new inhibitor discovery strategy based on peptides and nLc₄Cer will help in the future development of therapeutics against DENV infections. With effective antivirals, vaccination and vector controls such as fogging, dengue in endemic countries can be better controlled and the state of public health can be improved.

1.7 Thesis Organization

This thesis contains five chapters. Chapter 1 is the introduction to the whole study including problem statement, aim and objectives, scope and benefit of research.

Chapter 2 presents a literature review of the dengue phenomenon worldwide and especially in Malaysia, the dengue virus structure and its life cycle, the importance of dengue virus (DENV) envelope protein (E-protein) as target for viral entry as well as implementation of *in silico* study in this research.

Characterization of ligand binding site on DENV2 E-protein is discussed in Chapter 3. In this chapter, the usage of Discovery Studio 4.0 software in determining the binding site of E-protein will be discussed. Several regions of the E-protein are believed to be the target binding sites for inhibition activity. This chapter also presents data analysis of the binding pathway of neolactotetraosylceramide (nLc₄Cer) towards DENV2 E-protein. Docking pathway method was performed using AutoDock 4.2 to observe the movement of nLc₄Cer as it approached DENV E-protein prior to binding.

In Chapter 4, binding free energy calculations of peptides inhibiting DENV E-protein were evaluated by using MD simulation. In this study, the Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) and Molecular Mechanics Generalized Born Surface Area (MMGBSA) algorithms were used to assess the binding affinity of peptide-E-protein complexes. In addition of that, this chapter gives a description on how the peptides were synthesized using the automated peptide synthesizer. Purification of the synthesized peptides was done using liquid chromatography mass spectrometry (LCMS) and high performance liquid chromatography (HPLC).

Chapter 5 provides the conclusion of research studied and recommendation for further works. The flow chart below (Figure 1.1) shows the architecture of this study

towards finding potential inhibitors that can inhibit the formation of DENV E-protein trimeric conformation.

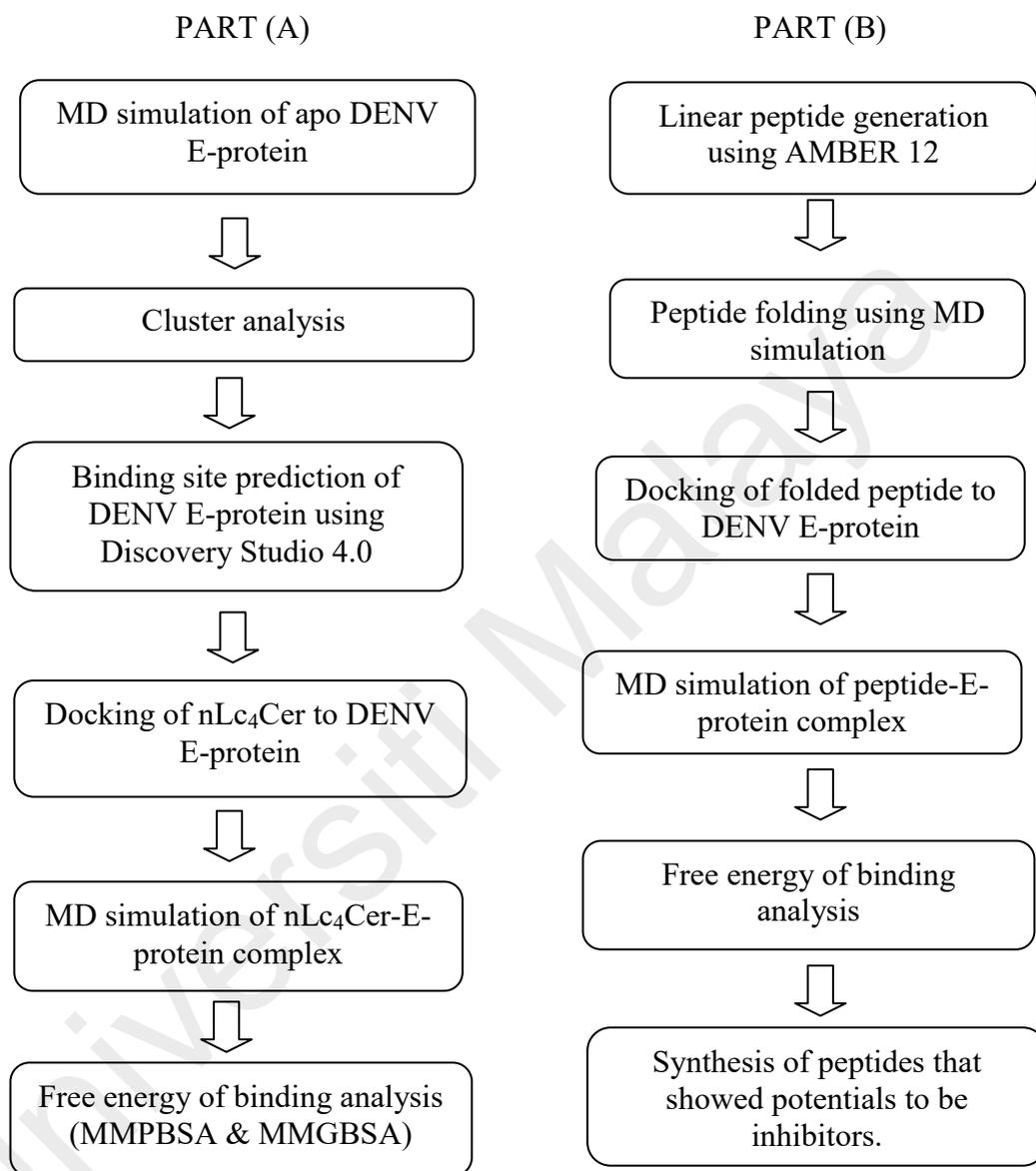


Figure 1.1: Workflow of this research project.

CHAPTER 2: LITERATURE REVIEW

2.1 Dengue

2.1.1 Global Burden of Dengue Fever

According to the World Health Organization (WHO), dengue fever is a disease spread by female *Aedes aegypti* mosquitoes (Figure 2.1). WHO has characterized the disease as one of the world's fastest-growing viral risk. Today, dengue ranks as the most important mosquito-borne viral disease in the world (Chang, Tien, & Lu, 2018). The dengue virus (DENV) passes through the mosquito gut into the mid gut; it then replicates in the mid gut and also in body tissues. After five days, the virus then infects the salivary glands. Upon feeding, the mosquito bites a human and delivers the dengue virus through the saliva.

Dengue fever is a febrile illness that affects infants, toddlers and adults with symptoms appearing 3-14 days after the pandemic bite. This acute illness of sudden onset usually follows with symptoms such as headache, fever, exhaustion, severe muscle and joint pain, swollen glands and rash. Other signs of dengue fever consist of bleeding gums, serious pain behind the eyes and red palms and soles. This disease can become lethal for people with low levels of immunity. Since there are four serotypes of dengue virus (DENV), which are DENV1, 2, 3 and 4, hence it is possible to get dengue fever several times. Nevertheless, a dengue infection produces immunity for a lifetime to that particular serotype to which the patient was exposed. On the contrary, it does not provide immunity to other dengue virus serotypes.

According to WHO, the year 2016 was characterized by large dengue outbreaks worldwide. The Region of the Americas reported more than 2.38 million cases in 2016, where Brazil alone contributed slightly less than 1.5 million cases, approximately 3 times higher than in 2014. There was also a record of 1,032 dengue deaths reported in

the region. Meanwhile, in 2017, the Region of Americas had reported 50,172 cases of dengue fever, a reduction as compared with corresponding periods in previous years (Adam et al., 2017). In general, majority of the dengue cases reported involved returning tourists that had been declared as disease endemic areas. Dengue cases were reported particularly in Northern Mexico, Virgin Islands, Guam and Samoa (Guarner & Hale, 2019). The Western Pacific Region had been reported dengue outbreaks in several Member States in the Pacific, as well as the circulation of DENV1 and DENV2 serotypes. An estimated 500,000 people with severe dengue require hospitalization each year, and about 2.5% of those affected die.

In the late 2015 and early 2016, the first dengue vaccine, Dengvaxia by Sanofi Pasteur, was registered in several countries for use in individuals 9 - 45 years of age, living in endemic areas (Bustamam, Aldila, & Yuwanda, 2018). Dengue vaccine development efforts aim for a vaccine which simultaneously provides long-term protection against all DENV serotypes (Schmitz, Roehrig, Barrett, & Hombach, 2011). Recent studies (Dans, Dans, Lansang, Silvestre, & Guyatt, 2018; Halstead, 2018) in Phase 3 trials in Asia and Latin America have shown that the licensed tetravalent vaccine, Dengvaxia, had variable efficacy depending on immune status prior to vaccination and the infection serotype. The Dengvaxia clinical trials had revealed that even individuals with detectable neutralizing antibodies to a particular serotype experienced vaccine break-through infections. In fact, based on long-term follow up data, Dengvaxia is no longer recommended for use in DENV-naive individuals as reported by Gallichotte et al. (2018).

Since there is no exact medicine to treat dengue fever available, thus prevention in terms of mosquito control is the main strategy in dengue management. Awareness of the geographical distribution and trouble of dengue is necessary in order to determine how to optimally allocate the limited resources available for dengue control

and in evaluating the impact of such activity universally. The global burden of dengue is terrifying and represents a growing challenge to public health. It is hoped that the evaluation of contemporary dengue risk distribution and burden will help to reduce dengue cases worldwide (Bhatt et al., 2013).



Figure 2.1: The *Aedes aegypti* mosquito in action (Dennis, Goldman, & Vosshall, 2019)

2.1.2 Dengue in South East Asia Region

Generally, 1.3 billion people living in Southeast Asia are at risk of dengue infection. Since 2003, only eight countries in the region had reported cases. According to the WHO, in 2004 Bhutan reported its first case, in 2005 Timor-Leste, then in 2006 Nepal. North Korea is the only country in the region that is free of this disease. The seasonal pattern of dengue vary across various countries: cases peak in India between August and November; in Indonesia during January and February; and in Myanmar and Sri Lanka between May and August. Basically, dengue fatality is not high, but costs including loss of productivity and the financial burden of health services have a large impact on economies and households (Ghosh & Dar, 2015). The disease can also affect revenue from tourism and foreign direct investment. According to health policy researchers at Brandeis University, Waltham, Massachusetts, USA, the annual

economic burden of dengue in Southeast Asia is nearly one billion US dollars with Indonesia and Thailand experiencing the highest costs. Thailand, Vietnam and Colombia have seen dramatic increase in dengue infections due to the absence of vaccination campaigns as well as lack of vector control activities (Lee et al., 2017). Meanwhile, in Philippines, which has the largest number of death, nevertheless had made some inroads as bringing numbers down from 499 deaths as they started a school-based dengue vaccination program in Manila as well as increase awareness in dengue understanding among public (Undurraga et al., 2017).

In Thailand, dengue hemorrhagic fever (DHF) poses a main problem to public health. Every year, DHF that can cause to organ impairment, severe bleeding and eventually death affects between 15,000 and 105,000 people. In fact, the prevention and treatment programs to reduce dengue cases become difficult as the number and location of cases differ drastically (Lauer et al., 2018). Meanwhile, according to Singapore National Environment Agency, in year 2017, there were 2,772 dengue cases had been reported and this is the lowest figure in the past 16 years. This decrement was believed due to the local population has built up immunity after high number of infected disease where there were dengue outbreaks (Ang et al., 2019). In other parts of Asia, Hong Kong, Japan and South Korea have seen increments in dengue cases, largely because of people coming back from visits in Southeast Asian countries (Ebi & Nealon, 2016).

Southeast Asian countries report heightened efforts this year to control this fatal disease. Without a preventive vaccines or treatment, the rigorous public-awareness campaigns are the only tools to encounter dengue. In Bangkok, teams of municipal officers in jumpsuits armed with mosquito-killer spraying machines and masks fan out every day at houses and communities where dengue patients have been reported. Their

responsibility is to prevent the possibility of dengue spread by killing adult mosquitoes and whipping out their breeding sites (Messina et al., 2019).

Until now there is no specific treatment for dengue fever. For serious dengue cases, the medication is only limited to intravenous drips to try to replenish fluids victims lose as they struggle against fevers reaching as high as 41°C. Maintenance of the patient's body fluid volume is extremely important especially for severe dengue care. Without good medication, victims may have bleeding and shock (Cucunawangsih & Lugito, 2017).

2.1.3 Dengue in Malaysia

In Malaysia, the wet season is already shaping up as potentially the worst on record for dengue fever with regional governments stepping up efforts to limit the spread of the potentially deadly disease. According to General Health Director, Dr. Noor Hisham Abdullah, the number of dengue fever cases for the period of February 18 to 24 increase by 52 compared to previous week. He claimed that a total of 1,224 cases were reported for that period compared to 1,172 cases earlier week. Twenty people have already died in Malaysia due to dengue compared with forty death were recorded in 2017, a drop of twenty cases by 50% (G.H. et al., 2019). Recent cases have been confirmed in Johor, Malacca, Negeri Sembilan, Kuala Lumpur, Selangor, Putrajaya, Perak, Penang, Pahang, Perlis, and Sarawak.

Health experts believe that rainy season that brought mosquitoes out in April, contributed to the seriousness of the dengue challenge. In addition, above average temperatures that many experts blame on global warming encourage early mosquito breeding. Meanwhile, dengue is thought to be mutating as a result of immunity that has

built up in the region. And if the virus is spread by travellers, more countries are expected to be affected (Pang & Loh, 2016).

2.1.4 Dengue Virus Structure and Life Cycle

Flaviviruses consist of single stranded positive sense RNA genomes that are approximately 11 kb in size. The viral genome is translated as a polyprotein in the cytoplasm. There are signal and stop-transfer sequences that direct the translocation of the polyprotein back and forth across the endoplasmic reticulum (ER) membrane. The polyprotein is subsequently co- and post-translationally modified by viral and host-encoded proteases to produce three structural and seven nonstructural proteins. The mature virion consists of three structural proteins which are the capsid, C; a membrane associated protein (which is produced from the precursor prM), M; and the envelope protein, E. The nonstructural (NS) proteins include large, highly conserved proteins, which are NS1, NS3, and NS5, and four small hydrophobic proteins, NS2A, NS2B, NS4A, and NS4B. The following diagram (Figure 2.2) depicts the flavivirus genome organization.

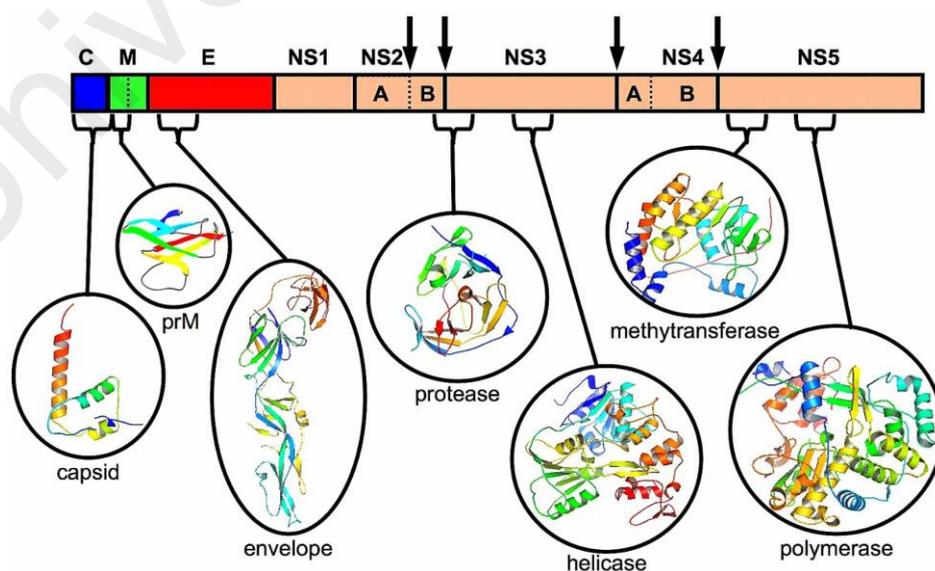


Figure 2.2: The dengue virus genome encodes three structural (capsid [C], membrane [M] and envelope [E]) and seven nonstructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) proteins (Tomlinson, Malmstrom, & Watowich, 2009).

As the major surface protein, many groups have concentrated on E-protein in their studies of viral particles. The crystal structure of E-protein revealed each that monomer consists of three domains that are believed to play crucial role in the virus life cycle (Hacker, 2009). Dengue E-protein with a molecular weight of 54.5 kDa is folded largely into beta-sheets and contain three distinctive domains: the N-terminus central domain (domain I); the fusion (or dimerization) domain that contains the fusion peptide (domain II); and the immunoglobulin (IgG) like domain (domain III). The domain III is mainly responsible for receptor recognition, which is essential for viral attachment to facilitate viral entry into host cells by receptor-mediated method (Watterson, Kobe, & Young, 2012). The mannose found at the glycosylated site appears to be critical for viral entry through receptor binding. Exposure of E-protein dimers to low pH in early endosomes after viral uptake results in irreversible conformational adjustment of E-protein to a trimer conformation. The crystallographic structures suggest that E-proteins in the dimer form experience a dramatic conformational adjustment through the interdomain linkers that results in irreversible change of the dimer E-protein to the homotrimer form. This pH-induced trimer structure is capable to ruin the strong packing on the outer virion surface.

DENV go into a cell through receptor-mediated endocytosis through binding of E-proteins and receptors. After the interaction of E-protein with the receptors, E-protein undergoes conformational adjustment to allow the fusion of plasma membrane with dengue viral envelope in a pH-dependent manner. The viral access can be stopped by using antibody against E-protein, or using protease to remove the cell surface receptors (Chen et al., 1997). DENV can also attach to macrophage or monocyte cells mediated by reactive immunoglobulin G (IgG) binding to Fc-receptors. The cells bearing Fc-receptors normally produce higher number of virus when infected with DENV in the presence of serologically cross-reactive but not neutralizing antisera. This

immunological phenomenon is known as “antibody-dependent enhancement” and is considered to play a role in the pathogenesis of dengue hemorrhagic fever and dengue shock syndrome (Flipse et al., 2016).

2.1.5 Dengue Fever & Dengue Treatment

Antiviral treatments for dengue infection still do not exist. Supportive care and hydration are the only treatments available for DENV infected patients especially for those who developed hypertension with dengue hemorrhagic fever (DHF). Infants and young children are often asymptotically infected with DENV with assorted clinical syndromes. They often present with an undifferentiated febrile illness accompanied by a maculopapular rash. For adults when infected, they are more likely to become sick with 80% of adults developing clinical symptoms of the disease. Dengue fever is commonly identified by an acute abrupt onset saddleback fever, serious headache, nausea and vomiting, myalgias, retro-orbital pain, an early maculopapular rash, low grade thrombocytopenia and hepatomegaly. Typically, patients will recover in two to seven days (John & Rathore, 2019; Soe et al., 2018).

Dengue patients may exhibit signs of dehydration due to fever, vomiting, or diarrhea and lost fluids should be replaced with oral electrolyte replacement solutions. The high dengue fever can be treated with Acetaminophen (Tylenol) for pain relief. Patients are encouraged to seek immediate medical attention if they develop symptom of serious dehydration or shock, where hospitalization and admission to the intensive care unit is highly endorsed. During hospitalization, doctors constantly monitor the patient’s blood pressure, hematocrit, platelet count, urinary output and mental status. Patients may also be transfused if they lose excessive amounts of blood due to hemorrhage. Admission into the ward is suggested for at least twenty-four hours following defervescence and patients should not be sent home until they meet the

following criteria: their appetite return, have clinical improvement, there is no respiratory distress, stable hematocrit, platelets greater than 50,000/mm³ and good urine output (Hacker, 2009).

In the late 2015 and early 2016, the first dengue vaccine, Dengvaxia by Sanofi Pasteur, was registered in several countries for the usage of individual 9-45 years of age. Outcomes from phase III clinical trials showed that the vaccine had successfully reduced dengue hospitalizations by 80%. Unfortunately, the average efficacy against DENV was low, especially against DENV1 at approximately 50% and against DENV2 at 39%. On the other hand, the average efficacy against DENV3 and DENV4 was slightly higher at 75% and 77%, respectively. In fact, clinical trials had confirmed that this vaccine caused elevated risks of hospitalization for children below nine years of age. The World Health Organization has therefore recommended the use of this vaccine only in endemic areas where high dengue cases had been reported (Fauci et al., 2019).

2.1.6 Dengue Virus Envelope (E) Protein as Target for Viral Attachment

Nowadays, the DENV entry step into host cell has become an interesting therapeutic strategy because it correlates to a barrier to disturb the beginning of the infection. Inhibition of DENV entry during the beginning stage of virus infection may weaken the viremia in infected human resulting in the interruption of dengue fever to the serious life threatening infection, reduce the infected vector number and finally breaking the transmission cycle. The envelope (E) protein which has 495 amino acids was reported to play an important role in viral infection through DENV attachment to the host cell receptors as it binds to receptors. It is also one of the structural proteins of DENV which is known to be a target of antiviral inhibitors and plays a special role in the cell membrane fusion process (Dubayle et al., 2015). A study by Zhang et al. (2004) revealed the high flexibility of E-protein as it undergoes considerable conformational

adjustment during maturation. The flexibility of E-protein is crucial for the maturation and fusion processes. Therefore, it is one of the most valuable candidate proteins for the development of DENV drugs.

The region of the E-protein that might bind to cell-specific receptors remain a big issue due to the virion surface glycoprotein containing no well-defined spike structures common to other enveloped viruses. With the development of bioinformatics, Ilyas and co-workers (2011) have developed some tools and methods to study the rules in the variation of E-protein based on the fact that, there must be some conservative regions to maintain its function and structure. Moreover by analysis of huge number of sequences of E-protein, it is possible to design new drugs for the prevention of DENV infection.

Structurally, the most attractive region of the E-protein to study is the domain III (DIII), which forms the virus surface and is involved in receptor binding with the host cells and fusion process. Moreover, DIII plays an important role both in spreading and inhibiting viral infection as well as stabilizing the E-protein structure (Elahi et al., 2014; Munoz et al., 2013). DIII is the main target of neutralizing antibodies against DENV (Valdes et al., 2011) and because it is relatively small, it is ideal for NMR and computational studies. Mouse monoclonal antibody (MAb) studies (Fibriansah et al., 2015) showed that the most potent antibodies bind to DIII and the most efficient inhibition on the viral attachment was observed when monoclonal antibody (MAb) bind with the epitopes on DIII (Chong et al., 2015). The study of potential inhibitor of DENV E-protein based on DIII was also done previously by Guzman and co-workers (2010). Additionally, Alhoot et al. (2013) successfully reported the inhibitory peptide that inhibited DENV entry by targeting the E-protein. This was in agreement with the study by Guardia and Leonart (2014) that reported the existence of small molecules that interfered with the DENV entry process mediated by class II fusion proteins.

2.2 *In Silico* Study

Computational drug discovery can accelerate the challenging process in designing and optimizing a new drug candidate. The impact of computational structure-based drug design on drug discovery has increased in the previous years due to the rapid development of faster architectures and improved algorithms for high-level computations in a time-affordable manner (Vivo et al., 2016). Receptor and ligand flexibilities are crucial in order to correctly predict ligand binding and related thermodynamic and kinetic properties. Therefore, molecular dynamics simulation and other computational techniques become common computational tools for drug discovery. A method famously known as computer aided drug design (CADD) has played an important role in the development of therapeutically important small molecules for over three decades (Sliwoski, Kothiwale, Meiler, & Lowe, 2014). This method can be divided into two which are structure-based and ligand-based drug designs. Structure-based method is related to high throughput screening in that both target and ligand structure information is necessary. This approach include ligands docking, pharmacophore and ligand design method. Meanwhile, ligand-based method uses only ligand information for predicting activity depending on its similarity or dissimilarity to previously known active ligands. In this study, the approach applied is more on structure-based drug design.

2.2.1 Automated Docking

Computational approach can be used for the categorizing of small molecule inhibitors against DENV replication through docking technique. Normally, molecular docking method is used to understand the drug-receptor interactions in modern drug design. This method rank molecules based on their binding affinities and further enhance the molecules to improve binding characteristics. There are several softwares

available for the docking process such as FlexX, Autodock and Gemdock. The aim of docking is to predict the structure of the complex formed between the target protein and the ligand. Once the protein model is built, ligand-docking algorithms predict the ligand protein interactions through searching for the best steric and energetically favourable fit. This involves the complex with the minimum binding free energy for a ligand bound to the receptor binding site. Softwares such as Autodock 4.2 can be used to dock a ligand into the receptor to identify the active binding sites and to study the interactions such as electrostatic interactions, hydrogen bonds, hydrophobic interactions, hydrophilic interactions and Van der Waal's interactions. The general idea of a docking program is to position the ligand in possible binding modes in the protein active site and eventually calculate score for the protein-ligand complex. A successful docking protocol depends on both the ability to accurately predict the binding pose and estimate their binding affinity. In AutoDock 4.2, the protein is preserved rigid while the ligand is allowed full flexibility in most of the docking methods (Peng, 2015).

Recently, studies that combine both experimental data and computational methods have increased dramatically because the techniques are very crucial in drug design. Identification and optimization of the best ligand on the structures of biomolecules are common scientific challenges. Docking studies enable researchers to determine the best position for a ligand to bind on a macromolecule, while molecular dynamics (MD) simulation describe the relevant interactions that retain this binding. MD simulation normally represent the macromolecule movements in more detail. In the case of a protein, the side chain, backbone and domain movements can show how ligands are trapped during varieties of conformational states. Besides that, it is likely to distinguish every binding site that would be able to accommodate different ligands through atomic movement. Furthermore, MD is used to examine the motion of the key catalytic residues side chains, which could give information about the formation of

protein transition states. All this information can be used in order to suggest the most possible site of binding interactions.

Basically, MD simulation method can be combined with docking method to predict possible protein-ligand complexes. The particularities of both techniques are complementary: the rigidity and driven strategy of some docking methods and the force-field-dependent flexibility of MD simulation can be combined to achieve the target. The location of a ligand inside a binding site is predicted by a docking calculation thereby yielding the energy-dependent location and conformation. Once the ligand is in the most probable site, the MD simulation models the atom motion involved in the interaction.

2.2.2 Molecular Dynamics Simulation Studies on DENV E-Protein Systems

Chemistry has been generally known to be an experimental science where no molecule could be investigated without being synthesized or found in nature. Nowadays with the advent of modern computers and advances in simulation techniques, the dynamics and structure of molecular systems can be studied theoretically without them being synthesized nor have to be naturally occurring. The development of a method that allows the relaxation dynamics and equilibrium fluctuations of protein systems sampled by molecular dynamics to be tested via a direct comparison with experiment provides a powerful tools in the validation of MD simulations (Ahmadi, 2013). Compatibility between the simulation results and the experimental data has encouraged scientists to apply molecular dynamics (MD) simulation approach widely.

The molecular dynamics (MD) method is the most typical method used for *in silico* studies of molecular motion and flexibility at the atomic level. Static models only give little information about the dynamics of the structure, thus structure and dynamics have been combined in order to give a good understanding of the receptor system (Peng,

2015). In reality, the molecules undergo thermal motions and constantly transform their geometry, and MD simulations aim towards providing a more realistic representation of this molecular behaviour (Heavner, 2004). The application of computational tools to biomolecular analysis has jumped recently due to advances in the quality of both software and hardware (Jambrina & Aldegunde, 2016). There is no single experimental technique that is able to effectively depict the dynamic structure of DNA, however, MD simulations can cater an exhaustive theoretical description. The knowledge of protein dynamics is crucial in order to justify protein folding, misfolding, molecular function, aggregation, signal transduction and allostery as well as applications in drug design and protein engineering (Katagi, 2013; Schaller, Connors, Oelmeier, Hubbuch, & Middelberg, 2015).

Recently, a number of simulations have been carried out in order to explore various effects in the protein-ligand complex binding free energy. Results from molecular dynamics simulations were found to be more reliable than results from molecular docking. Dubey and co-workers (2014) had proven via MD simulations that the binding free energy of drug R1 is better than the other drugs, and this pattern corroborated with the experimental observations. MD simulations has also become a prominent approach to generate an ensemble of receptor structures for docking propose (Wong, 2008). On the other hand, Lin et al. (2002) performed MD simulation on the ligand-free protein and then docked several ligands to the snapshot structures obtained from the MD simulation. The MD simulation was performed using AMBER software while docking was performed using Autodock software. The ligands were docked to 5000 snapshot structures from the first 50 ns segment of MD simulation, and it was found out that some of these dynamics snapshots were capable to bind the ligand and gave poses closest to the experimental structures. Parikesit et al. (2013) used molecular docking to predict the formation of peptide bonds between the inhibitor with DENV E-

protein cavity, followed by MD simulation to analyse the interactions between the protein and cyclic peptides at different temperature and time. In MD simulation, the starting structure may affect the simulation time to achieve the equilibrium. The closer the initial structure to the experimental one, the faster the simulation will reach the equilibrium state that agrees with the real system (Ahmadi, 2013).

2.2.3 Free Energy of Binding

Calculations of relative binding free energies have been attempted for carbohydrate protein complexes using the free energy perturbation method. Molecular Mechanics Generalized Born Surface Area (MMGBSA) method is used as a tool in the analysis of carbohydrates in aqueous solution and dynamically bound to protein receptors. This method has successfully ranked the binding affinity of a series of carbohydrate-protein complexes. Nevertheless, the Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) approach is a more accurate simulation and permits the comparison of quite disparate ligands, such as the mono-, tri- and pentasaccharide (Bryce, Hillier, & Naismith, 2001).

Normally MD simulations were used to generate an ensemble of binding conformations in the existence of explicit water, and ultimately the MMPBSA approach was used to estimate the binding energy. This method has been applied to estimate binding free energies and to figure out the relative stabilities of various biomolecular structures. Furthermore, this method has been used to understand biomolecular associations in detail by decomposing the total binding energy into a series of components. To decompose the binding energy, at first ΔE_{MM} , ΔG_{polar} , and $\Delta G_{nonpolar}$ are individually calculated for each residue and were then summed up to get the contribution of each residue to the binding energy. However, by extending the

simulation time, it does not necessary improve the correlations between the predicted binding free energies and the experimental values (Hou, Wang, Li, & Wang, 2011).

MMPBSA has also been combined with MD simulations to rescore a set of docked complexes, significantly improving the utility of the complexes identified. In this case, MD simulations were used to produce an ensemble of binding conformations in the existence of explicit water, and additionally, the MMPBSA approach was used to estimate the binding energy. This tool is convenient in order to count relative binding energies, such as to compare various ligands binding into the same receptor protein. However, MMPBSA rank the binding affinities of the candidate molecules rather than give accurate predictions of the absolute binding free energies (Hou et al., 2011). On the other hand, the calculation of the entropic contribution to the binding free energy is another challenging problem where it is time-consuming, and the magnitude of standard error is high compared to the other energetic terms. The net entropic contribution is usually small, and multiple studies have suggested that corrections for changes in the configurational free energy of the system lead to only a small improvement in the correlation with experiment (Kumari, Kumar, Consortium, & Lynn, 2014b).

To gain further insight into the contribution of individual residues to binding, free energy decomposition was performed by Peng (2015). In his study, it was reported eventhough acetylcholine (ACh) bound to the same binding site in all the muscarinic receptors, it seemed to flexibly move in the binding site and make contacts with the key residues with different strengths. Favourable contributions to the binding begin from vdW interactions and the non-polar part of the solvation free energy, as opposed to unfavourable total electrostatic contributions. Therefore, the total electrostatic contributions became the major reason for differences in the binding free energy.

In MMPBSA or MMGBSA, binding free energy (ΔG_{bind}) between a ligand (L) and a receptor (R) to form a complex RL is calculated as follows:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \approx \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S \quad (2.1)$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{internal}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdW}} \quad (2.2)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}} \quad (2.3)$$

where ΔE_{MM} , is the change of the gas phase MM energy, ΔG_{sol} is the solvation free energy and $-T\Delta S$ is the conformational entropy upon binding. ΔE_{MM} includes $\Delta E_{\text{internal}}$ (bond, angle and dihedral energies), electrostatic $\Delta E_{\text{electrostatic}}$ and van der Waals energies ΔE_{vdW} . ΔG_{sol} is the sum of electrostatic solvation energy (polar contribution), $\Delta G_{\text{PB/GB}}$, and non-electrostatic solvation component (non-polar contribution), ΔG_{SA} . The polar contribution is calculated using either GB or PB model, while the non-polar energy is estimated by solvent accessible surface area (SASA). The conformational entropy change ($-T\Delta S$) is usually computed by normal-mode analysis on a set of conformational snapshots taken from MD simulations (Hou et al., 2011).

2.2.4 Protein-Ligand Interactions

A detailed understanding of the protein–ligand interactions is essential to understand biology mechanism at the molecular level. Due to that reason, an in-depth understanding of the molecular interaction is very important in facilitating the discovery, design, and development of drugs.

2.2.4.1 Hydrogen Bonding Interactions

A hydrogen bond is the electrostatic attraction between polar molecules that exists when a hydrogen (H) atom is bound to a highly electronegative atom such as fluorine (F), oxygen (O) or nitrogen (N), which is named as the hydrogen bond acceptor

(HBA). When binding a ligand to its receptor, a hydrogen bond normally contributes between 0.5 and 4.7 kcal/mol to the binding energy. Ligand hydrogen-bond donors and acceptors are energetically favourable in solution, where they form hydrogen bonds with the surrounding water molecules, and ultimately form hydrogen bonds with receptor residues (Durrant & McCammon, 2011).

The main role of hydrogen bonds is to correctly position the ligand within the active site and then hold the protein active site in a ligand-friendly conformation. Hydrogen bond is also essential in determining the three-dimensional structures adopted by proteins and nucleic bases. Furthermore hydrogen bonds play a crucial role in the stability between subunits in multimeric proteins. These hydrogen bond attractions can occur between molecules (intermolecular) or within different parts of a single molecules (intramolecular). Basically, a hydrogen bond is stronger than a van der Waals interaction, and may exist in inorganic molecules such as water and in organic molecules such as DNA and proteins.

Peng (2015) in her hydrogen bond analyses on the MD trajectories highlighted that transmembrane (TM), TM1 and TM2 underwent much less movement than the other TMs. These regions were likely to be stabilized by a network of hydrogen bonding interactions. The hydrogen bond networks are crucial for molecular interactions preserving the individual helical structure and the overall architecture of the TM bundle in the ground and activated states. Upon ligand binding, hydrogen bond networks may be either strengthened, resulting in stabilization of the receptor conformation, or disrupted, facilitating receptor activation. In another study, Ahmadi (2013) suggested that one of the main characteristic of sugars was the ability to participate in hydrogen bonds as both acceptor and donor. Each sugar has hydrophilic and hydrophobic region, and the detailed interaction of the hydrophilic part is under the control of the carbohydrate head group which could involve in hydrogen bonding.

In a study by Deol et al. (2004), a detail picture of the interactions between lipid head groups and protein could also be achieved by analysing H-bonds. The monomer system exhibited more hydrogen bonding interactions with lipids in comparison of two monomers present in dimer-membrane system. This would be expected as monomer alone has larger surface area to interact with lipid molecules, whereas monomers in a dimer system (where one side interact with each other) only had the remaining one side to interact with lipid molecules. Moreover, higher amount of transient H-bonds (occupancy < 20%) occurred between donors and acceptors which could probably due to the thermal "breathing" motion of lipid and protein molecules.

Padariya et al. (2015) also observed the existence of hydrophilic polar side chains (Gln, Thr, Tyr and Lys) and non-polar (Gly) residues made very stable hydrogen bonds with lipids. The dynamic analysis of H-bond interactions of every residue in the protein revealed an important role of aromatic residues in drug binding, given their ability to stack with aromatic compounds and get involved in electrostatic interactions with charged compounds.

2.2.4.2 Van der Waals Interactions (vdW)

By definition, van der Waals (vdW) forces are driven by induced electrical interactions between two or more atoms or molecules that are very close to each other. Unfortunately, vdW interaction is the weakest of all intermolecular attractions between molecules. The role of vdW is very important, as its presence would enable the ranking of the ligands in the docking process, as agreed by Godoi et al. (2017) in their study on NS2B-NS3pro inhibitors. In fact, they reported that compound coded 12m demonstrated vdW interactions with Ser17 and Ser131, and π -interaction with His51 present in the catalytic cavity, indicating an important relationship to the enzyme inhibitory mechanism.

Meanwhile, an inspection of the thermodynamical parameters shows that vdW interactions played the most important role for binding affinity and the highest for ligand 5-(3-chlorophenyl)-N-(2-phenyl-2H-benzo[d][1,2,3]triazol)furan-2-carboxamide (R1), as reported by Dubey et al. (2017) in their study targeting domain III of dengue envelope protein. The unfavourable solvation energy and favourable vdW interactions of ligand R1 was expected due to high lipophilicity of the ligand which made it highly hydrophobic.

In the interaction of ligand with DENV E-protein, as reported by Lavanya et al. (2015), which considered the energetic contribution through vdW interactions, the compound Gedunin showed the highest score. However, the number of hydrogen bonds formed between Gedunin and the active site residues were less compared to other ligands studied which were Nimbin and Azadirone. Meanwhile, in the work done by Nasution et al. (2017), it was claimed that not only hydrogen bond interaction could be formed but also other non-covalent interaction, such as vdW interaction, resulting in the binding affinity between cyclic peptides and the NS2B-NS3 protease.

2.3 Inhibitors Targeting Envelope Proteins

2.3.1 Carbohydrate as a Small Molecule Inhibitor

Many important biological processes involve carbohydrate-protein interactions. Carbohydrates are generally viewed as highly polar molecules. Carbohydrates are characterized by their considerable conformational flexibility and dense polar functionality. Since carbohydrates are dynamic in nature and interact intimately with aqueous solvent, the prediction of a protein-carbohydrate complex is crucial in numerous aspects including antibody-antigen recognition, gene expression, cell-cell adhesion, enzyme-substrate specificity and in molecular transportation (Bryce et al.,

2001; Nivedha, Makeneni, Foley, Tessier, & Woods, 2014; Schmidt et al., 2012). A physical understanding of carbohydrate-protein interactions aids in the development of therapeutic agents designed to block such interactions.

The strength of this kind of interaction is also determined by the carbohydrate conformation and orientation with respect to the binding site. In order to probe protein-carbohydrate interactions, docking and molecular dynamics (MD) simulations techniques could be used. Automated molecular docking can be utilized to identify binding sites in carbohydrate-binding proteins, or to predict the bound orientation and conformation of a carbohydrate or drug mimic in a known binding site. Once a docked model is obtained, it can then be used as a starting point for further refinement and exploration using MD simulations. Molecular dynamics (MD) simulations are employed to refine docked models and to study the dynamic nature of binding.

A well-known characteristic of protein-carbohydrate interactions is the low affinity of binding, usually in the millimolar range. These interactions are driven by a favourable enthalpy offset by the multiple contact points (hydrogen bonds, van der Waals interactions and hydrophobic stacking) between the carbohydrate and the protein. Moreover, highly organized hydrogen-binding atoms and a higher frequency of hydrogen bonds per unit area result in a closely spaced protein-carbohydrate interface. In addition, it is not surprising that water plays an essential role in both protein-protein and protein-carbohydrate interactions as the human body has a water content of 70%. Water molecules are also specifically involved in the binding of proteins to carbohydrates by mediating ligation of the carbohydrate epitope to the binding cleft of the protein. Upon protein-carbohydrate complexation, water molecules will tend to escape to the bulk with a concomitant decrease or increase in energy depending on their pre-existing molecular interactions. After protein-carbohydrate binding has occurred,

the complex formed will be resoluted and the water molecules will arrange themselves according to the new surface exposed (Holgersson, Gustafsson, & Breimer, 2005).

DENV2 has also been shown to interact with a glycosphingolipid called neolactotetraosylceramide (nLc₄Cer) that is expressed on mammalian cells such as human erythroleukemia K562 cells and baby hamster kidney BHK-21 cells. The carbohydrate recognition domain (CRD) of DC-SIGN has been shown to interact with DENV by engaging carbohydrate moieties located on N67 of two neighbouring E-proteins (Gregory Donald Gromowski, 2008). All serotypes of DEN viruses, DENV1 to DENV4, reacted with nLc₄Cer, and non-reducing terminal disaccharide residue Galβ1-4GlcNAcβ1- was found to be a critical determinant for the binding of DENV2. These findings strongly suggested that multivalent nLc₄ oligosaccharide could act as competitive inhibitor against the binding of DENV2 to the host cells (Aoki et al., 2006).

Complex carbohydrates normally have a large number of rotatable bonds and as a consequent a large number of theoretically possible conformations can be generated. However, with the increase in computer processing power recently, MD simulations have become the standard method to study the conformations of carbohydrates (Frank, 2015). A major advantage of MD methods is that carbohydrates can be studied in explicit solvent as well as in their biological context such as glycoproteins, glycolipids, or protein-carbohydrate complex.

2.3.2 Peptide

2.3.2.1 Peptide as an Inhibitor

Peptide development as therapeutic drug against DENV is a promising field in drug discovery. Peptide consists of amino acid sequence which is derived from the small part of a pathogenic protein, and due to that reason, they are very vulnerable to hydrolysis and oxidation reaction. Generally, a peptide has unique pharmacological properties due to its higher bioactivity, high specificity and selective to their target, with low interaction and toxicity compared to other drugs, as well as low accumulation in tissues (Chew et al., 2017). Furthermore, the peptide-based drugs have been widely distributed in the market, and approved by the Food and Drug Administration (FDA), such as Buserelin and Leuprorelin (Usmani et al., 2017). Peptide drugs can also disrupt protein-protein interfaces that cannot be inhibited by the available small molecules or even act as allosteric modulators (Yang et al., 2015). Parikesit et al. (2013) in their research had screened 300 commercial cyclic peptides against two binding sites of NS5 methyltransferase in order to develop new hit inhibitors against DENV. Analysis of ligand-enzyme binding free energy found that, there were two best peptides that maintained stable complex conformations throughout the MD simulations. These two commercial cyclic peptides were suggested to be potential candidates to be developed into antiviral agents against DENV.

Alhoot et al. (2013) had designed four putative antiviral peptides to target the domain III (DIII) of DENV2 E-protein using BioMoDroid algorithm. This study found out that they were two peptides showing compelling DENV inhibition, when simultaneously incubated in the plaque formation assay, as well as in the reverse transcription polymerase chain reaction (RT-qPCR) and Western blot analysis. Furthermore, the transmission electron microscopy (TEM) images showed that the

inhibitory peptides caused structural abnormalities and alteration of the arrangement of the viral E-protein, which interfered with virus binding and entry. These peptides would have the potential to be active against all the serotypes of dengue and could be a promising therapeutic agent for attenuating dengue infection.

Peptides become novel recognition agents in bioassays as they can be easily synthesized chemically or in bacterial cells by recombination. They retain their binding affinity in various environmental conditions and can be simply adjusted to be used in different bioassays. Peptides binding selectively to the target molecules are mostly selected from peptide libraries that consist of a combination of a great number of various peptides (Schmidt, Lee, Yang, & Harrison, 2012). It is feasible to model more potent peptide ligands by understanding the structure-activity relationship of the current peptides. A good model peptide can precisely demonstrate greater selectivity and specificity with lower toxicity, compared to other conventional drug-like molecules. However, the use of peptides can be affected by deficiency such as low bioavailability, metabolic degradation and low conformational stability. To overcome these limitations, non-proteinogenic amino acids may be added in the sequence to produce peptides that are naturally stable to proteases and peptidases and eventually fold into well-ordered secondary structures (Maffucci, Pellegrino, Clayden, & Contini, 2014).

2.3.2.2 Peptide Folding in Peptide Inhibitor Design

The understanding of peptide folding is very crucial in order to understand how a protein folds. By knowing the 3D structure or the folding of the peptide, it would allow us to understand the functionality and how it interact with other molecules. This would lead in the designing of peptides with potential biotechnological or pharmaceutical purpose (Daura, Mark, & Gunsteren, 1999). To determine the 3D-structure of a peptide through experiment is not an easy task. In principal, the process of

peptide folding could be simulated directly on a computer using MD simulations method (Daura, Jaun, Seebach, Gunsteren, & Mark, 1998; Matthes & Groot, 2009). This way specific atomistic information can be obtained, which may be tough to get through experimental techniques. One of the major challenges in peptide folding simulations is to choose a correct force field due to possible biases different force fields have toward certain types of secondary structure (Cino, Choy, & Karttunen, 2012).

Understanding the dynamics and mechanism of peptide folding maintains to be one of the most challenging problems in molecular biology. Peptide folding simulations and experiments characterize the dynamics and molecular mechanisms in the early events of protein folding. In general, this type of simulations can be very long up to microseconds to stand a good opportunity to observe a single folding event and the force field being used must correctly depict the relative energies of a wide array of unfolded or misfolded conformations that appear during the folding process. Experimentally, peptide folds at very fast rates, requiring probing on the nanoseconds time resolution (Gnanakaran, Nymeyer, Portman, Sanbonmatsu, & Garcia, 2003). In addition, MD simulations also provide accurate information on the structure nature and relationships that takes place during peptide folding processes, as well as identify key intermediates and barrier to folding. Long simulations can be especially useful when performed near the peptide melting temperature, where the folded and unfolded states are equally populated and folding and unfolding occur on the same timescale (Piana, Klepeis, & Shaw, 2014).

Normally, the short peptides and protein fragments become the ideal model systems for the investigation because of their smaller size and structural simplicity. Of particular interest are α -helices and β -hairpins, which are two essential secondary structures in most of the proteins. Experiments display that the formation of β -hairpins, is generally more difficult compared to the formation of α -helices. In addition of its

structural simplicity, the β -hairpin is believed to fold in a manner which is similar to the folding of small proteins (Shao, Yang, & Gao, 2009). All this knowledge opens a wide range of possibilities to use MD simulations in order to understand the process of peptide folding and predicting possible folds of peptides in solution.

2.3.3 Small Molecule Inhibitors

This section briefly review on small molecule inhibitors that inhibit the viral attachment and membrane fusion of the DENV E-protein. There are no specific treatments available and development of the antiviral drugs for DENV is challenging due to the immunization and long lasting protection against all four serotypes. A study reported by Chao et al. (2018) had screen few compounds to detect small molecules that interfered in the conformational transition which at the end resulted in a series of cyanohydrazone compounds that bound to a soluble DI-DII fragment. *In vitro* result showed that these compounds blocked fusion, and infection in cell culture by attach to the E-protein conformation on the virion surface before the virus attached to a cell.

On the other hand, a series of compounds studied by Aarthy et al. (2018), which were quercetin, silymarin, dapagliflozin and fisetin, could be potential candidates to inhibit the DENV E-protein. Docking studies revealed that these compounds possessed strong interaction with good binding energies. In addition, molecular dynamics simulations also revealed that these compounds were highly stable as well as no weak interactions were observed between complexes.

Meanwhile, a study performed by Tambunan et al. (2016) had screened about 1,320 designed ligands which resulted in 3 best ligands that could form interaction with target protein and fusion peptide. These ligands showed good affinity with DENV E-protein based on free energy of binding values and hydrogen bond interactions. In

addition, natural product compounds were also used in the form of fragments could be potential inhibitor for the DENV E-protein (Tambunan and Alkaff, 2018). The ligands had a lower $\Delta G_{\text{binding}}$ values and better molecular interaction with the DENV E-protein β -OG pocket binder compared to the native substrate, β -OG. The result indicated that the fragment-based drug design employed in the study could be an important computer-aided drug design and discovery method in developing a new drug for various diseases.

Basically, the advancement of small molecules for DENV antiviral drugs, has been a passive process. At the moment, only four small molecule antiviral drugs, which are Chloroquine, Celgosivir, Balapiravir and UV-4B, have entered phase 1 or phase II clinical trials (Tian, Zhou, Takagi, Kameoka, & Kawashita, 2018). Pre-clinical and clinical trials on antiviral drugs development are still ongoing, and many more research need to be performed in order to improve the current study status. Hopefully in the future, many challenges towards finding effective drug candidates could be overcome, and the current research work would be able to yield a powerful and effective DENV antiviral therapy.

CHAPTER 3: STUDIES INVOLVING NEOLACTOTETRAOSYLCERAMIDE

(nLc₄Cer)

3.1 Introduction

The DENV E-protein is responsible for receptor recognition and attachment to the host cell. The E-protein comprises three domains and the receptor binding Domain III (DIII) has been pursued as drug targets. In general, DIII region in E-protein has appeared to be responsible for the initial contact and accumulation of DENV on the surface of host cells by binding to glycosaminoglycan receptors (Behnam, Nitsche, Boldescu, & Klein, 2016). Therefore, molecular dynamics (MD) simulation and docking methods were used to identify possible binding sites in DIII region. These methods allowed us to explore the various conformational states and protein flexibility and eventually gets the binding energy towards these multiple structures. In this study, the E-protein pre-fusion structure was analysed to identify the potential binding site that may bind neolactotetraosylceramide (nLc₄Cer), which eventually may interfere with the conformational transitions that mediate the attachment process. Previous study (Hidari & Suzuki, 2011) reported that nLc₄Cer was responsible for virus adsorption and virus attachment to the host cells. In mammalian cells, nLc₄Cer is a DENV receptor and has the non-reducing terminal disaccharide residue Galβ1-4GlcNAc of nLc₄Cer which is a critical determinant for the binding of DENV2. Therefore, MD simulation was performed to generate multiple conformations followed by docking to get best conformation, and eventually E-protein-nLc₄Cer complex was simulated in explicit solvent to understand the interactions involved in the complex.

Meanwhile, deep understanding of the association process is compulsory in order to design drugs that are capable to inhibit with disease-related protein interactions. Unfortunately, drug binding is a process that can be difficult to observe via experiment

and difficult to analyse using computational techniques. Until now, a consensus for the ligand entrance is still not clear with a growing interest among experimental researchers that computational simulations always show multiple pathways. In fact, modelling of ligand entrance from conventional molecular dynamics techniques has shown to be tough and costly (Huang & Wong, 2007; Le, 2012). Most computational models consider only the final conformation of the final docked protein-ligand complex and do not consider the dynamics of the ligands as it enters the protein binding site. In this study, a cost-effective method called “pathway docking” was used to reveal possible paths of the ligand approaching the protein binding pocket. This work demonstrates the power of pathway docking to show full binding pathway on how a drug find its target binding site. A docking approach that directly reveals the ensemble of pathways of nLc₄Cer to the binding pocket of dengue virus (DENV) envelope (E) protein is described.

3.2 Methods

3.2.1 Characterizing Ligand Binding Site In Dengue Virus Type 2 (DENV2) Envelope Protein

3.2.1.1 Materials

The three-dimensional structure of DENV E-protein was retrieved from the Protein Data Bank (<http://www.rcsb.org/pdb>; PDBid: 1OKE) (Berman et al., 2000). Chlorine atoms, water and glycerol molecules were removed. The structure of neolactotetraosylceramide (nLc₄Cer) was downloaded from PubChem page (<http://www.ncbi.nlm.nih.gov/pccompound>) (Wang, Xiao, Suzek, Zhang, Wong, et al., 2009) and was minimized using Hyperchem Pro 6.0 software (Young, 2004) with PM3 parameters using the steepest descent and conjugate gradient methods (termination conditions were set to a maximum of 500 cycles or 0.1 kcal/Å mol rms gradient). This

energy minimization guaranteed a low energy conformation with suitable bond lengths and angles.

3.2.1.2 Molecular Dynamics (MD) Simulation of DENV Envelope (E) Protein

MD simulation of DENV E-protein complex was run in explicit solvent with a total simulation time of 50 nanoseconds using AMBER software (Salomon-Ferrer, Case, & Walker, 2012). It is important to run a simulation long enough to allow the system to properly equilibrate. Explicit solvent was chosen as opposed to *in-vacuo*, so that the simulation can imitate the system as close to reality as possible. Generally, the results based on explicit solvent are closer to the experimental data than those based on implicit solvent simulations. Water was used as the solvent because most proteins operate in an aqueous environment and most *in vitro* studies are also performed under this condition. The SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977a) was used to constrain bonds containing hydrogen atoms. Simulation was initiated by heating the system from 0 K up to the final temperature of 310 K. The system was subsequently solvated with TIP3P water molecules in a truncated octahedral simulation box extending up to 12 Å from the solute in each direction. In this case, periodic boundary was applied to the system to obtain consistent behaviour. To ensure the overall neutrality of the system, Na⁺ and Cl⁻ ions were added. All minimization and MD simulations were conducted using the Particle Mesh Ewald Molecular Dynamics (PMEMD) program from AMBER 12 package (Masova & Kollman, 2000; Salomon-Ferrer et al., 2012). The system was first subjected to 500 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization. Minimization was performed in order to avoid any bad contacts prior to MD runs. To equilibrate the explicit solvent system, 200 picoseconds MD simulations were performed at a pressure of 1 atm using the Berendsen weak-coupling algorithm (Simmerling, Strockbine, &

Roitberg, 2002). A time step of 2 femtoseconds was used to integrate the equations of motions.

3.2.1.3 Clustering

The representative structures during the 100 ns simulation of apo E-protein were generated using the cluster option in AMBER 12 software. The obtained conformational MD trajectories were clustered into nine conformations in order to focus the protein conformational analysis on the binding site. The structures were clustered based on the similar conformations using the backbone atom-positional RMSD of all atoms. Cluster program from ptraj was used to get a representative and average structure of the MD run. Several snapshots were extracted from the trajectory to represent various conformations for DENV E-protein. This was done by clustering the trajectory using backbone dihedral and selecting the conformations closest to the centre of each cluster as a representative conformation. The hierarchical algorithm was used for clustering (J. Shao, Tanner, Thompson, & Cheatham, 2007). Theoretically, different conformations will give different binding site that eventually lead to variety of binding energies. MD simulation was performed on DENV E-protein to generate an ensemble of monomeric conformations to be used as docking targets. Thus, the ability of a ligand in changing the protein conformation may be effective in blocking the DENV from successfully attacking the human cells.

3.2.1.4 Binding Site Prediction

Protein binding sites are the regions where molecular interactions occur with ligand. Binding sites may change their sizes and shapes upon binding. Therefore, analysis of protein binding sites is highly crucial in order to understand the biological

processes that are involved following binding with ligands. Binding pockets were detected using Discovery Studio 4.0. The binding site was detected by using the “find sites from receptor cavities” in the Tools section under the “Define and Edit Binding Sites” in Discovery Studio Program. Binding site uses a CHARMM-based molecular dynamic scheme to seek for the optimal binding sites for docking (Yi et al., 2015). Based on the analysis of the geometry shape of the protein surface, the cavity which could bind to the substrate was identified, and the potential binding sites were predicted. Then, the optimal binding site was chosen based on the shape and location of the cavity, the location of the residue and the conserved amino acid. A site sphere radius was set to assign the entire binding pocket. Other parameters were set as default. By using this program, the binding sites and volume of each pocket and cavity of E-protein were analysed.

3.2.1.5 Docking of nLc₄Cer to DENV E-Protein

The E-protein structure used in the computational docking run was based on the selected snapshot extracted from clustered MD trajectories. Docking of nLc₄Cer to the DENV E-protein was performed using AutoDock 4.2 (Morris et al., 2010; Ruba, Arooj, & Naz, 2014). In AutoDock, all the missing side chain atoms of the target protein were checked and then repaired. Water molecules and ligand attached were removed. In order to use the AutoDock 4.2 force field properly, polar hydrogen atoms were added and nonpolar hydrogen atoms were merged, and solvation parameters were assigned by default using the graphical user interface of AutoDockTools (ADT). In this process, AutoDock 4.2 used the united atom model to represent molecules and the AutoDock scoring function was calibrated using Gasteiger partial charges on both the ligand and macromolecule (Zamri, Teruna, Rahmawati, Frimayanti, & Ikhtiarudin, 2019). Polar hydrogens are normally hydrogen atoms that are bonded to electronegative atoms such

as oxygen and nitrogen while non-polar hydrogens are hydrogens bonded to carbon atoms. All bonds were made rotatable and flexible by allowing the detection of root torsion. All docking calculations were performed using Lamarckian Genetic Algorithm (LGA) to determine the “globally” optimized conformation. The grid box, with spacing of 0.375 Å and a dimension of 128 x 100 x 116 points along the x,y and z axes, was centred on the molecule. A population size set up at 150; maximum number of energy evaluation, 2,500,000; maximum number of generations, 27 000; rate of gene mutation, 0.02; and cross rate, 0.8 were used for 100 search runs. Generally, several docking runs were required in order to identify the conformations of the ligand within binding site. The remaining parameters were set as default. The root mean square deviation (RMSD) tolerance for each docking was set at 4.0 Å. The best structure where the system had the lowest energy was chosen for further analysis.

3.2.1.6 Calculation of Free Energy of Binding

Spontaneous changes in a protein system, such as the binding of a ligand to receptor, are characterized by a decrease in free energy. The more negative the free energy, the stronger the binding occurs. By calculating free energy of binding, a correct description of the interaction between the protein and the ligand can be obtained (Donnini, 2007). In this study, the free energy of binding (ΔG_{bind}) of E-protein-nLc₄Cer complex was calculated based on Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) and Molecular Mechanics Generalized Born Surface Area (MMGBSA) procedures in AMBER 12 (Hou et al., 2011; Zhu, Beroza, & Artis, 2014). A total of 500 snapshots were collected for binding free energy analysis. The calculations, which were performed for each of these snapshots, were intended for the estimation of the free energy of binding of the protein-ligand complex (ΔG_{bind}) between

the nLc₄Cer and E-protein, using the following relationships (Kumari, Kumar, Consortium, & Lynn, 2014a) :

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \quad (3.1)$$

$$= E_{\text{MM}} + G_{\text{GB/PB}} + G_{\text{nonpolar}} - T\Delta S$$

$$E_{\text{MM}} = E_{\text{bonded}} + E_{\text{ele}} + E_{\text{vdW}} \quad (3.2)$$

where G_{complex} , G_{protein} and G_{ligand} are the free energies of the complex, protein and ligand, respectively; E_{MM} is the change of molecular mechanics potential energy upon ligand binding that includes van der Waals (E_{vdW}) and electrostatic (E_{ele}) energies (Equation 3.2); $G_{\text{GB/PB}}$ and G_{nonpolar} are the polar and nonpolar components of the desolvation free energy, respectively, and $-T\Delta S$ is the change of conformational entropy upon ligand binding (which was not considered in this study because of the high computational cost and it tends to have a large error margin leading to significant uncertainty in the result) (Chen, Zheng, & Zhang, 2015; Homeyer & Gohlke, 2012; McGee, Miller, & Swails, 2009). In general, binding is normally presumed to restrain the flexibility of receptor and resulting in increasing cost of conformational entropy. Presently, the conformational entropy seems to be the most difficult to calculate. The simulation calculation does not consider desolvation and the result considered to be inadequate. The large error, outlier trajectories and spurious correlations highlight that the sampling procedure still remains the weak point of entropy calculations (Grunberg, Nilges, & Leckner, 2006).

In order to apply the MMPBSA and MMGBSA formulations, a representative set of equilibrium conformations for the complex, free protein and free ligand were first obtained by atomistic MD simulations in explicit solvent. In the post-processing phase, the solvent was discarded and replaced by a dielectric continuum. Changes in the individual terms (E_{MM} , G_{sol} , $-T\Delta S$) between the unbound state and bound (complex) states were calculated, and contributed to the binding free energies.

These energy contributions were computed from the atomic coordinates of the protein, ligand, and complex using the (gas phase) molecular mechanics energy function (or force field). The solvation free energy term, G_{sol} , contains both polar and nonpolar contributions. The polar contributions are accounted for by the Generalized Born, Poisson-Boltzmann model and the non-polar are assumed proportional to the solvent-accessible surface area (SASA) (Equation 3.3).

$$G_{\text{sol}} = G_{\text{PB(GB)}} + G_{\text{SASA}} \quad (3.3)$$

3.2.1.7 Analysis of the Decomposition of Free Energy

Free energy decomposition for the E-protein-nLc₄Cer complex was examined in order to obtain information on important residues involved in the complex binding. The energy decomposition was carried out using the `mm_pbsa.pl` implemented in the AMBER12 package to calculate the per-residue decomposition. Per-residue decomposition separated the energy contribution of each residue from the association of receptor with the ligand into three terms: van der Waals contribution (ΔE_{vdW}), electrostatic contribution (ΔE_{ele}) and solvation contribution (Chu et al., 2015). The decomposition of free energy of E-protein-nLc₄Cer complex was calculated based on the MMPBSA and MMGBSA protocols. In this study, the decomposition of free energy of each system was calculated every 10 ns of the trajectories and 500 snapshots were extracted from simulation within this time range. The interaction energy profiles were generated by decomposing the total binding free energies into residue-residue interaction (free energies).

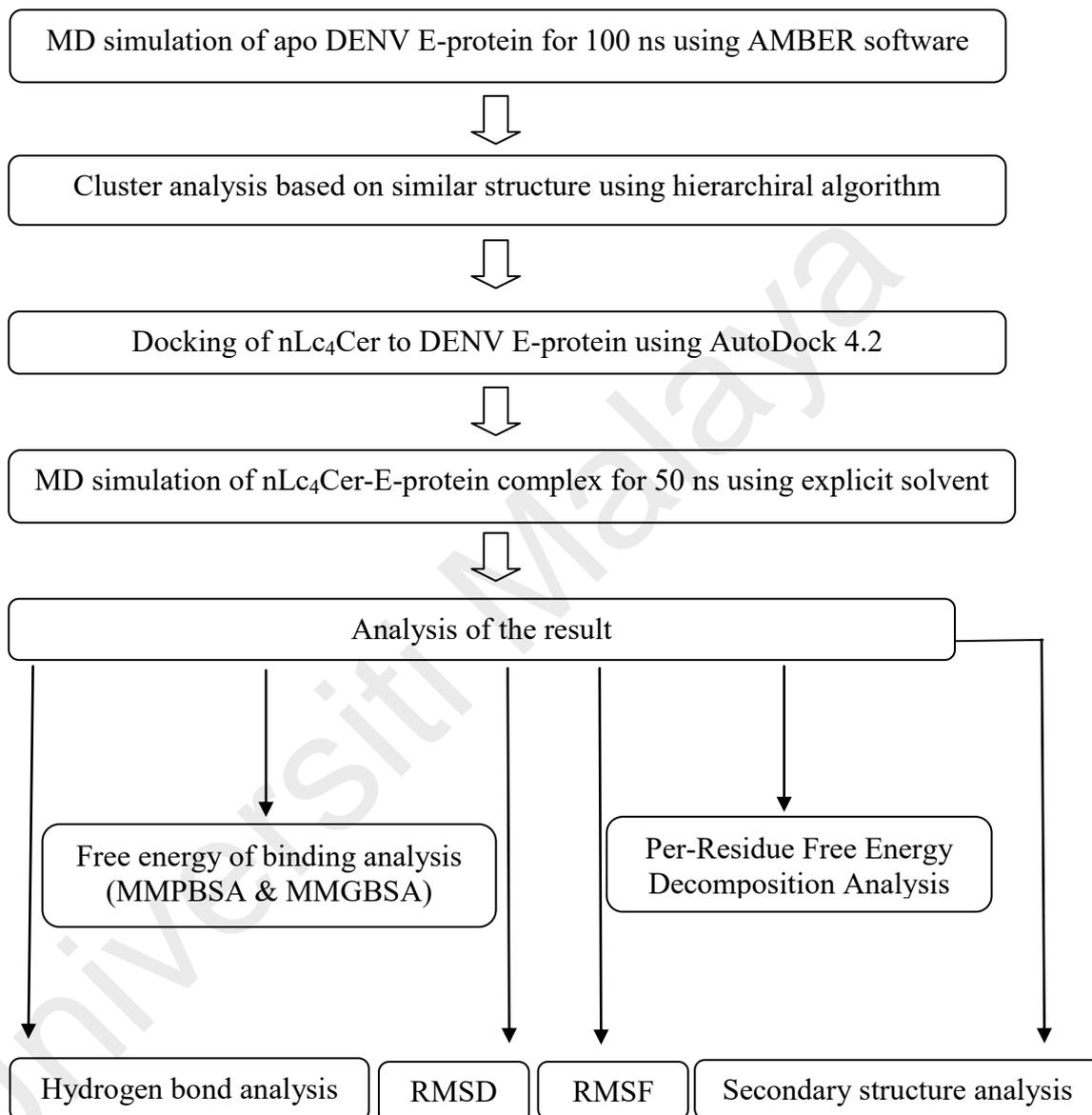
3.2.1.8 Analyses of Results

The AMBER package provides energy output in a text file, where energy is given as a function of time. The ptraj module of AMBER 12 was used to write the backbone RMSD of atomic positions relative to the starting structure. The convergence of the energies, temperature, pressure and root mean square deviation (RMSD) were checked to indicate the system stability. Selected conformations were analysed using Discovery Studio Visualizer (Isa et al., 2019) and Ligplot software (Wallace, Laskowski, & Thornton, 1995) which identified the hydrogen bond, hydrophobic and pi-pi interactions between DENV E-protein and nLc₄Cer. Meanwhile, Ligplot program was used to generate 2D schematic diagrams of protein-ligand interactions.

Universiti Malaysia

3.2.1.9 Flow Chart

The flow chart below shows method involved in this study towards finding potential inhibitors that can inhibit the formation of trimeric conformation of DENV E-protein.



3.2.2 Binding Pathway of Neolactotetraosylceramide towards Dengue Virus Type 2 (DENV2) Envelope Protein

3.2.2.1 Materials

The three-dimensional structure of DENV E-protein was retrieved from the Protein Data Bank (<http://www.rcsb.org/pdb>; accession code 1OKE) (Berman et al., 2000). Chlorine atoms, water and glycerol molecules were removed. The structure of neolactotetraosylceramide (nLc₄Cer) was downloaded from the PubChem page (<http://www.ncbi.nlm.nih.gov/pccompound>) (Wang, Xiao, Suzek, Zhang, Wang, et al., 2009) and was minimized using Hyperchem Pro 6.0 software with PM3 parameters using the steepest descent and conjugate gradient methods (termination conditions set to a maximum of 500 cycles or 0.1 kcal/Å mol rms gradient). This energy minimization guaranteed a low energy conformation with suitable bond lengths and angles.

3.2.2.2 Docking Pathway of nLc₄Cer towards DENV E-Protein

Docking of nLc₄Cer to the crystal structure of DENV2 E-protein was performed using AutoDock 4.2 (Morris et al., 2010). This is one of the most suitable softwares to perform molecular docking of ligand to their macromolecular receptor (Rizvi, Shakil, & Haneef, 2013). The graphical user interface program “AutoDock Tools” was used to prepare, run and analyse the docking simulation. For the protein input file, water molecules and ligand attached to the E-protein were removed. Polar hydrogen atoms were added and non-polar hydrogen atoms were merged, and Kollman charges and solvation parameters were assigned by default. For the ligand input file, Gasteiger charges were added, non-polar hydrogen atoms were merged and all bonds were made rotatable. All docking calculations were performed using the Lamarckian Genetic Algorithm (LGA) to determine the optimized conformation. In this study, domain III

was which contain the binding site. Therefore, the grid box with spacing of 0.375 Å and dimension of 112 x 96 x 118 points along the x, y and z axes was centred on the molecule. This procedure resulted in a total of 7 grid boxes for the binding of nLc₄Cer to E-protein. A population size was set up at 150; maximum number of energy evaluation, 2,500,000; maximum number of generations, 27 000; rate of gene mutation, 0.02; and cross rate, 0.8 were used for 100 search runs. In this study, several docking runs were required in order to identify the conformations of a ligand within a receptor pocket. The remaining parameters were set as default. A root mean square deviation (RMSD) tolerance for each docking was set at 4.0 Å. During the docking process, the E-protein was made rigid, while positions and torsional bonds of the ligand were kept free for flexible docking. For the purpose of pathway docking, the docking job was run in three different directions along the x, y and z axes. At each direction, the grid box was moved every 10 Å closer towards binding site. Since the starting position for each docking run was chosen based on the lowest energy of the docked structure, only one time docking was required to determine the docking pathway. A series of overlapping grid-boxes sliding along the x, y and z-directions were created at starting positions which were sufficiently high above the surface to its binding site. In this study, binding pathway was performed by sliding the docking grid box towards the binding site (Figure 3.1). The sliding box method allows a box at constant size to slide along the DENV E-protein structure. The initial slide of the grid-box covered some fraction of the protein surface to facilitate cluster analysis and ensure the docking scoring function would provide meaningful results (Tran, Le, & Truong, 2013). Docking simulations were performed for every possible box location; therefore producing a score for each potential docking site (Martins-Jose, 2013). Optimized pathway was then selected based on the lowest binding energy along the x, y and z axes. The flow of work in this part of the study is shown in Figure 3.2.

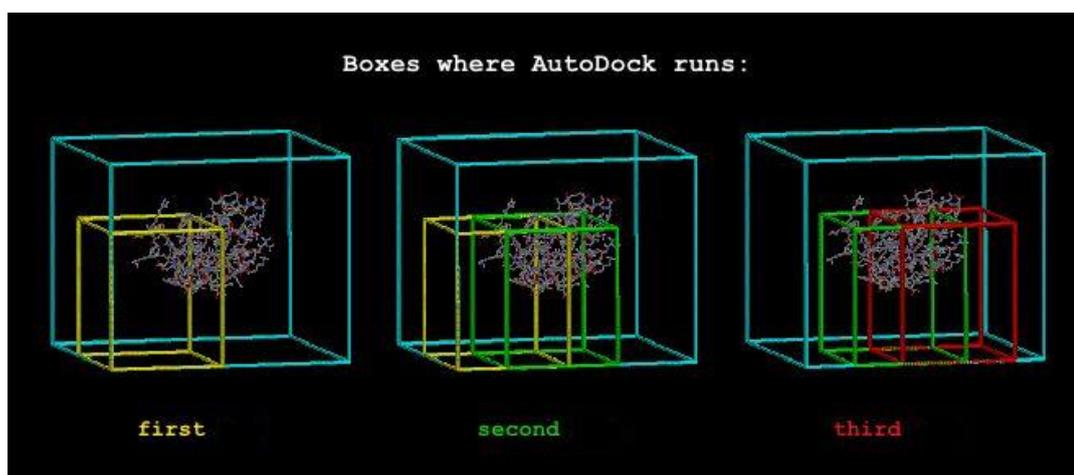


Figure 3.1: The sliding grid box location on the E-protein where docking has performed (Tran et al., 2013).

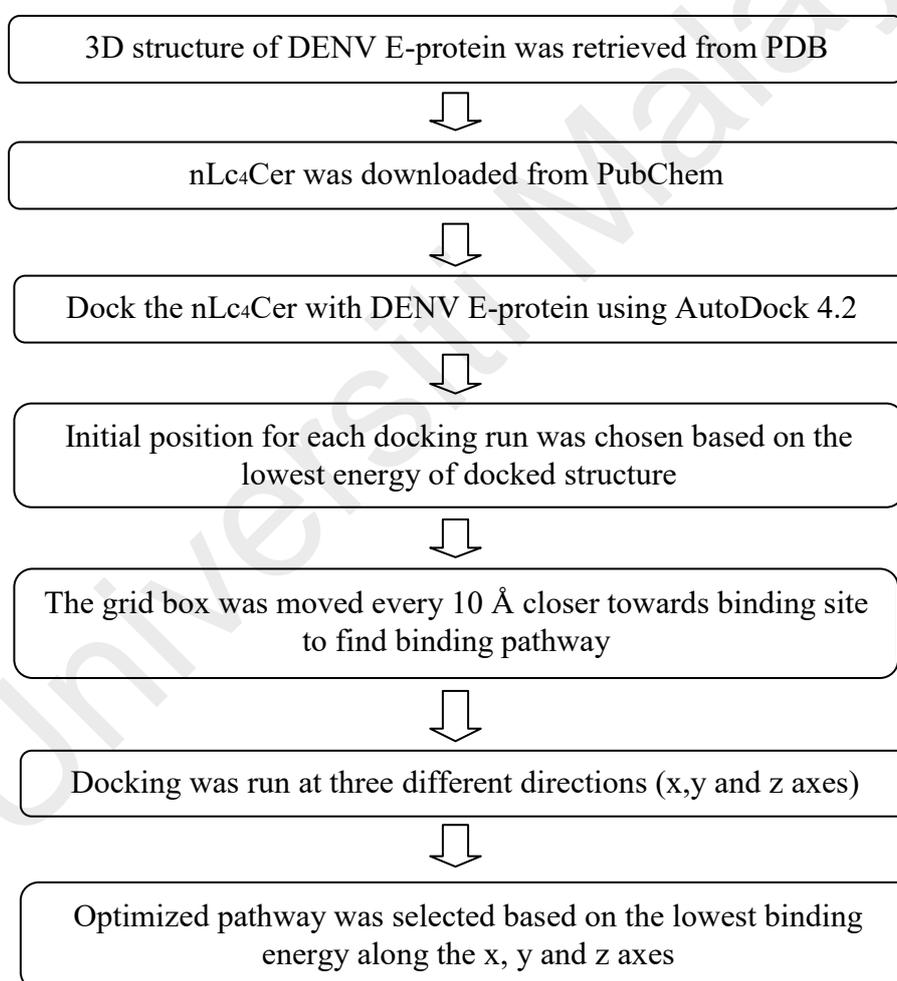


Figure 3.2: Flow chart showing summarized method involved in the simulation of binding pathway of nLc₄Cer towards DENV E-protein.

3.3 Results and Discussion

3.3.1 Characterizing Ligand Binding Site in Dengue Virus Type 2 (DENV2) Envelope Protein

3.3.1.1 Simulation of DENV E-Protein for 100 ns

Simulation of the DENV E-protein pre-fusion dimer was analysed to outline the effect that the conformational changes had on the interactions between residues, domains and the subunits. During MD simulations, significant structural changes that occurred could be related to the conformational changes towards the protein post-fusion state. Thus, it was important to consider the flexibility of the target protein by using multiple active conformational states. In this work, the DENV E-protein conformational states obtained from MD simulations were investigated to take into account the target protein flexibility, where the intention was to sample as closely as possible the landscape of the ligand binding region prior to inhibitor binding. The obtained conformational MD samples were then clustered into nine groups. A number of sets of protein conformational structures at different binding sites were chosen to be employed in the following ensemble docking study. Presently, ensemble docking is famously known as a powerful strategy to incorporate protein flexibility, notably when the ensembles were extracted from MD simulations (Campbell, Lamb, & Joseph-McCarthy, 2014).

3.3.1.2 Prediction of Binding Site on DENV E-Protein

In order for proteins to perform biological function successfully, interaction with a ligand molecule is often necessary. This interaction is specific, especially within the ligand binding location. The binding sites for small molecules are normally located in pockets or crevices on the protein surface or sometimes partially buried (Leis,

Schneider, & Zacharias, 2010). A receptor may have several potential binding sites, thus, it is crucial to know which is the active site that enables specific biological functions either via inhibition or catalysis. An active site is the region of the target protein that is responsible for its activity and it is made up of different kinds of amino acid residues. Since there is no prior knowledge of the binding site location, the identification of pockets and cavities can be a starting point for protein function annotation, protein ligand docking and protein structure-based drug design (Huang, 2009). Therefore, it is compulsory to perform a search on the whole protein surface or well known as blind docking in order to detect protein pockets.

This study is to find potential binding sites on the DENV E-protein, whereby, the virus membrane fusion activity can be inhibited by nLc₄Cer. This prediction usually requires the 3D structure of the target protein in order to search for the ligand binding site, and the E-protein ectodomain is the most suitable candidate (Degreve, Fuzo, & Caliri, 2012). Since DENV E-protein has a very shallow hydrophobic pocket (Harrison, 2008), thus, the binding site was recognized using the “Define and Edit Binding Site” program from Discovery Studio 4.0. This program utilizes several redundant already validated geometric algorithm to search for protein cavities (Koska et al., 2008).

Three possible binding sites were obtained and selected based on the size of pocket volume. Determination of pocket volume is based on the geometry shape of the protein surface analysed. Protein binding regions provide a microenvironment for substrates and inhibitors to interact and modulate the protein's activity (Tripathi & Kellog, 2010). In this software, the binding site is defined by a sphere that encloses part of the binding site volume. The predicted binding sites with the lowest binding energy are shown in Figure 3.3. The first binding site, BS1 (a) consists of 33 residues (Val24, Glu26, His27, Leu45, Lys51, Pro53, Ile129, Val130, Glu133, Pro166, Ser186, Thr189, Gly190, Leu191, Asp192, Met196, Val197, Leu198, Leu199, Gln200, Ala205, Leu207,

Val208, Glu269, Ile270, Gln271, Leu277, Leu278, Phe279, Thr280, Gly281, His282, and Ser284) having volume of 271.38 Å (Table 3.1). While the second binding site, BS2 (b) consists of 12 residues (Glu13, Gly14, Val15, Ser16, Gly17, Gly18, Ser19, Trp20, Val21, Met34, Ala35, and Lys36) having volume of 40.88 Å. The third binding site, BS3 (c) has 18 residues (Met1, Arg2, Cys3, Ile6, Asp42, Glu44, Ile46, Val140, Ile141, Thr142, Gly152, Asn153, Asp154, Thr155, Lys157, His158, Gly159, and Lys160) with occupied volume of 39.50 Å. These different binding pockets were determined based on the calculation of volume of cavities, and then were docked with nLc₄Cer. The cavities found on the protein surface are of special significance as they are potential sites of protein-ligand and protein-protein interactions (Bhinge et al., 2004). Table 3.1 show that BS1 has the lowest binding energy, followed by BS2 and BS3 with binding energy -2.80, -2.13 and -0.97 kcal/mol respectively. This shows that the bigger the volume of the binding site, the easier nLc₄Cer would be able to fit into the binding site. The data also revealed that, the bigger the volume, the lower binding energy of the complex, thus resulting in favourable interaction. This appropriate binding pocket can accommodate a drug molecule properly and at the same time, reduction in the volume of binding cavity may lead to the relative surface area of the interaction become smaller thus resulting in poor binding affinity (Li et al., 2015). Although they may be a number of voids inside the protein, it has been observed that the binding site is usually the largest cavity in a protein because a large pocket provides large surface area and hence increased the opportunity for molecule binding (Tripathi & Kellog, 2010). Determination of this binding cavity is highly important in order to predict protein structure, function, stability, mutation studies and identification of potential drug and lead optimization (Bhinge et al., 2004; Pandey et al., 2016).

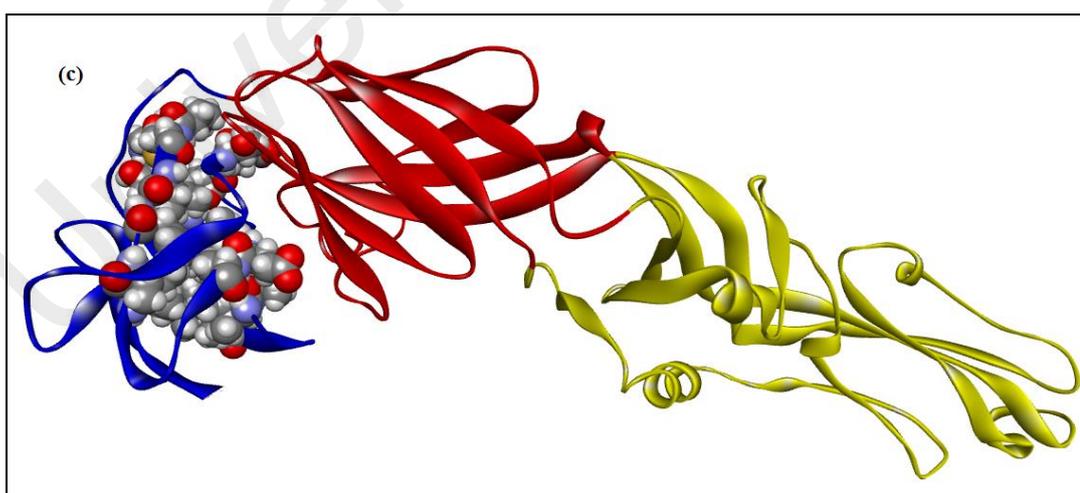
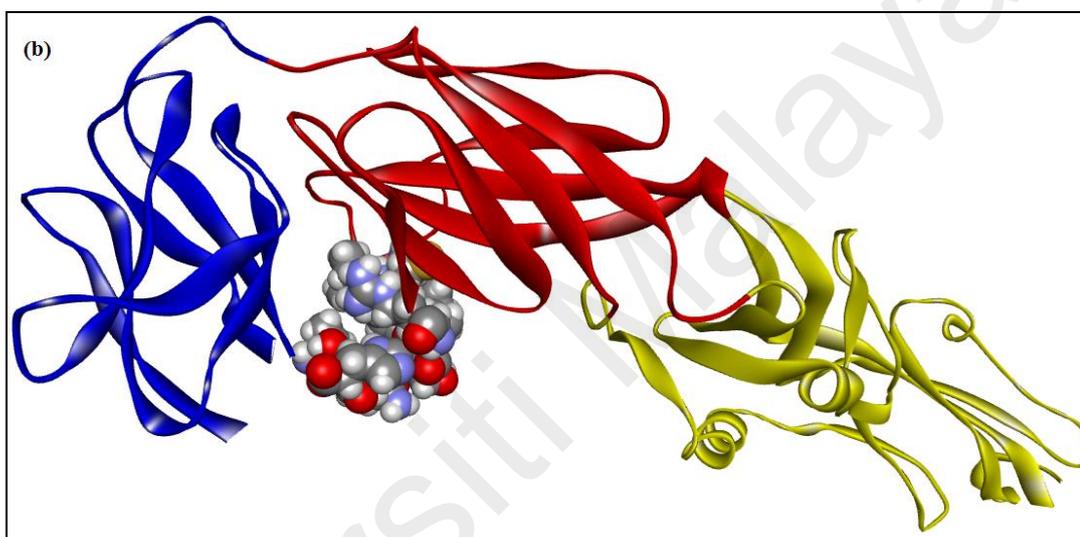
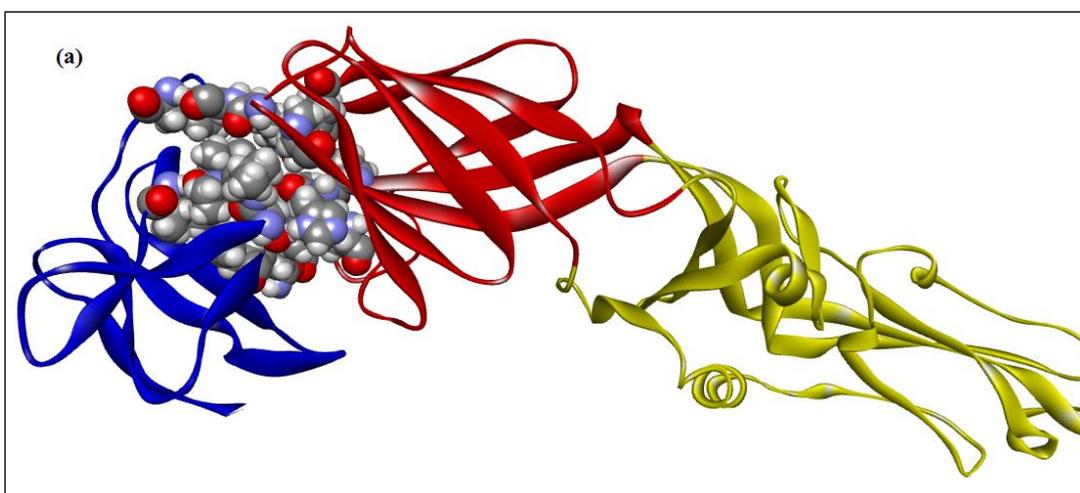


Figure 3.3: Positions of the three binding sites, (a) BS1, (b) BS2 and (c) BS3, as predicted by Discovery Studio 4.0. Ligand binding site is shown in CPK presentation. DI, DII and DIII are illustrated in red, yellow and blue coloured ribbons, respectively.

Table 3.1: Binding energy of E-protein-nLc₄Cer complex subjected to various binding sites defined by Discovery Studio 4.0 program.

Binding site (BS)	BS1	BS2	BS3
Binding energy (kcal/mol)	-2.80	-2.13	-0.97
Number of conformation	18	53	79
Volume (Å ³)	271.38	40.88	39.50

Binding site detection is a challenging process because proteins frequently undergo large structural changes upon ligand binding. Binding site also presents considerable plasticity that may be important in structure-based drug discovery. Study done by Almeida et al. (2013) showed that through their Ligplot studies of the protein-ligand binding modes revealed changes in both the protein sub-pockets conformation and in ligand interactions, indicating that the binding site was definitely not static and this plasticity could be explored for competitive structure-based drug design.

According to Ali et al. (2014), this binding position varieties would lead to conformational change that could interrupt the protein normal function. Alhoot et al. (2013) claimed that, a selection of suitable receptor binding site may lead to conformational changes in order to inhibit viral entry into host cells. Furthermore, E-protein binding sites required a decasaccharide for strong interaction (Marks et al., 2001). Meanwhile, a study reported by Hung et al. (2004) had suggested that E-protein mediated membrane fusion required post binding conformational changes that promoted a dimer-to-trimer transition. However, the displacement of the ligand away from the binding site was obviously due to the reduced interactions between the protein and the ligand in the intermediate subensembles as observed by Koppisetty et al. (2015).

Meanwhile, Yennamali et al. (2009) had found the cavity site between domain I-III on the DENV2 E-protein through docking and MD simulation methods. They found the existing of binding site only in dimeric structure. The presence of a ligand that can occupy this cavity would be able to stabilize the dimer structure and finally inhibits E-

protein trimerization, so that the fusion process can be stopped. On the other hand, Yan and Zou in their study (2014) that predicted the peptide binding sites on protein surfaces reported that the cluster with the largest size was predicted as the putative binding site. They also claimed that the binding site was located in the largest pocket on the protein surface. These data would be beneficial to gain insights into protein–drug interactions. Study focusing on binding sites and ligand that can occupy this space may help in search for potential DENV inhibitors in order to prevent the viral entry into host cell.

3.3.1.3 Refinement of Docked E-Protein-nLc₄Cer Complex using MD Simulation

The main objective of the MD simulation study is to investigate the positional and conformational changes of nLc₄Cer in the binding site that provides insight into the binding stability. Monitoring the complex's pressure, energy and density are the most basic metric to monitor a system's equilibration and stability. By checking the density, temperature, total energy, RMSF, RMSD and radius of gyration (R_g) (figures as presented in Figures 3.4-3.11) it is possible to determine whether the system is reasonable or not. Furthermore, the stability of the system also proved the credibility of docking results. Any damage to the complex structure such as denaturation will affect the above parameters. In order to get these parameters, ptraj tool was used and since the backbone of the protein is the main concern, only backbone atoms C, CA and N were considered. The backbone atoms were used for residues that were chemometrically predicted to form alpha helix. Figure 3.4 showed energies that exist during simulation where red line indicates the kinetic energy, black line is the potential energy and green line is the total energy. Total energy is the sum of potential energy and kinetic energy. The kinetic energy remained constant during the simulation implying that temperature thermostat, which acts on the kinetic energy, was working correctly. Besides that, the potential energy also remained constant indicating that the relaxation was completed

and reached equilibrium. This is normally expected, as during MD simulation, the complex become more relaxed, allowing the water to interact with the residues from the binding site. It is interesting to note that, Almeida et al. (2013) in their study observed a protein-ligand complex being diminished at the beginning of the simulation and then became stabilized after 10 ns. In fact, during the last 5 ns of simulation, the protein ligand interaction increases again when approaching the initial state.

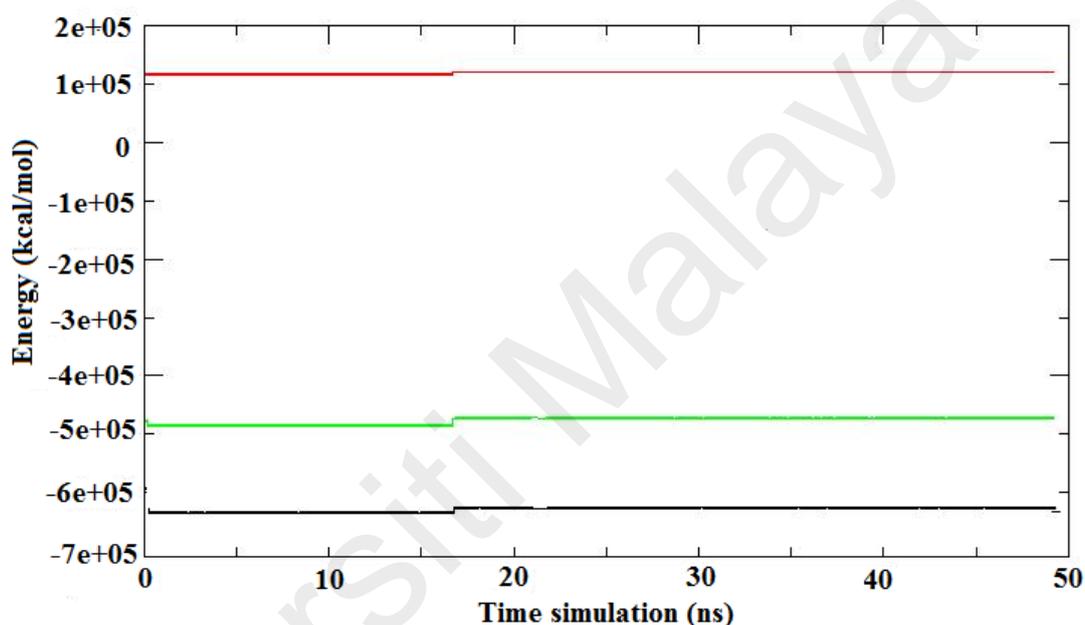


Figure 3.4: The energy of the system: red line (positive) shows the kinetic energy, and black line (negative) shows the potential energy. The green line shows the total energy of E-protein-nLc₄Cer complex during 50 ns simulation. The heating process is not shown.

The E-protein-nLc₄Cer complex has reached an equilibration volume (Figure 3.5). The smooth transitions in this plot followed by the oscillation above a mean value suggest that volume equilibration has been successful. This system has equilibrated at a density of approximately 0.98 g cm⁻³. This seems reasonable as the density of pure liquid water at 300 K is approximately 1.0 g cm⁻³. The water density remains stable throughout production run and the average is near the experimental neat value (1 gcm⁻³) which implies the reliability of the system improves (Figure 3.6). Meanwhile, the pressure graph seems to show that the pressure fluctuate wildly around 100 bar = 1 atm

during the simulation (Figure 3.7). This is sufficient to indicate successful equilibration. This large fluctuation may be due to the system size. Pressures are reported in bars in which negative pressures indicate that the system would like to contract its volume, i.e. that an outward force on the walls of the container would be required to keep the system at its current volume; positive pressure is the opposite: the system would like to increase its volume, and an inward-directed external force would be required to keep the system at its current volume.

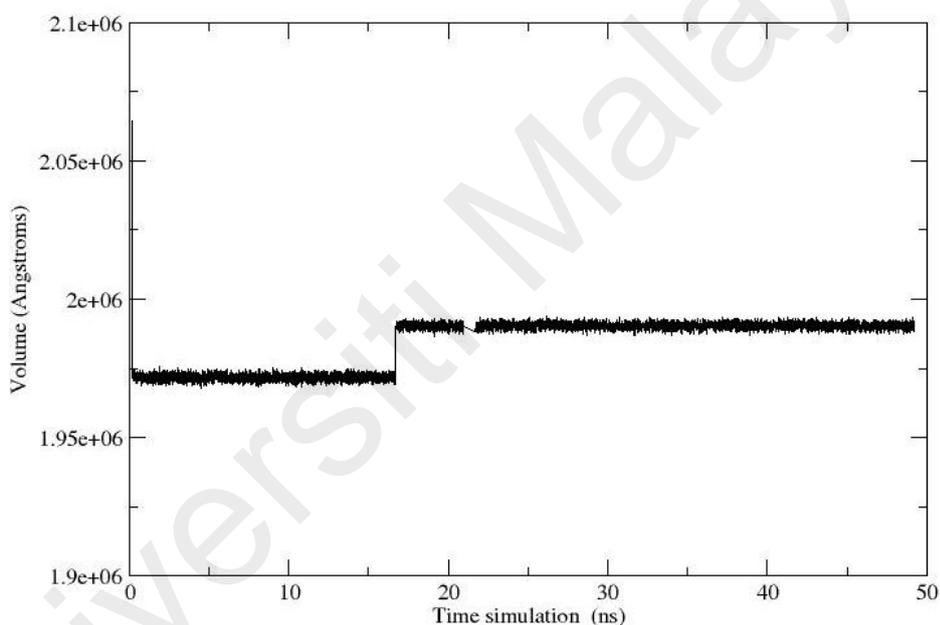


Figure 3.5: Volume of E-protein-nLc₄Cer complex during 50 ns MD simulation.

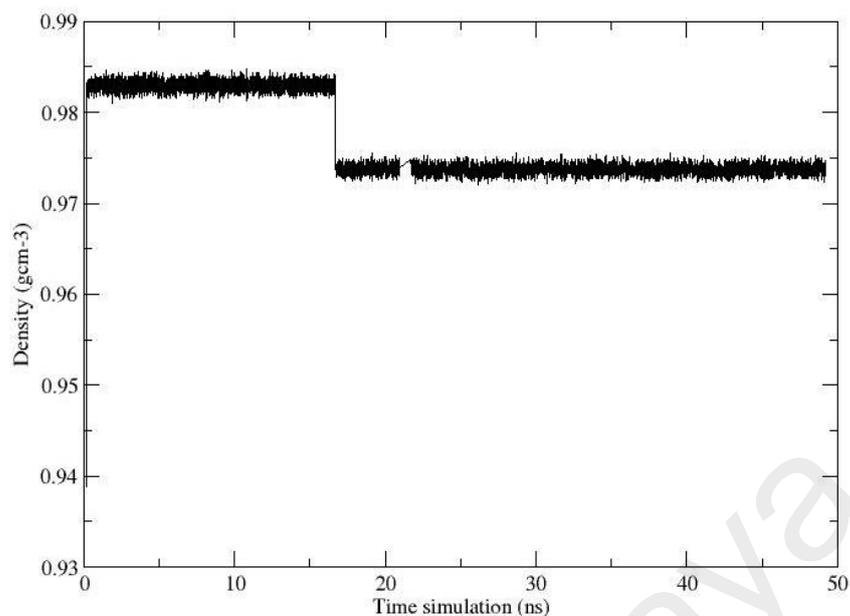


Figure 3.6: Density of E-protein-nLc₄Cer complex during 50 ns MD simulation.

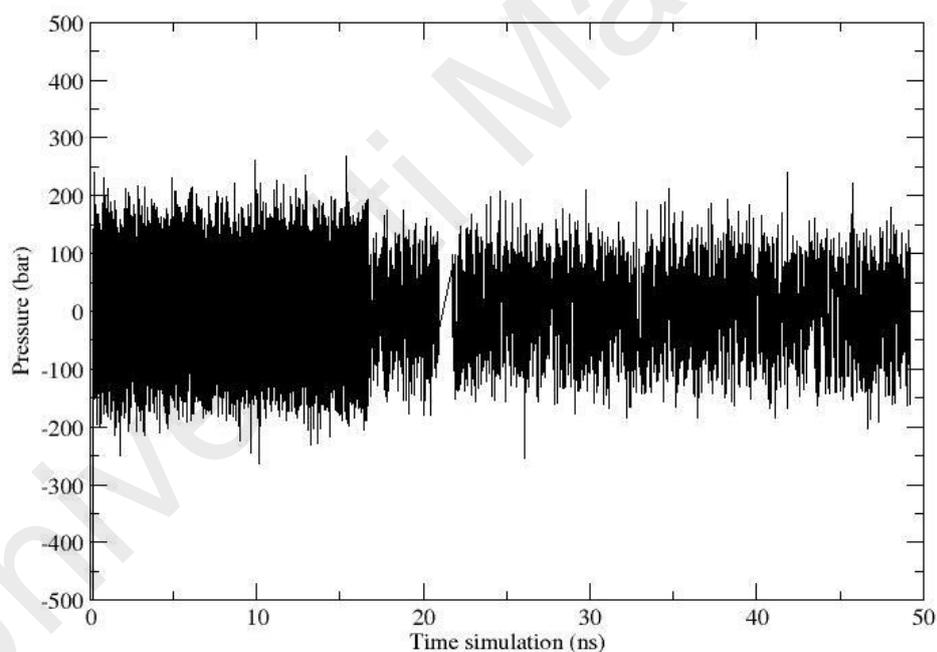


Figure 3.7: Pressure of E-protein-nLc₄Cer complex during 50 ns MD simulation.

On the other hand, Figure 3.8 showed stability in temperature during simulation at constant energy implying that the temperature regulation was applied correctly. In this study, the MD simulation was running up to 50 ns because the complex have the possibility to surpass greater conformational changes after longer simulation time,

revealing important motions that are hardly detected in a shorter simulation. These motions can be related to important conformational states during protein equilibrium (Almeida et al., 2013). Other researchers such as Behera et al. (2015) and Naseri et al. (2015) in their antibody and receptor binding modelling studies had performed simulations up to 10 ns at temperature of 310 K under 1 bar pressure.

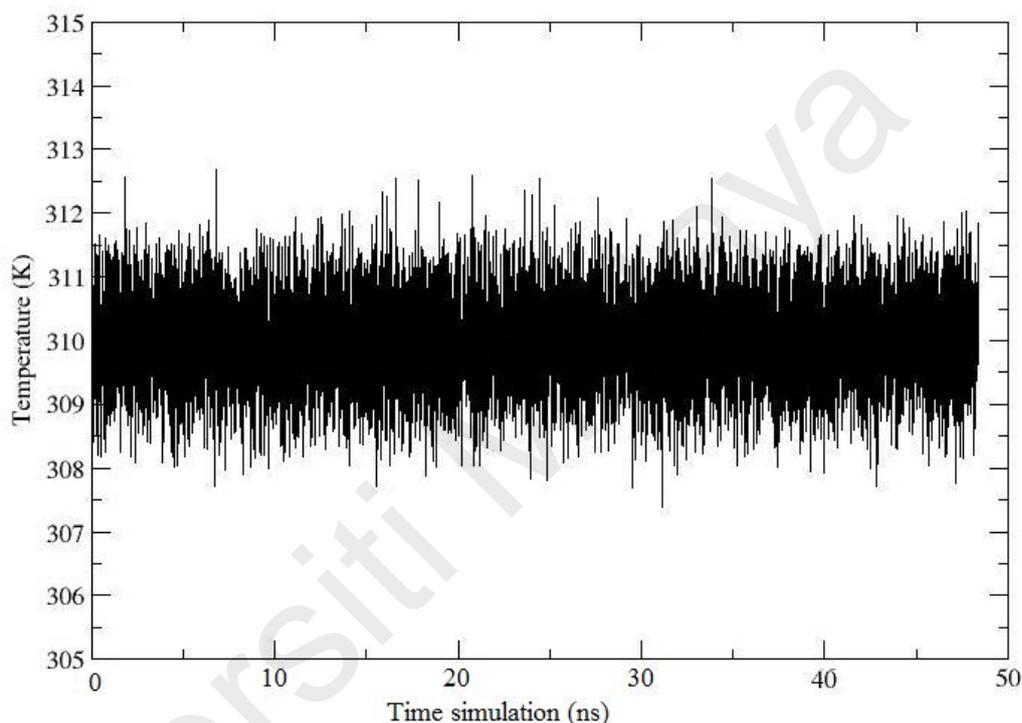


Figure 3.8: Temperature of E-protein-nLc₄Cer complex during 50 ns MD simulation.

The next step is to look at the root mean square deviation (RMSD) curves to observe the conformational changes of E-protein-nLc₄Cer complex. RMSD values are beneficial to describe the magnitude of the conformational changes between the two molecular structures. It is used to characterize the quality of biomolecular simulations, to cluster conformations and as reaction coordinate for conformational changes. RMSD can be calculated over all atoms of a conformation or a subset either to focus comparison on a region of interest like an enzyme active site or to exclude highly variable elements like surface loops or hydrogen atoms. It is normally calculated using only one atom per amino acid residue, normally the α -carbon in order to capture protein

conformational changes. The RMSD reflects the change of protein backbone over the simulation time. RMSD indicates the similarity of the proteins if their value is small such as few angstrom (Å). In Figure 3.9, there is a slight fluctuation in the RMSD of the E-protein-nLc₄Cer complex due kinetic energy of every atom that causes fluctuation and rise in RMSD. The average RMSD found is 3.5 Å. The RMSD was relatively higher at the initial stage of the simulation (almost till 10 ns of production dynamics) and probably corresponding to small relaxation of the protein following release from its crystallographic environment), followed by longer time scale and fluctuations. The high RMSD values are due mainly to large structural amino acid changes in the loop regions that are in contact with neighbouring proteins. Besides that, the large RMSD indicates that structures in the trajectories are significantly different from the initial structure as RMSD measures the overall change in conformation from the reference structure. The peak fluctuation then decreases and almost becomes steady for the rest of the simulation. These observations suggest that the complex structure is stable and retains its overall structure during simulation.

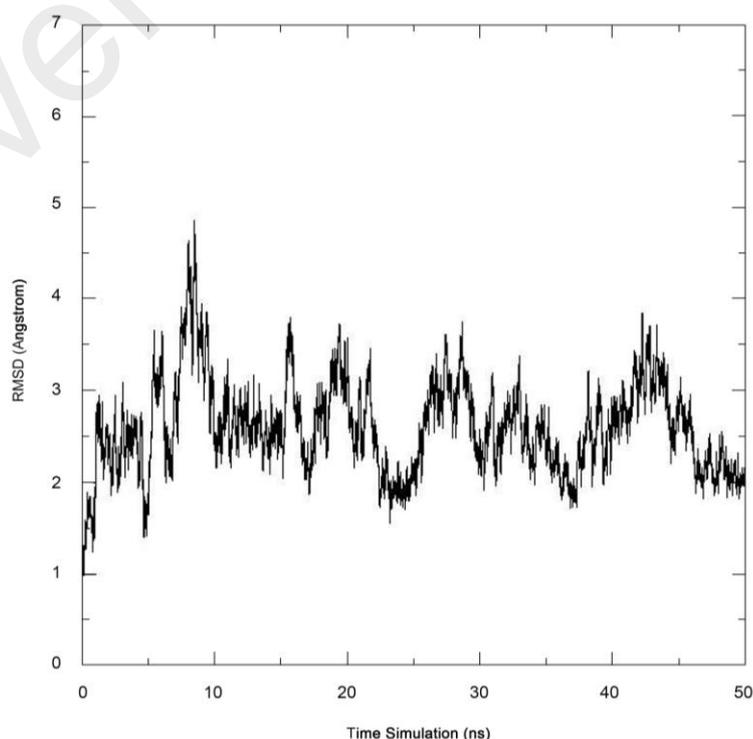


Figure 3.9: RMSD of E-protein-nLc₄Cer complex during 50 ns MD simulation.

Meanwhile, the flexibility of the E-protein-nLc₄Cer complex was investigated in terms of the atom-positional root-mean-square fluctuations (RMSF) for all non-hydrogen atoms. RMSF measures how much each individual atom moves around and is a totally different calculation from RMSD. It reflects the overall movement of atoms or residues over all frames. It is a qualitative measure of protein flexibility as it gives a view of relationship between protein conformational flexibility and dynamics. RMSF of C_α atom from the MD simulation structures against starting structures was calculated to identify the most flexible regions of the E-protein-nLc₄Cer region. The larger RMSF value conveys more flexible region while the lower RMSF value entails the more constrained region (Moonrin et al., 2015). The RMSF average value was 3 Å (Figure 3.10) and this moderate fluctuation of backbone deviation during simulation reflects the global structural rearrangement of complex is well equilibrated, maintained and not artificially altered by the simulation. The fluctuations during MD simulation are expected as a result of solvated environment and a fully flexible protein.

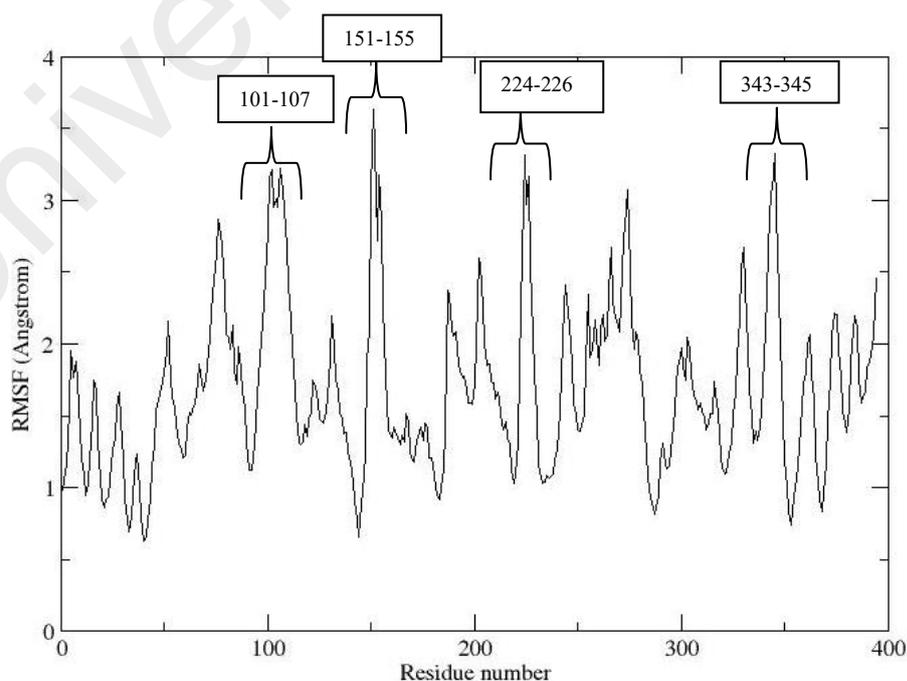


Figure 3.10: RMSF of E-protein-nLc₄Cer complex during 50 ns MD simulation.

The structure showed more flexibility with the RMSF average value of 3.5 Å at residue numbers 151-155. In addition, residues 343-345 which belong to DIII region also showed flexibility due to high RMSF values. Based on the results, the residues located in following regions: 101-107, 151-155, 224-226, 343-345 were among the most fluctuated and mobile residues. Therefore, this region could be critical regions to thermostability and can be a potential target for stability enhancement through rational design. Normally, the region with high fluctuations is related to connecting loops, turns and bends that connect secondary structure elements as well as helices compared to other regions (Ganoth, Friedman, Nachliel, & Gutman, 2006). The same observation was also reported by Pang et al. (2003) where regions correspond to surface exposed loops have been postulated to play an important role in receptor binding. They also added that, the region with large fluctuation form „jaws“ of the protein and contain the so-called doorkeeper residues thought to be important in maintaining access to the ligand binding site. Large fluctuation was also observed by Wichapong et al. (2014), where they reported that high RMSF values of the human-D-amino acid oxidase (h-DAAO) monomer are likely to result from the absence of contact between the chains and therefore had more space to fluctuate. This trend suggested that, the flexibility of ligand and receptor may enhance affinity between them. Ganoth et al. (2006) also reported high RMSF of the protein in the M1c1p-IQ4 simulation and this was due to the structural modification process. However, the reduction in RMSF value was due to rigid region in the complex with only a limited space to flex in. Other study reported by Panigrahi et al. (2013) also observed minimal fluctuation of the RMSF at the end of MD equilibration. The RMSF curves revealed a pattern similar to that calculated for the entire simulation time, although the extent of the fluctuations is smaller.

Meanwhile, the compactness of the DENV E-protein-nLc₄Cer complex throughout the simulation can be monitored using radius of gyration (R_g). R_g is another indicator of structure stability as well as estimation of the molecule size during simulation (Medvedev et al., 2014). It measures a distance of all atoms to the centre of mass (COM) of the structure and is calculated according to Equation 3.4 (Mueller, 2010).

$$R_g = \left(\frac{\sum_i r_i m_i}{\sum_i m_i} \right)^{1/2} \quad (3.4)$$

Where r_i is the position of atom i with respect to the COM of the structure, and m_i the mass of atom i .

A stable complex normally maintains a relatively steady R_g , whereas unfolding or collapse will be indicated by R_g change over time. From Figure 3.11, initially, a fast collapse at 13 ns (34.5 Å) was observed and continued to have a modest decrease which indicate the stability of the complex structure throughout simulation. And again, the immediate collapse of the complex's structure at 38 ns (34.2 Å) to form a more compact structure was observed. R_g analysis revealed that the system remained compact during the 50 ns simulation time and produced average R_g value of 35 Å. This may be due to strong affinity of nLc₄Cer to the E-protein (Jana, Cahturvedi, & Robine, 2014). On the other hand, the complex that showed smaller R_g , could have less molecular surface area, thus resulting in a decrease in protein exposure to water as the structure become more compact (Mozafari, Tazikeh-Lemeski, & Saboury, 2016). In this study, no significant changes were observed for R_g , implying a sustained stability and compactness of the complex. However, fluctuations in R_g values were recorded suggesting a loss of compactness for its complex. Overall, this study reveals that some critical points of the complex are liable to make extensive moves.

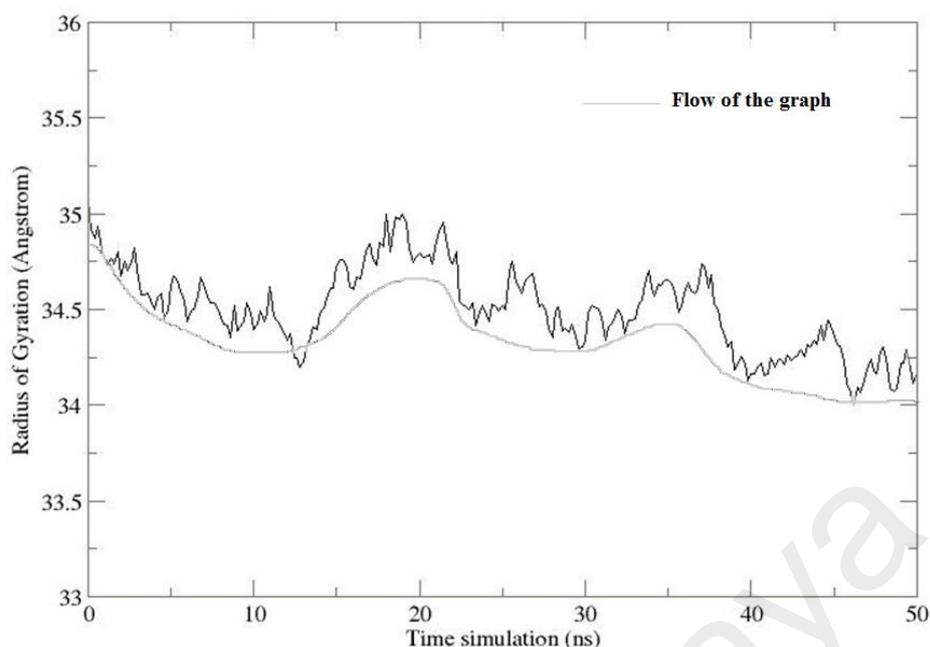


Figure 3.11: Radius of gyration (R_g) of E-protein-nLc₄Cer complex for backbone heavy atoms during 50 ns MD simulation.

3.3.1.4 Cluster Analysis

Conformational changes are extremely important in molecular function where it has been proven experimentally that E-protein undergoes extensive conformational change during infection process (Zhang, Sun, & Rossmann, 2015; Zhang et al., 2004). In order to accomplish its diverse functions, the E-protein must exhibit some conformational and dynamical peculiarities at single molecular level, which cannot be investigated with the employment of traditional experimental techniques. Degreve et al. (2012), employed MD simulation to characterize extensive structural changes in E-protein under distinct intensive thermodynamic parameters. A comprehensive knowledge of the relationship between sequence, structure and the associated dynamics is compulsory in order to understand the protein molecular function and to design proteins with novel functions (Katagi, 2013).

Cluster analysis was done in order to classify large number of conformations based on the similar structures. The notion of the importance of receptor flexibility has fostered the usage of computational technique such as MD simulations to generate

ensembles of energetically accessible conformations (Tautermann, Seeliger, & Kriegel, 2014) as well as to investigate the conformational and positional changes of ligand that provide insights into the binding stability (Verma, Jatav, & Sharma, 2014). Clustering method relies on similarity or dissimilarity measures between structures. This method has emerged as a useful, automated technique to extract conformational states from MD simulation data (Philips, Colvin, & Newsam, 2011). In this study, MD simulation of DENV E-protein was performed to generate an ensemble of conformations for docking purpose. The representative structure of the most populated clusters for each conformer was generated using the Cluster trajectory option in ptraj tool. In most cases, the most representative is the most populated conformation and are shown in Figure 3.12. The lowest binding energy conformation in the first cluster was considered as the most favourable docking pose (Heavner, 2004). The representative structures were then used in docking study to find the exact binding energy values.

Each trajectory was divided into nine clusters using hierarchical clustering algorithm. Hierarchical clustering algorithm was chosen due to their apparent good performance in analysing MD trajectory data and its frequent availability in MD analysis packages (Wolf & Kirschner, 2013). All nine cluster groups were chosen to be representative structures and named as follows: c0, c1, c2, c3, c4, c5, c6, c7 and c8. The cluster with the most population, indicated that this cluster conformation was more favourable for ligand binding and located in the first rank. Low $\Delta G_{\text{binding}}$ values signify that nLc₄Cer was in the most stable conformation when bound with E-protein. Table 3.2 reveals that cluster rank, **c0**, represents the conformation group with high number of conformations in the cluster and with the low binding energy which is -2.83 kcal/mol. Low binding energy value signifies that the nLc₄Cer was in the most stable conformation when bound to E-protein with the binding chance of 80-90% (Usman Sumo Friend Tambunan, Noors, Parikesit, Elyana, & Ronggo, 2011). This cluster

conformation with low binding energy was chosen as a starting structure for the subsequent MD simulation.

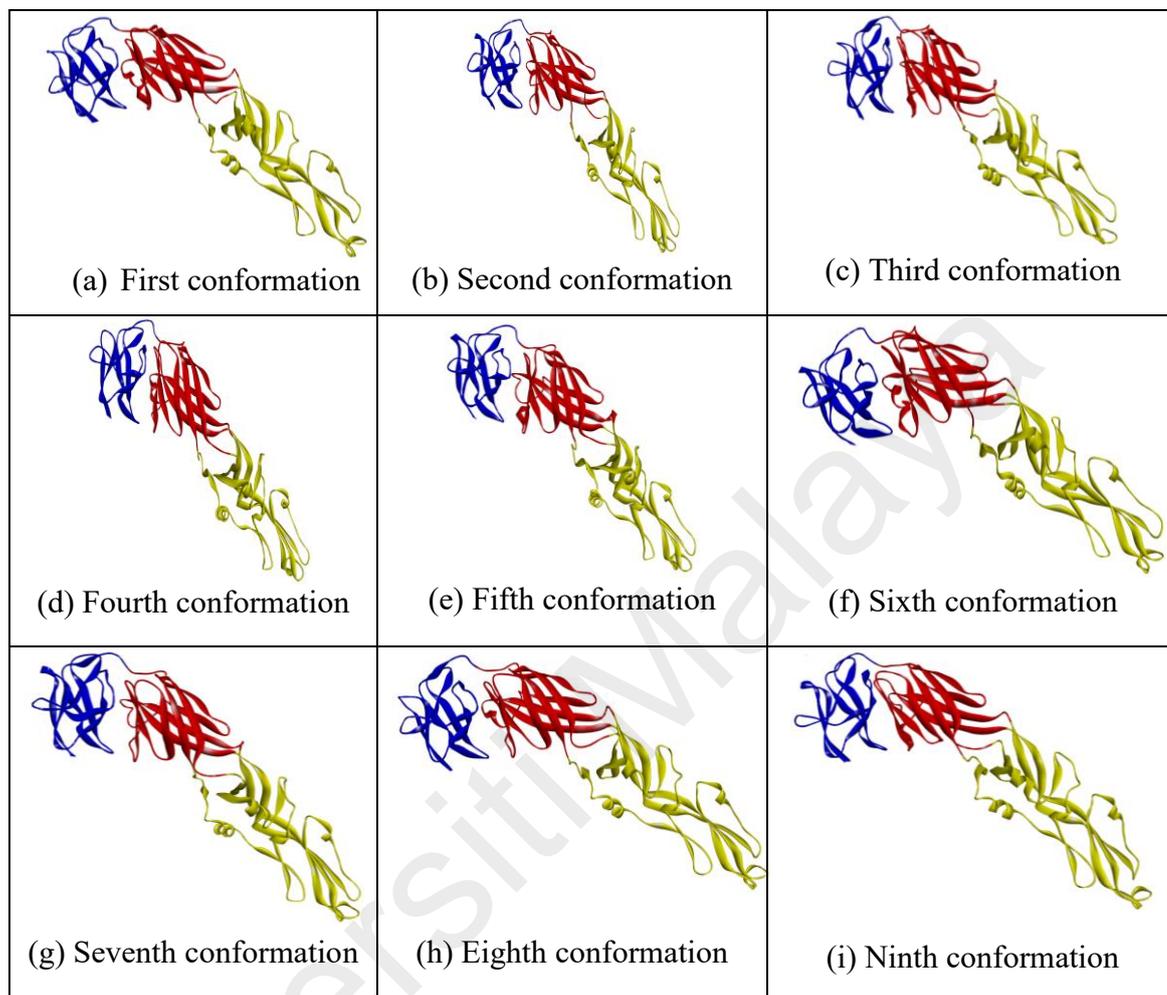


Figure 3.12: The conformational structures of DENV E-protein identified by clustering technique. DI, DII and DIII are illustrated in red, yellow and blue colours, respectively.

Table 3.2: Binding energy and RMSD of E-protein-nLc₄Cer complex subjected to different conformations generated from MD trajectories.

Conformation	Binding energy (kcal/mol)	Number in cluster	RMSD at DIII
c0	-2.83	35	1.2013
c1	-1.51	27	1.2635
c2	-1.56	13	1.0599
c3	-2.53	15	1.0748
c4	-2.89	23	1.3953
c5	-1.16	23	1.0792
c6	-0.97	20	1.1992
c7	-2.24	23	1.1391
c8	-2.22	11	1.1454

Each docking conformation was then manually inspected to emphasise the common clustering pattern. Normally, the same protein can populate multiple binding modes. Meanwhile, a ligand would be able to bind to a target with a variety of conformations due to ligand or receptor protein symmetries. The use of these multiple MD configurations can significantly improve docking results since there was an improvement in docking performance compared to docking without MD run. Therefore, nLc₄Cer was docked against a variety of E-protein conformations. Table 3.2 shows that binding energy between each cluster is between -1 to -2 kcal/mol with the highest binding energy is conformation **c6** (-0.97 kcal/mol) and the lowest binding energy is **c0** (-2.83 kcal/mol). This data suggests that the complex can appear in different conformations even though they may have identical energy. This is in line with the study reported by Guruprasad et al. (2014) where different conformations may have the same energy even though each conformation has a specific energy.

Furthermore, the negative binding energy indicates that nLc₄Cer has shown binding affinity towards E-protein. In fact, the similarity in binding energy revealed that the structure of E-protein-nLc₄Cer complex is not differing so much between each representative structure. This is supported by the RMSD values where the values between each cluster group is not changing so much (Table 3.2). In this case, the ligand acts by disrupting the movement of E-protein required for inter-conversion between the "open" and "closed" conformations. This binding process invokes large conformational changes in both the E-protein and nLc₄Cer. The conformational change during binding directly impacts on how the process must be simulated in order to give accurate results (Treesuwan et al., 2009).

These varieties of conformations are important as ligands would be able to retard or block viral entry if they bind selectively to any conformations of the fusion protein. These varieties of conformation revealed a holistic picture of protein plasticity as well

as protein function. The plasticity of the binding site environment and in the binding modes is evidenced by the differences observed during the MD simulations. In fact, one way to predict the protein function is based on their dynamics. As a result, this requires the conformational space to be sufficiently explored by a suitable method which, in this case, is MD simulations.

3.3.1.5 Binding Free Energy of E-Protein-nLc₄Cer Complex

Carbohydrates are characterized by their considerable conformational flexibility and dense polar functionality. Because of their dynamic nature and the fact that they interact intimately with aqueous solvent, free energy analysis on MD trajectories of a receptor-ligand complex in solution is chosen. Although experimentally inaccessible, determination of the free energy components associated with these structural features was feasible through theoretical analysis. MMPBSA and MMGBSA studies were performed in order to estimate the binding free energy for the interaction between nLc₄Cer and E-protein during MD simulation. According to Koppisetty et al. (2015), this binding energy calculation has a potential application in understanding the underlying phenomena as well as to design novel inhibitors of molecular interaction pathways.

In the interaction free energy calculations, the major contribution to the stabilization of the E-protein-nLc₄Cer complex was attributed to the molecular mechanics component of the interaction energy and the nonpolar component of the solvation energy. Various energy contributions to the calculation are shown in Table 3.3 which consist of electrostatic (polar) and van der Waals (non-polar) interactions, polar solvation energy and non-polar solvation energy (Duan, Feng, & Zhang, 2016). The favourable formation of E-protein-nLc₄Cer complex is driven by the electrostatic ($\Delta E_{\text{electrostatic}}$) and the vdW (ΔE_{vdW}) terms of the molecular mechanics energy.

Table 3.3: Binding free energies predicted using the MMPBSA and MMGBSA methods at different simulation times (All data are given in kcal/mol).

Method	Contribution	10 ns	20 ns	30 ns	40 ns	50 ns
MM	ELE	-59.11	-102.41	-70.31	-67.97	-79.50
	VDW	-5022.71	-5012.03	-5080.32	-5113.47	-5149.76
	GAS	-5062.45	-5095.31	-5131.24	-5161.94	-5209.54
PBSA	PBSUR	-6.42	-6.91	-6.60	-6.53	-6.30
	PBCAL	74.84	106.53	78.65	81.08	81.26
	PBSOL	68.42	99.62	72.06	74.55	74.96
	PBELE	15.73	4.12	8.34	13.11	1.75
	PBTOT	-4994.03	-4995.69	-5059.18	-5087.39	-5134.59
GBSA	GBSUR	-3.80	-4.37	-3.95	-4.03	-3.77
	GBCAL	63.06	100.15	69.75	70.74	76.36
	GBSOL	59.26	95.78	65.80	66.71	72.59
	GBELE	3.95	-2.26	-0.56	2.77	-3.15
	GBTOT	-5003.19	-4999.53	-5065.44	-5095.23	-5136.96

The individual energy contributions: ELE = electrostatic energy as calculated by the molecular mechanics (MM) force field; vdW = van der Waals contribution from MM; GAS = total gas phase energy ELE+vdW+INT; INT = internal energy arising from bond, angle and dihedral terms in the MM force field (this term always amounts to zero in the single trajectory approach); PBELE/GBELE = sum of the electrostatic solvation free energy and MM electrostatic energy; PBSUR/GBSUR = non-polar contribution to the solvation free energy calculated by an empirical model; PBCAL/GBCAL = the electrostatic contribution to the solvation free energy calculated by PB or GB respectively; PBSOL/GBSOL = sum of non-polar and polar contributions to solvation free energy (PBSUR + PBCAL); PBTOT = final estimated binding free energy calculated by MMPBSA method; GBTOT = final estimated binding free energy calculated by MMGBSA method (all energies are in kcal/mol).

In this study, binding free energies represent the sum of the total intermolecular energy, total internal energy and torsional free energy minus the energy of the unbound system (Gautam et al., 2012). The negative value of electrostatic energy (ELE) during the 50 ns simulation (-79.50 kcal/mol) signifies favourable interaction and this energy is compulsory for the interactions of all biological macromolecules (Zhang, Witham, & Alexoy, 2011). These interactions generate attractive and repulsive forces, which cause motion of molecules (Pacholczyk & Kimmel, 2011). This is due to the fact that most biological macromolecules are highly charged and the existence of this charge distribution may optimize the electrostatic binding free energy when ligands bind to their protein targets (Francis, 2002; Kukic & Nielsen, 2010). Thus, favourable

electrostatic interactions are produced by complementary charge distribution between binding partners. Furthermore, electrostatic complementary between two binding components resulting in steering effect which enhances their association rate (Schlick, 2012). Study reported by Ganoth et al. (2006) also agreed that as the electrostatic interactions of the solute with the solvent are taken into account, it appeared that the total solvation energy was in favour of the bound state of the protein-peptide complex. The current finding suggests that electrostatic energy could be responsible for the binding affinity of nLc₄Cer to E-protein.

Meanwhile, the effect of van der Waals (vdW) cannot be simply ignored as it is closely related to the hydrophobic interaction energy. This vdW force is the sum of the attractive and the repulsive non-bond forces between atoms or molecules. During the first 10 ns, the vdW energy was -5022.71 kcal/mol, then, slightly increased to -5012.03 kcal/mol at 20 ns simulation and then underwent continuous constant decrease until the end of simulation (-5080.32, -5113.47 and -5149.76 kcal/mol at 30 ns, 40 ns and 50 ns, respectively). This could be explained where at the beginning of the simulation (20 ns), the attraction between nLc₄Cer and E-protein was being diminished, and the increment in vdW energy was expected during the MD simulations, where the complex became more relaxed, allowing water to interact with the residues from the binding site, thus decreased interactions between nLc₄Cer and E-protein. Meanwhile, the decreased in vdW energy at 30 ns was due to the vdW forces between two particles, resulting in very strong interactions. Besides that, low values of these energy components can be associated with strong ligand binding (Sindhu & Srinivasan, 2015). This major role played by the vdW interactions demonstrated how inter-residue contacts, where the tight binding of ligand to the surface exposed amino acid residues on the protein occurred (Ganoth et al., 2006). Thus, contribute to the energy stabilization that provide stability to the complex and favourable interaction was observed.

Nevertheless, the unfavourable contribution to the complex stability was observed in GBCAL, GBSOL and GBELE calculations where the interaction energies were positive throughout the 50 ns time simulation. The same unfavourable contribution was also observed in MMPBSA calculation. This unfavourable binding may be due to the weaker electrostatic interactions between ligand and receptor compared to stronger force between ligands and solvents (Ma et al., 2015). Even though some internal energy terms (GBCAL, GBSOL and GBELE) values are positive, disavouring the complex formation, contribution from van der Waals interaction, one of the main components favouring the complex formation. According to AMBER developers (McGee et al., 2009), favourable electrostatic energy and unfavourable solvation free energy are expected during MD simulation. This symbolizes the energy that one has to use to desolvate the binding particles and to align their binding interfaces. Yet, the molecular mechanics energy component of the interaction energy strongly favours the complexes over the unbound molecules.

The total energy GBTOT is negative throughout simulation which energetically favours the complex stability. Initially, the GBTOT was -5003.19 kcal/mol during the first 10 ns simulation, and then gradually decreased to -5136.96 kcal/mol as the simulation was extended up to 50 ns. On the other hand, the values from PB method (PBTOT) were slightly higher where -4994.03, -4995.69, -5059.18, -5087.39 and -5134.59 kcal/mol were observed at 10, 20, 30, 40 and 50 ns respectively. Both results revealed that the longer the simulation, the better the binding interaction between the nLc₄Cer and E-protein.

In term of the calculation models that had been used in this study, the Generalized Born (GB) model had been proven to successfully calculate accurate solvation energies of small molecules, score protein conformations, evaluate protein-ligand binding, predict pK_a and simulate implicit solvent molecular dynamics. Study by

Feig et al. (2003) found that MMGBSA model was the most successful model in ranking the binding affinities as it gave lower binding energy compared to MMPBSA. In addition, molecular mechanics based on GB model performed well compared to PB (Poisson Boltzmann) in order to rank the binding affinities of systems without metal ions in the binding sites. MMGBSA methods are typically more stable and accurate than MMPBSA methods in a study of the accuracy of continuum solvation models for drug-like molecules. This is in line with our observation where MMGBSA gave better results than MMPBSA. Therefore, the data represented here describe the ligand with best binding ability in term of MMGBSA total binding. Previous study had also shown that the MMGBSA method was a useful tool in the analysis of carbohydrate, bound and in aqueous solution. The method has successfully ranked the binding affinity of a series of carbohydrate-protein complexes as reported by Bryce et al. (2001). They also chose to use the MMGBSA model in seeking to capture the important effects of conformational plasticity and aqueous solvent on carbohydrate-protein binding thermodynamics and structure. In contrast, MMPBSA calculation was used to evaluate electrostatic potentials for static structures of biomolecules in solution. Additionally, all in all, the application of MMPBSA method for MD simulations is limited due to high computational costs and technical difficulties (Feig et al., 2003).

3.3.1.6 Per-Residue Free Energy Decomposition (DC) Analysis

Per-residue free energies of nLc₄Cer-E-protein complex were examined in order to obtain the information on important residues involved in complex binding. The characterization of protein-ligand interfaces has revealed that the binding energy is dominated by only a few important residues that have been termed as hot spots. It has been suggested by Gromowski et al. (2008) that hot spots of binding energy on a protein are predefined, such that particular residues on the protein can be designated as

“warm” sites for binding interactions. The binding free energy was then decomposed to study the contribution of each residue in the receptor and ligand interactions. This method aims to address the contributions of each residue responsible for the protein ligand interactions at the interface, which then contribute to the full description of the energetic influence on the binding affinity (Chen et al., 2015; Moonrin et al., 2015).

The plots of the decomposed energies (kcal/mol) versus amino acids residues are shown in Figures 3.13 and 3.14. This plot gives an overview of the general evolution of the molecular interactions between DENV E-protein and nLc₄Cer. The residues were found to exhibit positive and negative influences on binding to the E-protein. Negative (favourable) binding free energies indicated that the equilibrium was tilted in favour of the bound complex, while positive values indicated that the equilibrium was tilted towards the unbound protein and ligand (Woods et al., 2014). It can be seen from the graph that, nine residues are mainly responsible for the E-protein-nLc₄Cer complex binding, namely Asn37, Leu294, Met301, Ile335, Pro336, Phe337, Leu351, Val354 and Asn355 (calculated per residue free energies are -2.94 kcal/mol, -1.2 kcal/mol, -1.43 kcal/mol, -1.57 kcal/mol, -1.34 kcal/mol, -1.16 kcal/mol, -1.68 kcal/mol, -3.61 kcal/mol and -3.59 kcal/mol, respectively, subjected to MMGBSA). The analysis indicated that residues Asn37, Val354 and Asn355 contributed the most to the binding between nLc₄Cer and E-protein. Most of the residues showed consistent negative value which clearly showed that nLc₄Cer interacted successfully with E-protein. The hydrophobicity and polarity of the above residues made them crucial in structural stability and in binding to the nLc₄Cer. Moreover, these residues formed strong interactions and these interactions made major contributions to the binding free energy. However, energy coming from residues Ala35, Glu338 and Arg350 (0.05 kcal/mol, 0.7 kcal/mol and 0.08 kcal/mol respectively) had less favourable interactions with E-protein due to their low binding affinity. The free energy decomposition analysis showed that the contributions

from these residues were very small, even unfavourable for nLc₄Cer binding. This lower energy is also due to the charged residues that energetically made favourable contact with water compared with the nLc₄Cer itself. Although these residues showed less contribution to receptor-ligand interactions, they might function as stabilizing and thus facilitate binding for the nLc₄Cer.

In terms of time simulation, there was no significant trend in residues involved in the interaction as the same residues also took part in the binding. Both polar and hydrophobic residues had the ability to form hydrogen bonds by donating or accepting electron from an electronegative atom, thus resulting in stabilization of the E-protein-nLc₄Cer complex. Furthermore, the presence of charged residues stabilised the binding of nLc₄Cer to E-protein indicating that these residues were crucially involved in the binding. Residue Asn355 remained the largest contributor in the binding energy up to 50 ns which was -4.28 kcal/mol. Thus, it is believe that several residues were verified to have significant effective contributions to the complex. A study reported by Wichmann et al. (2010) had suggested that the stability of protein ligand complex interfaces were highly dependent on critical amino acids (hot spots), which contributed to a large fraction of the binding energy at a particular interface and were often surrounded by energetically less important residues. Consequently, disruptors or inhibitors of protein-ligand interaction did not necessarily target the entire interacting surface but rather could be designed to address only those residues located at the hot spots. Based on these interactions, it was found that nLc₄Cer had the potential to block DENV attachment to its candidate receptor. In order to confirm this, the exact binding location of the nLc₄Cer in the E-protein, co-crystallization of the E-protein with these compounds would be able to highlight their exact location and can be compared with the results obtained in this study.

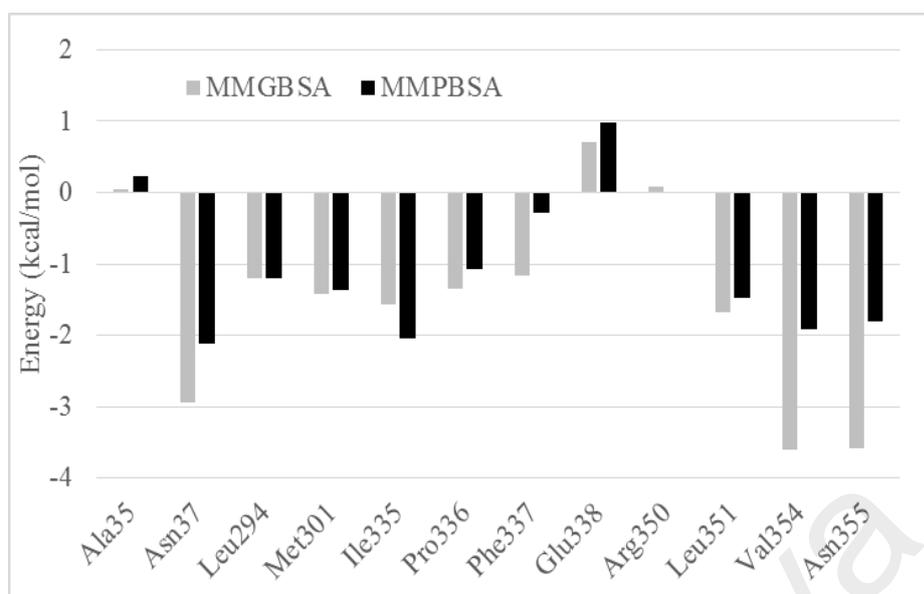


Figure 3.13: Histogram showing the calculated per-residue free energy decomposition using MMPBSA and MMGBSA approaches for E-protein-nLc₄Cer complex at 50 ns simulation.

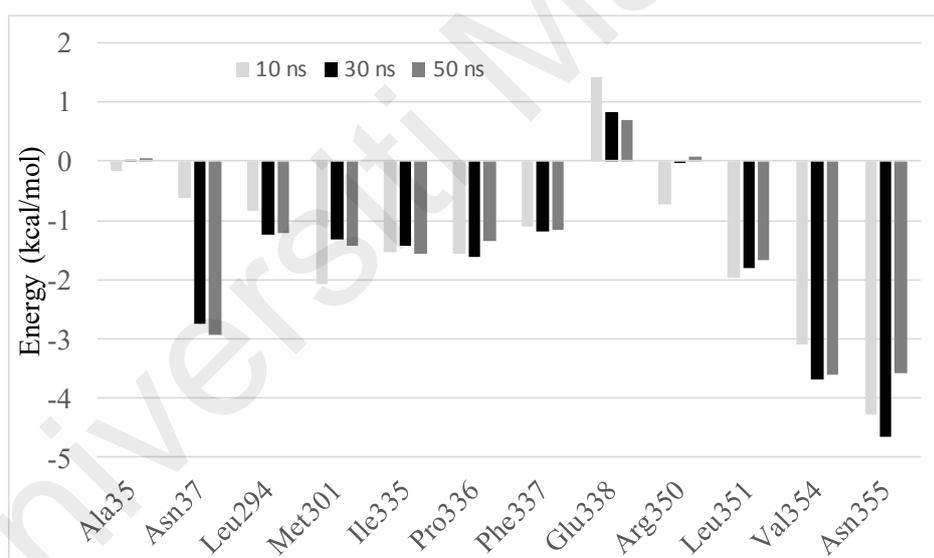


Figure 3.14: Histogram showing the calculated per-residue free energy decomposition using MMGBSA approach for E-protein-nLc₄Cer complex at different time simulation.

3.3.1.7 Analysis of E-Protein-nLc₄Cer Complex Interactions

In order to visualize hydrophobic and H-bond interactions between E-protein and nLc₄Cer, LigPlot program was employed to describe the protein-ligand interaction (Huang, Chen, Chang, & Chen, 2015). The interactions mediated by hydrogen bonds and hydrophobic contacts are shown in Figure 3.16. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by arcs with spokes radiating towards the ligand atoms they contact. Since hydrogen bonding is one of the major contributors for protein-ligand interaction, the formation of stable H-bonds between E-protein and nLc₄Cer molecule were analysed from MD simulation trajectories. This ligand was docked in positions near the hydrophobic residues in the binding pocket and formed stable hydrogen bonding interactions with the binding site residues. It formed hydrogen bonding interactions mainly with residues Ala35, Asn37, Lys38, Phe337 and Asn355 with distance 3.17 Å, 2.84 Å, 2.93 Å, 3.03 Å and 2.98 Å, respectively. The presence of functional groups made these highly polar residues very easy to form hydrogen bond, which eventually has significant effect on the protein stability. These results are also in line with a study by Almeida et al. (2013) where they reported that hydrogen bonds provided strong attractive forces sufficient to stabilize the binding of doxycycline to the hydrophobic residues thereby preventing the conformational rearrangement necessary for DENV replication.

Meanwhile, McAtamney (2008) confirmed that tetrasaccharide bound to DIII envelope glycoprotein (EGP), involving each of its carbohydrate moieties in his investigation of nLc₄Cer ligand binding to EGP. Epitope mapping by STD NMR spectroscopy also revealed that the H-1 protein of the N-acetyl-D-glucosamine (GlcNAc) made closest contact with DIII via its N-acetyl group. There were also other factors such as polarization and induced dipoles that could contribute to the binding. Water was found to play a significant role in protein-carbohydrate interactions. They

revealed that hydrogen bonds mediated by water were as strong as those without the intervening water bridges (Wacowich-Sgarbi, 2000).

On the other hand, nine hydrophobic interactions (Figure 3.16) were found with residues Lys36, Leu351, Val354, and Pro356. These residues were found to be conserved and stable throughout the simulations. The existence of these hydrophobic residues resulted in a favourable binding energy that stabilized the complex. Despite their distinctive polarity, carbohydrates also contain hydrophobic patches on their surfaces, due to the steric disposition of hydroxyl groups. These patches form contacts with the hydrophobic zones in the protein's binding site, which generally involve aromatic, as well as non-polar amino acids (Wacowich-Sgarbi, 2000). Therefore, the presence of these hydrogen bonding and hydrophobic interactions are crucial for binding affinity of E-protein-nLc₄Cer complex.

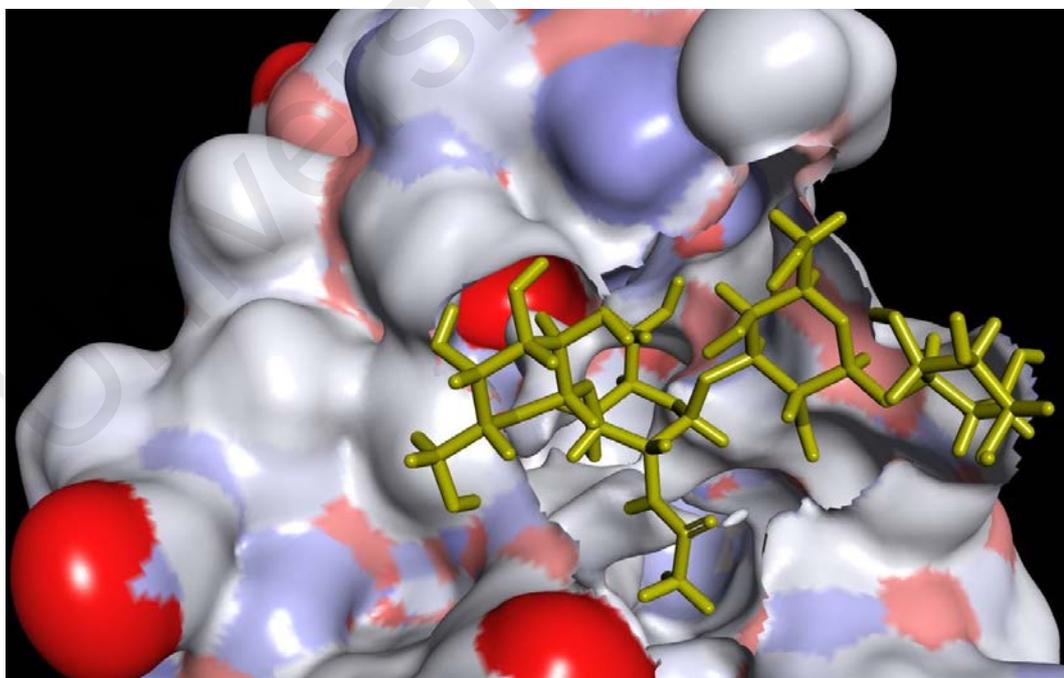


Figure 3.15: Connolly surface representation of nLc₄Cer at the binding site.

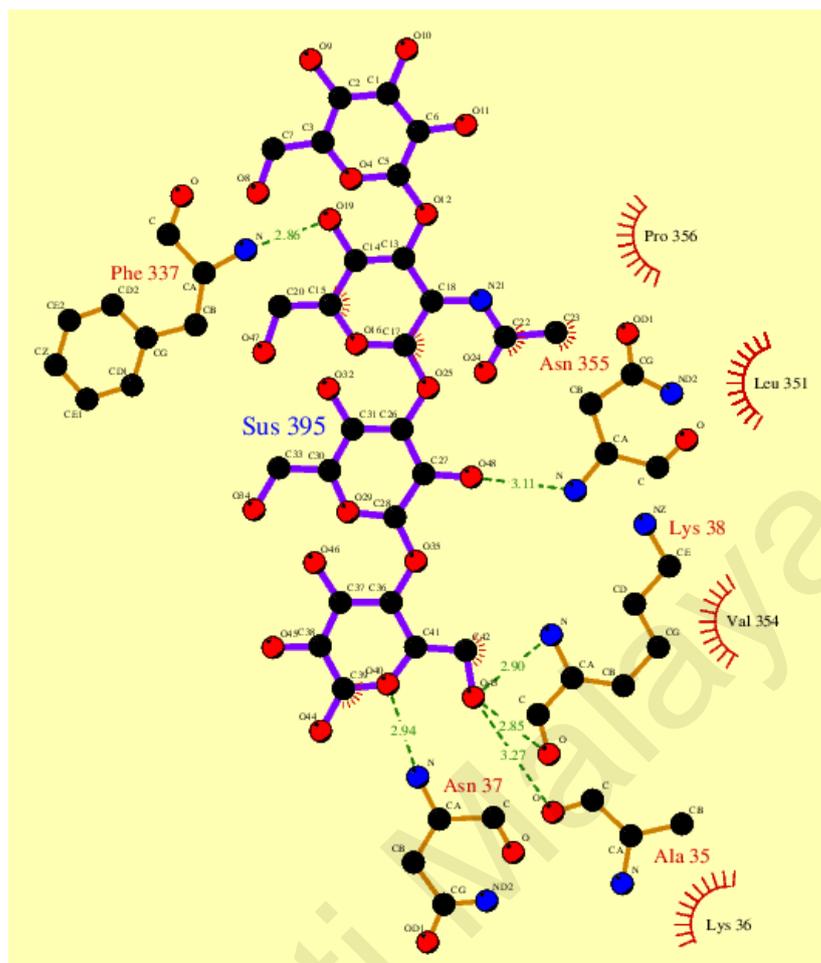


Figure 3.16: Schematic diagram (2D) of residues in the binding site which exhibit interactions with nLc₄Cer at 50 ns MD simulation, obtained using the Ligplot program. Keys for the plot are as follows: Ligand nLc₄Cer is shown in stick and ball, H-bonds are shown as dashed green lines and hydrophobic contacts are indicated with spoked red arcs.

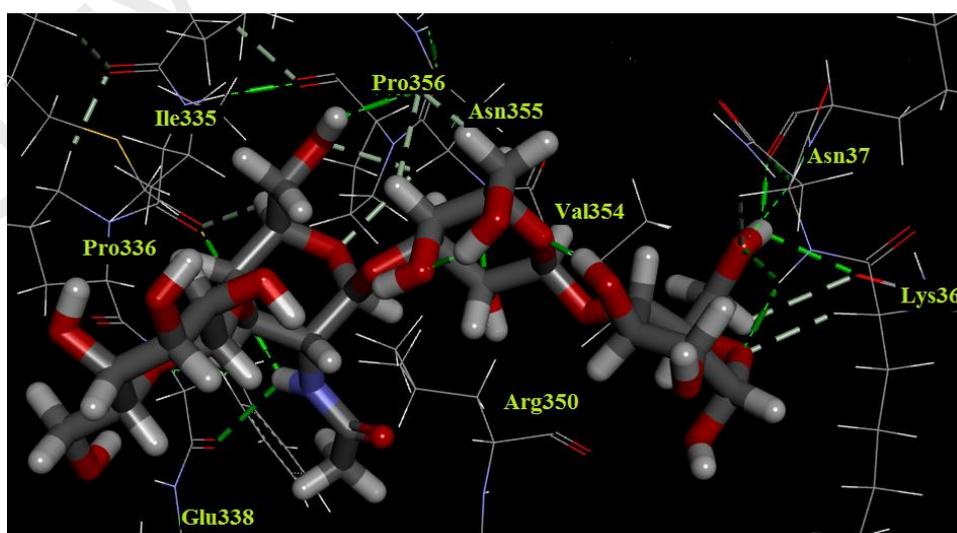


Figure 3.17: View of nLc₄Cer (stick) at the binding site of DENV E-protein (line). Residues interacting with ligand are shown as sticks.

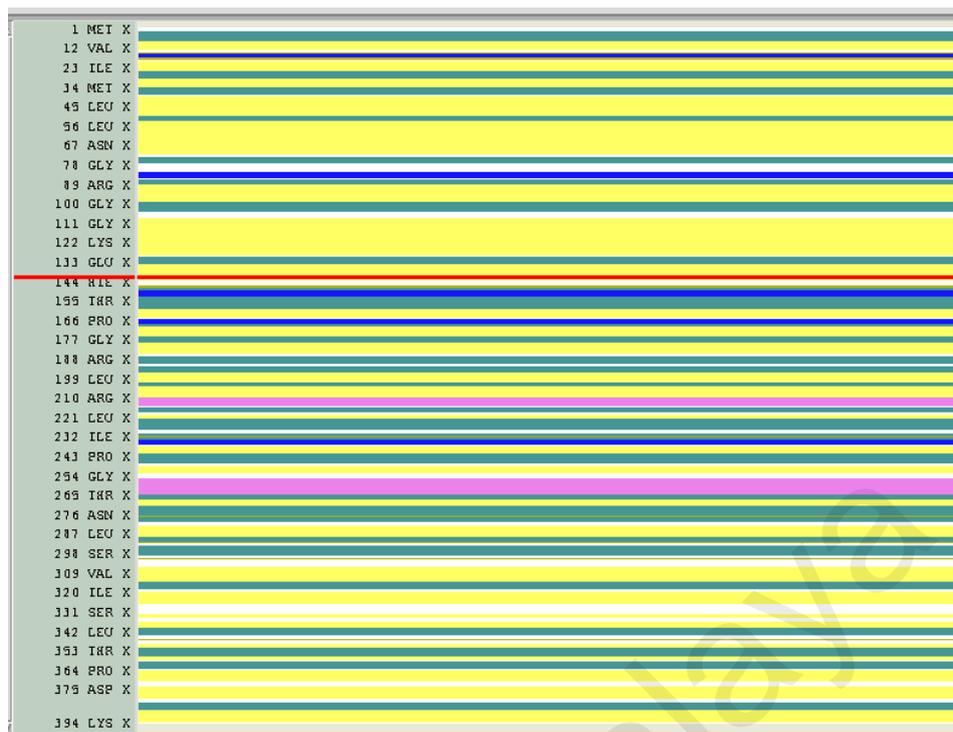
3.3.1.8 Secondary Structure Analysis

Secondary protein structure is important as it represents the local conformation of amino acids into regular structures. A secondary structure analysis was performed using the VMD program, which determines the existence of hydrogen bonds as criteria for the presence of secondary structure. In Figure 3.18, the secondary structure element of each residue as a function of time was plotted. Colours were used to distinguish between secondary structure types. From the figure, the most important colours found was yellow for extended configuration and turquoise for turn configuration. These results indicated that the extended conformation played an important role in maintaining the protein structure and became the predominant secondary structure for the nLc₄Cer ligand with E-protein. The overall secondary structure pattern was maintained during the 50 ns MD simulation, although there was slight change at some points of time. The timeline analysis of the secondary structure during the 50 ns of MD simulations showed that the protein domains were stable during the trajectory. All the rest of the residues were in isolated bridge, α -helix, 3-10 helix and coil regions, constantly alternating during simulation, as shown in the Figure 3.18. The simplest secondary structure element is the β -turn. MD simulation of E-protein-nLc₄Cer complex in water had demonstrated that turn-like structures could rapidly form, disappear and reform, with coils and turns working as the mobile regions. Meanwhile, α -helices are the commonest secondary structural elements because they are generated by local hydrogen bonding between C=O and N-H groups close together in the sequence (Richardson & Richardson, 1990).

Secondary structure contributes significantly to the stabilization of the overall complex fold. The arrangement of secondary structure elements provides a convenient way to classify types of folds. This prediction is beneficial because the pattern of secondary structure elements along the chain can be characteristic of certain overall

protein folds. In general, secondary structure elements such as α -helices and β -strands are highly important for protein's structure and protein's function. In fact, hydrogen bond plays an important role in the formation of these two common secondary structures in proteins (Kukkurainen, 2009). Since secondary structures are stabilised by H-bonds between backbone carbonyl oxygen atom and the amide hydrogen atom, the breaking of backbone H-bonds can lead to large conformational change in protein structure and thus could drive the protein away from its native state in MD simulation (Duan, Mei, Zhang, Tang, & Zhang, 2014). Therefore, the collapse in E-protein-nLc₄Cer complex during simulation was probably accompanied by changes to its secondary structure.

This secondary structure study contributes significantly to the understanding in the stabilization of the overall protein fold. The arrangement of secondary structure elements provided a convenient way to classify types of folds. Such prediction is basically useful because the trend of secondary structure elements along the chain can be characteristic of certain overall protein folds. Therefore, it is reasonable to examine the development of secondary structures as a first approximation to total folding properties of an amino acid sequence.



- T Turn
- E Extended Configuration
- B Isolated Bridge
- H Alpha Helix
- G 3-10 Helix
- I Pi-helix
- C Coil (none of the above)

Figure 3.18: Secondary structure of E-protein-nLc₄Cer complex at 50 ns.

3.3.2 Binding Pathway of Neolactotetraosylceramide towards Dengue Virus Type 2 (DENV2) Envelope Protein

3.3.2.1 Analysis of E-Protein-nLc₄Cer Interactions

Docking method monitors the changes in E-protein-nLc₄Cer interactions as the ligand moves along the association pathways. Association processes are usually under kinetic, rather than thermodynamic control, especially, when several macromolecules compete for the same binding site or when a protein encounters alternative pathways (Zhou & Bates, 2013). Major questions include which factors control the formation of the protein-ligand complexes and whether association pathways really occur. To answer these questions, a computational approach was consistently used to analyse the complete ensemble of the association pathways. In this work, a docking approach that directly revealed the ensemble of neolactotetraosylceramide (nLc₄Cer) pathways to the binding pocket is presented (Figure 3.19). Preliminary results suggested that protein-ligand association method would be able to identify surface regions in which binding is most likely to occur (Camacho & Vajda, 2002). By using this method, multiple pathways through which the ligand approached the entrance of the binding pocket could be observed from the protein surface. In order to identify the docking pathway of nLc₄Cer towards DENV E-protein, three possible docking pathways connecting to the ligand to the binding pocket were identified. Figures 3.20 to 3.22 show the direction and conformation of nLc₄Cer to bind at different directions of E-protein which are along the x, y and z axes. There were three pathways leading to the binding pocket. These major pathways led to these docking poses after the ligand entered the mouth of the binding pocket. Starting from the ligand outside the receptor, the molecular docking simulations led to a ligand docked inside the binding pocket resulting in ligand binding to the binding pocket. The ligand was forced to proceed along these pathways to associate. These figures revealed that, there are multiple pathways (x, y and z axes) leading from the E-protein surface to the binding pocket entrance. From these variety of pathways

they was only one pathway with the lowest binding energy that led to successful docking after the ligand entered the pocket, as shown in Figure 3.23. The different starting structures yielded a variety of exit pathways.

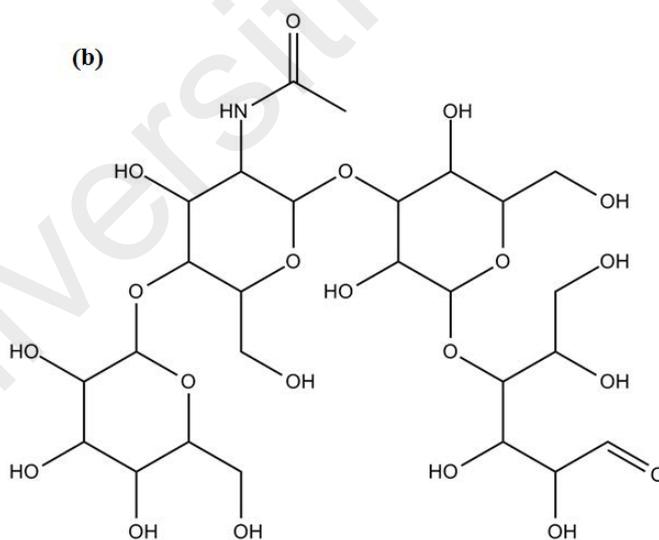
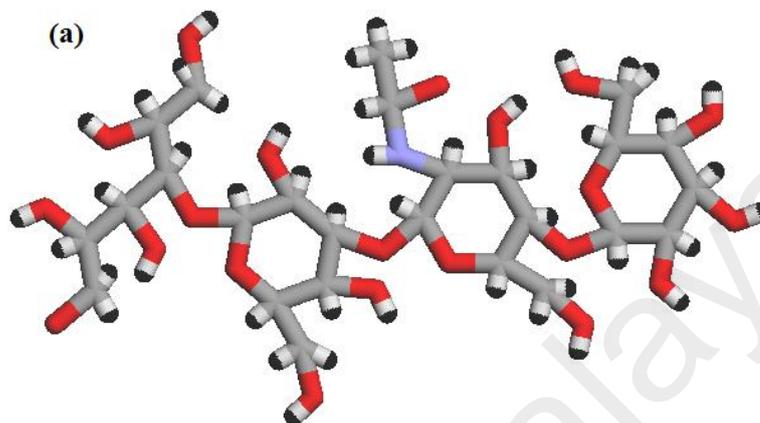


Figure 3.19: Chemical structure of neolactotetraosylceramide (nLc₄Cer), C₂₆H₄₅NO₂₁ in (a) 3D and (b) 2D structures.

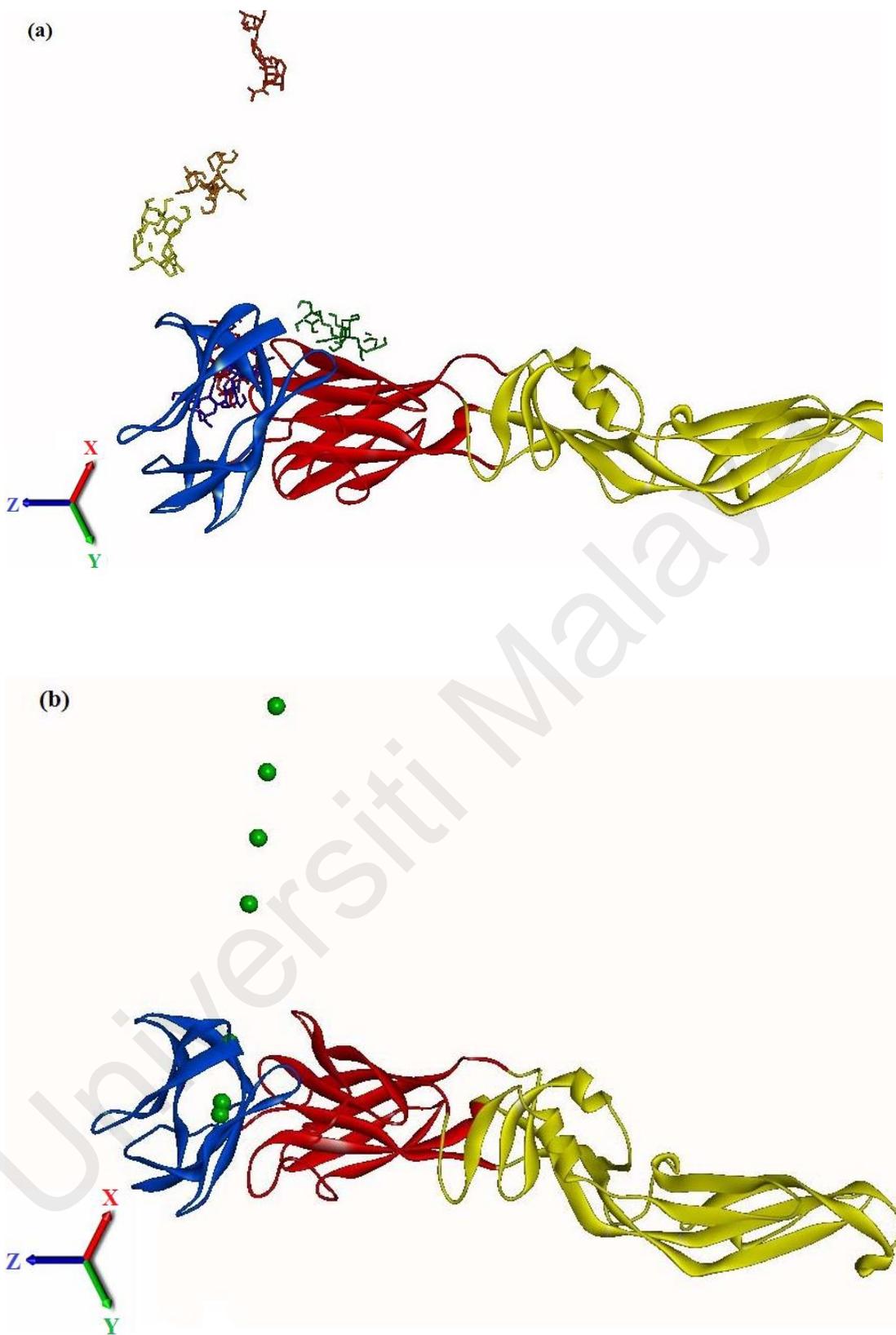


Figure 3.20: Ligand pathway along x-axis shown in stick (a) and CPK (b).

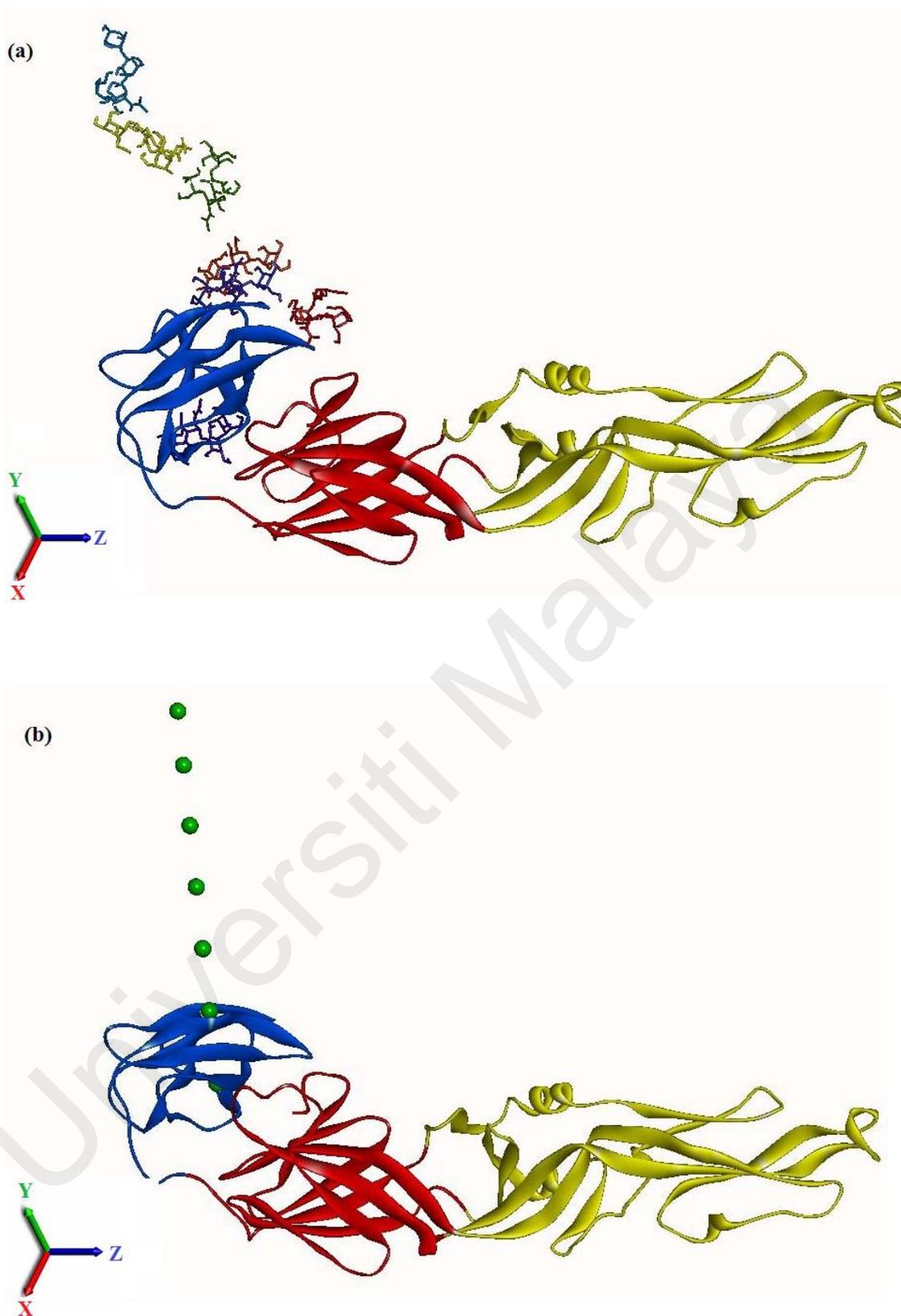


Figure 3.21: Ligand pathway along y-axis shown in shown in stick (a) and CPK (b).

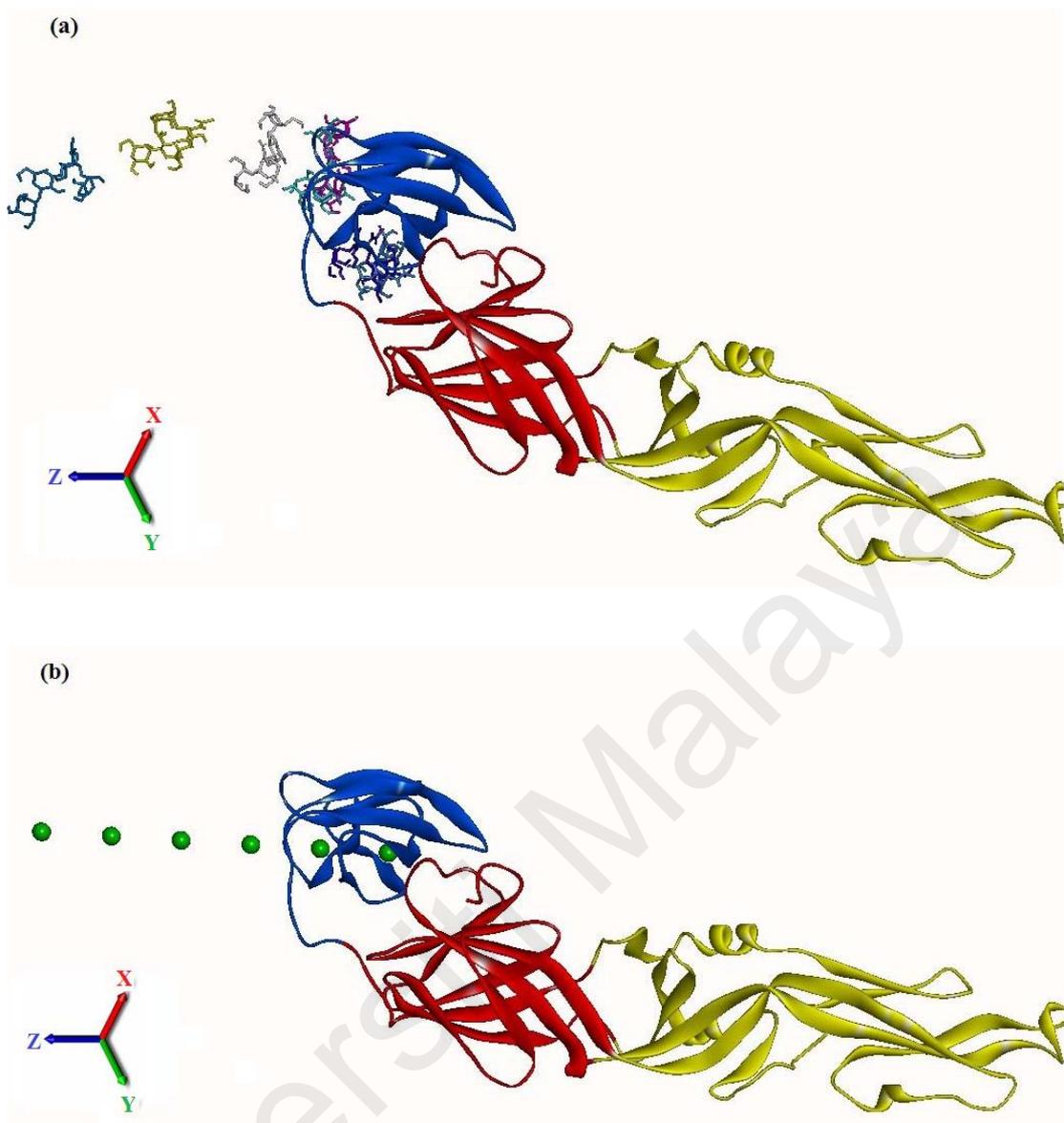


Figure 3.22: Ligand pathway along z-axis shown in stick (a) and CPK (b).

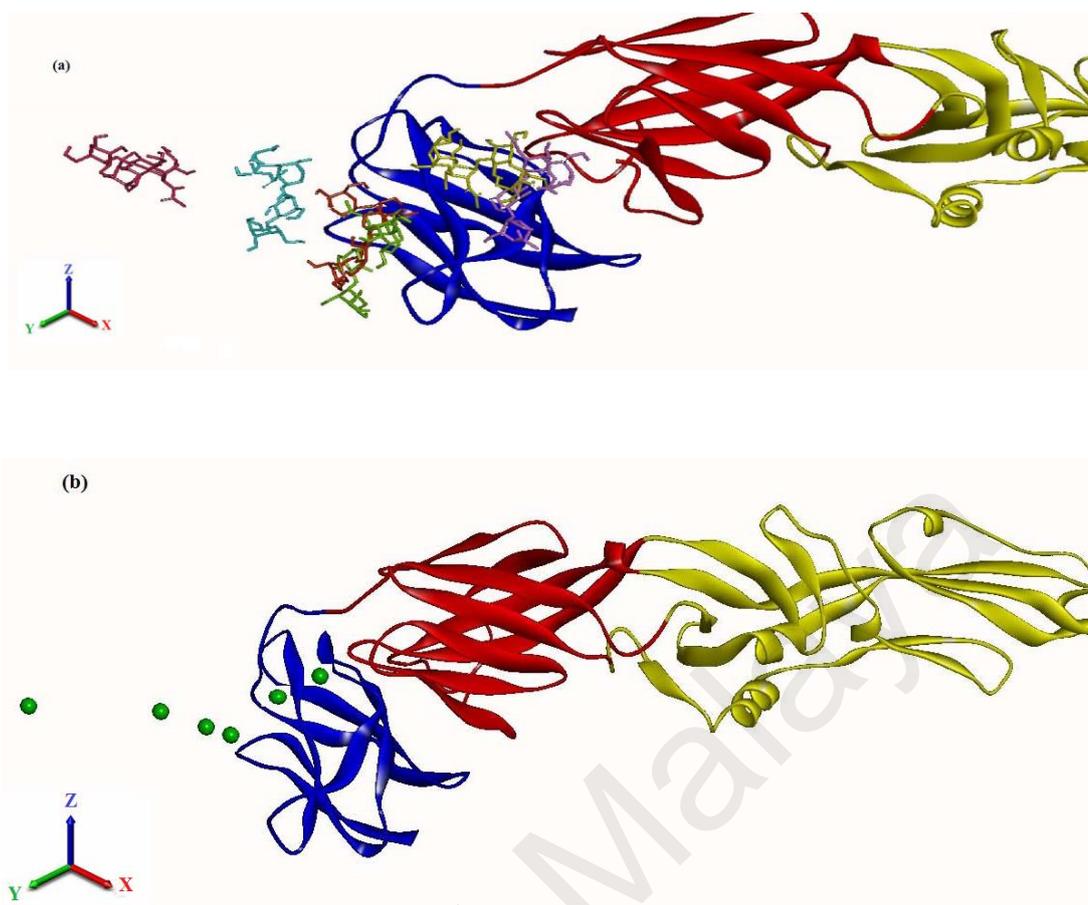


Figure 3.23: Optimized ligand pathway shown in stick (a) and CPK (b).

The ligand diffusion through the protein matrix is a complex dynamical process involving a large number of different pathways. Docking simulation had successfully revealed multiple pathways for ligand binding as this approach treat the ligand as fully flexible molecules (Sonoda, Martinez, Webb, Skaf, & Polikarpov, 2008). The ligand entered the ligand-binding pocket predominantly along a groove on the protein surface. In this study, nLc₄Cer followed three major paths since it moved along the groove toward the entrance. However, not many ligands had been reported to be able to enter the pocket directly from the solution. Wallach and co-workers (2010), in their study, had mapped each drug to several potentially affected pathways by docking the ligand into a set of various pathways in human proteins. Meanwhile, in a study by Huang and Wong (2012), it was revealed that two major docking pathways connected the binding pocket and the protein surface. The movement of the peptide along the pathways was found to couple with residues in the protein activation loop. Even though these residues might not affect binding affinity significantly, they could influence the kinetics of peptide entrance and release. Path I was suggested to represent a main pathway for ligand dissociation. They also claimed that Paths II and III were also likely to be ligand escape routes. They believed that different escape paths were favoured in particular situations, suggesting that it was possible to design ligands that only associated stably with their receptor. Their study also revealed that, proteins that could participate in various pathways would be able to offer multiple biological processes.

Beside of binding following different pathways, a ligand can also bind via one pathway such as in ligand flux analysis, which was demonstrated in a study by Negami et al. (2014) that where the ligand entered the ligand-binding pocket through a specific pathway. Once the ligand entered the pocket, it moved toward the ligand-binding site through a narrow tunnel. Flux analysis result that compared between the two protein systems clearly indicated that the ligands tend to move through grooves on the protein surface and not over protrusions. The groove was mainly formed by hydrophobic

residues. This movement was due to the ligand effectiveness to interact with more protein atoms in a groove rather than on the surface of a protrusion. In fact, the longer ligand binding pathway observed in the levansucrase-sucrose system compared to other system was due to the stronger interactions of sucrose with protein atoms.

In order to move within protein receptor, ligand uses the protein intrinsic flexibility, which was found to play an important role in facilitating ligand entry into or exit from the binding pocket (Huang & Wong, 2007). This is specifically important for a ligand that enters the protein from the proximal side, because it gives the ability to reach its binding site from this region. The large movement of some residues observed during docking suggested the protein flexibility played an important role in facilitating docking. The flexibility of Ile223 residue could also aid binding with ligand by moving away from the binding pocket to open up space for the ligand (Huang & Wong, 2007).

Docking simulation had successfully revealed multiple pathways for ligand binding as this approach treated the ligand as fully flexible molecules (Sonoda et al., 2008). Besides that, the polarity of the ligand also plays a predominant role in its migration within the protein (Mouawad, Tetreau, Abdel-Azeim, Perahia, & Lavalette, 2007). As shown in the study by Aci-Seche (2011) starting with the ligand outside the receptor, the simulation led to a ligand docked inside the binding pocket resulting in a structure very close to the holo-form of the complex. This entry process was also guided by hydrophobic interactions, and the most interesting part was that the entry pathways were very identical to the exit pathways. Therefore, knowledge of ligand entry or exit pathways of a receptor would give useful information on how to discriminate between different ligand pathways that might have favourably docked in the binding pocket.

Table 3.4 illustrates the binding energy of the complex as the ligand approached the binding site and Table 3.5 shows the binding energy of the complex following the optimized ligand pathway. The result revealed that the binding energy values slowly

decreased as nLc₄Cer moved from the E-protein surface towards the binding pocket. Spontaneous structural changes occurred to the protein system when the ligand bound to the receptor and this was characterized by free energy reduction. The change in the system's free energy will determine the extent of protein-ligand association. This process of a ligand docking to a binding site mimics the natural course of interaction between the ligand and its receptor via the lowest binding energy pathway. In this way, the ability of biological system to reach stability and vitality could be obtained through the "communication" between the system's components. All putative drug pathway interactions are inferred by protein-ligand docking. The drug-pathway interaction is the sum of docking scores over all proteins belonging to the pathway. This pathway information is highly recommended to predict protein-drug interactions especially in situation where multiple drugs modulate the same protein target. Moreover, it also allows to identify cases in which the modulation of a pathway via different proteins affects the same biological process (Wallach et al., 2010). Zhao et al. (2013) demonstrated the power of pathway docking to predict the novelty of *in vitro* enzymatic activities and *in vivo* physiological functions. Meanwhile, Negami et al. (2014) agreed that docking simulations produce acceptable results and are useful to study the protein-ligand association process. In their study, 100 times of docking simulation with different initial placements of the ligands were done. They observed that the ligand molecules entered into the correct ligand binding pocket and were stable inside the binding pocket. These results suggested that molecular docking technique provides the adequate and complementary alternative to describe this protein ligand association process.

Table 3.4: Binding energy (kcal/mol) of nLc₄Cer towards E-protein from different pathways.

	Path 1	Path 2	Path 3	Path 4	Path 5	Path 6	Path 7
x-axis	-2.47	-1.39	-2.17	0.08	6.76	7.15	7.16
y-axis	-2.47	-1.19	-1.47	0.25	7.15	7.16	7.16
z-axis	-2.47	-2.41	-1.90	-1.77	3.20	7.16	7.16

Table 3.5: Binding energy (kcal/mol) of optimized pathway for nLc₄Cer-E-protein complex selected from x, y and z axes.

Optimized pathway	Path 1	Path 2	Path 3	Path 4	Path 5	Path 6	Path 7
Binding energy (kcal/mol)	-2.47	-2.41	-2.17	-1.77	3.20	7.15	7.16

Ligand binding pathway represent an important component of ligand affinity, since high energy barrier may block the ligand from entering the binding site (Aci-Seche et al., 2011). Furthermore, determination of ligand entry is a challenging task both experimentally and computationally, especially when experimental structures of some components do not exist due to experiment ligand pathways do not give rise to any spectroscopic signal (Kurcinski, Jamroz, & Kolinski, 2011). Wei et al. (2004) reported, in their study that, the calculated electrostatic energy increased (become less favourable) as charged residues were brought closure together upon inhibitor binding. During the closure of the active site, a favourable reduction of non-polar surface was balanced by the unfavourable electrostatic energy increment. The energy differences among the different conformations were meant to reflect the general trend that ligand-bound conformations might have higher energy than the apo conformation does.

According to the study by Hsu et al. (2013), the pathway-based screening strategy was important to identify multi target inhibitors and to elucidate protein-ligand binding mechanisms. The concept of the pathway-based screening technique was to simultaneously screen various proteins pathway and extract conserved binding environments of these proteins to identify different target inhibitors. This pathway-

based screening strategy could be used to boost the hit rate because the pathway anchors were often highly conserved and basically crucial to perform biological functions. This proves that the pathway anchors regularly play important roles for ligand binding. Thus, the compounds that match the pathway anchors were normally potential inhibitors of the target proteins. This study had proven that, even though there were four proteins that had different functions, their pathway anchor residues had similar physicochemical properties to interact with their substrates and cofactors. Meanwhile, Held and Noe (2012) had reviewed approaches to study the probability distribution of association pathways. It was reported that association might occur via pre-binding sites outside the native binding site that metastably associated ligands, and also from binding or dissociation that had the potential to occur.

As for the current study, the binding of nLc₄Cer ligand to E-protein involved a number of different forces, which in the end contributed to the binding affinity of the complex. This receptor-ligand association was driven by electrostatic interaction and desolvation process resulting in the possibility of finding the approximate binding region. The role of electrostatics in binding pathways had been particularly well studied (Holgerrsson et al., 2005). As the interaction partners approach each other, electrostatic forces becomes significant for favourably interacting molecules (Held, Metzner, Prinz, & Noe, 2011). Fast association is achieved by favourable electrostatic interactions between the proteins, which enhance their association rate. Long range electrostatic interactions have been shown to play a critical role in the binding pathways of tightly-bound complexes. The association process is occasionally broken down into two steps. First, the binding partners encounter each other through a random collision. Second, as the binding surfaces are correctly aligned, the molecules then dock to form the native stereospecific complex. Interactions that stabilize the transition state would accelerate molecular association rates (Meneses & Mittermaier, 2013).

The relatively high association enthalpies (-2.47 kcal/mol) (Table 3.5) of E-protein-nLc₄Cer interactions could also be due to higher degree of hydrogen bonds per unit area as well as more favourable angles and shorter distances between the molecules involved in the hydrogen binding. Upon E-protein-nLc₄Cer complexation, water molecules tend to escape to the bulk with a concomitant decrease or increase in energy, depending on their pre-existing molecular interactions. After E-protein-nLc₄Cer binding had occurred, the water molecules then arranged themselves according to the new surface exposed. In addition, the inner core structure of an oligosaccharide had been shown to influence the binding affinity of the E-protein-nLc₄Cer complex. In order to allow rapid rates of association and dissociation, slight modifications are compulsory to maximize the number of favourable interactions in the ligand bound, as well as in the free, protein structures. For instance, many ligands bind between domains that move together to include the ligand. Hence, the ligand would be able to associate and dissociate and the interactions between protein and ligand are maximized (Ehrlich, Nilges, & Wade, 2005).

Generally, the communications in biological system are mediated through this association process between molecules. In most cases, the approaches of rigid-body docking of two or more structures are quite successful. Normally, docking pathway is accompanied by conformational changes that result in the reduction of the conformational entropy. However, this conformational changes are limited and only involve the side chains at the association interface (Ahmad, 2012). There are also alternate methods such as steered and biased molecular dynamics (MD) approaches, where biases are applied to steer a ligand to enter or be released from a protein that can speed up the ligand simulation of loading into or unloading from a protein. MD simulation may provide alternative way in finding pathways for ligand entry and exit from receptor. Unfortunately, these simulations are expensive to be done especially when many runs have to be carried out in order to identify representative pathways and

to estimate activation barrier as well as to perform long runs to minimize the artificial effects of the applied biases. In addition, some of the simulation such as ligand dissociation dynamics is extremely slow in comparison to the timescales accessible to present simulation techniques and computer resources. This does not mean that the actual event of ligand dissociation takes too long, but it is clear that conformational sampling cannot be done effectively in conventional MD simulation (Martinez et al., 2005). Therefore, it is useful to search for alternative ways that utilize different approximations. LaBute et al. (2014), had successfully predicted association of off-target effects by using molecular docking scores for drug-protein matrices.

3.4 Conclusion

In this study, the binding of nLc₄Cer to E-protein had proven effective through prediction of binding sites based on protein cavity and MD simulation methods. Several parameters such as energy, temperature, RMSD, RMSF and radius of gyration that can be used to check the MD simulation accuracy had shown consistency during the 50 ns simulation time. Meanwhile, cluster analysis had presented nine conformational clusters, with each cluster not showing much difference in binding energy. Nevertheless, the cluster group with the lowest binding energy was chosen for further MD study and was confirmed to have reasonable binding mode. In addition, post analysis on the simulated DENV E-protein-nLc₄Cer complex was done by calculating the binding free energy and it was found that electrostatic and vdW interactions seemed to operate mostly during the long-range attraction between the E-protein and nLc₄Cer before the complexes were formed. Meanwhile, through binding energy decomposition analysis, several important residues that contributed to the complex binding energy have been identified. In addition, it was found that hydrophobic and hydrogen bonding interactions were responsible for the high affinity of nLc₄Cer towards E-protein.

Meanwhile, docking approach was performed to systematically investigate E-protein-nLc₄Cer association pathways. This method provides an extensive analysis of association pathways by which a ligand approaches its target protein. The result analysis demonstrated three association pathways leading from the surface of the E-protein to the correct docking pathway. The ligand flexibility was found to play an important role in ligand entry and exit from the binding pocket. Knowledge of association pathway would be useful in discriminating different pathways that could have been favourably docked the ligand into the binding pocket. In addition, this method allow us to efficiently study the ligand binding processes and may help those who study drug discovery to find optimal association pathways and to design those ligands with the best binding kinetics (Chang, Trylska, Tozzini, & McCammon, 2007). The most important thing is the information gained from this study is very beneficial for drug design as it compare the relative ease of different ligands to enter their target protein and the durations in which different drug candidates would stay in the binding pocket once they get there (Huang & Wong, 2007). How long a ligand is likely to remain bound to its receptor, thus generating the desired pharmacological effects. Ligands that remain bound to their receptor for a longer time are pharmacologically more appealing than those characterized by a short-lived complex (Vivo et al., 2016).

CHAPTER 4: STUDIES INVOLVING PEPTIDES

4.1 Introduction

Peptides form a major part in the biological signalling mechanism. They carry information to cells and ultimately become main regulators of life. Normally the peptides are used in skin care to increase the amount of collagen being produced in the skin, or in other words, peptide is used to minimize wrinkles and relax muscular activity (Ahsan, 2019). Peptide synthesis is also become a main factor to confirm the structure of natural proteins in medical research and to investigate how protein structure and function are controlled by the amino acid sequence (Brummelhuis, Wilke, & Borner, 2018). Moreover, peptide-based drugs are becoming an important aspect in the pharmaceutical drug market, especially when many methods have been developed to improve production and to reduce metabolic breakdown (Aneja, & Chaiken, 2013).

In previous study (Xu, Rahman, Othman, Hu, & Hu, 2012), a series of peptides were designed to target the domain III (DIII) of envelope (E) protein with the aim to block dengue virus (DENV) entry into target cells. DIII is a crucial part of E-protein as it is the receptor binding domain. It is also very antigenic and antibodies to this domain are able to efficiently neutralize virus infection (Guardia, Quijada, & Lleonart, 2017). Due to this important role of E-protein and its DIII, the search for peptides with binding activity to this protein is compulsory in order to find peptides that are able to retard the infection process. These designed peptides (Xu et al., 2012) should have the ability to block virus infectivity and not become toxic to the target cell. This is supported by *in vitro* drug-screening assay study performed by Baharuddin et al. (2014) that was developed to assess the ability of the peptides to interact with DIII. Their results observed the peptides had shown binding with DIII region of E-protein and the binding between them appeared to be cooperative. The affinity of peptide with code DS36opt for the DIII protein (K_d of $9.31 \pm 0.15 \mu\text{M}$) was slightly higher than that for DN58opt

(K_d of $9.44 \pm 0.28 \mu\text{M}$). Meanwhile, the peptide coded DN58wt exhibited weaker binding affinity for the DIII protein (K_d of $10.79 \pm 0.17 \mu\text{M}$) compared to DS36opt and DN58opt peptides.

Binding affinities of these peptides towards E-protein were predicted by using a combination of docking and explicit solvent molecular dynamics (MD) simulation methods. Blind docking was performed to investigate the possible binding modes. The interactions within the bound complexes were further investigated by using the Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) and Molecular Mechanics Generalized Born Surface Area (MMGBSA) methods. Specifically, the peptides encoded as DN57opt, DN81opt, DS27opt, DS04opt, DS03opt, DS10wt, DS36opt, DS36wt, DN58opt and DN58wt, which were designed by Xu et al. (2012) using a Monte Carlo method, were used. These peptides comprised of short sequences of 20-29 amino acids in length.

Meanwhile, peptide synthesis is the process of making short sequences of polypeptides by adding one amino acid at a time and linking them via peptide bonds. In this study, solid phase peptide synthesis (SPPS) method was used, which include peptide synthesis, cleavage and precipitation. Due to the possibility of unintended reactions and in order to control the coupling reaction, protecting groups are compulsory (Isidro-Llobet, Alvarez, & Albericio, 2009) in peptide synthesis. The synthesized peptides are often purified using reverse-phase high performance liquid chromatography (RP-HPLC). After the desired sequence of amino acids has been obtained, the peptides are then removed from the polymeric support. The advantage of solid phase peptide synthesis (SPPS) is that, it allows scientists to control the order of reactions, such as to control which amino acids are made to react first, which ones come second and so on. This approach ensures that only the desired amino acid sequence in the peptide is synthesized.

4.2 Methods

4.2.1 Binding Free Energies of Peptides Inhibiting Dengue Virus (DENV) Envelope Protein by using Docking & Molecular Dynamics Simulation

4.2.1.1 Materials

The three-dimensional structure of pre-fusion dengue virus (DENV) E-protein was retrieved from the Protein Data Bank (<http://www.rcsb.org/pdb>; accession code 1OKE). The pre-fusion structure of DENV E-protein was used in this study as the target structure for the computational method applied in designing peptide entry inhibitors (Alhoot et al., 2013). For the protein molecule, all heteroatoms which included the drugs, chlorine atoms, water, glycerol and other molecules originating from the crystallization buffers were removed. The peptides used in this study were comprised of 20-29 amino acids in length (Table 4.1).

Table 4.1: Peptides used in this study (Xu et al., 2012).

Peptide	No. of aa	Sequence	Location
DN57opt	28	WYFIRKEFFERIRFLPQRNPHRRDDEWD	205-232
DN81opt	19	WIFIRYEFFRSFKFLWRGN	205-223
DS36opt	20	RHWEQFYFRRRERKFWLFFW	351-370
DS27opt	20	KEYFRFFHCHNHQREWHWH	261-280
DS04opt	20	IWWRPDWPWFIFYFREW RW	31-50
DS03opt	20	FPDFHHD RY YHFHWKRYQH	21-40
DS10wt	20	VCKHSMVDRGWGNGCGLFGK	91-110
DS36wt	20	LITVNPIVTEKDSPVNIEAE	351-370
DN58opt	21	TWWCFYFCRRHHPFWFFYRHN	374-394
DN58wt	21	GDSYIIIGVEPGQLKENWFKK	414-447

aa: amino acid.

4.2.1.2 *In Silico* Peptide Folding

AMBER 12 program (Salomon-Ferrer et al., 2012) was used to model the 3D structures of the peptides. Peptide folding was performed to simulate the possible starting structures of the peptides prior to docking. Wild type peptides with codes DS36wt, DS10wt and DN58wt have 100% identity to a part of DENV E-protein amino acid sequence, while the new optimized peptides with codes DN57opt, DN81opt, DS27opt, DS04opt, DS03opt, DS36opt and DN58opt are totally different compared to wild type amino acid sequence (Xu et al., 2012).

All the peptide folding simulations started with extended conformations of the peptides to demonstrate that these peptides could be folded from an arbitrary starting structure (Simmerling, Strocbine, & Roitberg, 2002). Topology and parameter files of the unfolded peptides (wild type and optimized peptides) were generated via the xLEAP program of the AMBER 12 simulation package (Obiol Pardo, 2008). The peptide folding simulation was performed using an all-atom classical simulation with AMBER ff99SB force field. The best 3D model was selected according to the lowest energy generated by ptraj program, considering that the lowest energy model indicated best peptide stability, and thus, best quality for the peptide structure.

Energy minimization and MD simulations were subsequently performed using PMEMD.CUDA module in AMBER 12 (Salomon-Ferret, Gotz, Poole, Grand, & Walker, 2013) on NVIDIA GPU Quadro 600, which sped up the simulation time with a factor of 22.6 to 31.7 improvements compared to regular CPU, to obtain the trajectory files from each simulation. A short minimization (total of 1000 steps) of the starting structure was performed which comprised 500 steps of steepest descent followed by 500 steps of conjugate gradient. This process did not cause the peptide to completely fold since minimization only drive the starting structure to a local minimum. It reduced steric clashes of the structure, fix hydrogen atoms positions so that the system was stabilised

prior to MD simulation. Since the system was small, non-bond energy cutoff was employed in the calculation of the full Generalised Born solvation energy.

For the peptide folding simulation, the system was heated up over time step of 50 picoseconds in a total of 7 stages. Heating in stages reduced the chances that the system would blow up by allowing it to equilibrate at each temperature change. Normally, MD simulation would be run at 300 K. In this study, the system was heated up to 325 K to avoid being kinetically trapped in local minima, leading to a faster folding pathway (Simmerling, Strocbine, et al., 2002). A time step of 2 femtoseconds was used to integrate the equations of motion. The SHAKE algorithm was used to fix the bond distances consisting of hydrogen and heavy atoms, and the Berendsen thermostat was used for temperature control. Finally, peptide folding was simulated for 100 ns at a constant temperature of 325 K and pressure (1 atm) by the Berendsen weak-coupling algorithm (Simmerling, Strocbine, et al., 2002).

4.2.1.3 Docking of Peptides to DENV E-Protein

For each peptide, the folded structure with the lowest energy along the 100 ns simulation was chosen as the starting structure for docking to DENV E-protein (PDB id: 1OKE). Docking was done by using the SwarmDock server (bmm.cancerresearchuk.org/~SwarmDock/), which performed flexible protein-protein docking using the SwarmDock algorithm, where the peptide and side chains of the E-protein were kept flexible throughout the docking runs. This docking program was used to predict and assess the interactions between DENV E-protein and peptides. The server started with the pre-processing of input structures where the structures were checked for structural correctness, modelling of missing and non-standard residues, and structure minimization using the CHARMM molecular mechanics package. This was then followed by docking of the peptides using a hybrid particle swarm optimisation where a

set of approximately 120 starting positions were generated so that they evenly spaced around the DENV E-protein. The whole docking process was repeated four times at each starting position and the best structure found during the optimisation was kept for the final post-processing stage. The job was then submitted to the server as a full blind docking run. After the docking had completed, all the poses underwent minimization using CHARMM force field and the minimised structures were then re-ranked. Finally, these structures were clustered and the results for this submission were then returned in PDB formatted structures, allowing for some visualisation of the clustered solutions. The SwarmDock docking algorithm had been selected in this study because this server was widely used and had shown considerable success in the critical assessment of predicted interactions (CAPRI) experiments. Therefore, this type of docking protocol is believed to perform comparatively well when evaluated against other top docking algorithm (Torchala, Moal, Chaleil, Fernandez-Recio, & Bates, 2013).

4.2.1.4 Molecular Dynamics Simulations of Peptide-E-Protein Complexes

Docked peptide-E-protein complex structures with the best pose and lowest binding energy were selected for MD studies using AMBER 12. Each complex was solvated with a truncated octahedral box of TIP3P water with the box boundaries of at least 12 Å from the complex, and was neutralized with Na⁺ and Cl⁻ ions. The fully solvated system was then minimized in two stages. The first stage involved minimization of the peptide with 500 steps of steepest descent followed by 500 steps of conjugate gradient with a force constant of 500 kcal/mol. The second stage of minimization was run on the whole peptide-E-protein complex with 1000 steps of steepest descent followed by 1500 steps of conjugate gradient minimization without restraints. Minimization was followed by simulation with parameters comprised of the time step of 2 femtoseconds, Langevin thermostat set to 310 K and electrostatic interactions were treated using the particle mesh Ewald method. The SHAKE algorithm

(Ryckaert, Ciccotti, & Berendsen, 1977b) was applied to constrain all hydrogen-heavy atom bonds. Then, the complex was heated from 0 to 310 K for 200 picoseconds and submitted to molecular dynamics in NPT (constant number of particles, pressure and temperature) ensemble for 50 ns. During this phase, structural coordinates of the system were taken at 0.1 picosecond intervals to build a trajectory of the system dynamics. Time-dependent properties were calculated from the production trajectory. The convergence of the energies, root mean square deviation (RMSD) and radius of gyration (R_g) were checked to indicate the system stability.

4.2.1.5 Analyses of Results

The structures of the peptide-E-protein complexes were verified using Ramachandran plot, Verify3D and ERRAT programs. A Ramachandran plot was generated to check for the stereochemical quality of each peptide-E-protein complex. The Ramachandran plot was generated from the online RAMPAGE server at <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php> (Lovell et al., 2003). The Verify3D software was used to determine the compatibility of the 3D atomic models with their own 1D amino acid (Luthy, Bowie, & Eisenberg, 1992). Incorrect segments of a model normally have Verify3D scores below 0.1 which reveal a major problem to the model. This program (Laskowski, MacArthur, Moss, & Thornton, 1993) is available in the online server version at the Structural Analysis and Verification Server of UCLA (University of California, Los Angeles; <http://services.mbi.ucla.edu/SAVES/>). Interactions between the peptides and E-protein in the docked complexes were visualized using the Visual Molecular Dynamics (VMD) program (Humphrey, Dalke, & Schulten, 1996), analysed using Discovery Studio Visualizer (Koska et al., 2008) and Ligplot (Wallace et al., 1995) softwares which identified the hydrogen bonding and hydrophobic interactions between E-protein and the peptides.

4.2.1.6 Calculation of Free Energy of Binding

The binding free energy (ΔG_{bind}) of each complex was calculated based on Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) and Molecular Mechanics Generalized Born Surface Area (MMGBSA) procedures in AMBER 12 program. In this study, the binding free energy of each system was calculated when the complexes appeared to gain stable configurations. A total of 500 snapshots were collected for estimation of the binding free energy (ΔG_{bind}). The free energy of binding of a protein-ligand complex is calculated following relationship shown below (Kumari et al., 2014a):

$$\begin{aligned}\Delta G_{\text{bind}} &= G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \\ &= E_{\text{MM}} + G_{\text{GB/PB}} + G_{\text{non-polar}} - T\Delta S\end{aligned}\quad (4.1)$$

Where G_{complex} , G_{protein} and G_{ligand} are the free energies of the complex, protein (monomer) and ligand (monomer), respectively; E_{MM} is the change of the molecular mechanics potential energy upon peptide binding that includes van der Waals (E_{vdW}) and electrostatic (E_{ele}) energies; $G_{\text{GB/PB}}$ and $G_{\text{non-polar}}$ are the polar and non-polar components of the desolvation free energy, respectively; and $-T\Delta S$ is the change of conformational entropy upon peptide binding (which was not considered in this study because of the high computational cost required and the tendency for a large error margin that will lead to significant uncertainty in the results) (Homeyer & Gohlke, 2012; Hou et al., 2011; Tue-nguen et al., 2013).

4.2.1.7 Analysis of the Decomposition of Free Energy

Free energy decomposition for each peptide-E-protein complex was examined to obtain information on important residues involved in the complex binding. The energy decomposition was carried out using the `mm_pbsa.pl` implemented in the AMBER12 package to calculate the per-residue decomposition. Per residue decomposition calculates the energy contribution of single residue by integrating the interactions of each residue over all residues in the system. The decomposition of free energy of peptide-E-protein complex was calculated based on the MMPBSA and MMGBSA protocols. By using the MMPBSA and MMGBSA programs in AMBER 12, the energy contribution of each residue ($\Delta G_{\text{residue}}$) was divided into three parts: van der Waals energy (ΔG_{vdW}), intermolecular electrostatic energy (ΔG_{ele}), and solvation energy (ΔG_{sol}) due to solvent effect, which was a sum of the polar solvation energy (ΔG_{GB}) and the non-polar solvation energy (ΔG_{SA}).

$$\Delta G_{\text{residue}} = \Delta G_{\text{vdW}} + \Delta G_{\text{ele}} + \Delta G_{\text{solvation}} = \Delta G_{\text{vdW}} + \Delta G_{\text{ele}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}} \quad (4.2)$$

Where ΔG_{vdW} and ΔG_{ele} are nonbonded van der Waals and electrostatic interactions between two residues, respectively. In this study, the decomposition of free energy of each system was calculated from the last 10 ns of the trajectories from where 500 snapshots were extracted. The interaction energy profiles were generated by decomposing the total binding free energies into residue-residue interaction.

4.2.1.8 Calculation of Hydrogen Bond (H-bond) Occupancies

The hydrogen bonds between the peptide and the DENV E-protein were computed using the trajectory of a production run. The criterion for the hydrogen bond formation X-H...Y, where X, Y is either an O or an N atom, must satisfy two conditions: (1) the distance between X and Y atoms must be $< 3.5 \text{ \AA}$; and (2) the angle of X-H-Y must be $> 120^\circ$. Snapshots of trajectories forming H-bonds were calculated and are presented for the total 50 ns simulation.

4.2.1.9 Assessment of the Secondary Structure of Folded Peptides

The secondary structures of the peptides folded were assessed over the course of 100 ns simulations at 325.15 K using the Virtual Molecular Dynamics (VMD) program. VMD plugin Timeline was used to analyse and identify events in MD trajectories (Isralewitz, 2011). Timeline creates an interactive 2D box-plot-time versus structural component that can show detailed structural events of an entire system over an entire MD trajectory. Events in the trajectory appear as patterns in the 2D plot. In this study the secondary structures were analysed both in peptide folding and docked complex.

4.2.1.10 Flow Chart

The work flow of the steps involved in this section is shown in Figure 4.1.

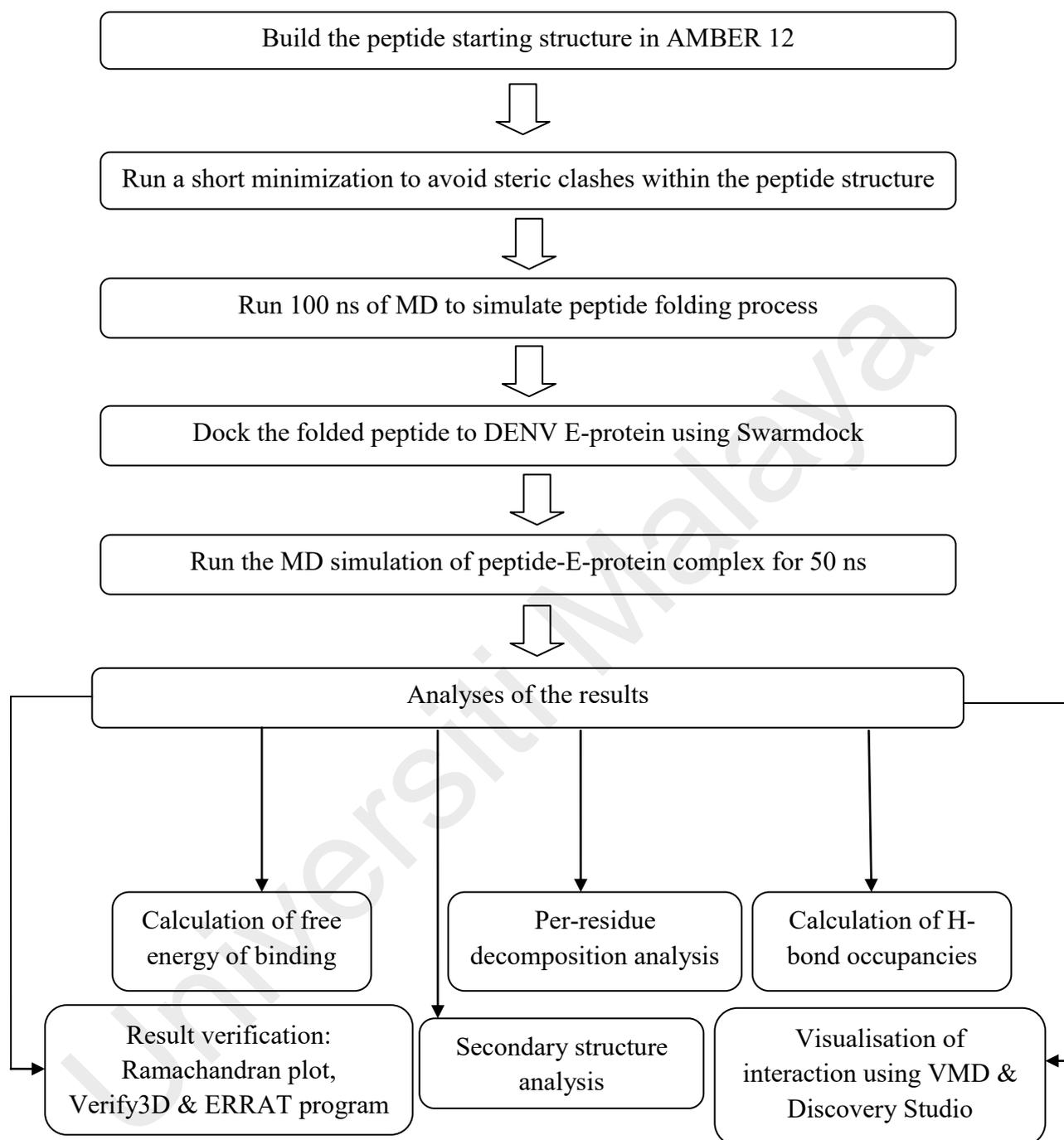


Figure 4.1: Flow chart showing the methods involved in the study of peptide-E-protein complex interactions.

4.2.2 Peptide Synthesis

The steps involved in this part of the study are shown in the flowchart in Figure 4.2.

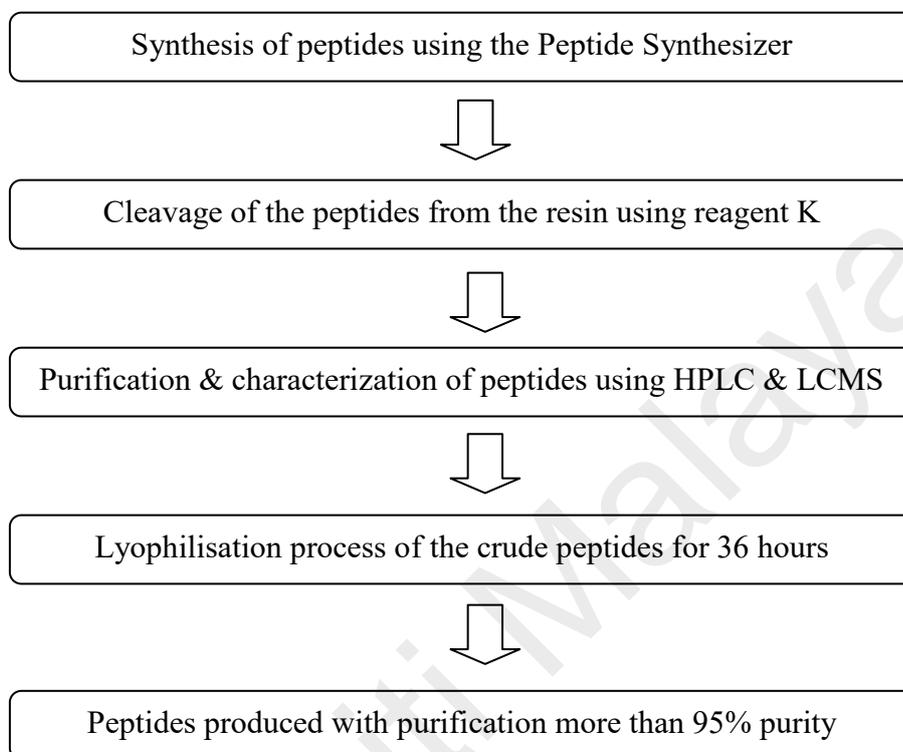


Figure 4.2: The flow chart shows summarized method involved in the peptide synthesis.

4.2.2.1 Solid Phase Peptide Synthesis (SPPS) Method

The linear peptides were prepared by automated peptide synthesis on a Symphony parallel synthesizer (Protein Technologies, Tucson, AZ, USA) via standard solid phase peptide synthesis (SPPS) method on chlorotrityl chloride (CTC) resin (0.5 mmol/g) with Fluorenylmethyloxycarbonyl (Fmoc) protection (Leronymaki et al., 2015). Peptides used in this study are listed in Table 4.2. In the SPPS method, the CTC resin was used as a support to which the growing peptide was anchored. The Fmoc-protected amino acids were coupled using diisopropylcarbodiimide (DIC, 6.0 equiv) and 1-hydroxybenzotriazole (HOBt, 6.0 equiv) as coupling reagent and additive, respectively, with coupling period of 2 hours. The Fmoc group was removed using a solution of 20% piperidine in dimethylformamide (DMF) for 3 minutes, followed by 20% piperidine in DMF at room temperature for 17 minutes. First amino acid with temporary protecting group is attached to the resin via its C-terminus. After the addition of an amino acid, the protection group was removed and the resin washed prior to subsequent addition of an amino acid. After another 5 minutes, the resin was washed with DMF for five times and treated with capping solution [2% acetic anhydride + 2% *N*-Ethyl-diisopropylamine (DIPEA) in DMF] twice for 5 minutes. In preparation for the next amino acid coupling, the resin was washed with DMF and treated with 20% piperidine in DMF twice at 5 minutes intervals, which were then followed by three times DMF washing steps. The coupling of next amino acids were performed in a similar fashion to the first coupling procedure except the reaction time before the addition of dichloromethane (DCM) was extended to 20 minutes and the repeat of the coupling was carried out using a reaction mixture with 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU). Additionally, the piperidine treatment was extended to 6 minutes. For the coupling of amino acid 21-30, the coupling procedure was performed three times, twice with *O*-(Benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium tetrafluoroborate (TBTU) and once

with HATU. The treatment time with piperidine was extended to 8 minutes. For further coupling, the coupling was carried out once with TBTU and twice with HATU activation. The process is repeated until the sequence is completed. The peptide synthesis steps are summarized in Figure 4.3.

Table 4.2: List of peptides used in this study. The peptide in bold indicates the peptides that have been selected for synthesis.

Code	No. of aa	Sequence	Chemical formula	Molecular weight (g/mol)
DN57opt	28	WYFIRKEFFERIRFLPQRNPHRR DDEWD	$C_{179}H_{255}N_{53}O_{44}$	3853.29
DN81opt	19	WIFIRYEFFRSFKFLWRGN	$C_{132}H_{178}N_{32}O_{25}$	2613.04
DS36opt	20	RHWEQFYFRRRERKFWLFFW	$C_{150}H_{198}N_{42}O_{27}$	3021.46
DS27opt	20	KEYFRRFFHCHNHQREWHWH	$C_{134}H_{174}N_{44}O_{28}S$	2881.17
DS04opt	20	IWWRPDWPWFIFYFREW RW	$C_{150}H_{191}N_{37}O_{27}$	2944.37
DS03opt	20	FPDFHHDRIYHFHWKRYQH	$C_{140}H_{171}N_{39}O_{29}$	2864.11
DS10wt	20	VCKHSMVDRGWGNGCGLFGK	$C_{92}H_{143}N_{29}O_{25}S_3$	2151.51
DS36wt	20	LITVNPIVTEKDSPVNIEAE	$C_{96}H_{161}N_{23}O_{34}$	2181.45
DN58opt	21	TWWCFYFCRRHHPFWFFYRHN	$C_{151}H_{182}N_{40}O_{26}S_2$	3037.46
DN58wt	21	GDSYIIIIGVEPGQLKENWFKK	$C_{113}H_{173}N_{27}O_{32}$	2421.76

aa: amino acid.

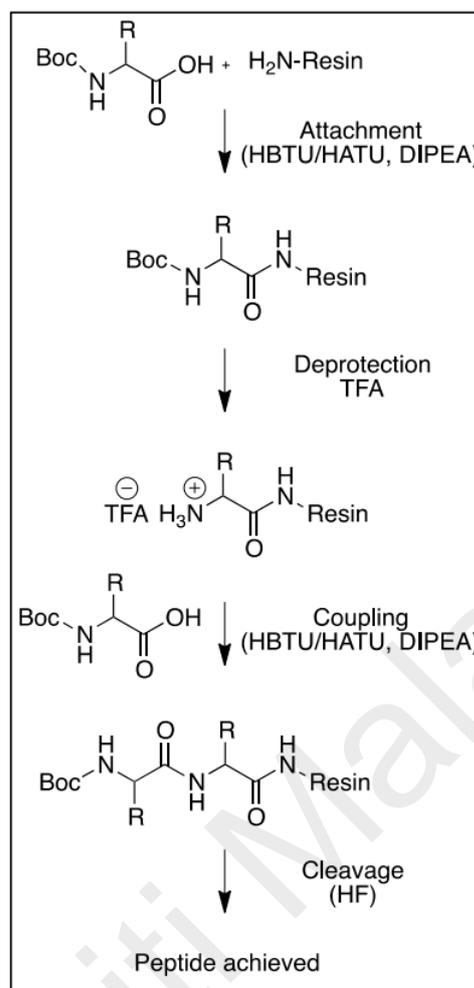


Figure 4.3: Flow chart of solid support peptide synthesis (Made, Els-Heindl, & Beck-Sickinger, 2014).

After completion, the peptide side chains were deprotected twice (10 minutes) with 20% piperidine in DMF to remove Fmoc group from the resin-bound polyamide and simultaneously cleaved from the solid support by treatment with reagent K (trifluoroacetic acid (TFA)/water/phenol/thioanisol-1,2-ethanedithiol) (82.5:5:5:2.5) for 2 hours. The TFA solutions were concentrated by nitrogen flow and the compounds were precipitated with diethyl ether (Et₂O) to yield the crude materials as white powders, which were subsequently centrifuged and Et₂O decanted.

The peptide chain was released by treatment with TFA/H₂O/TIS (95/2.5/2.5, v/v/v) for 1 hour at room temperature. The colourless solution was filtered and the resin was washed with dichloromethane (DCM). The solvent and residues from the cleavage cocktail were concentrated under nitrogen. The crude peptide was precipitated with cold

Et₂O and lyophilized using Savant AES 2000 Automatic Environmental SpeedVac system. The crude peptides were purified using preparative reverse-phase high performance liquid chromatography (RP-HPLC) and characterized using liquid chromatography mass spectrometry (LCMS) (Shimadzu LC/MS 2020). The final products were obtained with >95% purity (Xu, Lam, Zhang, & Li, 2013). Quantification and characterization data are given in Table 4.3.

Table 4.3: List of synthesized peptides chosen for LCMS analysis.

Peptide Code	Monoisotopic Mass	Charge State	Mass-to-charge ratio (m/z)
DS36opt	3021.50	+3	1007.7
		+4	755.5
		+5	604.8
DS36wt	2181.49	+2	1090.4
DN81opt	2613.04	+5	523.3

4.2.2.2 Purification of Peptides

Peptide purification was done via RP-HPLC using a 4.6 mm×150 mm, 5- μ m particle diameter, 100-Å pore size, Kromasil C8 column. The mobile phases were prepared in gradients from 0.1% TFA (trifluoroacetic acid) in acetonitrile (ACN) (solution A) and 0.1% aqueous TFA (solution B). For peptides to come out, the solvent gradient range was 30% to 50% in 15 minutes. The flow rate was 1.2 mL min⁻¹.

4.2.2.3 Lyophilisation

The ACN/H₂O solution was removed using a rotary evaporator. The peptide film on the side of the flask was re-dissolved in H₂O and transferred to a 15 ml conical flask. The peptides were frozen in dry ice and then transferred to a lyophilisation vessel, which was then attached to the lyophilizer for 36 hours to get dry powder of the peptides.

4.3 Results and Discussion

4.3.1 Binding Free Energies of Peptides Inhibiting Dengue Virus (DENV) Envelope Protein by using Docking & Molecular Dynamics Simulation

4.3.1.1 *In Silico* Peptide Folding

Peptide folding simulations were performed at three different temperatures which are 325 K, 343 K and 373 K (equivalent to 52°C, 70°C and 100°C, respectively). For docking purpose, peptide folding conformations at lowest potential energy were chosen to dock with DENV E-protein (Table 4.4). This is due to the reason that the conformation that had the lowest Gibbs free energy should be thermodynamically most stable. Tambunan et al. (2009) in their work also used three different temperatures to get peptide folding conformation at the lowest energy. They claimed that by increasing the temperature, the movements of the peptide would also increase which resulted in high flexibility and dynamics. In addition, higher temperature is associated with a collapse of the peptide from extended coils into more compact structures (Hansmann, Masuya, & Okamoto, 1997). In fact, by increasing the temperature, the protein unfolds and reaches a much more complex and heterogeneous ensemble of conformations, termed as “denatured”, resulting in difficulties to characterize the structural properties of the denatured states.

In this study, implicit solvent model was used to fold the peptide because it would be able to accelerate folding due to its low viscosity that eases in chain diffusion (Nguyen, Maier, Huang, Perrone, & Simmerling, 2014). Peptide folding is basically important because amino acids interact with each other to produce a precise three-dimensional structure in order for the peptide to carry out a particular function, and only became biologically active when it folded into specific, complex shape (Ho & Dill, 2006). However, incorrectly folded proteins may result in many serious and fatal

neurodegenerative diseases such as mad cow disease, Alzheimer's disease and Creutzfeldt-Jakob disease.

Table 4.4: Potential energy for peptide folding at different temperatures.

	325 K (52°C)	343 K (70°C)	373 K (100°C)
Peptide	Energy (kcal/mol)	Energy (kcal/mol)	Energy (kcal/mol)
DN57opt	-1596.0140	-1568.8126	-1545.8778
DN81opt	-754.2383	-746.7346	-712.6002
DS36opt	-1150.8756	-1135.4784	-1100.4549
DN58opt	-662.2767	-636.9502	-622.5598
DS27opt	-873.6834	-842.2912	-821.3698
DS04opt	-833.8357	-809.8625	-781.4046
DS03opt	-646.7372	-628.8458	-602.2527
DS10wt	-516.5984	-494.4233	-469.3584
DS36wt	-556.6981	-533.9942	-505.9983
DN58wt	-561.3614	-529.4702	-492.2771

4.3.1.2 Docking of Peptides to DENV E-Protein

The binding energy (Table 4.5) of each peptide-E-protein complex was calculated using the Swarmdock algorithm, where flexible peptide was docked into the E-protein binding site with flexible side chains. Docking analysis was employed to get the best orientation and binding affinity of a peptide to its protein receptor. The negative binding energies for all these peptides indicated strong binding between peptide and E-protein and demonstrated that the protein was in a favourable conformation. This concluded that the peptides were able to find the binding site within the protein receptor. In this case, the E-protein underwent induced fitting because it continuously changed the conformation and the shape in response to peptide binding. Overall, Table 4.5 showed that the complex containing DN58opt peptide to give the lowest binding energy (-34.18 kcal/mol), meanwhile DN58wt peptide showed highest binding energy (-20.00 kcal/mol), for monomer E-protein. This revealed the less stability of the wild type peptide when in a complex, compared to optimized peptide. This is in line with the study by Costin and co-worker (2010) which reported that their optimized peptide had

the best RAPDF scores for structural stability and binding compared to wild type sequences.

Table 4.5: Binding energy for both monomer and dimer E-protein docked with peptides generated from Swarmdock docking server.

Peptide code	Monomer	Dimer
	Binding energy (kcal/mol)	Binding energy (kcal/mol)
DN58opt	-34.18	-29.54
DN57opt	-33.38	-30.96
DS36opt	-32.18	-31.06
DN81opt	-31.21	-31.36
DS03opt	-31.20	-30.52
DS27opt	-28.60	-24.33
DS04opt	-27.90	-29.00
DS36wt	-24.41	-22.68
DS10wt	-21.95	-21.57
DN58wt	-20.00	-22.58

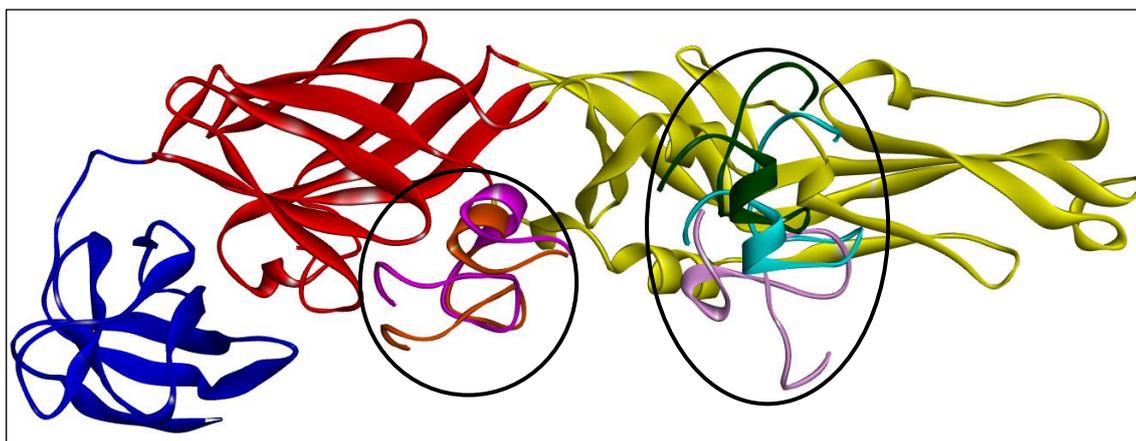
In this study, the peptides were docked to both the monomer and dimer structures of E-protein. Eventhough in the crystal structure, E-protein existed as a dimer, due to the need to save computational time, the monomer structures were used for further MD simulation study. Furthermore, the monomer structures give the lower binding energy compared to dimer structures indicating more stable structure. This is supported by Das et al. (2011) that perform docking both to the monomer and dimer structures of Glutathione transferase-P1P1 (hGSTP1-1) and their result revealed that the dimer only showed one binding sites while the monomer revealed three additional binding sites for ligand binding. Furthermore, the monomer structure was found to provide a wider range of binding sites for naturally occurring ligands and these docking results suggested that the enzyme subunit interface might be important for hGSTP1-1 interactions with the ligands. On the other hand, Padariya and co-workers (2015) observed more hydrogen bond interactions with lipids in monomer system compared to dimer membrane system. This was expected as monomer alone had larger surface to interact with lipid molecules,

whereas monomer in dimer system has only a fraction of its surface exposed for interaction with lipid molecule.

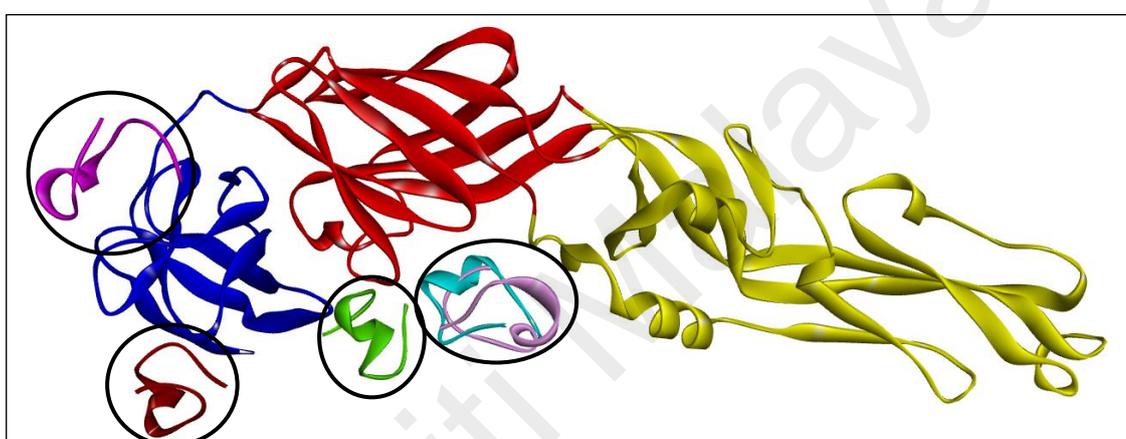
In this study, the trend of the binding energy for some of the peptides such as DS36opt, DS36wt, DN58opt and DN58wt were in agreement with the results reported for the tryptophan fluorescence quenching assay performed by Baharuddin et al. (2014). Good correlations between the docking scores and the dissociation constant (K_d) values could be observed as shown in (Table 4.6). Figure 4.4 illustrate the five best conformations of the peptides in each peptide-E-protein complex. In general, the peptides were able to dock in the binding pocket at various locations. In this study, the peptides were docked in the largest or second largest pockets in the protein as suggested by Dagliyan et al. (2011). They caused steric hindrance due to their large structures and interacted with the chosen residues via their functional groups. Three possible binding sites were located in a) between Domains I and II, b) between Domains II and III, and c) Domain III as illustrated in Figure 4.5. The best docking pose with the lowest binding energy was recorded to be for Domain III, which was also the potential site for antibody binding as well as other small molecules (Guardia & Lleonart, 2014; Poggianella et al., 2015).

Table 4.6: Relationship between the binding activities from fluorescence quenching assay (Baharuddin et al., 2014) and the binding energies calculated using the Swarmdock docking algorithm.

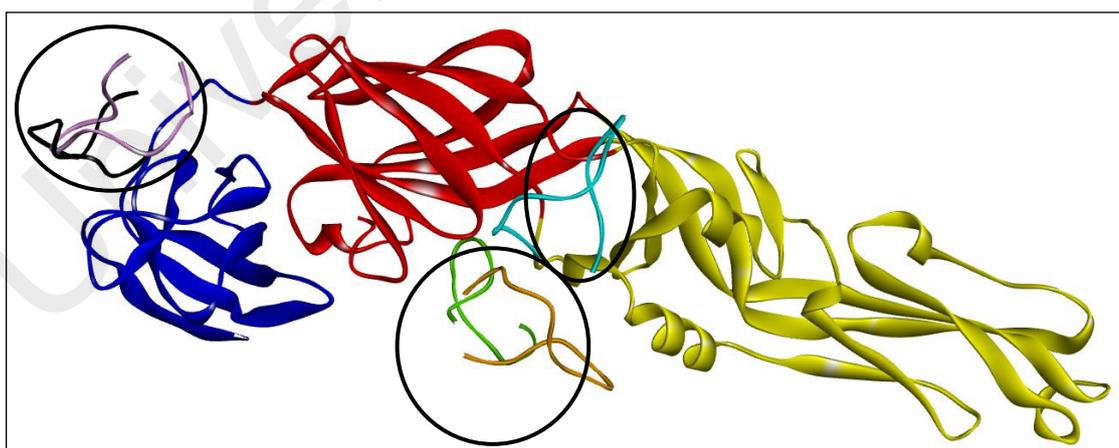
Peptide code	Peptide sequence	Binding Free Energy (ΔG_{bind}) (kcal/mol)	K_d (experiment) (μM)
DS36opt	RHWEQFYFRRRERKFWLFFW	-32.18	9.31 ± 0.15
DS36wt	LITVNPIVTEKDSPVNIEAE	-24.41	N/A
DN58opt	TWWCFYFCRRHHPFWFFYRHN	-34.18	9.44 ± 0.28
DN58wt	GDSYIIIIGVEPGQLKENWFKKG SSIGQMF	-20.00	10.79 ± 0.17



(a) DN57opt-E-protein complex.

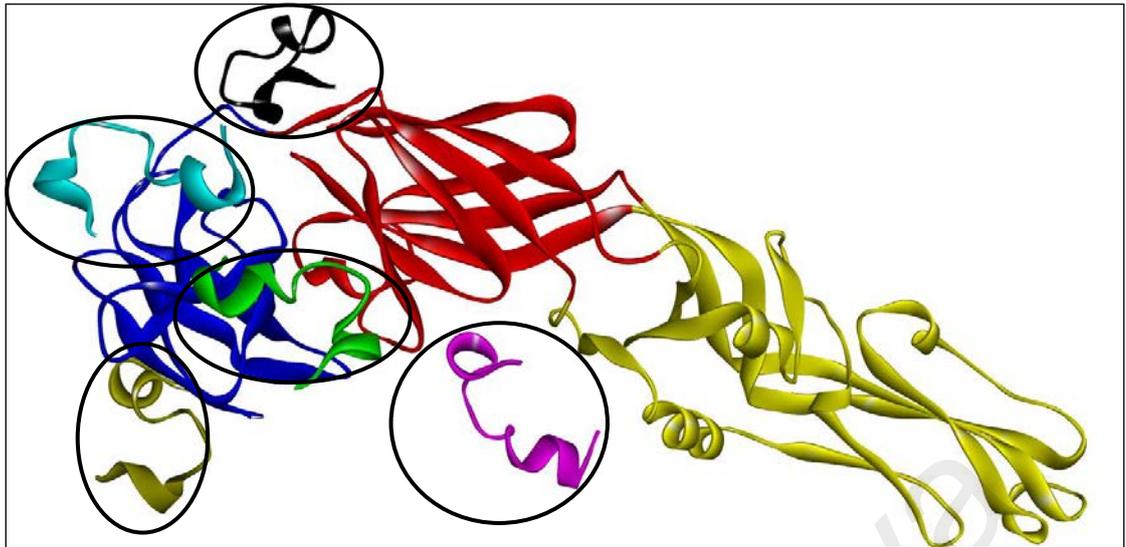


(b) DN81opt-E-protein complex.

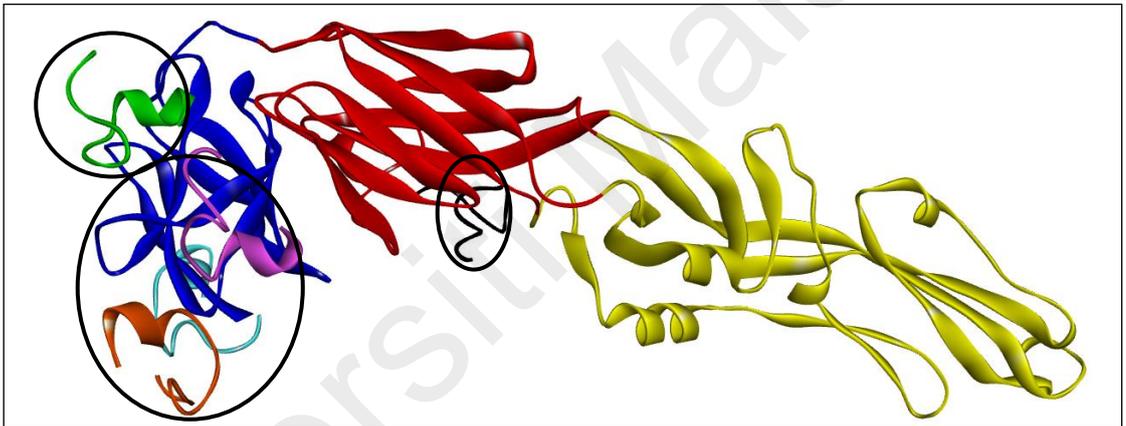


(c) DS360pt-E-protein complex.

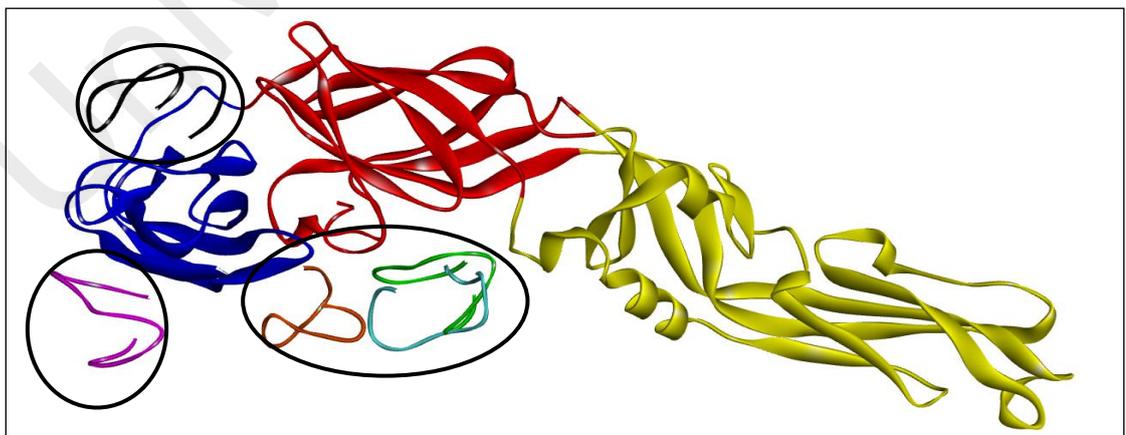
Figure 4.4: The top five docked conformation of peptides (a-j) generated from SwarmDock docking server (DI-red, DII-yellow, DIII-blue and peptide-colourful). Peptides (a-j) are illustrated in various colours and are in circles.



(d) DN58opt-E-protein complex.

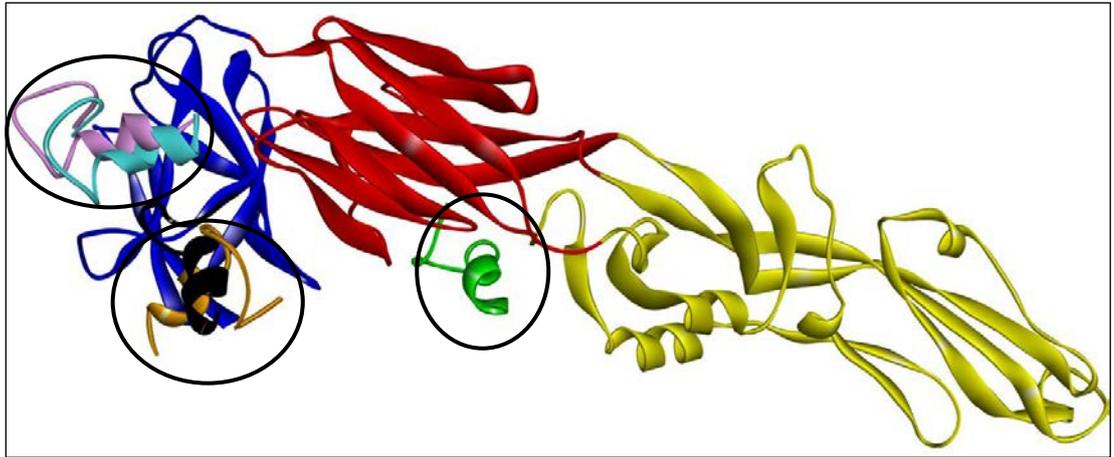


(e) DS27opt-E-protein complex.

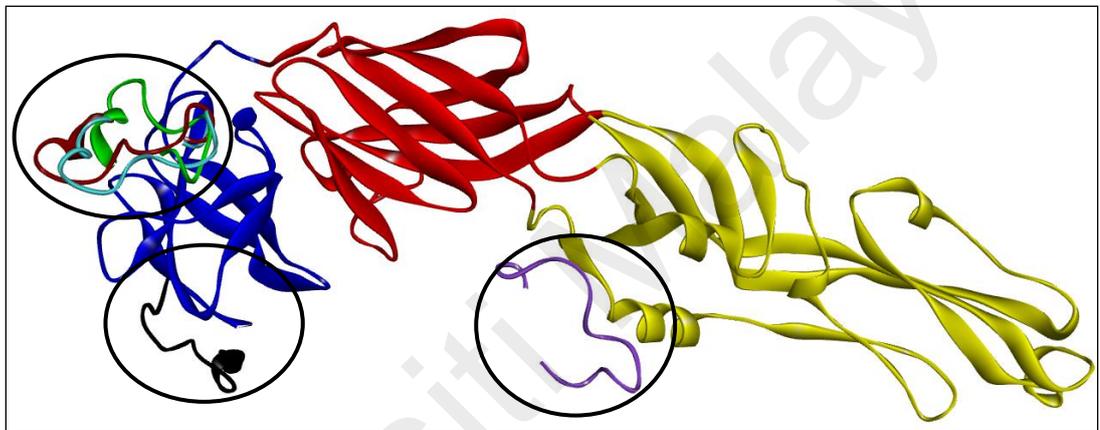


(f) DS04opt-E-protein complex.

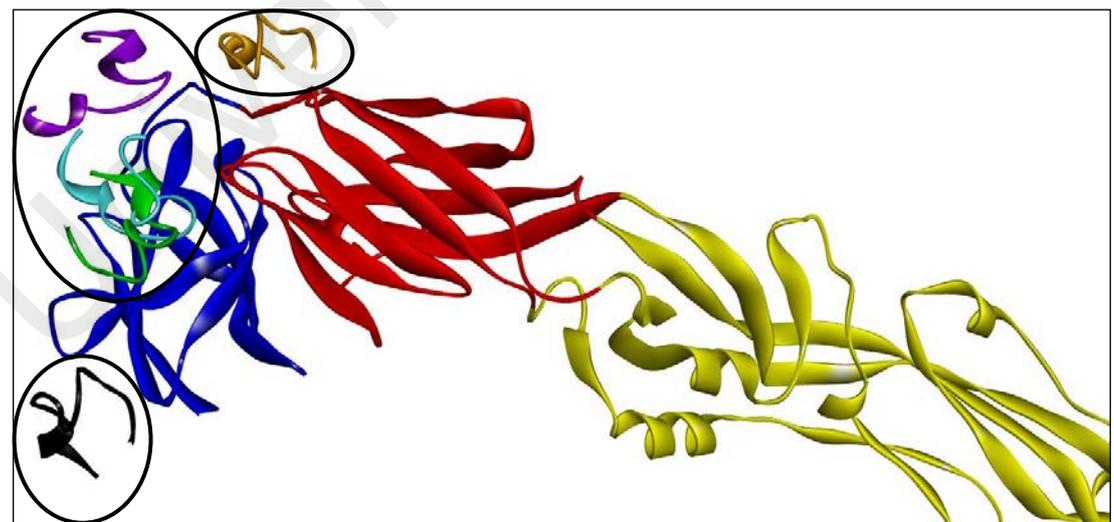
Figure 4.4: (Continued)



(g) DS03opt-E-protein complex.

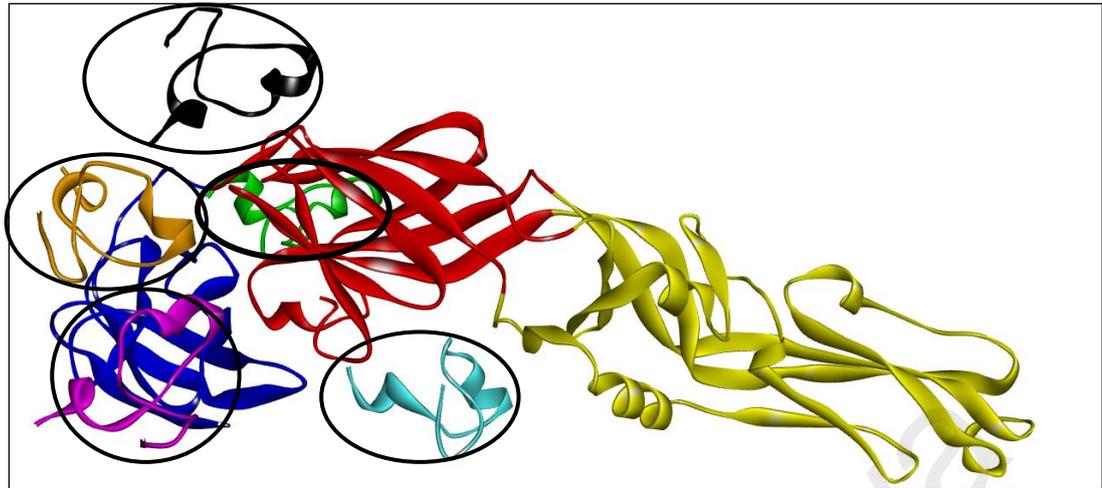


(h) DS10wt-E-protein complex.



(i) DS36wt-E-protein complex.

Figure 4.4: (Continued)



(j) DN58wt-E-protein complex.

Figure 4.4: (Continued)

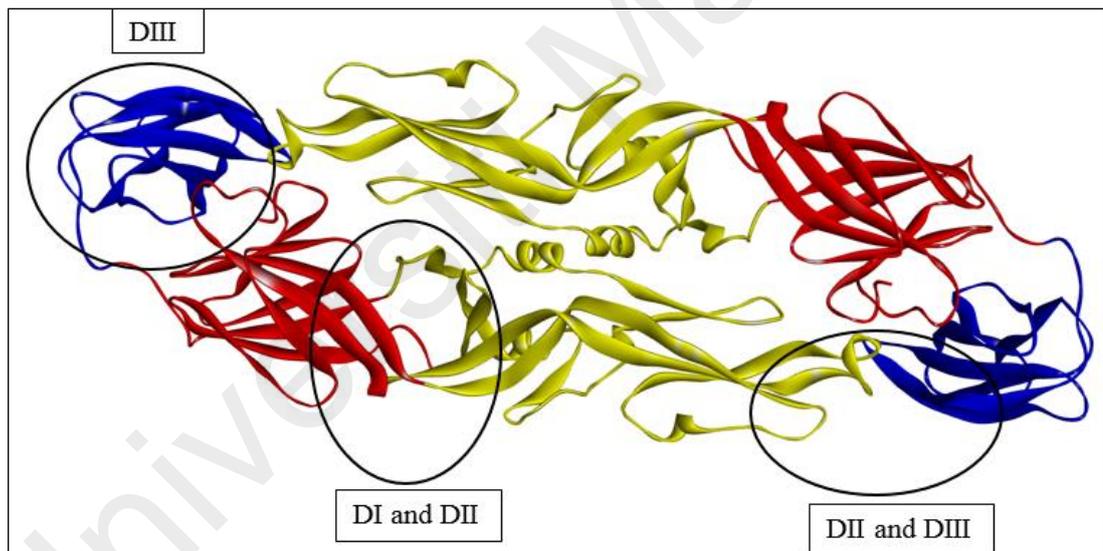


Figure 4.5: Three possible sites for peptide binding (in circles) towards DENV E-protein. The DENV E-protein is shown as dimer (DI-red, DII-yellow, DIII-blue).

4.3.1.3 Refinement of Docked Peptide-E-Protein Complexes using MD simulation

MD simulation was performed on the best docked complex structure to investigate the role of receptor flexibility on the binding of these peptides in aqueous solution as well as to assess the stability of peptide-E-protein complex throughout the simulation time. The best docking pose was further refined using explicit solvent MD simulations. In this study, explicit solvent was applied in the MD simulation as we would like to create the system to be as close to the reality as possible.

4.3.1.4 Root Mean Square Deviation (RMSD)

To validate the dynamic stability of the complexes, total potential energy and root mean square deviation (RMSD) of the backbone atoms along the 50 ns MD trajectories were monitored. Figures 4.6-4.9 depict the pressure, temperature, density and energies for the peptide-E-protein complexes over the course of 50 ns simulation. Examination of the pressure variable revealed seemingly large fluctuations in instantaneous pressure. The phenomenon was expected where the pressure fluctuated around 100 bar = 1 atm during the MD simulation. The negative pressure corresponded to a “force” acting to decrease the box size and the positive pressure to a “force” acting to increase the box size. The important point here is that while the pressure graph seemed to show that the pressure fluctuated wildly during the simulation the mean pressure stabilized around 1 atm. This was sufficient to indicate successful equilibration of the system. Meanwhile, the temperature remained more or less constant at 310 K during the simulation indicating the use of Langevin dynamics for temperature regulation was successful. On the other hand, the simulation had equilibrated at a density of approximately 0.975 gcm^{-3} . This seemed reasonable as the density of pure liquid water is approximately 1.00 gcm^{-3} . In order to validate the dynamic stability of the complexes, the total potential energy and root mean square deviation (RMSD) for

the backbone atoms along the 50 ns MD trajectories were monitored. Figures 4.9 and 4.10 illustrate the total potential energy and RMSD profiles for the best five peptides studied. The potential energy remained constant during simulation implying that the relaxation was completed and equilibrium had been reached (Heavner, 2004).

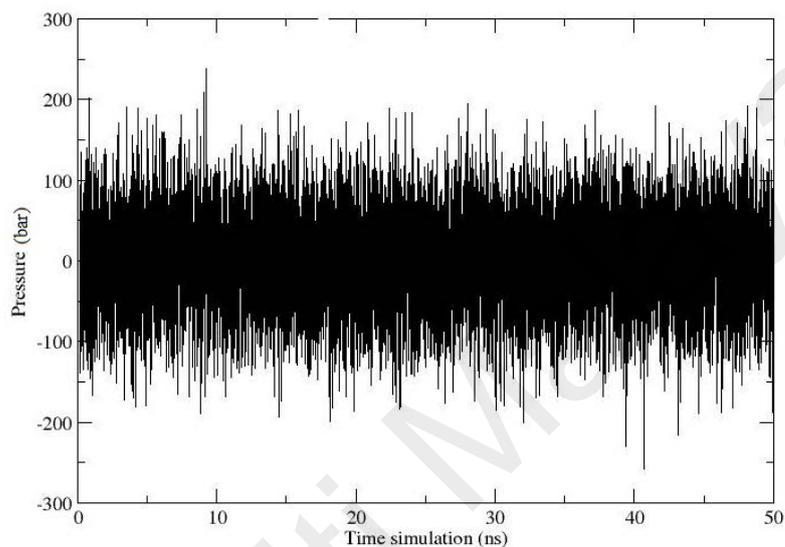


Figure 4.6: Pressure of peptide-E-protein complexes during 50 ns MD simulation.

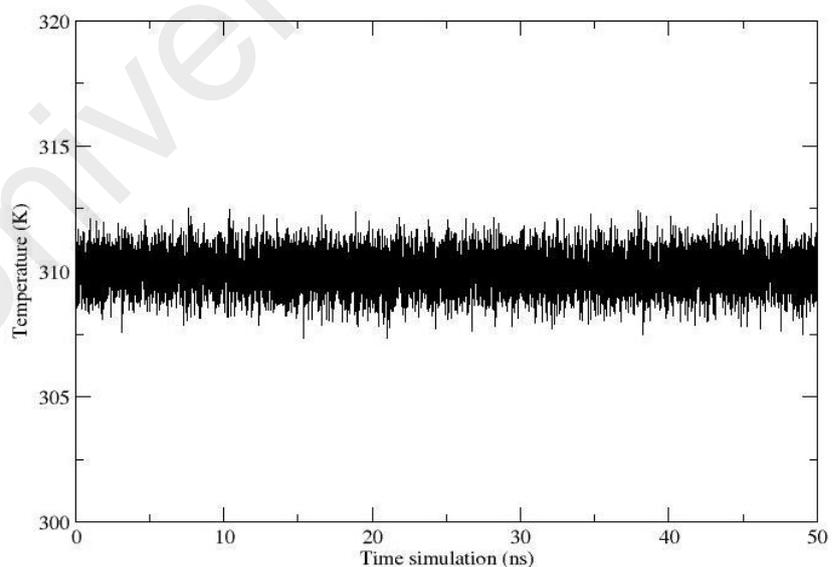


Figure 4.7: Temperature of peptide-E-protein complexes during 50 ns MD simulation.

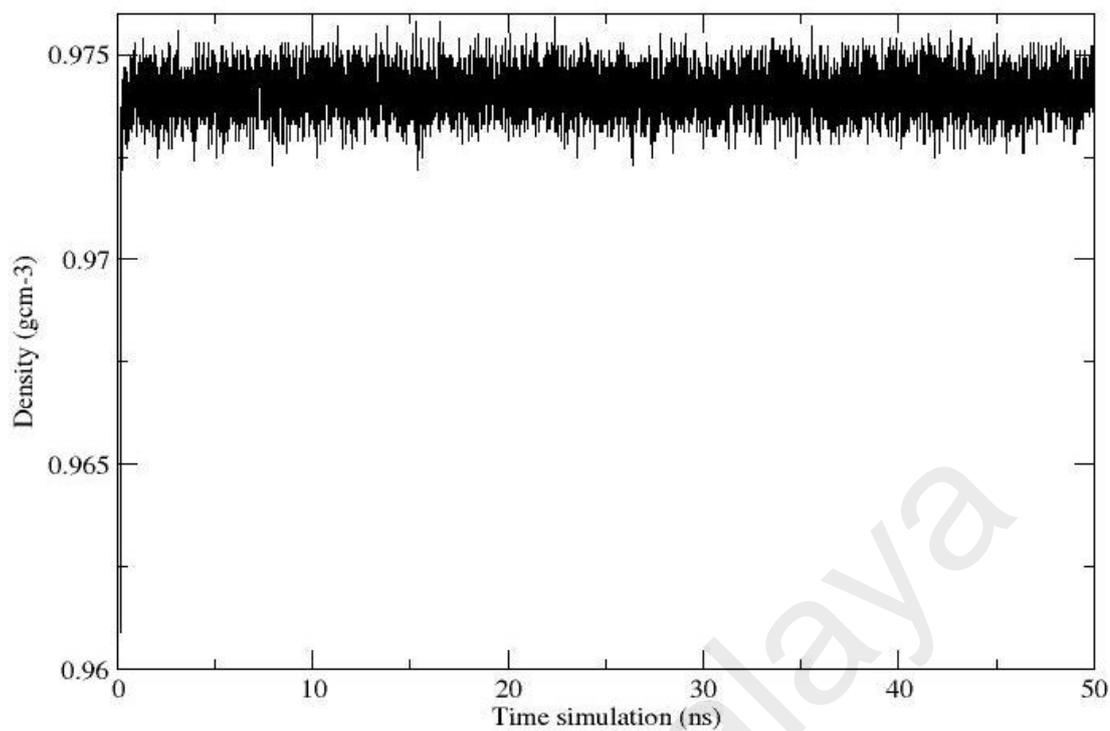


Figure 4.8: Density of peptide-E-protein complexes during 50 ns MD simulation.

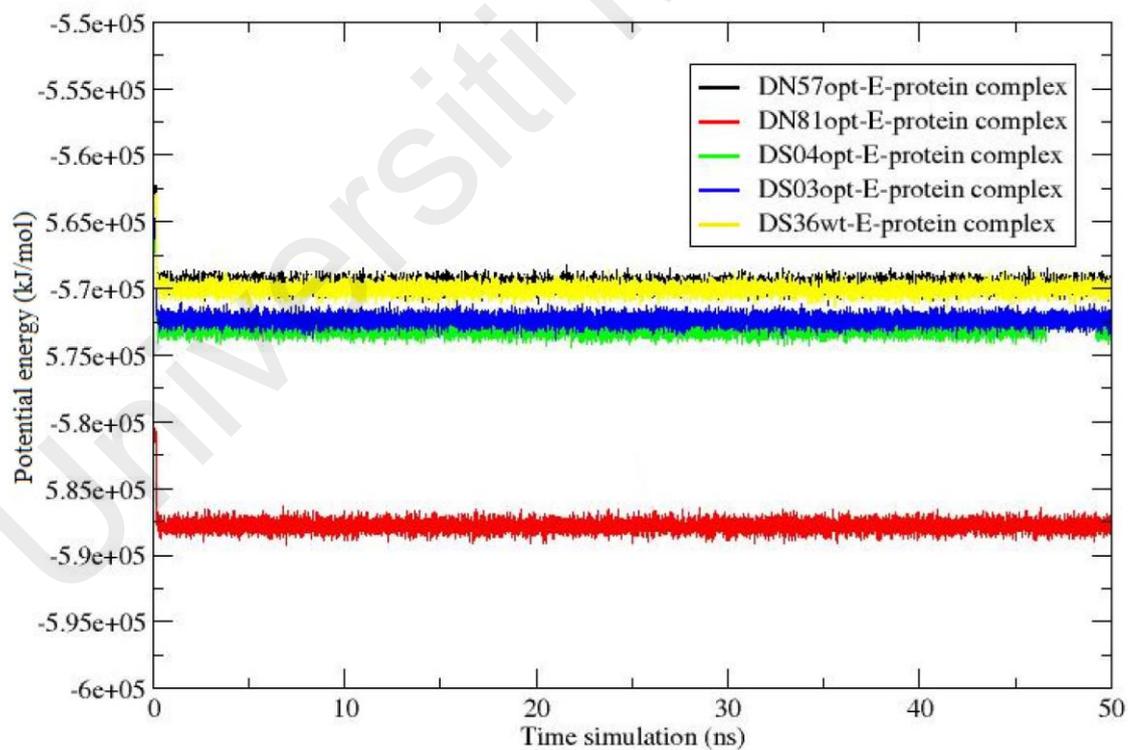


Figure 4.9: Potential energy of top five peptide-E-protein complexes during 50 ns simulation. The heating process is not shown.

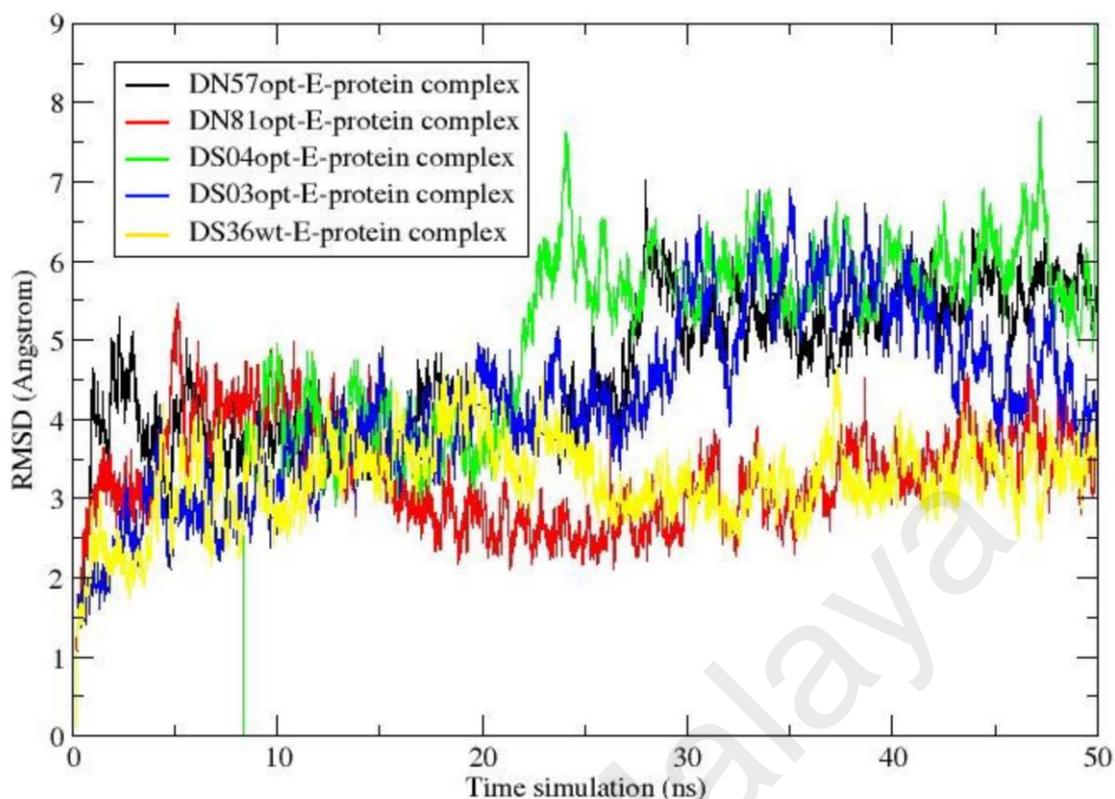


Figure 4.10: RMSD of top five peptide-E-protein complexes during 50 ns MD simulation.

In order to study conformational changes and structural properties over the simulation time, RMSD values were calculated for the peptide-E-protein complexes (Figure 4.10). For this purpose, ptraj tool was used and the backbone atoms of C, C_α and N were considered. For the complex containing DN81opt (red line) and DS36wt (yellow line) RMSD values were observed to oscillate steadily throughout the simulation which suggested that there were no large conformational changes in the complexes and the peptides remained in the protein binding pocket. However, the complexes containing DS04opt, DN57opt and DS03opt gave higher RMSD fluctuations around 5-7 Å after 25 ns due to large structural changes of the complexes. The increase in RMSD might be due to flexible regions such as large loops of the peptides. This is supported by Landon et al. (2008) that also observed an increment in RMSD values, which was due to large structural changes with respect to the positions of the amino acids in the loop regions of the protein. It is interesting to note that the longer the simulation runs, the more these molecules diffuse, resulting in larger RMSD values. The

increase in RMSD also associated with the random movement of protein atoms to stabilize the protein structure while a lower RMSD normally indicates that the bound complex is stable. The overall RMSD value of the complexes under study was 6.5 Å, indicating the reliability of the MD simulation protocol had been successfully applied. This result is in line with the study reported by Dagliyan et al. (2011) on protein-peptide recognition which observed the presence of peptide in the native binding site when RMSD from the native pose was lower than 10 Å.

4.3.1.5 Radius of Gyration (R_g)

The compactness of the peptide-E-protein complex throughout the simulation can be monitored using radius of gyration (R_g). The compactness has been defined as a ratio of the accessible surface area of a protein to the surface area of the ideal sphere of the same volume. It measures a distance of molecule size from its centre of mass. In this study, R_g was calculated for the backbone heavy atoms and was calculated according to Equation 4.3:

$$R_g = \left(\frac{\sum_i r_i^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \quad (4.3)$$

where m_i is the mass of atom i , and r_i is the position of atom i with respect to the centre of mass of the peptide (Spoel, Vogel, & Berendsen, 1996).

The variation of radius of gyration (R_g) as function of time is presented in Figure 4.11 and from this figure it is clear that R_g showed significant fluctuations along the trajectory which revealed the flexibility of the protein. Normally, a stable complex maintains a relatively steady R_g value, whereas unfolding or collapse is indicated by R_g change over time. Nevertheless, these observed fluctuations were concentrated around an average, which indicated that not all the complexes unfolded along the simulation time. There was also a fast collapse of the complex structure, which then became stable

and again, the immediate collapse of the complex's structure to form a more compact structure. The gradual R_g decrease would be due to the loops at both ends of the α -helix that sometimes curled inward. By contrast, the large fluctuations in R_g could be due to unfolding and refolding of parts of the α -helix. The collapse in peptide-E-protein complex was probably accompanied by changes to its secondary structure which is important for protein's structure and function.

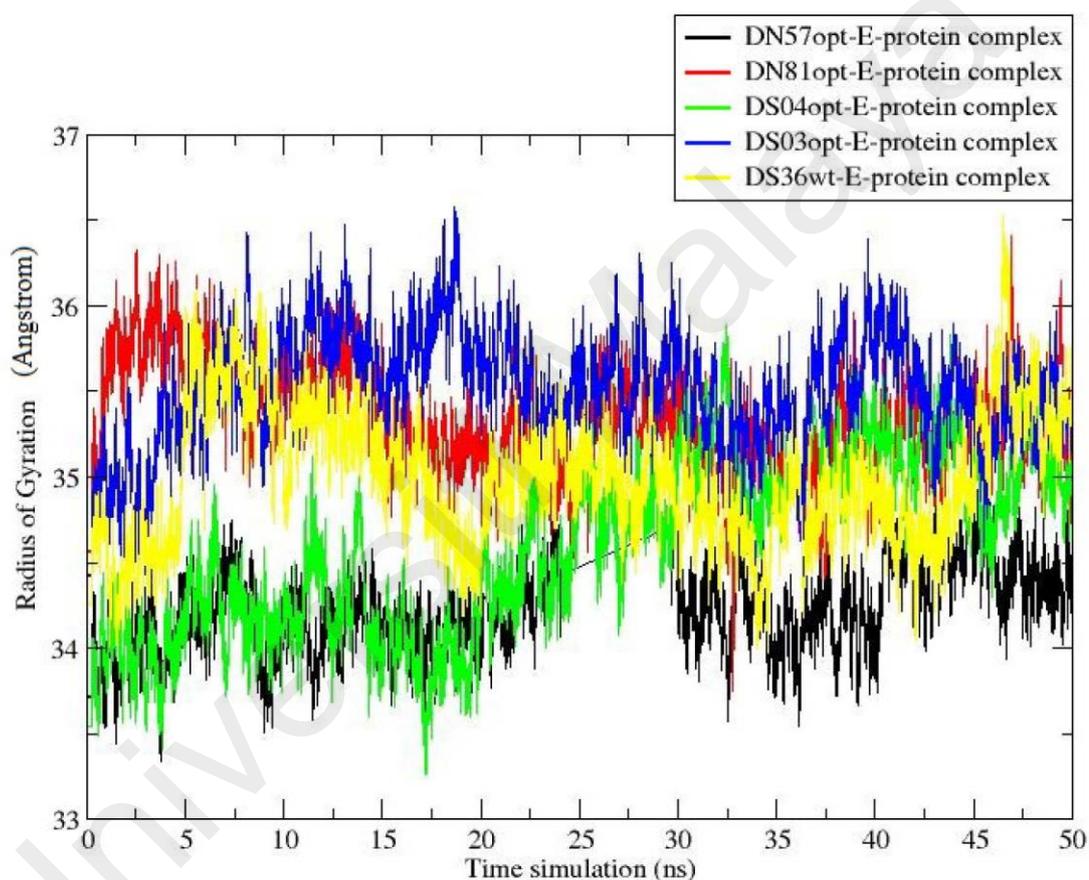


Figure 4.11: Radius of gyration (R_g) of top five peptide-E-protein complexes for backbone heavy atoms during 50 ns MD simulation.

4.3.1.6 Root Mean Square Fluctuation (RMSF)

The flexibility of the peptide-E-protein complex was investigated in terms of the atom-positional root-mean-square fluctuations (RMSF) for all non-hydrogen atoms. RMSF measures how much each individual atom moves around and is totally a different calculation than RMSD. It reflects the overall movement of atoms or residues over all frames. It is also a qualitative measure of protein flexibility as it gives a view of the relationship between protein conformational flexibility and dynamics. The fluctuations during MD simulation are expected as a result of solvated environment and a fully flexible complex.

The RMSF values of these best five peptides-E-protein complex ranged from 3-8 Å, showing a large degree of movement. Most of these flexible residues were located at both C and N terminals of the protein. This fluctuation of backbone deviation reflected the global structure rearrangement of peptide-E-protein complex that was well equilibrated, maintained and not artificially altered by the simulation. The structure showed more flexibility with the RMSF average value of 7.5-8 Å at residue numbers 55-100 and 200-250. In addition, residues 350-400 which belong to domain III region also showed flexibility due to high RMSF values. As seen in Figure 4.12, for both DN57opt (black line) and DS04opt (green line), the Gly349, Glu370, and Asn390 residues were among the most fluctuated and mobile residues. Therefore, these regions could be critical to thermostability and could be potential target sites for stability enhancement through rational design. The region with high fluctuations is normally related to connecting loops, turns and bends with helices compared to other regions. In addition, these high RMSF values of the monomer were likely resulting from the absence of contact between the chains and therefore had more space to fluctuate (Wichapong et al., 2014). The increment in RMSF value also suggested that the flexibility of peptides and E-protein might enhance binding affinity between them. The

flexibility of the loop regions also influenced the mobility of complex. However, lower RMSF values for DN81opt and DS03opt-E-protein complexes were observed due to the peptides rigidly held in the complex with only a limited space to flex. The minimal fluctuation of RMSF was also observed at the end of MD equilibration as reported by (Panigrahi et al., 2013).

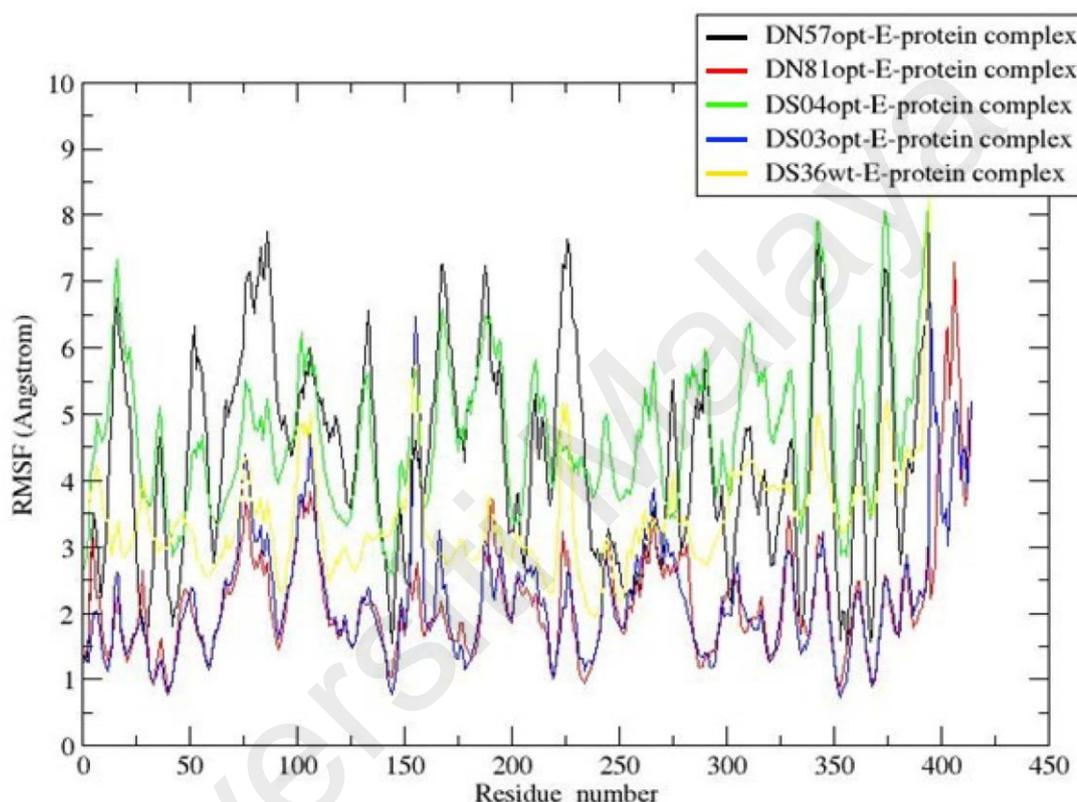


Figure 4.12: RMSF profile of top five peptide-E-protein complexes during 50 ns MD simulation.

4.3.1.7 Validation of 3D Model Structure

In this section, the quality of the peptide-E-protein complex, generated by MD trajectories was evaluated. Ramachandran plot program was used to examine geometry and stereochemistry of all peptide-E-protein complex structures. This plot also included a representation of the favourable and unfavourable regions for residues, so that the correctly built structure could be determined. The Ramachandran plots of all the ten complexes showed that most of the residues were in the sterically allowed region (Table

4.7). Nevertheless, few amino acids that were in the disallowed region belonged to the loop regions corresponding to the structurally variable region.

Table 4.7: Validation of peptide-E-protein complex structures using different verification programs.

Complex	Ramachandran plot		Verify3D (%)	ERRAT (%)
	Residues in Most Favoured Region (%)	Residues in Disallowed Region (%)		
DN57opt	92.3	1.7	91.00	77.44
DN81opt	93.4	0.7	89.59	74.19
DS36opt	93.4	0.5	86.47	72.65
DS27opt	93.2	0.2	90.82	75.89
DS04opt	94.4	0.7	88.41	73.88
DS03opt	93.2	1.0	88.41	81.40
DS10wt	92.4	0.7	92.51	71.43
DS36wt	95.9	0.5	89.86	84.68
DN58opt	93.7	0.5	91.81	65.71
DN58wt	95.0	0.7	87.62	79.70

All the peptide-E-protein complexes had shown that more than 85% of the residues were located in the most favoured region with few residues ($\leq 2\%$) in the disallowed region. The results (Table 4.7) suggested that the peptide-E-protein complex structures were acceptable and possessed sufficient stereochemical qualities and hence these structures were used for further MD simulation studies.

Meanwhile, less than 2% residues in the disallowed region were due to certain amino acids in each complex, which were Lys202 (for DN57opt, DS03opt, DS36wt and DN58wt respectively), Ala245 (DN81opt), Glu383 (for DS04opt and DS10wt) and Arg403 (DS36opt). These amino acids were not located in any of the peptide binding region. The result established that only a very small fraction of residues lie in appreciably disallowed regions with small polar residues having a significantly greater probability of adopting unusual backbone conformations. Disallowed conformations are mostly found for polar or charged residues (Gunasekaran, Ramakrishnan, & Balaram, 1996). In this study, Lys202, Ala245, Glu383 and Arg403 are bulky charged residues

which have a relatively low propensity for backbone distortions. Disallowed regions basically involve steric hindrance between the side chain C methylene group and main chain atom. Hence it frequently occurs in turns regions of proteins where any other residue would be sterically hindered. Nonetheless, this disallowed residue does not occur frequently in a protein (Pandey et al., 2015).

The complex structures were also validated by other structure verification servers such as Verify 3D and ERRAT to check the quality of the complex structures. Verify 3D determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, nonpolar, etc) and compare the results to good structures. Then, a database generated from good structures is used to obtain a score for each of the 20 amino acids in this structural class. The vertical axes in the plots (Figure 4.13) represent the average 3D-1D profile score for each residue in a 21-residue sliding window. The scores range from -1 for bad score to +1 for good score. In this analysis, it was found that none of the amino acids had a negative score. Therefore, the predicted models were compatible with its amino acid sequence with more than 85% of the residues with an averaged 3D to 1D score of more than 0.2. In addition, the compatibility scores above zero correspond to an acceptable side chain environment (Elengoe, Naser, & Hamdan, 2014). Thus, the results proved that all the complex structures were of reasonable quality and were expected to be satisfactory.



Figure 4.13: Verify3D plot for DS36wt at 50 ns simulation. 89.86% of the residues had an average 3D-1D score of more than 0.2.

On the other hand, ERRAT is a program that is used to verify protein structures, as well as to obtain a stable structure. A stable structure is used in MD simulations in order to gain the binding energy, intermolecular energy, electrostatic energy and total internal energy (Bhargava, Nath, Seth, Pant, & Dixit, 2014). ERRAT works by analysing the statistics of non-bonded interactions between nitrogen, carbon and oxygen atoms, with higher scores indicating higher quality (Elengoe et al., 2014; Wallner & Elofsson, 2006). The highest ERRAT score obtained for the peptide-E-protein complex was 84.68% (DS36wt) with the generally accepted range of >50, indicating a high quality model (Khor, Tye, Lim, Noordin, & Choong, 2014). Thus, this analysis revealed that the backbone conformation and non-bonded interactions of the peptide-E-protein complexes fit well within the range of a high quality model (Elengoe et al., 2014). The validated complex structures used in this study may be used further to understand the potential interactions between peptides and DENV E-protein.

4.3.1.8 Binding Free Energy of Peptide-E-Protein Complexes

Docking results obtained in this study revealed the binding affinities of the peptides to E-protein to be in the order of DN58opt > DN57opt > DS36opt > DN81opt > DS03opt > DS27opt > DS04opt > DS36wt > DS10wt > DN58wt (Table 4.5). However, only five peptide-E-protein complexes were able to run up to 50 ns of MD simulations. In this work, MMPBSA and MMGBSA methods were used to calculate the binding free energy (ΔG_{bind}) of the peptide-E-protein complexes over the course of 50 ns trajectory. Even though the simulation time was quite short, longer MD simulation does not necessarily mean better predictions. Even a simulation with a length on the order of 10 ns would give high precision and accuracy values of binding free energy. In this study, as the simulation was extended to 60 ns, the peptide was found outside the

binding pocket, providing no indication that association would occur if the simulations were to be run longer. Due to this reason, the MD simulation was run up to 50 ns only.

To characterize the binding affinities of peptide-E-protein complexes, the MMPBSA and MMGBSA methods were used to calculate the binding free energy. Both Poisson-Boltzmann (PB) and Generalized Born (GB) models were used to evaluate the electrostatic component of the solvation free energy and had been successfully applied to various protein-ligand or protein-peptide complexes. The PB approach is well established and used widely in the literature (Wang et al., 2016). It has gained acceptance as a convenient and reliable alternative to explicit solvent in energy calculations. Nevertheless, the GB method is still useful to utilize since it saves computational time (Wang, Morin, Wang, & Kollman, 2001). Therefore, in this study, the MMPBSA and MMGBSA methods were used to compare binding free energies from the various types of peptides.

The individual energies that contributed to the calculation of binding free energy on the basis of MMPBSA and MMGBSA are listed in Table 4.8. The estimated total binding free energy is negative for all of the complexes, indicating favourable binding contacts. Generally, the MMPBSA gave lower PBTOT value compared to MMGBSA for all the peptides studied. Even though MMGBSA approach gave slightly higher binding energy, the results obtained in this study still suggested that the complexes were in their favourable bound states. This proved that MMPBSA method gives better performance in ranking the binding affinities since this technique employs a more rigorous algorithm than MMGB method (Heavner, 2004; Wang et al., 2001). This is supported by Hou and co-workers (2011) where in their study it was shown

Table 4.8: Binding free energies predicted using the MMPBSA and MMGBSA methods at different time simulation (All data are given in kcal/mol).

Method	Contribution	DS36wt		DN81opt		DN57opt		DS03opt		DS04opt	
		10 ns	50 ns								
MM	ELE	-22.40	-24.70	-173.22	-232.47	-21.05	-151.34	-45.44	-120.86	-8.21	-164.37
	vdW	-81.77	-87.59	-53.86	-42.55	-37.19	-42.46	-36.01	-32.81	-21.45	-19.93
	GAS	-104.17	-112.29	-227.08	-275.02	-58.24	-193.80	-81.45	-153.67	-29.66	-184.30
PBSA	PBSUR	-11.97	-11.83	-9.59	-8.13	-7.27	-8.61	-7.20	-6.72	-4.70	-4.02
	PBCAL	55.71	63.65	209.36	253.57	44.79	165.61	73.93	126.69	23.22	161.32
	PBSOL	43.74	51.82	199.77	245.44	37.51	157.01	66.73	119.97	18.52	157.30
	PBELE	33.31	38.95	36.14	21.10	23.74	14.27	28.49	5.84	15.01	-3.05
	PBTOT	-60.43	-60.46	-27.31	-29.58	-20.72	-36.80	-14.72	-33.70	-11.13	-27.00
GBSA	GBSUR	-7.67	-7.61	-4.68	-4.01	-3.48	-4.92	-3.43	-3.57	-1.91	-2.36
	GBCAL	59.04	67.64	213.05	264.70	53.01	176.51	84.93	136.66	30.37	166.51
	GBSOL	51.37	60.03	208.36	260.69	49.53	171.59	81.51	133.09	28.46	164.16
	GBELE	36.64	42.94	39.82	32.23	31.97	25.17	39.49	15.80	22.16	2.14
	GBTOT	-52.80	-52.26	-18.72	-14.33	-8.71	-22.21	0.06	-20.57	-1.19	-20.14

The individual energy contributions: ELE = electrostatic energy as calculated by the molecular mechanics (MM) force field; vdW = van der Waals contribution from MM; GAS = total gas phase energy ELE+vdW+INT; INT = internal energy arising from bond, angle and dihedral terms in the MM force field (this term always amounts to zero in the single trajectory approach); PBELE/GBELE = sum of the electrostatic solvation free energy and MM electrostatic energy; PBSUR/GBSUR = non-polar contribution to the solvation free energy calculated by an empirical model; PBCAL/GBCAL = the electrostatic contribution to the solvation free energy calculated by PB or GB respectively; PBSOL/GBSOL = sum of non-polar and polar contributions to solvation free energy (PBSUR + PBCAL); PBTOT = final estimated binding free energy calculated by MMPBSA method; GBTOT = final estimated binding free energy calculated by MMGBSA method (all energies are in kcal/mol).

that MMPBSA performed better in calculating absolute binding free energies. Therefore, MMPBSA method is preferable and results used using this calculation method will be used in further discussions. After all, both MMPBSA and MMGBSA models gave quantitatively very similar trend and this indicated the favourable binding of the peptides to DENV E-protein.

Basically, the binding mode that has the lowest binding free energy is expected to be the most favourable binding. The complex containing DS36wt had the most favourable binding energy, followed by DN81opt, DN57opt, DS03opt, and DS04opt where the binding energies were -60.43/-60.46 kcal/mol, -27.31/-29.58 kcal/mol, -20.72/-36.80 kcal/mol, -14.72/-33.70 kcal/mol and -11.13/-27.00 kcal/mol, for 10 ns and 50 ns simulation respectively. The major contributors to the binding free energy were both coming from the van der Waals (vdW) and electrostatic (ELE) energies, as calculated in this study using the molecular mechanics (MM) force field. The role of vdW and ELE interactions upon peptide binding were computed to study the forces that led to the most stable conformation.

The gas-phase electrostatic value (ELE) of DN81opt showed highly negative value (-173.22/-232.47 kcal/mol), followed by DS03opt (-45.44/-120.86 kcal/mol), DS36wt (-22.40/-24.70 kcal/mol), DN57opt (-21.05/-151.34 kcal/mol) and DS04opt (-8.21/-164.37 kcal/mol) indicating that ELE interactions could contribute to the binding specificity. In this case, ELE interactions are important forces in the primary approach of the peptides and receptor to each other. These types of interactions are of long-range types and can be determinative in the final protein-ligand complex stability (Ebadi, Razzaghi-Asl, Khoshneviszadeh, & Miri, 2013). These electrostatic interactions can also drive processes such as protein folding, protein-ligand binding, protein-protein interaction, electron transfer, protein binding release and enzyme reaction. Since electrostatic interactions play a major role in protein interaction, the lack of protein

polarization could be reflected in deficiencies in some structural details and dynamical properties extracted from MD simulation. This is extremely important since electrostatic interaction play a significant role in protein functions and structures.

On the other hand, the effect of vdW could not be ignored as it is closely related to the hydrophobic interaction energy. This vdW interaction is a non-specific interaction and is a common name for the attractive and repulsive forces between non-bonded atoms. The total binding free energy (PBTOT) is the sum of ELE, vdW and PBSOL and these energies contributed favourably to the peptide binding. This is in line with the study by Obiol (2008) that showed vdW interaction being the main energy term favouring the peptide binding affinity. In this study, DS36wt showed the lowest vdW energy value (-81.77/-87.59 kcal/mol) followed by DN81opt (-53.86/-42.55 kcal/mol), DN57opt (-37.19/-42.46 kcal/mol), DS03opt (-36.01/-32.81 kcal/mol) and DS04opt (-21.45/-19.93 kcal/mol). In this case, DS36wt fitted more snugly within the binding cavity leading to a tighter binding to E-protein. This favourable fit between the peptide and E-protein could also be due to shape complimentary, in addition to the comparatively stronger vdW interaction. This proved that the vdW interaction was an important factor to the binding affinity. Besides that, the high value of this energy component has been associated with strong ligand binding (Mayuri, Bauve, & Kuhn, 2010), thus providing stability to the complex, and therefore a favourable solvation contribution was observed. The effect of solvation also plays an important role where the existing interactions between water molecules that form hydrogen bonds with the complex also contribute to the complex stability (Bianco, Iskrov, & Franzese, 2012). Therefore, both the intermolecular vdW and the ELE interactions are important contributions to the binding. The presence of these interactions that contributed to the binding free energy plays a crucial role in distinguishing the bioactivity of the peptides.

Meanwhile, the PB/GBSUR terms refer to the non-electrostatic contribution of the solvation effects and are proportional to the solvent-accessible surface area of the molecule. Results showed that energy contributions from PBSUR for DS36wt, DN81opt, DN57opt, DS03opt and DS04opt were -11.97/-11.83 kcal/mol, -9.59/-8.13 kcal/mol, -7.27/-8.61 kcal/mol, -7.20/-6.72 kcal/mol and -4.70/-4.02 kcal/mol, respectively. These negative values indicated that they were favourable components, even though the absolute values were relatively small. On the other hand, the electrostatic contribution to solvation (PB/GBCAL) led to a significant loss of the total binding free energy. PB/GBCAL were shown to be unfavourable to the binding, as indicated by the positive values, which were 55.71/63.65 kcal/mol, 209.36/253.57 kcal/mol, 44.79/165.61 kcal/mol, 73.93/126.69 kcal/mol and 23.22/161.32 kcal/mol for DS36wt, DN81opt, DN57opt, DS03opt, and DS04opt, respectively.

4.3.1.9 Per-Residue Free Energy Decomposition (DC) Analysis

Besides ranking the binding free energies correctly, another advantage of the MMPBSA and MMGBSA models is that they allow for the decomposition of the total binding free energy into individual components, thus enabling the understanding of the complex binding process in detail. To identify the key residues (hot spot) related to the binding process, the binding free energy between the E-protein and peptide inhibitor was decomposed into the contribution of each residue (Ge et al., 2012; Yang, Shen, Liu, & Yao, 2011). These hot spot residues are considered as one of the possible ways to disturb the peptide-protein interaction. Normally, these hot spots are surrounded by moderately conserved and energetically less important residues. They appear to be clustered in tightly packed regions in the centre of the interface (Grosdidier & Fernandez-Recio, 2008). Figures 4.14-4.18 show the energy contribution of each residue to the binding energy. Each amino acid residue was found to exhibit positive or

negative influence on the binding to the E-protein which demonstrated the favourable and unfavourable interactions of individual residue (Yang et al., 2011).

From the per-residue binding free energy decomposition analysis, some residues of the E-protein binding site that contributed to the ligand binding were identified. Figure 4.14 shows that 12 residues on the E-protein were mainly responsible for the peptide DS36wt binding to the protein, namely Met34, Asn37, Leu294, Met297, Ile335, Pro336, Phe337, Met340, His346, Val354, Asn355 and Phe373. This combination of polar and hydrophobic residues clearly showed that they might form side chain or side chain-main chain hydrogen bonds with the peptide. It can be noted that Val354 showed a highly favourable interaction with DS36wt (PBTOTAL = -3.61 kcal/mol) indicating that this charged amino acid was energetically making favourable contact in the binding. For DN81opt peptide, residues of the binding site such as Gly17, Asn37, Met301, Val347, Leu348, Leu351, Val354 and Pro356 were found to be important amino acids in terms of their energetic contributions to the complex formation (Figure 4.15). This polar amino acid (Asn37) may form side chain with polar amine carbonyl group, thus capable to form hydrogen bond. It also gave rigidity to the protein structure by imposing certain torsion angles on the segment of the peptide chain. It is notable that Val347 was the key residue that had the lowest binding free energy (PBTOTAL = -1.6 kcal/mol) and this energy contribution was mainly driven by hydrophobic interaction. Even though the overall contribution to the binding free energy by these residues was small, this finding would have consequences to explain the function of the systems.

On the other hand, for the complex containing DN57opt peptide, residues Gln52, Ala54, Leu56, Ile129, Gln131, Pro132, Pro187, Asn194, Arg210, Gly223 and Ala224 showed significant and favourable contributions to the binding due to their negative energy values. As can be seen from Figure 4.16, the major favourable energy contributions originate predominately from the residue Arg210 (PBTOTAL = -5.01

kcal/mol). Other observed E-protein binding site residues that contributed to the per residue binding energy were Ala54, Leu56, Ile129, Pro132, Pro187 and Ala224. Normally, Gly223 is found at the surface of proteins, usually within loop, resulting in high flexibility to this region. On the contrary, Pro132 provide rigidity to the protein structure by imposing certain torsion angles on the segment of the polypeptide chain. These two residues are often highly conserved in protein families since they are essential for preserving a particular protein three-dimensional fold (Lodish et al., 2000).

For the complex containing DS03opt peptide (Figure 4.17), ten residues were found to form a hot spot on the E-protein. Interestingly residues participated in hydrophobic interactions such as Met301, Pro336, Ile339, Met340, Val347, Ile379 and Val382 contributed to important hydrophobic contact with the E-protein with Val347 (-2.73 kcal/mol) had the highest contribution to the binding energy. These hydrophobic interactions had minimum coulombic interaction energies (Ebadi et al., 2013). It also further illustrated that charged residue such as Lys344 and polar residue such as His346 were keys residue in selective binding. This favourable contribution could partially be due to their important role in stabilising the protein-peptide binding (Betts & Rusell, 2003).

There were six peptide residues for DS04opt-E-protein complex that firmly bound to E-protein namely Ile129, Gln131, His209 and Gln211 (Figure 4.18). Our result showed that Ile129 played the most critical role in the binding with per residue free energy of -0.45 kcal/mol. The presence of this charged residue had shown to be stabilising the binding of the peptide to the E-protein. In addition, the favourable contribution of residues Ile129 in a deep hydrophobic pocket (Figure 4.19) proved to be important for the E-protein and this residue was responsible for the increase in hydrophobic interactions.

The hydrophobicity and polarity of the residues involved made them crucial in structural stability and binding to the peptide. In fact, the stability of protein-protein interfaces are dependent on these important amino acids which contributed to a large fraction of the binding energy at a particular interface, and were surrounded by energetically less important residues. Consequently, a peptide inhibitor of peptide-protein interaction does not necessarily need to target the entire surface but rather could be designed to address only those residues located at the hot spots (Wichmann et al., 2010).

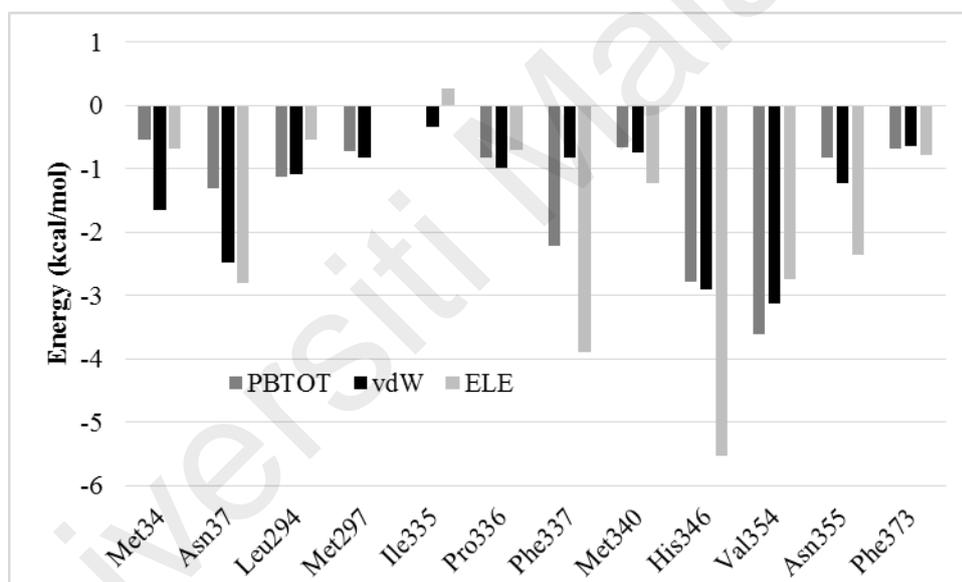


Figure 4.14: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DS36wt-E-protein complex.

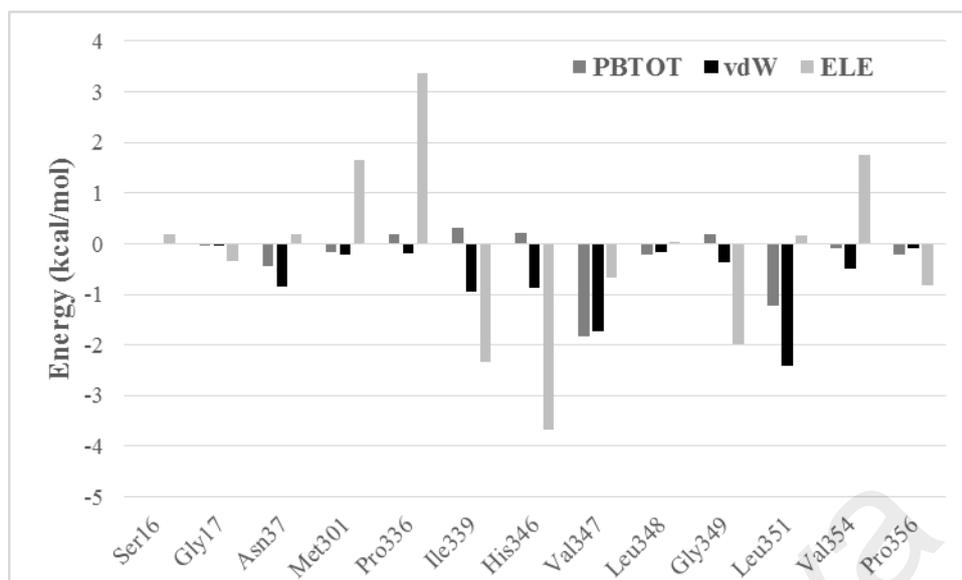


Figure 4.15: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DN81opt-E-protein complex.

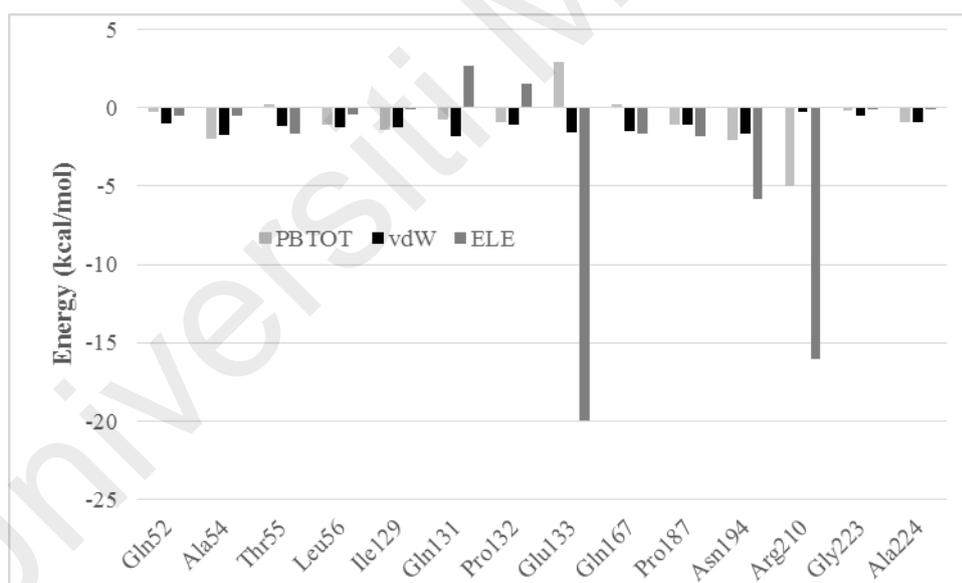


Figure 4.16: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DN57opt-E-protein complex.

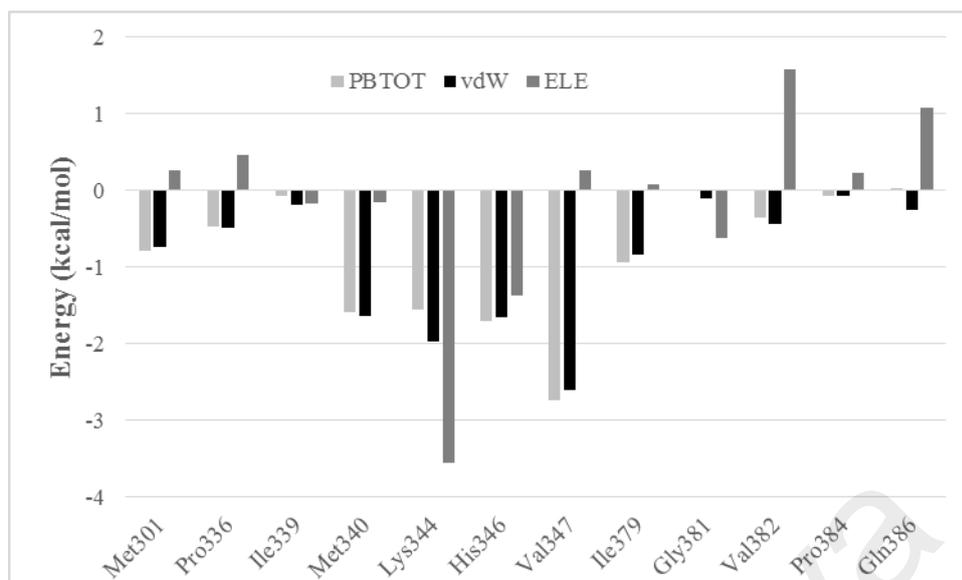


Figure 4.17: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DS03opt-E-protein complex.

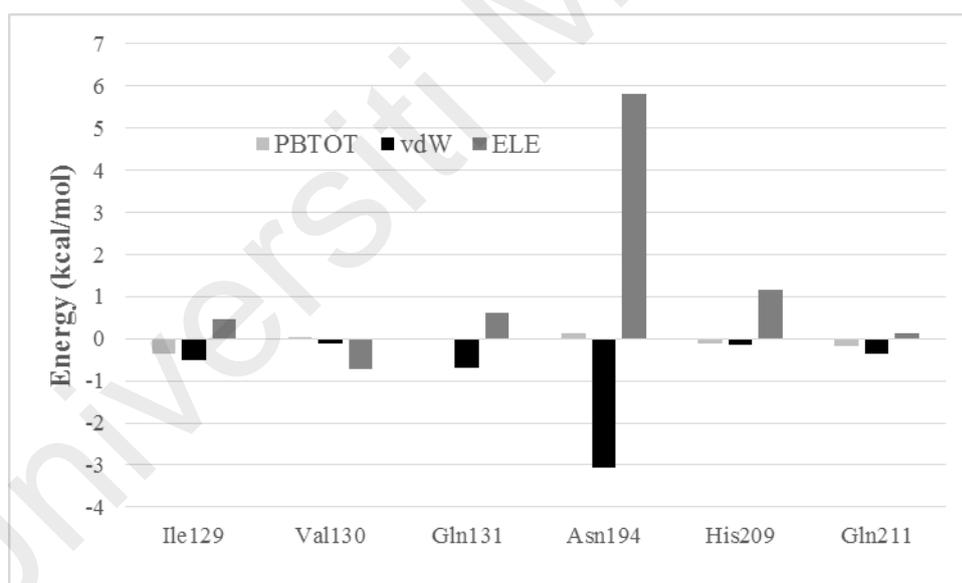


Figure 4.18: Histogram showing the calculated per-residue free energy decomposition using MMPBSA approach for DS04opt-E-protein complex.

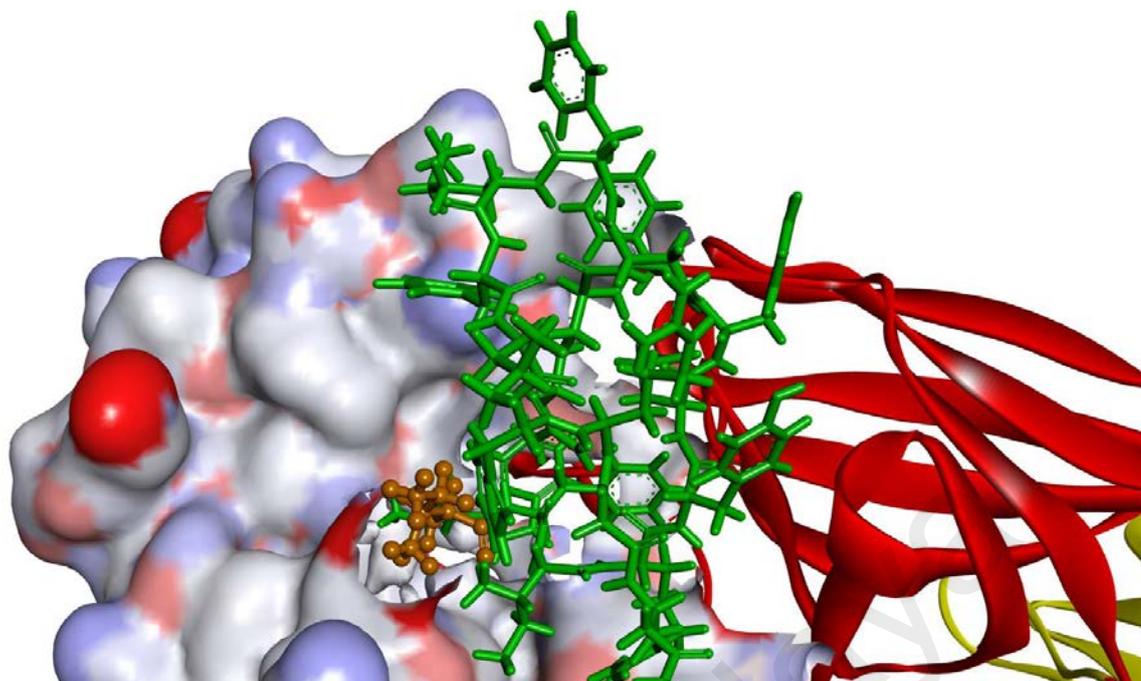


Figure 4.19: Connolly surface representation of residue Ile129 (brown) at the binding site.

4.3.1.10 Analysis of Peptide-E-Protein Interaction

Interactions between the peptides and DENV E-protein were analysed using the LIGPLOT program (Wallace et al., 1995), which portrayed the H-bond interaction patterns and hydrophobic contacts between the peptides and active site residues. The criteria for H-bond interaction used was when the distance between the hydrogen and the heteroatom was within the range of 2.5-3.5 Å and the bond angle was within 109°-110° (Stojanovic & Zaric, 2009). Meanwhile, hydrophobic interactions are particularly important in stabilizing a protein conformation. In this study, analyses of H-bonds and the hydrophobic interactions before and after simulations were performed to indicate the importance of amino acids that contributed to the interactions (Chaudhary & Prasad, 2014).

As seen in Table 4.9, hydrophobic interactions were accounted by interactions between the peptide DN57opt with nine residues in the binding site, which were Ala54,

Leu56, Lys88, Ile129, Trp220, Pro222, Gly223, Ala224 and Ile232. However, after 50 ns simulation, only seven residues, which were Gln52, Ala54, Leu56, Gln131, Pro132, Gln167 and Gly223, were involved in hydrophobic interactions. On the other hand, before simulation, H-bond interactions were recorded involving three residues (Arg57, Arg210 and Asp225), and after 50 ns four residues (Thr55, Asn133, Asn194 and Arg210) were involved. As could be anticipated, side chains of the positively charged amino acids, such as Arg, were the most frequently involved in hydrogen bond formation during the 50 ns simulation. Moreover, there were several interactions with side chains of Asp, where Asp was defined as a hydrogen donor. This was as expected since the side chain of Asp is usually negatively charged. The side chains of polar residues, such as Thr, also participated in hydrogen bond formation, while hydrophobic residues were involved in hydrogen bonds via their backbones (Stojanovic & Zaric, 2009). Thus, this directionality of the hydrogen bond is very important since it contribute favourably to protein stability (Pace et al., 2014).

For DN81opt-E-protein complex, before simulation, there were thirteen residues involved in hydrophobic interaction, whereas after 50 ns simulation the number of residues participating in hydrophobic interaction reduced to five. The same was with H-bond interaction where before simulation there were three residues involved and reduced to two residues after 50 ns. Thus, the interaction study showed more complex stability from initial simulation compared to after simulation. The increase in the number of hydrophobic residues may increase the biological activity of the drug lead. In fact, it has been reported that the binding affinity and drug efficacy could be associated with hydrophobic interaction (Patil et al., 2010). On the other hand, complex containing DS04opt peptide formed H-bonds with Arg57, Asn194 and Arg210 before simulation, and formed H-bonds with Asp192, Asn194, Glu195 and Arg210 after simulation. On top of that, hydrophobic contacts were also recorded with fifteen residues before

simulation, and reduced to four residues after 50 ns simulation. The same trend was also observed for DS03opt-E-protein complex where there were fifteen and four residues involved in hydrophobic contacts before and after 50 ns, respectively. The peptide formed H-bond with Lys36, Glu338 and Asn355 before simulation and with Lys344 and Glu383 after 50 ns. The involvement in both hydrophobic and H-bond interactions by these amino acids is an additional strong argument for their importance.

Meanwhile, complex containing DS36wt peptide shows H-bond contacts with Val347, Arg350 and Asn355 while surrounding residue, Lys36, Met297, Met301, Ile335, Pro336, Glu338, Ile339, Leu348, Gly349, Leu351, Val354 and Pro356 displayed hydrophobic interactions with the peptide before simulation. However, after 50 ns, DS36wt peptide did not reveal any H-bond nor hydrophobic interaction. The loss of H-bond during simulation results in decrease in complex stability. Hence, the absence of hydrogen bonding between the peptide and the surrounding amino acid residues at the binding site may cause its low binding affinity, thus resulting in the low inhibition activity and the high $K_{i,exp}$. The result obtained also revealed the importance of hydrophobic interactions in stabilizing the peptide at the binding interface. According to Patil et al. (2010), the increase in binding affinity of a complex molecule due to optimization of hydrophobic interaction at the target-drug interface, comparatively demonstrate better efficacy of drug leads. Thus, the results presented herein demonstrated that hydrogen bonding and optimized hydrophobic interactions both stabilized the peptides at the target site and helped alter binding affinity. This proved the importance of both interactions in drug design.

Table 4.9: H-bond and hydrophobic interactions analyses using Ligplot, for the top five peptide-E-protein complexes.

Peptide	Before simulation		After 50 ns simulation		
	Hydrophobic interactions	H-bond interactions	Hydrophobic interactions	H-bond interactions	
DN57opt	Ala54	Arg57	Gln52	Thr55	
	Leu56	Arg210	Ala54	Asn133	
	Lys88	Asp225	Leu56	Asn194	
	Ile129		Gln131	Arg210	
	Trp220		Pro132		
	Pro222		Gln167		
	Gly223		Gly223		
	Ala224				
	Ile232				
	DN81opt	Asn37	Ile335	Phe337	Lys36
Leu294		Phe337	Val347	Glu338	
Met301		Gln358	Leu351		
Lys334			Val354		
Pro336			Asn355		
Glu338					
Lys344					
Arg345					
His346					
Val347					
Leu351					
Asn355					
Pro356					
DS04opt		Ala54	Arg57	Ile129	Asp192
		Thr55	Asn194	Val130	Asn194
	Leu56	Arg210	Gln131	Glu195	
	Ile129		Gln211	Arg210	
	Gln131				
	Glu133				
	Glu195				
	Gln211				
	Leu214				
	Asp215				
	Trp220				
	Pro222				
	Gly223				
	Ala224				
	Ile232				
DS03opt	Asn37	Lys36	Met340	Lys344	
	Leu294	Glu338	Arg345	Glu383	
	Met301	Asn355	His346		

Table 4.9: (Continued)

	Lys334		Val347	
	Ile335			
	Pro336			
	Phe337			
	Met340			
	Arg345			
	His346			
	Leu348			
	Arg350			
	Val354			
	Pro356			
	Gln386			
DS36wt	Lys36	Val347	N/A	N/A
	Met297	Arg350		
	Met301	Asn355		
	Ile335			
	Pro336			
	Glu338			
	Ile339			
	Leu348			
	Gly349			
	Leu351			
	Val354			
	Pro356			

4.3.1.11 Hydrogen Bond Occupancy in Molecular Dynamics Simulation

Peptide-E-protein complex system has the ability to participate in H-bond as both donors and acceptors. Since the formation of internal H-bonds in protein polarizes the donors and acceptors, the H-bonds in proteins are basically more stable during MD calculations. The variation of donor-acceptor distance using MD simulation could be used to evaluate the forming and breaking of H-bonds. A H-bond is characterized according to the distance between two heavy atoms, such as nitrogen and oxygen atoms, is in the range of 2.5-3.2 Å and the angle N-H-O is in the range of 130°-180° (Schaeffer, 2008). In this section, the H-bond angle setting was setting slightly different with the Ligplot program (section 4.3.1.10) due to different algorithm used.

In this work, MD simulations of peptide-E-protein complexes were performed to study hydrogen bonds between the peptide and E-protein. In order to gain insight into the efficiency of the peptide binding into E-protein binding pocket, the percentage occupancy and number of hydrogen bond formation were examined. A summary of the percentage of H-bonds between the peptides and the E-protein in the trajectories is given in Table 4.10. The computational result shows that occupancy percentages of hydrogen bonds averaged over simulation time were 91.44%, 90.68%, 72.48% and 41.24% for each complex containing DS04opt, DS36wt, DS03opt and DN57opt peptides, respectively. For DS04opt-E-protein complex, additional H-bond with Asp192 with occupancy of 35.44% present in the trajectory. Complex containing DS36wt peptide also interacted with His144 with low percentage (35.48%) of H-bond. For DS03opt and DN57opt-E-protein complexes, the percentage of H-bond with Glu383 and Asp418 were 46.16% and 23.4% respectively also show low H-bond patterns. The lower occupancy means that the hydrogen bond is sometimes broken in the course of the simulation whereas higher occupancy indicates additional H-bonds between the amino acids residue and the ligand molecule are created during simulation

(Mukhametov, Newhouse, Aziz, Saito, & Alam, 2014). Hence, it could be suggested that H-bonds were frequently formed and broken during the MD simulations in this study. The breaking of intra-protein hydrogen bonds might also be responsible for deformation or denaturation of some local structures of the complexes during MD simulation. The breaking of these bonds could result in changes in the secondary structures since the backbone H-bonds lie inside either helices or sheets or other local secondary structures in the protein (Ebadi et al., 2013). Therefore, this study provides strong evidence that hydrogen bonds are dynamically stable and majority of H-bonds appearing in a structure remain intact in MD simulation.

Table 4.10: Hydrogen bond analysis between peptides and E-protein at 50 ns.

System	Donor	Acceptor	Occupancy (%)	H-bond distance (Å)	H-bond angle (°)
DS04opt	Glu195@OE1	Arg413@NH2	91.44	2.783	133.9
	Asp192@OD2	Arg413@NH2	35.44	2.807	168.9
	Asp192@OD1	Arg413@NH2	32.56	2.811	174.3
DS36wt	Pro400@O	Arg350@NH2	90.68	2.801	135.9
	Pro400@O	Arg350@NE	42.92	2.876	168.0
	His144@ND1	Thr353@OG1	35.48	2.889	139.5
DS03opt	Lys344@O	His401@NE2	72.48	2.819	167.6
	Glu383@OE2	Arg403@NH2	46.16	2.795	175.3
	Glu383@OE2	Arg403@NH1	45.44	2.808	161.3
DN57opt	Glu401@OE2	Asn194@ND2	41.24	2.862	135.8
	Glu401@OE1	Asn194@ND2	35.72	2.870	133.6
	Asp418@OD2	Arg188@NH1	23.40	2.798	170.5

4.3.1.12 Secondary Structure Analysis

Molecular dynamics (MD) simulations technique is one of the most popular computational methods to investigate the secondary structures of proteins and peptides. The secondary structure of a protein indicates the local conformation of the polypeptide backbone, normally focusing on common regular folding patterns. A few types of secondary structure are particularly stable and occur widely in proteins such as α -helices, β -sheets, turns and coil and are normally distinguished by the backbone torsional angles (ϕ , Ψ) and the types of non-covalent interactions. Thus, in order to have an insight on how overall peptide folding representation looks like, we investigated the folding behaviour and secondary structure formation of the peptide-E-protein complex at the nanosecond timescale. In the previous study done by Demir et al. (2014), the secondary structure of selected peptides had been characterised via circular dichroism (CD) spectroscopy and conventional MD simulations. The results obtained from MD were found to be consistent with the CD results.

In order to get insight into secondary structural elements of peptide folding, in this study, secondary structural elements were depicted as a function of simulation time for both peptide folding and peptide-E-protein complexes. As seen in peptide folding secondary structures (Figure 4.20), the conformational ensembles of peptide folding are dominated by “turn structures” followed by β isolated bridges, 3-10 helices, α -helices, and with small amount of extended configurations and pi-helices. Turn structures are the mobile regions of the molecule and is the simplest secondary structure element. The presence of turn structures indicated that some hinge residues may be flexible due to their inherent conformation properties or because of cross-correlated motions. Analysis also demonstrated that turn structures can rapidly form, disappear and reform. These folded peptides had a random secondary structure in solution, which may increase their capability to interact with the target protein easily and this could be the main reason for

increasing affinities of the peptides. Meanwhile, β isolated bridges showed an equal propensity in both rigid as well as flexible mobile residues and these structures could appear or disappear during simulation. However, any extended configuration structures were not observed. Analysis of secondary structure evolution throughout the MD trajectories also showed that α -helices in the peptide folding remained intact for most of the systems.

On the other hand, for all the peptide-E-protein complexes, the extended configurations, turn configurations and α -helices, were dominant conformational structures while 3-10 helices and coils were minor structures (Figure 4.20). Extended structures and coils frequently appear in flexible regions rather than in rigid regions. Interestingly, most amino acids were observed to be present in the extended configuration within the proteins. However, not all peptides were able to form stable α -helices. This could be due to amino acid sequence that affected the stability of α -helices. Additional interactions should occur between amino acid side chains in order to stabilize this structure. The presence of residues such as Glu, Asn, Ser, Thr and Leu tend to prevent formation of α -helices once they occur close together in the chain. The presence of many of these residues, with positively charged R groups, will repel each other and prevent formation of the alpha helices. Thus, this secondary structure study is significant in generating information on the factors that ensure stabilization of the peptide folding as well as providing a convenient way to classify types of folds.

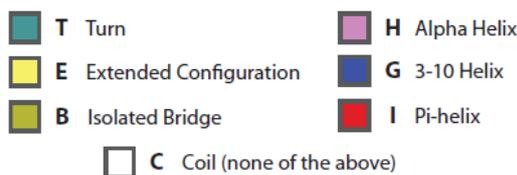
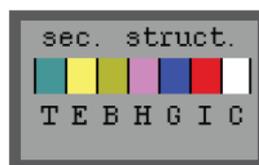
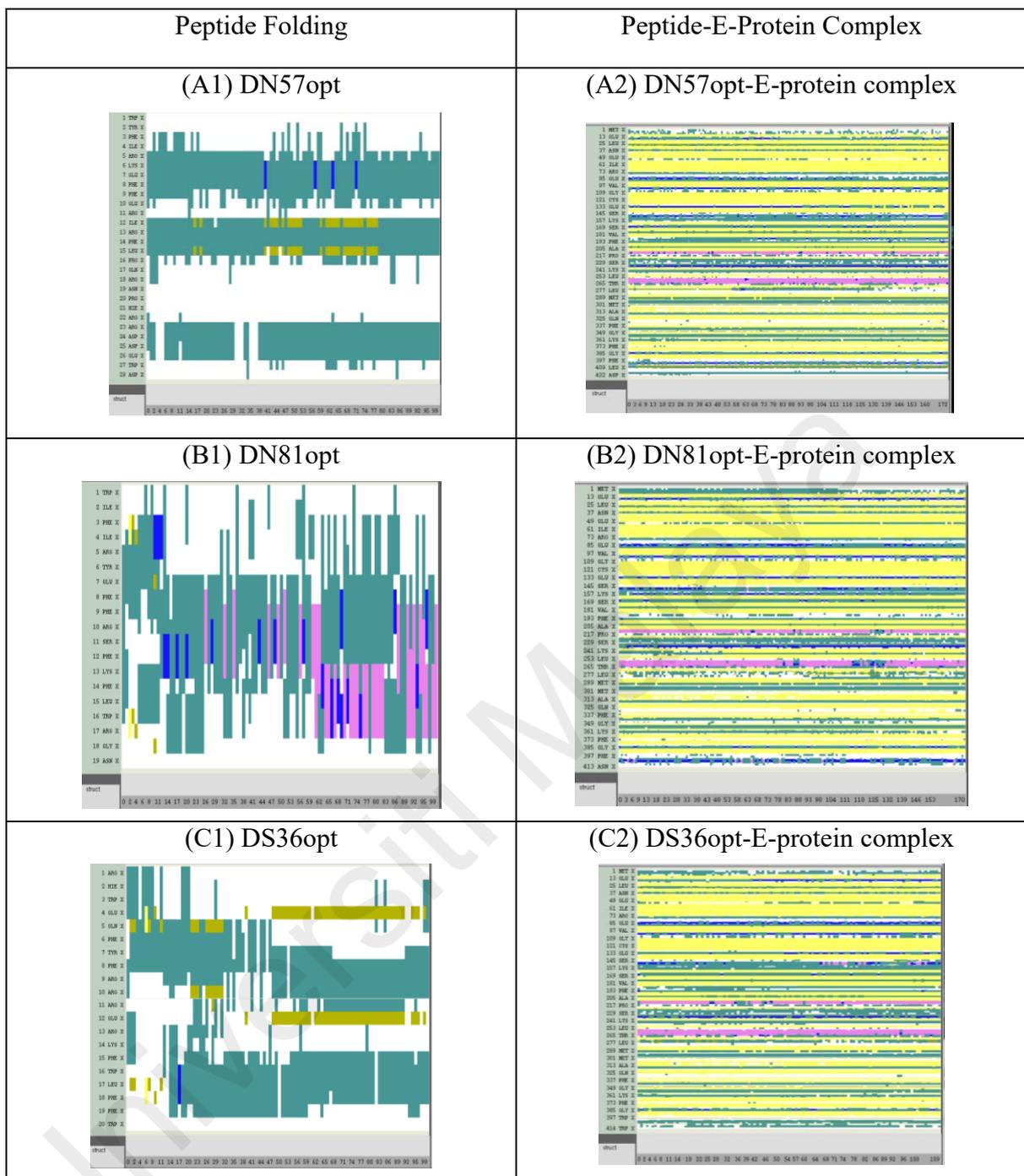


Figure 4.20: Secondary structure plots of peptide-E-protein complexes using Visual Molecular Dynamics (VMD).

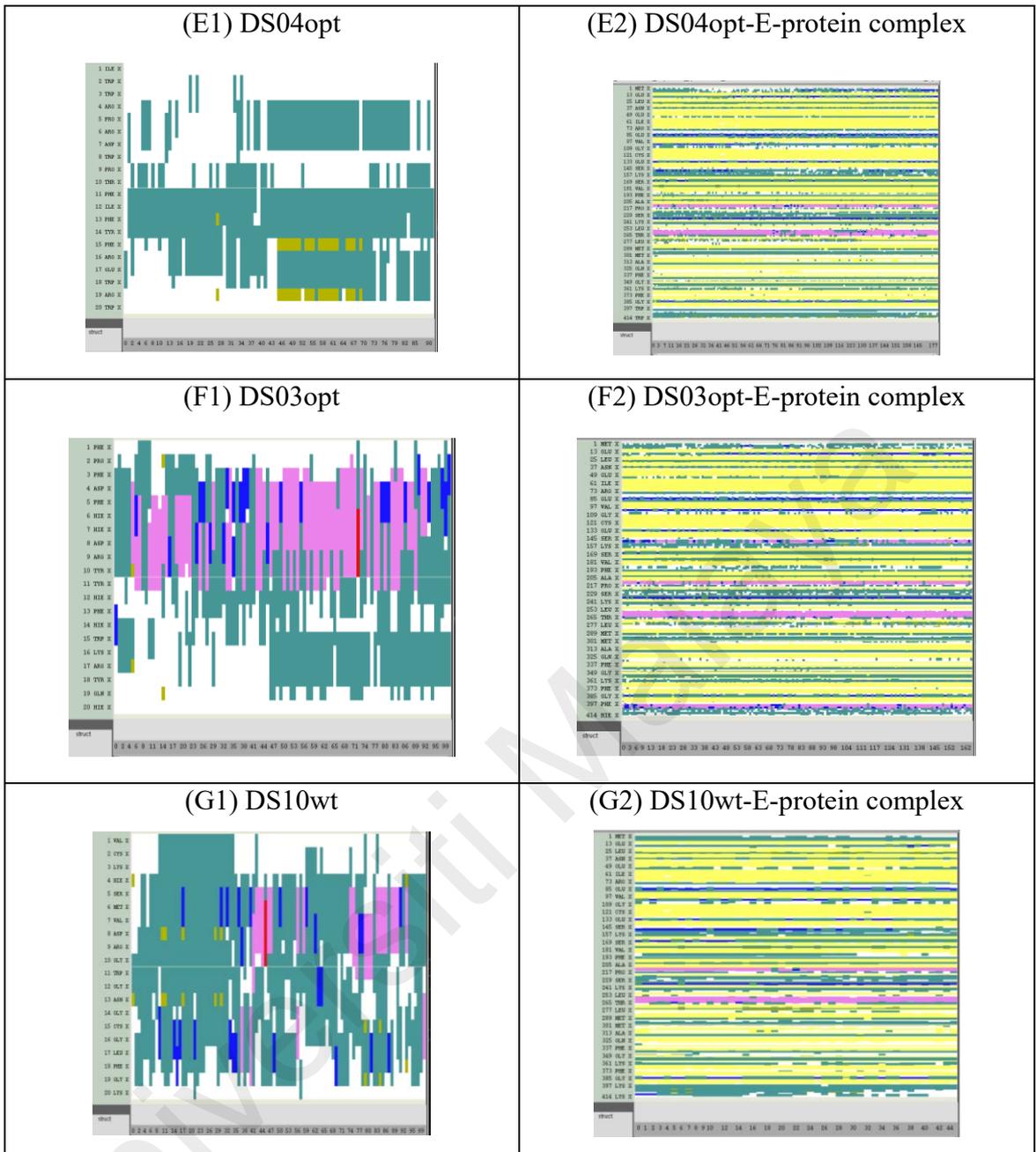


Figure 4.20: (Continued)

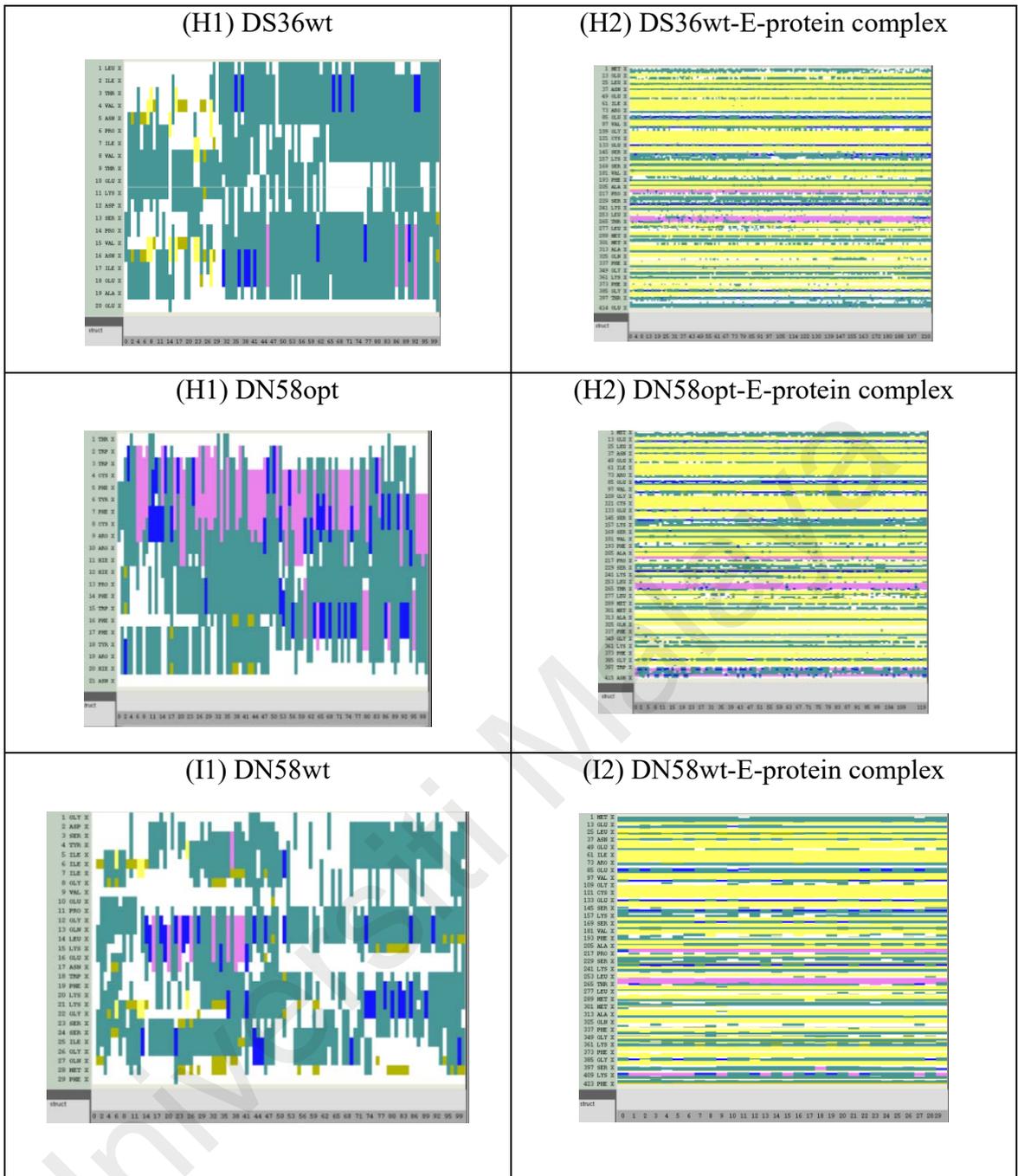


Figure 4.20: (Continued)

4.3.2 Peptide Synthesis

In this study, the peptides with codes DN81opt, DS36wt, and DS36opt were selected for synthesis due to their ability to interact with DIII and the binding between them appeared to be cooperative (Baharuddin et al., 2014). In addition, docking and molecular dynamics simulation that had been performed in this study proved the potential of these peptides as inhibitors. This is also in line with the study performed by Xu et al. (2012) which successfully identified that these peptides correspond to a region of the dengue virus envelope protein and were designed against target DIII of DENV2 envelope protein. Peptides derived from the protein-protein interface have been shown to be able to mimic the modes of binding of its original domain to its specific partner protein. Thus, they may serve as promising leads for drug development. In future, variety of assay tests will be performed on these synthesized peptides to confirm the inhibition effect.

4.3.2.1 Solid Phase Peptide Synthesis (SPPS)

Solid phase peptide synthesis (SPPS) is defined as a process in which a peptide anchored by its C-terminus to an insoluble polymer is assembled by the successive addition of the protected amino acids constituting its sequence. In this study, the C-to-N direction was used to synthesize the peptides as the sequences are long (19-28 amino acids). Attachment of the amino acid can be extended with the peptide chain anchored through the backbone amide of the C-terminal residue. This allows an attachment peptide to be synthesized in the common C-to-N direction, with the free C-terminus available for elaboration into suitable peptide mimetic functional groups. Furthermore, in (C-to-N) peptide synthesis, an excess of the activated carboxyl component is used to drive the reaction to completion. However, in N-to-C synthesis, the carboxyl group is anchored and could not be generated in excess (Bodanszky, 1993). In addition, this

reverse strategy (N-to-C peptide synthesis) allows the generation of C-terminal-modified peptide, which have potential application in therapeutic field as well as in fragment condensations for the assembly of large peptides. Nevertheless, this method is only applicable to peptide aldehydes and chloromethyl ketones (Jakubke & Sewald, 2008; Lebl & Houghten, 2001). Since in this study, no modified in C-terminal is required. Thus, the common C- to N- terminal direction of peptide synthesis is used.

In general, each amino acid addition is referred to a cycle consisting of cleavage of the Fmoc protecting group, washing step, coupling of a protected amino acid and washing steps. In this study, the automated peptide synthesizer was used to synthesize the peptides (Figure 4.21). As the growing chain is bound to an insoluble support, the excess of reagents and soluble by-products can be removed by simple filtration. Washing steps with appropriate solvents ensure the complete removal of cleavage agents after the de-protection step as well as the elimination of excesses of reagents and by-products resulting from the coupling step. In this study, the peptide synthesis and purification steps were quite challenging because of poor peptide solubility, which led to low coupling and/or deblocking reactivity (Viau, Letourne, Sirois-Deslongchamps, Boulanger, & Fournier, 2007) as well as due to difficulty in attaching the amino acids peptides with long sequences (28 amino acids).



Figure 4.21: Automated peptide synthesizer used in this study.

In this study, chlorotriyl chloride (CTC) is used as polymeric support as it is one of the most widely used resin for the solid-phase synthesis of C-terminal peptides. The advantage of CTC resin is, it can be used for the preparation of both protected and unprotected peptides (Garcia-Martin, Bayo-Puxan, Cruz, Bohling, & Albericio, 2007). In addition, the CTC resin is preferred due to its ability to reduced formation of side reactions such as recemization and diketopiperazine. This is due to the characteristic hindered structure of chlorotriyl group. Moreover, CTC resin can be applied to prepare protected peptide acids under mild cleavage conditions, which can be used in the fragment condensation for larger polypeptides (Lee, Ryoo, & Lee, 2007). Due to its great properties as a support for solid phase peptide synthesis, the CTC resin was chosen.

Meanwhile, Flourenylmethyloxycarbonyl (Fmoc) protecting group usage is compulsory to prevent undesirable side reactions with various amino acid side chains. In fact, Fmoc protection allows for a milder deprotection scheme compared to tert-Butyloxycarbonyl (Boc) protection group. It has the ability to be cleaved under mild basic conditions and has stability towards acid. Fmoc deprotection utilize a base which is in this study, 20% piperidine in DMF which resulting in neutralization of exposed amine. Thus, the neutralisation of the peptide-resin is not necessary as compared to Boc approach.

In order to facilitate peptide formation with minimum side reactions, chemical groups that would be able to bind to the amino acid reactive groups and eventually block the functional group from nonspecific reaction has been used. Coupling reagent and additives such as Diisopropylcarbodiimide (DIC) and 1-Hydroxybenzotriazole (HOBt) has been used for this purpose. DIC is particularly useful reagent for SPPS as it easily handled as a liquid, and the urea byproduct formed is soluble in most organic solvents, allowing facile removal during resin washes (Isidro-Liobet et al., 2019). The

presence of these coupling reagents would form a less-reactive intermediate that reduces the risk of racemization and eventually shorten the length and cause branching of the peptide chain (Valeur & Bradley, 2009). In this study, racemization is referring to the one of the main side reaction when activating carboxyl groups of amino acid. Presently, this combination of HOBt and DIC is one of the best methods for amino acid coupling.

Meanwhile, TBTU is another coupling reagent used in SPPS that also showed resistance against racemization in spite of their high coupling rates. It has frequently used in peptide synthesis due to its mild activating properties. Previous studies (Gutheil & Xu, 2002) had successfully proved the high purity (80%) and low racemization (5%) of synthesized peptides using combination of HBTU/DIEA and HATU/TMP coupling methods. HBTU and TBTU differ only in the choice of anion. The same result was also observed by Bodanszky (1993) which claimed that the HATU/TMP coupling method gave low levels of epimerization.

4.3.2.2 Lyophilisation

After the cleavage of the peptides from the polymeric support, the crude materials were treated with glacial acetic acid and lyophilized. This treatment significantly improved the solubility of the peptides thus allowing their dissolution in aqueous conditions (Viau et al., 2007). This was then followed by purification using reverse phase high performance liquid chromatography.

This lyophilisation process is important for long-term storage of peptides. In this study, freeze dryer was used (Figure 4.22) and the process to lyophilised takes around 36 hours. Lyophilized peptides can be stored for years at temperatures of -20°C or lower with little or no degradation (Hoofnagle et al., 2016). In contrast, peptides in

solution form are much less stable. Since peptides are susceptible to degradation by bacteria, they should dissolve in sterile and purified water.



Figure 4.22: Freeze dryer used in this study for lyophilisation process.

4.3.2.3 Purification of Peptides

The crude peptides obtained from solid phase peptide synthesis (SPPS) contained many by-products resulting from the side products stemming from cleaved side chains or oxidation during the cleavage and deprotection process. Purification of peptides requires the removal of deletion peptides resulting from incomplete coupling/deprotection steps, from racemization or side-chain rearrangement, and from various chemical substances introduced during the deprotection or cleavage stages of SPPS procedure.

In this study, reverse phase high performance liquid chromatography (RP-HPLC) was used to purify the peptides at minimal cost. The combination of RP-HPLC and liquid chromatography mass spectrometry (LCMS) provides a powerful tool for peptide analysis. By measuring the molecular weight of the peptides present in each of the peaks, LCMS provides useful information that assists in identifying the peaks

separated by RP-HPLC. LCMS also confirms the purity of each peak and reports their molecular weights (Kang, 2012).

Theoretically, peptides are separated by interacting with the hydrophobic surface of particles packed in columns. The particles in the column are usually made of silica because silica is physically robust and stable under most solvent conditions (except at pH greater than 6.5). When separating peptides by using RP-HPLC, detection is normally by UV absorption at 214-215 nm. The peptide bond absorbs well in this wavelength range and provides the most sensitive detection for all types of peptides (Anthis & Clore, 2013).

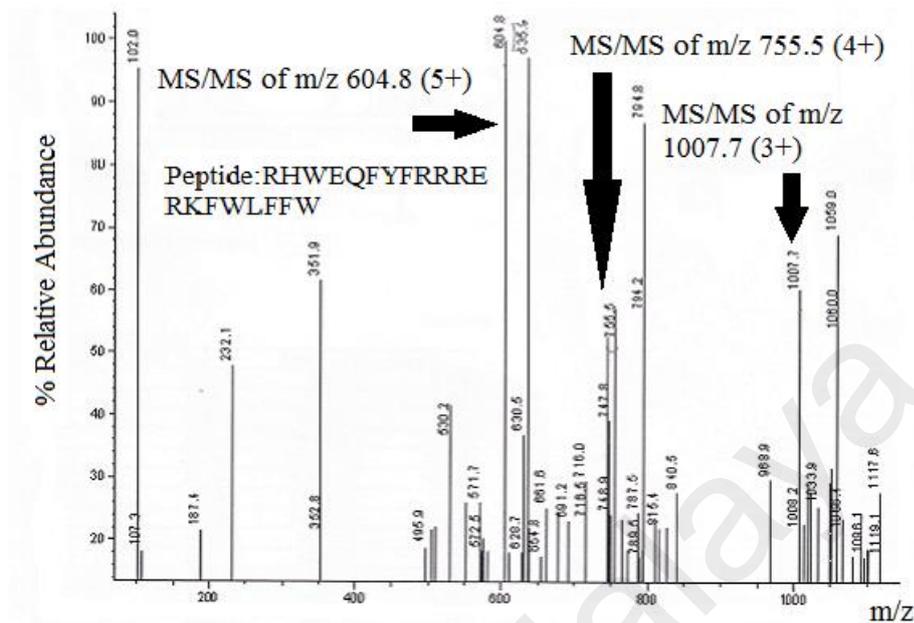
RP-HPLC conditions are essentially the reverse of normal phase chromatography. The peptides bind on the column through hydrophobic interactions and are eluted by decreasing the ionic strength. Generally, the column supports are composed of hydrocarbon alkane chains which are covalently attached to silica. These chains range from C4 to C18 carbon atoms in length. Since elution from the column is a function of the hydrophobicity, the longer chain hydrocarbon columns are better for small, highly charged peptides (Carr, 2002).

Mass spectrometry has become one of the well-known methods of detection for high-performance liquid chromatography (HPLC) analysis of biopolymers for varieties of applications. When liquid chromatography is interfaced directly to mass spectrometry (LCMS), molecular weight information of peptide can be obtained from the mass spectra. Peptides produce varieties of product ions depending on the quantity of vibrational energy they possess and the time-window allowed for dissociation. The benefit of using high resolution mass analysers is the ability to predict the charge state of a peptide ion by calculating the m/z difference between its isotopic peaks. This

information will lead us to calculate the monoisotopic mass or molecular weight of the peptide being ionised (HuiSong, 2010).

The basic concept of LCMS is to form ions from a sample, to separate the ions based on their m/z ratio and to measure the abundance of the ions. During electrospray ionization, peptide acquires multiple charges (Samson, Rentsch, Minuth, Meier, & Loidl, 2019). This allows it to be analysed by a mass spectrometer with a relatively limited mass (mass-to-charge) range. It is common to see 2+, 3+, 4+ and 5+ charge states depending on the size of the peptide. Peptides that are greater than 2,500 Da tend to multiply charge under electrospray conditions, thus reducing their m/z into analytical range. Most peptides below 4,000 Da predominantly form the $m+2$ charge under electrospray conditions. Larger peptides and proteins tend to form additional; $m+3$ and $m+4$ charge states, resulting in the signal being distributed between several different ions. Thus, multiple charges produced by peptides are crucial in determining molecular weights of synthesized peptides.

For peptide with code DS36opt (Figure 4.23), the spectrum is dominated by three peaks, one at m/z 604.8, one at m/z 755.5 and another at m/z 1007.7. These three peaks correspond to the same peptide. Since most electrospray spectrometers have good resolution, it is possible to look at the isotope of singly, doubly or multiple charged ions. By looking at the isotopic distribution of the above ions, the isotopes are separated by three mass units at m/z 1007.7, m/z 755.5 and at m/z 604.8 with each pattern corresponding to the 3+, 4+ and 5+ charges, respectively. This is due to the fact that mass spectrometer measures the mass-to-charge ratio. Another advantage of generating multiply charged ions with electrospray is that multiply charged ions tend to give more complete fragmentation spectra. This is particularly important for de novo sequencing of peptides (Trauger, Webb, & Siuzdak, 2002).



To calculate the peptide ion's mass, the m/z for the monoisotopic peak was multiplied by the charge state. The calculations below show how the mass of the multiple charge of peptide was determined:

DS36opt:

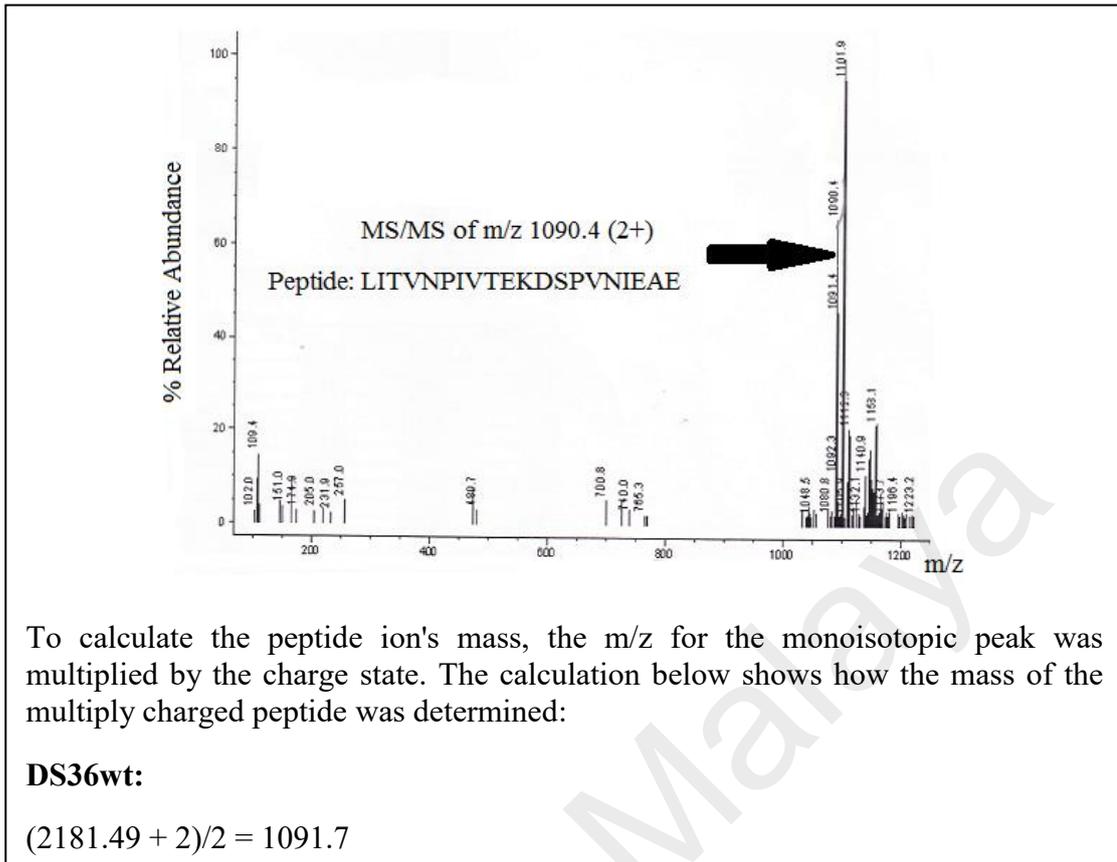
$$(3021.50 + 3)/3 = m/z 1008.17$$

$$(3021.50 + 4)/4 = m/z 756.38$$

$$(3021.50 + 5)/5 = m/z 605.3$$

Figure 4.23: LCMS spectrum of peptide with code DS36opt with $M_w = 3021.46$ g/mol. ESI calcd for peptide code DS36opt $[M+3H]^{3+}$ $m/z = 1007.7$, $[M+4H]^{4+}$ $m/z = 755.5$, $[M+5H]^{5+}$ $m/z = 604.8$

On the other hand, for peptide with code DS36wt (Figure 4.23), LCMS analysis conformed that only one peak in the chromatogram corresponds to the peptide with the same molecular weight which is at m/z 1090.4, corresponding to doubly charged ion (2+). The same pattern is observed with peptide DN81opt (Figure 4.24), where only one peak at m/z 523.2878 which corresponds to multiply charged ion (5+) is observed in the spectrum. LCMS spectrum of this multiply charged ion matches a 19-amino acid of the DN81opt peptide (WIFIRYEFFRSFKFLWRGN).

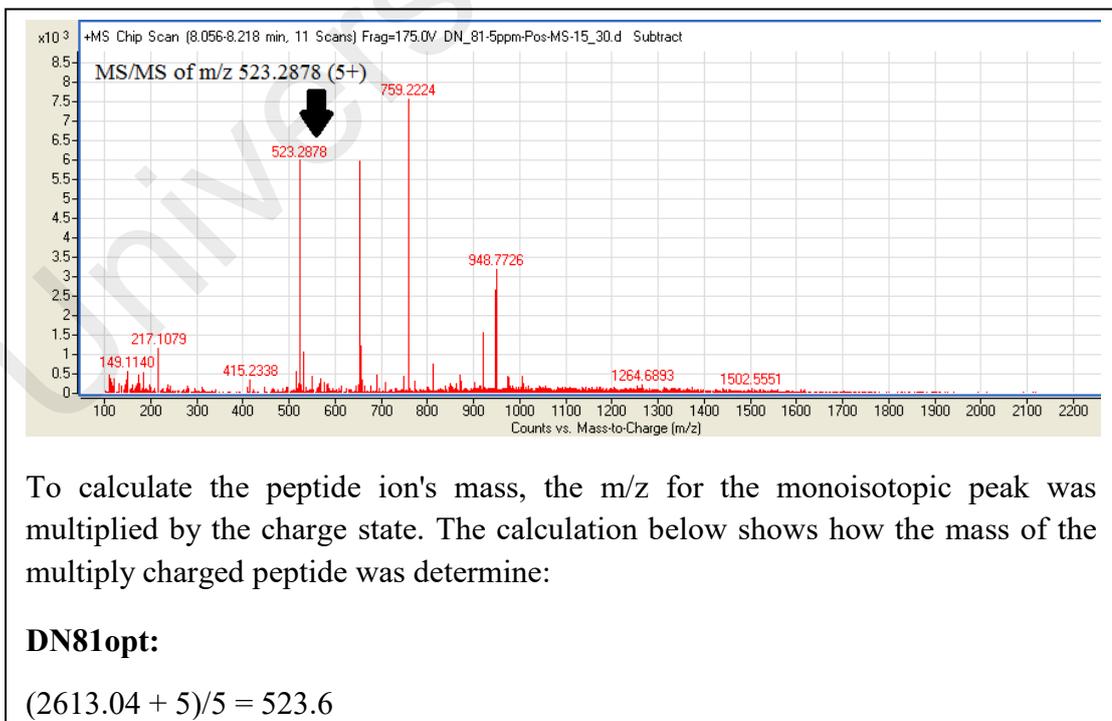


To calculate the peptide ion's mass, the m/z for the monoisotopic peak was multiplied by the charge state. The calculation below shows how the mass of the multiply charged peptide was determined:

DS36wt:

$$(2181.49 + 2)/2 = 1091.7$$

Figure 4.24: LCMS spectrum of peptide with code DS36wt with $M_w = 2181.45$ g/mol. ESI calcd for peptide code DS36wt $[M+2H]^{2+}$ m/z = 1090.4



To calculate the peptide ion's mass, the m/z for the monoisotopic peak was multiplied by the charge state. The calculation below shows how the mass of the multiply charged peptide was determine:

DN81opt:

$$(2613.04 + 5)/5 = 523.6$$

Figure 4.25: LCMS spectrum of peptide with code DN81opt with $M_w = 2613.04$ g/mol. ESI calcd for peptide code DN81opt $[M+5H]^{5+}$ m/z = 523.2878

4.4 Conclusion

In this study, MD simulations were performed on the peptides to evaluate their interactions with DENV E-protein in aqua, and to investigate the binding free energy by means of MMPBSA and MMGBSA methods and energy decomposition analyses. Results showed that the calculated binding free energies concurred well with the tryptophan fluorescence quenching assay previously report, thus confirming the binding affinities of these peptides. Meanwhile, RMSD values indicated the overall structural stability of the peptide-E-protein complexes. Electrostatic and van der Waals interactions contributed mostly during the interactions between the E-protein and the peptides. Additionally, several amino acids had been identified for their major contributions to the binding efficiency towards E-protein. It can be suggested that these amino acids can later be modified by mutation to improve the binding activity.

The production of peptides via automated peptide synthesis on solid support provided a great variety of benefits. Even though SPPS method offered excellent purity and yield standards, impurities and imperfections still occurred along the way. This was due to the increase in the length of the peptide sequences, as more steps are needed to complete the synthesis. Therefore, purification step was necessary in order to ensure optimal quality of the peptide.

In this study, selected peptides had been chosen for synthesis using automated peptide synthesizer. In the LCMS spectrum, the signal of the desired products were observed and the products were successfully purified. Purity and identity of the products were confirmed by HPLC. All the m/z values were consistent with the calculated ones based on the chemical formulae of the expected peptides. Based on the obtained analytical data, it could be concluded that the desired peptides were successfully obtained.

CHAPTER 5: GENERAL DISCUSSION & CONCLUSION

5.1 General Discussion

From Chapter 3, the concepts of identifying and characterizing protein binding pocket computationally using Discovery Studio 4.0 software had been discussed. In this study, nLc₄Cer which belongs to a carbohydrate group, may become a potential lead for drug design. There is always the possibility of the molecule binding to a site other than the predicted location on the E-protein. Parameters such as energy, temperature, RMSD, RMSF and radius of gyration (R_g) had shown consistency throughout the 50 ns simulation time. Meanwhile, cluster analysis had presented nine conformational clusters, with each cluster not showing much difference in binding energy. The cluster group with the lowest binding energy was chosen for further MD study and was confirmed to have reasonable binding mode. In addition, post analysis on the simulated DENV E-protein-nLc₄Cer complex was done by calculating the binding free energy and decomposed energy using MMPBSA and MMGBSA methods. Thus, several important residues that contributed to the complex binding energy had been identified. It was also found that, hydrophobic and hydrogen bonding interactions were responsible for the high affinity of nLc₄Cer to bind to the target. From this study, it can be concluded that the replication of DENV could be interrupted by suitable compounds targeting specific viral protein which is, in this case, the E-protein. This compound may serve as the basis for the development of new drugs to combat the DENV infection.

The study was then continue by applying docking approach to systematically investigate E-protein-nLc₄Cer association pathways. This study has demonstrated three docking pathways leading from the surface of the E-protein to the correct docking pathway. The ligand flexibility was found to play an important role in nLc₄Cer entry to the E-protein binding pocket. Knowledge of association pathway would be useful in

discriminating between different ligand pathways that could have been favourably docked in the binding pocket.

In Chapter 4, in order to evaluate the peptide interactions with E-protein and to investigate the binding free energy by means of MMPBSA and MMGBSA methods as well as energy decomposition analyses, the molecular dynamics simulation was performed. Results obtained showed that the calculated binding free energies concurred well with the tryptophan fluorescence quenching assay previously reported, thus confirming the binding affinities of these peptides. Electrostatic and van der Waals interactions had become major contributors to the interactions between E-protein and the peptides. The Ramachandran plot of the peptide-E-protein complexes had shown that less than 15% of protein residues were inside the disallowed region. Meanwhile, RMSD was used to ensure the overall structural stability of peptide-E-protein complexes. Also, several hot-spot residues had been identified through pairwise energy decomposition method for their major contributions to the binding efficiency towards E-protein.

Besides that, selected peptides had been chosen for synthesis using peptide synthesizer. The peptide purity was confirmed via LCMS and the signal of the desired products were observed. Furthermore, the masses of the successfully synthesized peptides had been confirmed with HPLC where all the m/z values were found to be consistent with the calculated expected peptides. Hopefully, this study may contribute to the search for antiviral substances, which interfere in the attachment process of the DENV to the host cell, or effectively disrupting the process of membrane fusion in the entry process of the virus into the host cell.

5.2 Conclusion

In conclusion, the aim and all objectives of this study have been achieved. Both nLc₄Cer and peptides used in this study could have huge therapeutic potentials which make them become attractive candidates for the development of drugs. Results obtained from this study indicate that these compounds could behave as inhibitors which bind across E-protein and act by inhibiting the conformational changes that occur during viral fusion with endosomal membranes. Anti-dengue drugs may potentially shorten the duration of illnesses and reduce the risk of disease progression, and this would be a significant advancement for both patients and health systems in countries where dengue become endemic. Anti-dengue drugs could also potentially lower viral loads and reduce dengue severity, thus resulting in dengue cases in endemic areas can significantly be managed and the state of public health can be improved. Last but not least, structural information derived from this study may aid in the selection of appropriate drug candidates for further development and would be able to assist in pharmacokinetic studies.

5.3 Significance of the Study

This study has yielded more information regarding peptides as DENV potential inhibitors as there were limited information available from the previous studies (Chew et al., 2017). The results obtained can be used as references in planning for more effective solutions in terms of effective vaccination and antiviral drugs to health problems in relation to dengue fever. This study will help to increase the awareness and knowledge as well as to provide more effective, sustainable educational resources regarding to dengue diseases all around the world. This study also highlights the importance of drug design and development since only symptomatic and supportive treatments are available to manage the dengue diseases. Moreover, steps to develop a

tetravalent anti-dengue therapeutic capable of providing protection against all four dengue serotypes can be planned to prevent sequential serotype infections.

5.4 Suggestion for Future Studies

Due to resource and technical limitations, only three peptides were successfully synthesized. In future, more peptides will be synthesized and tested for their inhibition study. Beside of using long sequence of amino acid, short peptides can be designed because shorter sequences are basically easy to synthesize compared to longer peptides. Secondary structures of the designed peptides can also be characterized by using experimental techniques such as circular dichroism (CD) spectroscopy. The outcome will highlight the accurate dominant structural components which are random coil, bend and turn for the target peptides.

Synthesized peptides can be tested via plaque reduction assay to confirm their inhibition effect. Similar reduction effects can also be observed using quantitative real-time polymerase chain reaction (PCR) where decrease in transcription level will prove the peptides to function as inhibitors.

Since analyses of the important residues that interacted during the binding had been discussed, therefore in future studies, mutagenesis study can be done by replacing the important amino acid residues with the other amino acids that may enhance the free energy of binding of inhibitor towards E-protein. In addition, quantum mechanics method can be used as a tool to study the breakage and formation of hydrogen bond.

Last but not least, steered MD technique can also be used to study the association pathways of ligands towards the binding site at constant velocity and force. Through this method, the pathway of nLc₄Cer approaching the E-protein binding site will be clearer.

REFERENCES

- Aarthy, M., & Singh, S. K. (2018). Discovery of potent inhibitors for the inhibition of dengue envelope protein: An *in silico* approach. *Current Topics in Medicinal Chemistry*, 18(18), 1585-1602.
- Aci-Seche, S., Genest, M., & Garnier, N. (2011). Ligand entry pathways in the ligand binding domain of PPAR-gamma receptor. *FEBS Letters*, 585, 2599-2603.
- Adam, J. K., Abeyta, R., Smith, B., Gaul, L., Thomas, D. L., Han, G., Tomashek, K. M. (2017). Clinician survey to determine knowledge of dengue & clinical management practices. *The American Journal of Tropical Medicine & Hygiene*, 96(3), 708-714.
- Ahmad, M. (2012). *Mechanisms of protein-protein association*. (PhD), Universitat des Saarlandes.
- Ahmadi, S. (2013). *Modelling and computer simulation studies of sugar based amphiphilic systems*. (Doctor of Philosophy), University of Malaya, Kuala Lumpur, Malaysia.
- Ahsan, H. (2019). Immunopharmacology and immunopathology of peptides and proteins in personal products. *Journal of Immunoassay & Immunochemistry*, 40(4), 439-447.
- Alhoot, M. A., Rathinam, A. K., Wang, S. M., Manikan, R., & Sekaran, S. D. (2013). Inhibition of dengue virus entry into target cells using synthetic antiviral peptides. *International Journal of Medical Sciences*, 10(6), 719-729.
- Ali, A., Rehman, T., Ahmad, N., Ali, R., & Shah, S. A. A. (2014). Active sites & potential inhibitors prediction for nonstructural 4A protein of dengue virus. *International Journal of Computational Bioinformatics & In Silico Modeling*, 3(2), 340-346.
- Almeida, H. d., Bastos, I. M. D., Ribeiro, B. M., Maigret, B., & Santana, J. M. (2013). New binding site conformations of the dengue virus NS3 Protease accessed by molecular dynamics simulation. *PLOS ONE*, 8(8), 1-15.
- Ang, L. W., Thein, T.-L., Ng, Y., Boudville, I. C., Chia, P. Y., Lee, V. J. M., & Leo, Y.-S. (2019). A 15-year review of dengue hospitalizations in Singapore: Reducing admissions without adverse consequences, 2003 to 2017. *PLOS Neglected Tropical Diseases*, 13(5). doi:<https://doi.org/10.1371/journal.pntd.0007389>

- Anthi, N. J., & Clore, G. M. (2013). Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Science*, 22, 851-858.
- Aoki, C., Hidari, K. I. P. J., Itonori, S., Yamada, A., Takahashi, N., Kasama, T., Suzuki, Y. (2006). Identification and characterization of carbohydrate molecules in mammalian cells recognized by dengue virus type 2. *Journal of Biochemistry*, 5(2), 605-618.
- Baharuddin, A., Hassan, A. A., Othman, R., Xu, Y., Huang, M., Tejo, B. A., Othman, S. (2014). Dengue envelope domain III-peptide binding analysis via tryptophan fluorescence quenching assay. *Chemical and Pharmaceutical Bulletin*, 62(10), 947-955.
- Behera, A. K., Basu, S., & Cherian, S. S. (2015). Molecular mechanism of the enhanced viral fitness contributed by secondary mutations in the hemagglutinin protein of oseltamivir resistant H1N1 influenza viruses: Modeling studies of antibody and receptor binding. *Gene*, 557(1), 19-27.
- Behnam, M. A. M., Nitsche, C., Boldescu, V., & Klein, C. D. (2016). The medicinal chemistry of dengue virus. *Journal of Medicinal Chemistry*, 59, 5622-5649.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235-242.
- Betts, M. J., & Russell, R. B. (2003). *Bioinformatics for geneticists*. New York City, United States of America: John Wiley & Sons, Ltd.
- Bhargava, K., Nath, R., Seth, P. K., Pant, K. K., & Dixit, R. K. (2014). Molecular docking studies of D2 Dopamine receptor with Risperidone derivatives. *Bioinformation*, 10(1), 8-12.
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Hay, S. I. (2013). The global distribution and burden of dengue. *Nature*, 496(7446), 504-507.
- Bhingre, A., Chakrabarti, P., Uthamallian, K., Bajaj, K., Chakraborty, K., & Varadarajan, R. (2004). Accurate detection of protein: ligand binding sites using molecular dynamics simulations. *Structure*, 12(11), 1989-1999.
- Bianco, V., Iskrov, S., & Franzese, G. (2012). Understanding the role of hydrogen bonds in water dynamics and protein stability. *Journal of Biological Physics*, 38(1), 27-48.

- Bodanszky, M. (1993). *Peptide chemistry*. New York: Springer-Verlag Berlin Heidelberg.
- Brummelhuis, N. t., Wilke, P., & Borner, H. G. (2018). *Sequence-controlled polymers*. Germany: Wiley-VCH Verlag GmbH & Co.
- Bryce, R. A., Hillier, I. H., & Naismith, J. H. (2001). Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of Oligosaccharide binding to Concanavalin A. *Biophysical Journal*, *81*, 1373-1388.
- Bustamam, A., Aldila, D., & Yuwanda, A. (2018). Understanding dengue control for short and long term intervention with a mathematical model approach. *Journal of Applied Mathematics*. doi:<https://doi.org/10.1155/2018/9674138>
- Camacho, C. J., & Vajda, S. (2002). Protein-protein association kinetics and protein docking. *Current Opinion in Structural Biology*, *12*(1), 36-40.
- Campbell, A. J., Lamb, M. L., & Joseph-McCarthy, D. (2014). Ensemble-based docking using biased molecular dynamics. *Journal of Chemical Information and Modeling*, *54*(7), 2127-2138.
- Carr, D. (2002). *The handbook of analysis and purification of peptides and proteins by reversed-phase HPLC*. United States of America: Grace Vydac.
- Chang, C.-E. A., Trylska, J., Tozzini, V., & McCammon, J. A. (2007). Binding pathways of ligands to HIV-1 Protease: coarse-grained and atomistic simulations. *Chemical Biology & Drug Design*, *69*(1), 5-13.
- Chang, C.-J., Tien, J., & Lu, M.-R. (2018). Epidemiological, clinical and climate characteristics of dengue fever in Kaohsiung City, Taiwan with implication for prevention and control. *PLOS ONE*, *13*(1), 1-15. doi:[10.1371/journal.pone.0190637](https://doi.org/10.1371/journal.pone.0190637)
- Chao, L. H., Jang, J., Johnson, A., Nguyen, A., Gray, N. S., Yang, P. L., & Harrison, S. C. (2018). How small-molecule inhibitors of dengue-virus infection interfere with viral membrane fusion. *eLIFE*, *7*. doi:[10.7554/eLife.36461](https://doi.org/10.7554/eLife.36461)
- Chaudhary, K. K., & Prasad, C. V. S. S. (2014). Virtual Screening of compounds to 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) from Plasmodium falciparum. *Bioinformatics*, *10*(6), 358-364.

- Chen, L., Zheng, Q.-C., & Zhang, H.-X. (2015). Insights into the effects of mutations on Cren7-DNA binding using molecular dynamics simulations and free energy calculations. *Physical Chemistry Chemical Physics*, 17(8), 5704-5711.
- Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R., Esko, J. D., Linhardt, R. J., & Marks, R. M. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nature Medicine*, 3, 866-871.
- Chew, M.-F., Poh, K.-S., & Poh, C.-L. (2017). Peptides as therapeutic agents for dengue virus. *International Journal of Medical Sciences*, 14(13), 1342-1359.
- Chong, W. L., Zain, S. M., Rahman, N. A., Othman, R., Othman, S., Nimmanpipug, P., Lee, V. S. (2015). Exploration of residue binding energy of potential ankyrin for dengue virus II from MD simulations. *Atlantis Press*, 100-103.
- Chu, W.-T., Zheng, Q.-C., Wu, Y.-J., Zhang, J.-L., Liang, C.-Y., Chern, L., Zhang, H.-X. (2015). Molecular dynamics (MD) simulations and binding free energy calculation studies between inhibitors and type II dehydroquinase (DHQ2). *Molecular Simulation*, 39(2), 137-144.
- Cino, E. A., Choy, W.-Y., & Karttunen, M. (2012). Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 8, 2725-2740.
- Costin, J. M., Jenwitheesuk, E., Lok, S.-M., Hunsperger, E., Conrads, K. A., Fontaine, K. A., Michael, S. F. (2010). Structural optimization and De Novo of dengue virus entry inhibitory peptides. *PLOS Neglected Tropical Diseases*, 4(6), 1-12.
- Cucunawangsih, & Lugito, N. P. H. (2017). Trends of dengue disease epidemiology. *Virology: Research and Treatment*, 8. doi:10.1177/1178122X17695836
- Dagliyan, O., Proctor, E. A., D'Auria, K. M., Ding, F., & Dokholyan, N. V. (2011). Structural and dynamic determinants of protein-peptide recognition. *Structure*, 19(12), 1837-1845.
- Dans, A. L., Dans, L. F., Lansang, M. A. D., Silvestre, M. A. A., & Guyatt, G. H. (2018). Controversy and debate on dengue vaccine series - paper 1: review of a licensed dengue vaccine: inappropriate subgroup analyses and selective reporting may cause harm in mass vaccination programs. *Journal of Clinical Epidemiology*, 95, 137-139.

- Das, A. J., Chalil, S., Nigam, P., Magee, P., Janneh, O., & Owusu-Apenten, R. (2011). Glutathione transferase-P1-1 binding with naturally occurring ligands: assessment by docking simulations. *Journal of Biophysical Chemistry*, 2(4), 401-407.
- Daura, X., Jaun, B., Seebach, D., Gunsteren, W. F. v., & Mark, A. E. (1998). Reversible peptide folding in solution by molecular dynamics simulation. *Journal of Molecular Biology*, 280, 925-932.
- Daura, X., Mark, A. E., & Gunsteren, W. F. v. (1999). Peptide folding simulations: no solvent required? *Computer Physics Communications*, 123, 97-102.
- Degreve, L., Fuzo, C. A., & Caliri, A. (2012). Extensive structural change of the envelope protein of dengue virus induced by a tuned ionic strength: conformational and energetic analyses. *Journal of Computer-Aided Molecular Design*, 26, 1311-1325.
- Demir, K., Kilic, N., Dudak, F. C., Boyaci, I. H., & Yasar, F. (2014). The investigation of the secondary structure propensities and free-energy landscape of peptide ligands by replica exchange molecular dynamics simulations. *Molecular Simulations*, 40(10), 1015-1025.
- Dennis, E. J., Goldman, O. V., & Vosshall, L. B. (2019). *Aedes aegypti* mosquitoes use their legs to sense DEET on contact. *Current Biology*, 29(6), 1551-1556.
- Deol, S. S., Bond, P. J., Domene, C., & Sansom, M. S. P. (2004). Lipid-protein interactions of integral membrane proteins: A comparative simulation study. *Biophysical Journal*, 87(6), 3737-3749.
- Donnini, S. (2007). *Computing free energies of protein ligand association*. (PhD), University of Oulu, Finland.
- Duan, L. L., Feng, G. Q., & Zhang, Q. G. (2016). Large-scale molecular dynamics simulation: Effect of polarization on thrombin-ligand energy. *Scientific Reports*, 6. doi:10.1038/srep31488
- Duan, L. L., Mei, Y., Zhang, Q. G., Tang, B., & Zhang, J. Z. H. (2014). Protein's native structure is dynamically stabilized by electronic polarization. *Journal of Theoretical & Computational Chemistry*, 13(3). doi:https://doi.org/10.1142/S0219633614400057

- Dubayle, J., Vialle, S., Schneider, D., Pontvianne, J., Mantel, N., Adam, O., Talaga, P. (2015). Site-specific characterization of envelope protein N-glycosylation on Sanofi Pasteur's tetravalent CYD dengue vaccine. *Vaccine*, 33, 1360-1368.
- Dubey, K. D., Tiwari, G., & Ojha, R. P. (2014). Molecular docking, MD simulations, and free energy calculations of DENV inhibitors for a novel binding site. *Journal of Biomedicine & Biotechnology*, 1-14.
- Dubey, K. D., Tiwari, G., & Ojha, R. P. (2017). Targeting domain-III hinging of dengue (DENV-2) protein by MD simulations, docking and free energy calculations. *Journal of Molecular Modeling*, 23(102). doi:https://doi.org/10.1007/s00894-00017-03259-00892
- Durrant, J. D., & McCammon, J. A. (2011). HBonanza: A computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. *Journal of Molecular Graphics and Modelling*, 31, 5-9.
- Ebadi, A., Razzaghi-Asl, N., Khoshneviszadeh, M., & Miri, R. (2013). Comparative amino acid decomposition analysis of potent type 1 p38a inhibitors. *Journal of Pharmaceutical Sciences*, 21(1). doi:10.1186/2008-2231-21-41
- Ebi, K. L., & Nealon, J. (2016). Dengue in a changing climate. *Environmental Research*, 151, 115-123.
- Ehrlich, L. P., Nilges, M., & Wade, R. C. (2005). The impact of protein flexibility on protein-protein docking. *Proteins: Structure, Function and Bioinformatics*, 58(1), 126-133.
- Elahi, M., Islam, M. M., Noguchi, K., Yohda, M., Toh, H., & Kuroda, Y. (2014). Computational prediction and experimental characterization of a "size switch type repacking" during the evolution of dengue envelope protein domain III *Biochimica et Biophysica Acta*, 1844, 585-592.
- Elengoe, A., Naser, M. A., & Hamdan, S. (2014). Modeling and docking studies on novel mutants (K71L and T204V) of the ATPase domain of human heat shock 70kDa protein 1. *International Journal of Molecular Sciences*, 15, 6797-6814.
- Fauci, A., Erbedding, E., Whitehead, S., Cassetti, M. C., Handley, F. G., & Gupta, R. (2019). Dengue vaccine clinical trials in India - An opportunity to inform the global response to a re-emerging disease challenge. *International Journal of Infectious Diseases*, 84, S4-S6

- Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., & Brooks, C. L. (2003). Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry*, 25(2), 265-284.
- Fibriansah, G., Tan, J. L., Smith, S. A., Alwis, R. d., Ng, T.-S., Kostyuchenko, V. A., Lok, S.-M. (2015). A highly potent human antibody neutralizes dengue virus serotypes 3 by binding across three surface proteins. *Nature Communications*. doi:10.1038/ncomms7341
- Flipse, J., Diosa-Toro, M. A., Hoornweg, T. E., Pol, D. P. I. v. d., Urcuqui-Inchima, S., & Smit, J. M. (2016). Antibody-dependent enhancement of dengue virus infection in primary human macrophages; balancing higher fusion against antiviral responses. *Scientific Reports*, 6. doi:10.1038/srep29201
- Francis, D. (2002). *Optimization of electrostatic binding free energy applications to the analysis and design of ligand binding in protein complexes*. (PhD), Massachusetts Institute of Technology.
- Frank, M. (2015). Conformational analysis of Oligosaccharides and Polysaccharides using molecular dynamics simulations *Glycoinformatics*, 1273, 359-377.
- G.H., T., N., Y., A.N., J., Mutalip, A., F., P., M.H., H., T., A. (2019). Prolonged dengue outbreak at a high-rise apartment in Petaling Jaya, Selangor, Malaysia: A case study. *Tropical Biomedicine*, 36(2), 550-558.
- Gallichotte, E. N., Baric, T. J., Yount, B. L., Widman, D. G., Durbin, A., Whitehead, S., Silva, A. M. d. (2018). Human dengue virus serotype 2 neutralizing antibodies target two distinct quaternary epitopes. *PloS Pathogens*, 14(2). doi:https://doi.org/10.1371/journal.ppat.1006934
- Ganoth, A., Friedman, R., Nachliel, E., & Gutman, M. (2006). A molecular dynamics study and free energy analysis of complexes between the Mlc1p protein and two IQ motif peptides. *Biophysical Journal*, 91(7), 2436-2450.
- Garcia-Martin, F., Bayo-Puxan, N., Cruz, L. J., Bohling, J. C., & Albericio, F. (2007). Chlorotriyl Chloride (CTC) resin as a reusable carboxyl protecting group. *Molecular Informatics*, 26(10), 1027-1035.
- Gautam, B., Singh, G., Wadhwa, G., Farmer, R., Singh, S., Singh, A. K., Yadav, P. K. (2012). Metabolic pathway analysis and molecular docking analysis for identification of putative drug targets in *Toxoplasma gondii*: novel approach. *Biomedical Informatics*, 8(3), 134-141.

- Ge, Y., Wu, J., Xia, Y., Yang, M., Xiao, J., & Yu, J. (2012). Molecular dynamics simulation of the complex PBP-2x with drug Cefuroxime to explore the drug resistance mechanism of *Streptococcus suis* R61. *PLOS ONE*, 7(4). doi:10.1371/journal.pone.0035941
- Ghosh, A., & Dar, L. (2015). Dengue vaccines: Challenges, development, current status and prospects. *Indian Journal of Medical Microbiology*, 33(1), 3-15.
- Gnanakaran, S., Nymeyer, H., Portman, J., Sanbonmatsu, K. Y., & Garcia, A. E. (2003). Peptide folding simulations. *Current Opinion in Structural Biology*, 13, 168-174.
- Godoi, I. P., Lima, W. G., Junior, M. C., Alves, R. J., Ferreira, J. M. S., Kong, D.-X., & Taranto, A. G. (2017). Docking and QM/MM studies of NS2B-NS3pro inhibitors: A molecular target against the dengue virus. *Journal of the Brazilian Chemical Society*, 28(5), 895-906.
- Gromowski, G. D. (2008). *Structure-function of dengue 2 virus envelope protein domain III with neutralizing antibodies*. (PhD), The University of Texas Medical Branch.
- Gromowski, G. D., Barrett, N. D., & Barrett, A. D. T. (2008). Characterization of dengue virus complex specific neutralizing epitopes on envelope protein domain III of dengue 2 virus. *Journal of Virology*, 82(17), 8828-8837.
- Grosdidier, S., & Fernandez-Recio, J. (2008). Identification of hot-spot residues in protein-protein interactions by computational docking. *BMC Bioinformatics*, 9(447). doi:10.1186/1471-2105-9-447
- Grunberg, R., Nilges, M., & Leckner, J. (2006). Flexibility and conformational entropy in protein-protein binding. *Structure*, 14, 683-693.
- Guardia, C. D. L., & Leonart, R. (2014). Progress in the Identification of Dengue Virus Entry/Fusion Inhibitors. *BioMed Research International*, 2014. doi:http://dx.doi.org/10.1155/2014/825039
- Guardia, C. d. l., Quijada, M., & Leonart, R. (2017). Phage-displayed peptides selected to bind envelope glycoprotein show antiviral activity against dengue virus serotype 2. *Advances in Virology*. doi:https://doi.org/10.1155/2017/1827341

- Guarner, J., & Hale, G. L. (2019). Four human disease with significant public health impact caused by mosquito-borne flaviviruses: West Nile, Zika, dengue and yellow fever. *Seminars in Diagnostic Pathology*, 36(3), 170-176.
- Gunasekaran, K., Ramakrishnan, C., & Balaram, P. (1996). Disallowed ramachandrann conformations of amino acid residues in protein structures. *Journal of Molecular Biology*, 264, 191-198.
- Gutheil, W. G., & Xu, Q. (2002). N-to-C solid-phase peptide and peptide Trifluoromethylketone synthesis using amino acid tert-Butyl Esters. *Chemical and Pharmaceutical Bulletin*, 50(5), 688-691.
- Guzman, M. G., Hermida, L., Bernardo, L., Ramirez, R., & Guillen, G. (2010). Domain III of the envelope protein as a dengue vaccine target. *Expert Review of Vaccines*, 9(2), 137-147.
- Hacker, K. E. (2009). *Characterization of dengue virus interactions with host cells*. (Doctor of Philosophy), University of North Carolina.
- Halstead, S. B. (2018). Safety issues from a phase 3 clinical trial of a live-attenuated chimeric yellow fever tetravalent dengue vaccine. *Journal Human Vaccines & Immunotherapeutics*, 14(9), 2158-2162.
- Hansmann, U. H. E., Masuya, M., & Okamoto, Y. (1997). Characteristic temperatures of folding of a small peptide. *Biophysics*, 94, 10652-10656.
- Harrison, S. C. (2008). Viral membrane fusion. *Nature Structural & Molecular Biology*, 15(7), 690-698.
- Heavner, S. E. (2004). *Molecular modeling and experimental determination of the structure of C8-Arylguanine modified oligonucleotides that preferentially adopt the Z-DNA conformation*. (Doctor of Philisophy), West Virginia University, Morgantown, West Virginia.
- Held, M., Metzner, P., Prinz, J.-H., & Noe, F. (2011). Mechanisms of protein-ligand association and its modulation by protein mutations. *Biophysical Journal*, 100(3), 701-710.
- Held, M., & Noe, F. (2012). Calculating kinetics and pathways of protein-ligand association. *European Journal of Cell Biology*, 91(4), 357-364.

- Hidari, K. I. P. J., Abe, T., & Suzuki, T. (2013). Carbohydrate-related inhibitors of dengue virus entry. *Viruses*, 5, 605-618.
- Hidari, K. I. P. J., & Suzuki, T. (2011). Dengue virus receptor. *Tropical Medicine and Health*, 39(4), 37-43.
- Ho, B. K., & Dill, K. A. (2006). Folding very short peptides using molecular dynamics. *PLOS Computational Biology*, 2(5), 228-237.
- Holgersson, J., Gustafsson, A., & Breimer, M. E. (2005). Special feature characteristic of protein-carbohydrate interactions as a basis for developing novel carbohydrate based antirejection therapies *Immunology and Cell Biology*, 83, 694-708.
- Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics poisson-boltzmann surface area method. *Molecular Informatics*, 31(2), 114-122.
- Hoofnagle, A. N., Whiteaker, J. R., Carr, S. A., Kuhn, E., Liu, T., Massoni, S. A., & Thomas, S. N. (2016). Recommendations for the generation, quantification, storage and handling of peptides used for mass spectrometry-based assays. *Clinical Chemistry*, 62(1), 48-69.
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MMPBSA and MMGBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 51(1), 69-82.
- Hsu, K.-C., Cheng, W.-C., Chen, Y.-F., Wang, W.-C., & Yang, J.-M. (2013). Pathway-based screening strategy for multitarget inhibitors of diverse proteins in metabolic pathways. *PLOS Computational Biology*, 9(7). doi:<https://doi.org/10.1371/journal.pcbi.1003127>
- Huang, B. (2009). MetaPocket: A meta approach to improve protein ligand binding site prediction. *A Journal of Integrative Biology*, 13(4), 325-330.
- Huang, H.-J., Chen, H.-Y., Chang, Y.-S., & Chen, C. Y.-C. (2015). Insight into two antioxidants binding to the catalase NADPH binding site from traditional Chinese medicines. *RSC Advances*, 5, 6625-6635.

- Huang, Z., & Wong, C. F. (2007). A mining minima approach to exploring the docking pathways of p-Nitrocatechol Sulfate to YopH. *Biophysical Journal*, 93, 4141-4150.
- Huang, Z., & Wong, C. F. (2012). Simulation reveals two major docking pathways between the hexapeptide GDYMNM and the catalytic domain of the insulin receptor protein kinase. *Proteins: Structure, Function and Bioinformatics*, 80(9), 2275-2286.
- HuiSong, P. (2010). *Developing a sensitive and selective method for peptide detection by mass spectrometry*. (Master Dissertation), University of Geneva.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14, 33-38.
- Hung, J.-J., Hsieh, M.-T., Young, M.-J., Kao, C.-L., King, C.-C., & Chang, W. (2004). An external loop region of domain III dengue virus type 2 envelope protein is involved in serotype-specific binding to mosquito but not mammalian cells. *Journal of Virology*, 78(1), 378-388.
- Ilyas, M., Rahman, Z., Shamas, S., Alam, M., Israr, M., & Masood, K. (2011). Bioinformatics analysis of envelope glycoprotein E-epitopes of dengue virus type 3 *African Journal of Biotechnology*, 10(18), 3528-3533.
- Isa, D. M., Peng, C. S., Chong, W. L., Zain, S. M., Rahman, N. A., & Lee, V. S. (2019). Dynamics & binding interactions of peptide inhibitors of dengue virus entry. *Journal of Biological Physics*, 45(1), 63-76.
- Isidro-Llobet, A., Kenworthy, M. N., Mukherjee, S., Kopach, M. E., Wegner, K., Gallou, F., Roschangar, F. (2019). Sustainability challenges in peptide synthesis & purification: from R&D to production. *The Journal of Organic Chemistry*, 84(8), 4615-4628.
- Isidro-Llobet, A., Alvarez, M., & Albericio, F. (2009). Amino acid-protecting groups. *Chemical Reviews*, 109(6), 2455-2504.
- Isralewitz, B. (2011). Timeline: A VMD plugin for trajectory analysis. In U. o. Illinois (Ed.), *Computational Biophysics Workshop*.
- Jakubke, H.-D., & Sewald, N. (2008). *Peptides from A to Z: A Concise Encyclopedia*. Germany: WILEY-VCH Verlag GmbH & Co.

- Jambrina, P. G., & Aldegunde, J. (2016). Chapter 20 - Computational tools for the study of biomolecules. *Computer Aided Chemical Engineering*, 39, 583-648.
- Jana, C., Cahturvedi, I., & Robine, O. (2014). Finding natural inhibitor for NS3 Protease enzyme from dengue virus via molecular modeling. *International Journal of Basic and Applied Biology*, 2(3), 124-129.
- John, A. L. S., & Rathore, A. P. S. (2019). Adaptive immune responses to primary & secondary dengue virus infections. *Nature Reviews Immunology*, 19, 218-230.
- Kang, J.-S. (2012). *Principles & applications of LC-MS/MS for the quantitative bioanalysis of analytes in various biological samples, tandem mass spectrometry - applications & principles*. Shanghai, China: InTech China.
- Katagi, G. M. (2013). *Analysis of molecular dynamics trajectories of proteins performed using different forcefields and identification of mobile segments*. (Master of Science (Engineering) Dissertation), Indian Institute of Science, Bangalore, India.
- Khor, B. Y., Tye, G. J., Lim, T. S., Noordin, R., & Choong, Y. S. (2014). The structure and dynamics of BmR1 protein from *Brugia malayi*: *In silico* approaches. *International Journal of Molecular Sciences*, 15, 11082-11099.
- Koppisetty, C. A. K., Frank, M., Lyubartsev, A. P., & Nyholm, P.-G. (2015). Binding energy calculations for hevein-carbohydrate interactions using expanded ensemble molecular dynamics simulations. *Journal of Computer-Aided Molecular Design*, 29(1), 13-21.
- Koska, J., Spassov, V. Z., Maynard, A. J., Yan, L., Austin, N., Flook, P. K., & Venkatachalam, C. M. (2008). Fully automated molecular mechanics based induced fit protein-ligand docking method. *Journal of Chemical Information & Modeling*, 48(10), 1965-1973.
- Kukic, P., & Nielsen, J. E. (2010). Electrostatics in proteins and protein-ligand complexes. *Future Medicinal Chemistry*, 2(4), 647-666.
- Kukkurainen, S. (2009). *A molecular dynamics study of ligand binding in mutant avidins*. (Master Dissertation), University of Tampere.
- Kumari, R., Kumar, R., Consortium, O. S. D. D., & Lynn, A. (2014a). g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, 54(7), 1951-1962.

- Kumari, R., Kumar, R., Consortium, O. S. D. D., & Lynn, A. (2014b). g_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*, 1951-1962.
- Kurcinski, M., Jamroz, M., & Kolinski, A. (2011). *Multiscale approaches to protein modeling*. Warsaw, Poland: Springer.
- LaBute, M. X., Zhang, X., Lenderman, J., Bennion, B. J., Wong, S. E., & Lightstone, F. C. (2014). Adverse drug reaction prediction using scores produced by large-scale drug-protein target docking on high-performance computing machines. *PLOS ONE*, *9*(9). doi:<https://doi.org/10.1371/journal.pone.0106298>
- Landon, M. R., Amaro, R. E., Baron, R., Ngan, C. H., Ozonoff, D., McCammon, J. A., & Vajda, S. (2008). Novel druggable hot spots in Avian Influenza Neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chemical Biology & Drug Design*, *71*, 106-116.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283-291.
- Lauer, S. A., Sakrejda, K., Ray, E. L., Keegan, L. T., Bi, Q., Suangtho, P., Reich, N. G. (2018). Prospective forecasts of annual dengue hemorrhagic fever incidence in Thailand, 2010-2014. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), 175-182.
- Lavanya, P., Ramaiah, S., & Anbarasu, A. (2015). Computational analysis reveal inhibitory action of nimbin against dengue viral envelope protein. *Virus Disease*, *26*(4), 243-254.
- Le, L. (2012). Incorporating molecular dynamics simulations into rational drug design: A case study on influenza a Neuraminidases. In H. Perez-Sanchez (Ed.), *Bionformatics* (pp. 160): In Tech.
- Lebl, M., & Houghten, R. A. (2001). *Peptides: The wave of the future*. United State of America: American Peptide Society.
- Lee, J.-S., Mogasale, V., Lim, J. K., Carabali, M., Lee, K.-S., Sirivichayakul, C., Farlow, A. (2017). A multi-country study of the economic burden of dengue fever: Vietnam, Thailand and Colombia. *PLOS Neglected Tropical Diseases*, *11*(10). doi: <https://doi.org/10.1371/journal.pntd.0006037>

- Lee, T.-K., Ryoo, S.-J., & Lee, Y.-S. (2007). A new method for the preparation of 2-chlorotriptyl resin and its application to solid-phase peptide synthesis. *Tetrahedron Letters*, *48*, 389-391.
- Leis, S., Schneider, S., & Zacharias, M. (2010). *In silico* prediction of binding sites on proteins. *Current Medicinal Chemistry*, *17*(15), 1550-1562.
- Leronymaki, M., Androutsou, M. E., Pantelia, A., Friligou, I., Crisp, M., High, K., Thelios, T. (2015). Use of the 2-chlorotriptyl chloride resin for microwave-assisted solid phase peptide synthesis. *Biopolymers*, *104*(5), 506-514.
- Li, D.-Z., Yu, G.-Q., Yi, S.-C., Zhang, Y., Kong, D.-X., & Wang, M.-Q. (2015). Structure-based analysis of the ligand-binding mechanism for DhelOBP21, a C-minus odorant binding protein, from *Dastarcus helophoroides*. *International Journal of Biological Sciences*, *11*(11), 1281-1295.
- Li, H., Aneja, R., & Chaiken, I. (2013). Click chemistry in peptide-based drug design. *Molecules*, *18*, 9797-9817.
- Lin, J.-H., Perryman, A. L., Schames, J. R., & McCammon, J. A. (2002). Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *Journal of the American Chemical Society*, *124*(20), 5632-5633.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular cell biology*, 4th edition. New York: W.H. Freeman and Company.
- Lovell, S. C., Davis, I. W., III, W. B. A., Bakker, P. I. W. d., Word, J. M., Prisant, M. G., & Richardson, J. S. (2003). Structure validation by C- C- α Geometry: ϕ , ψ and C β Deviation. *Proteins: Structure, Function and Bioinformatics*, *50*(3), 437-450.
- Luthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*, 83-85.
- Ma, S., Zeng, G., Fang, D., Wang, J., Wu, W., Xie, W., Zheng, K. (2015). Studies of N9-arethenyl purines as novel DFG-in and DFG-out dual Src/Abl inhibitors using 3D-QSAR, docking and molecular dynamics simulations. *Molecular BioSystem*, *11*, 394-406.

- Made, V., Els-Heindl, S., & Beck-Sickinger, A. G. (2014). Automated solid-phase synthesis to obtain therapeutic peptides. *Beilstein Journal of Organic Chemistry*, *10*, 1197-1212.
- Maffucci, I., Pellegrino, S., Clayden, J., & Contini, A. (2014). Mechanism of stabilization of helix secondary structure by constrained C alpha-tetrasubstituted alpha amino acids. *The Journal of Physical Chemistry B*, *119*, 1350-1361.
- Marks, R. M., Lu, H., Sundaresan, R., Toida, T., Suzuki, A., Imanari, T., Linhardt, R. J. (2001). Probing the interaction of dengue virus envelope protein with heparin: assessment of glycosaminoglycan-derived inhibitors. *Journal of Medicinal Chemistry*, *44*(13), 2178-2187.
- Martinez, L., Sonoda, M. T., Webb, P., Baxter, J. D., Skaf, M. S., & Polikarpov, I. (2005). Molecular dynamics simulations reveal multiple pathways of ligand dissociation from thyroid hormone receptors. *Biophysical Journal*, *89*(3), 2011-2023.
- Martins-Jose, A. (2013). Sliding box docking: a new stand-alone tool for managing docking-based virtual screening along the DNA helix axis. *Biomedical Informatics*, *9*(14), 750-751.
- Masova, I., & Kollman, P. A. (2000). Combined molecular mechanical and continuum solvent approach (MMPBSA/GBSA) to predict ligand binding. *Perspective in Drug Discovery and Design*, *18*, 113-135.
- Matthes, D., & Groot, B. L. d. (2009). Secondary structure propensities in peptide folding simulations: A systematic comparison of molecular mechanics interaction schemes. *Biophysical Journal*, *97*, 599-608.
- Mayuri, Bauve, E. L., & Kuhn, R. J. (2010). *Frontiers in dengue virus research*. Norfolk, United Kingdom: Caister Academic Press.
- McAtamney, S. (2008). *Investigation of dengue fever virus envelope glycoprotein carbohydrate-ligand recognition events essential for mammalian cell infection*. (Doctor of Philosophy), Griffith University.
- McGee, D., Miller, B., & Swails, J. (2009). Amber advanced tutorial 3 - Section 3.1.
- Medvedev, K. E., Alemasov, N. A., Vorobjev, Y. N., Boldyreva, E. V., Kolchanov, N. A., & Afonnikov, D. A. (2014). Molecular dynamics simulations of the Nip7 proteins from the marine deep- and shallow-water *Pyrococcus* species. *BMC Structural Biology*, *14*(23), 1-21.

- Meneses, E., & Mittermaier, A. (2013). Electrostatic interactions in the binding pathway of a transient protein complex studied by NMR and isothermal titration calorimetry. *The Journal of Biological Chemistry*, 289, 27911-27923.
- Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. R. W., Ray, S. E., Hay, S. I. (2019). The current and future global distribution and population at risk of dengue. *Nature Microbiology*. doi:<https://doi.org/10.1038/s41564-019-0476-8>
- Moonrin, N., Songtawee, N., Rattanabunyong, S., Chunsriviro, S., Mkmak, W., Tongsima, S., & Choowongkomon, K. (2015). *BMC Bioinformatics*, 16(103). doi:10.1186/s12859-015-0528-x
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2010). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785-2791.
- Mouawad, L., Tetreau, C., Abdel-Azeim, S., Perahia, D., & Lavalette, D. (2007). CO Migration pathways in cyto studied by molecular dynamics simulations. *Protein Science*, 16(5), 781-794.
- Mozafari, E. S., Tazikeh-Lemeski, E., & Saboury, A. A. (2016). Isothermal titration calorimetry & molecular dynamics simulation studies on the binding of indometacin with human serum albumin. *Biomolecular Journal*, 2(1), 34-43.
- Mueller, D. S. (2010). *The effects of urea and of pH on protein structure*. (PhD), University of Groningen, Netherlands.
- Mukhametov, A., Newhouse, E. I., Aziz, N. A., Saito, J. A., & Alam, M. (2014). Allosteric pocket of the dengue virus (serotype 2) NS2B/NS3 protease: *In silico* ligand screening and molecular dynamics studies of inhibition. *Journal of Molecular Graphics and Modelling*, 52, 103-113.
- Munoz, M. d. L., Limon-Camacho, G., Tovar, R., Diaz-Badillo, A., Mendoza-Hernandez, G., & Black, W. C. (2013). Proteomic identification of dengue virus binding proteins in *Aedes aegypti* mosquitoes and *Aedes albopictus* cells. *BioMed Research International*, 2013. doi:<http://dx.doi.org/10.1155/2013/875958>

- Naseri, A., Hosseini, S., Rasoulzadeh, F., Rashidi, M.-R., Zakery, M., & Khayamian, T. (2015). Interaction of norfloxacin with bovine serum albumin studied by different spectrometric methods; displacement studies, molecular modeling and chemometrics approaches. *Journal of Luminescence*, *157*, 104-112.
- Nasution, M. A. F., Aini, R. N., & Tambunan, U. S. F. (2017). Virtual screening of commercial cyclic peptides as NS2B-NS3 protease inhibitor of dengue virus serotype 2 through molecular docking simulation. *IOP Conference Series: Materials Science and Engineering*, *188*. doi:10.1088/1757-899X/188/1/012017
- Negami, T., Shimizu, K., & Terada, T. (2014). Coarse-grained molecular dynamics simulations of protein-ligand binding. *Journal of Computational Chemistry*, *35*, 1835-1845.
- Nguyen, H., Maier, J., Huang, H., Perrone, V., & Simmerling, C. (2014). Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *Journal of the American Chemical Society*, *136*, 13959-13962.
- Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B., & Woods, R. J. (2014). Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *Journal of Computational Chemistry*, *35*(7), 526-539.
- Obiol Pardo, C. (2008). *Disrupting the protein-protein recognition in cancer pathways by molecular modelling*. (PhD), Universitat de Barcelona.
- Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Schell, D., Grimsley, G. R. (2014). Contribution of hydrogen bonds to protein stability. *Protein Science*, *23*(5), 652-661.
- Pacholczyk, M., & Kimmel, M. (2011). Exploring the landscape of protein-ligand interaction energy using probabilistic approach. *Journal of Computational Biology*, *18*(6), 843-850.
- Padariya, M., Kalathiya, U., & Baginski, M. (2015). Structural and dynamic changes adopted by EmrE, multidrug transporter protein-Studies by molecular dynamics simulation. *Biochimica et Biophysica Acta*, *1848*(10), 2065-2074.
- Pandey, B., Grover, S., Tyagi, C., Goyal, S., Jamal, S., Singh, A., Grover, A. (2016). Molecular principles behind pyrazinamide resistance due to mutations in panD gene in Mycobacterium tuberculosis. *Gene*, *581*(1), 31-42.

- Pandey, B., Sharma, P., Tyagi, C., Goyal, S., Grover, A., & Sharma, I. (2015). Structural modeling and molecular simulation analysis of HvAP2/EREBP from barley. *Journal of Biomolecular Structure and Dynamic*, 34(6), 1159-1175.
- Pang, A., Arinaminpathy, Y., Sansom, M. S. P., & Biggin, P. C. (2003). Interdomain dynamics and ligand binding: molecular dynamics simulations of glutamine binding protein. *FEBS Letters*, 550(2003), 168-174.
- Pang, E. L., & Loh, H.-S. (2016). Current perspectives on dengue episode in Malaysia. *Asian Pacific Journal of Tropical Medicine*, 9(4), 395-401.
- Panigrahi, R., Adina-Zada, A., & Vrieland, J. W. A. (2013). Ligand recognition by the TPR domain of the import factor Toc64 from *Arabidopsis Thaliana*. *PLOS ONE*, 8(12). doi: <https://doi.org/10.1371/journal.pone.0083461>
- Parikesit, A. A., Kinanty, & Tambunan, U. S. F. (2013). Screening of commercial cyclic peptides as inhibitor envelope protein dengue virus (DENV) through molecular docking & molecular dynamic. *Pakistan Journal of Biological Sciences*, 16(24), 1836-1848.
- Patil, R., Das, S., Stanley, A., Yadav, L., Sudhakar, A., & Varma, A. K. (2010). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLOS ONE*, 5(8). doi:10.1371/journal.pone.0012029
- Peng, C. S. (2015). *Molecular modelling of Muscarinic Acetylcholine receptors: structural basis of ligand-receptor interactions*. (Doctor of Philosophy), University of Malaya, Kuala Lumpur.
- Philips, J. L., Colvin, M. E., & Newsam, S. (2011). Validating clustering of molecular dynamics simulations using polymer models. *BMC Bioinformatics*, 12(445). doi:10.1186/1471-2105-12-445
- Piana, S., Klepeis, J. L., & Shaw, D. E. (2014). Assessing the accuracy of physical models used in protein-folding simulations: quantitative from long molecular dynamics simulations. *Current Opinion in Structural Biology*, 24, 98-105.
- Poggianella, M., Campos, J. L. S., Chan, K. R., Tan, H. C., Bestagno, M., Ooi, E. E., & Burrone, O. R. (2015). Dengue E-protein domain III-based DNA immunisation induces strong antibody responses to all four serotypes. *PLOS Neglected Tropical Diseases*, 9(7). doi: <https://doi.org/10.1371/journal.pntd.0003947>

- Richardson, J. S., & Richardson, D. C. (1990). *Principles and pattern of protein conformation in Prediction of Protein Structure and the Principles of Protein Conformation*. New York: Plenum Press.
- Rizvi, S. M. D., Shakil, S., & Haneef, M. (2013). A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians *EXCLI Journal*, 12, 831-857.
- Ruba, S., Arooj, M., & Naz, G. (2014). *In silico* molecular docking studies and design of dengue virus inhibitors. *Journal of Pharmacy and Biological Sciences*, 9(2), 15-23.
- Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977a). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3), 327-341.
- Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977b). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3), 327-341.
- Salomon-Ferrer, R., Case, D. A., & Walker, R. C. (2012). An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science*, 3(2), 198-210.
- Salomon-Ferret, R., Gotz, A. W., Poole, D., Grand, S. L., & Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Mesh Ewald. *Journal of Chemical Theory and Computation*, 9(9), 3878-3888.
- Samson, D., Rentsch, D., Minuth, M., Meier, T., & Loidl, G. (2019). *Journal of Peptide Science*, 25(7), 1-11. doi:10.1002/psc.3193
- Schaeffer, L. (2008). *The practice of medicinal chemistry (fourth edition)*. San Diego, United State of America: Academic Press.
- Schaller, A., Connors, N. K., Oelmeier, S. A., Hubbuch, J., & Middelberg, A. P. J. (2015). Predicting recombinant protein expression experiments using molecular dynamics simulation. *Chemical Engineering Science* 121, 340-350.
- Schlick, T. (2012). *Innovations in biomolecular modeling & simulations: Volume 2*. London, United Kingdom: Royal Society of Chemistry

- Schmidt, A. G., Lee, K., Yang, P. L., & Harrison, S. C. (2012). Small-molecule inhibitors of dengue-virus entry. *PloS Pathog*, 8(4). doi:https://doi.org/10.1371/journal.ppat.1002627
- Schmitz, J., Roehrig, J., Barrett, A., & Hombach, J. (2011). Next generation dengue vaccines: A review of candidates in preclinical development. *Vaccine*, 29, 7279-7284.
- Shao, J., Tanner, S. W., Thompson, N., & Cheatham, T. E. (2007). Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. *Journal of Chemical Theory and Computation*, 3(6), 2312-2334.
- Shao, Q., Yang, L., & Gao, Y. Q. (2009). A test of implicit solvent models on the folding simulation of the GB1 peptide. *Journal of Chemical Physics* 130(19). doi:https://doi.org/10.1063/1.3132850
- Simmerling, C., Strockbine, B., & Roitberg, A. E. (2002). All-atom structure prediction and folding simulations of a stable protein. *Journal of the American Chemical Society*, 124(38), 11258-11259.
- Simmerling, C., Strockbine, B., & Roitberg, A. E. (2002). All-atom structure prediction and folding simulations of a stable protein. *Journal of the American Chemical Society*, 124(38), 11258-11259.
- Sindhu, T., & Srinivasan, P. (2015). Exploring the binding properties of agonists interacting with human TGR5 using structural modeling, molecular docking and dynamics simulations. *RSC Advances*, 5, 14202-14213.
- Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological Reviews*, 66, 334-395.
- Soe, H. J., Yong, Y. K., Al-Obaidi, M. M. J., Raju, C. S., Gudimella, R., Manikam, R., & Sekaran, S. D. (2018). Identifying protein biomarkers in predicting disease severity of dengue virus infection using immune-related protein microarray. *Medicine*, 97(5). doi:10.1097/MD.00000000000009713
- Sonoda, M. T., Martinez, L., Webb, P., Skaf, M. S., & Polikarpov, I. (2008). Ligand dissociation from estrogen receptor is mediated by receptor dimerization: Evidence from molecular dynamics simulations. *Molecular Endocrinology*, 22(7), 1565-1578.

- Spoel, D. v. d., Vogel, H. J., & Berendsen, H. J. C. (1996). Molecular dynamics simulations of N-terminal peptides from a nucleotide binding protein. *PROTEINS: Structure, Function and Genetics*, 24, 450-466.
- Stojanovic, S. D., & Zaric, S. D. (2009). Hydrogen bonds and hydrophobic interactions of Porphyrins in Porphyrin-containing proteins. *The Open Structural Biology Journal*, 3, 34-41.
- Tambunan, U. S. F., & Alkaff, A. H. (2018). *Identification of natural products as an inhibitor of B-OG pocket binder of dengue virus envelope protein using fragment based drug design & molecular docking approach*. Paper presented at the Proceedings of the 3rd International Symposium on Current Progress in Mathematics & Sciences
- Tambunan, U. S. F., Chua, W., Parikesit, A. A., & Kerami, D. (2016). Designing disulfide cyclic peptide as fusion inhibitor that targets dengue virus envelope protein. *Journal of Technology*, 78(4), 95-103.
- Tambunan, U. S. F., Noors, R. S., Parikesit, A. A., Elyana, & Ronggo, W. (2011). Molecular dynamics simulation of DENV RNA-dependent RNA-Polymerase with potential inhibitor of disulfide cyclic peptide. *Online Journal of Biological Sciences*, 11(2), 48-62.
- Tambunan, U. S. F., Parikesit, A. A., Hendra, Taufik, R. I., Amelia, F., & Syamsudin. (2009). *In silico* analysis of envelope dengue virus 2 and envelope dengue virus 3 protein as the backbone of dengue virus tetravalent vaccine by using homology modeling method. *Journal of Biological Sciences*, 9(1), 6-16.
- Tautermann, C. S., Seeliger, D., & Kriegl, J. M. (2014). What can we learn from molecular dynamics simulations for GPCR drug design? *Computational and Structural Biotechnology Journal*, 13(2015), 111-121.
- Tian, Y.-S., Zhou, Y., Takagi, T., Kameoka, M., & Kawashita, N. (2018). Dengue virus & its inhibitor: a brief review. *Chemical & Pharmaceutical Bulletin*, 66(3), 191-206.
- Tomlinson, S. M., Malmstrom, R. D., & Watowich, S. J. (2009). New approaches to structure-based discovery of dengue protease inhibitors. *Infectious Disorders - Drug Targets*, 9(3), 327-343.
- Torchala, M., Moal, I. H., Chaleil, R. A. G., Fernandez-Recio, J., & Bates, P. A. (2013). SwarmDock: A server for flexible protein-protein docking. *Bioinformatics*, 29(6), 807-809.

- Tran, D.-T. T., Le, L. T., & Truong, T. N. (2013). Discover binding pathways using the sliding binding-box docking approach: application to binding pathways of oseltamivir to avian influenza H5N1 neuraminidase. *Journal of Computer-Aided Molecular Design*, 27(8), 689-695.
- Trauger, S. A., Webb, W., & Siuzdak, G. (2002). Peptide and protein analysis with mass spectrometry. *Spectroscopy*, 16, 15-28.
- Treesuwan, W., Wittayanarakul, K., Anthony, N. G., Huchet, G., Alniss, H., Hannongbua, S., Mackay, S. P. (2009). A detailed binding free energy study of 2 : 1 ligand-DNA complex formation by experiment and simulation. *Physical Chemistry Chemical Physics*, 11, 10682-10693.
- Tripathi, A., & Kellog, G. E. (2010). A novel & efficient tool for locating & characterizing protein cavities & binding sites. *Proteins: Structure, Function and Bioinformatics*, 78(4), 825-842.
- Tue-nguen, P., Kodchakorn, K., Nimmanpipug, P., Lawan, N., Nangola, S., Tayapiwatana, C., Lee, V. S. (2013). Improved scFv Anti-HIV p17 binding affinity guided from the theoretical calculation of pairwise decomposition energies and computational alanine scanning. *BioMed Research International*, 2013. doi:10.1155/2013/713585
- Undurraga, E. A., Edillo, F. E., Erasmo, J. N. V., Alera, M. T. P., Yoon, I.-K., Largo, F. M., & Shepard, D. S. (2017). Disease burden of dengue in the Philippines: Adjusting for underreporting by comparing active & passive dengue surveillance in Punta Princesa, Cebu City. *The American Journal of Tropical Medicine & Hygiene*, 96(4), 887-898.
- Usmani, S. S., Bedi, G., Samuel, J. S., Singh, S., Kalra, S., Kumar, P., Raghava, G. P. S. (2017). THPdb: Database of FDA-approved peptide and protein therapeutics. *PLOS ONE*, 12(7). doi:10.1371/journal.pone.0181748
- Valdes, I., Gil, L., Romero, Y., Castro, J., Puente, P., Lazo, L., Hermida, L. (2011). The chimeric protein domain III-capsid of dengue virus serotype 2 (DEN-2) successfully boots neutralizing antibodies generated in monkeys upon infection with DEN2. *Clinical and Vaccine Immunology*, 18(3), 455-459.
- Valeur, E., & Bradley, M. (2009). Amide bond formation: beyond the myth of coupling reagents. *Chemical Society Reviews*, 38, 606-631.

- Verma, R., Jatav, V. K., & Sharma, S. (2014). Identification of inhibitors of dengue virus NS2B/NS3 Serine Protease: A molecular docking and simulation approach. *Asian Journal of Pharmaceutical & Clinical Research*, 8(1), 287-292.
- Viau, M., Letourne, M., Sirois-Deslongchamps, A., Boulanger, Y., & Fournier, A. (2007). Study of solid-phase synthesis and purification strategies for the preparation of polyglutamine peptides. *Peptide Science*, 88(5), 754-763.
- Vivo, M. D., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry*, 59, 4035-4061.
- Wacowich-Sgarbi, S. A. (2000). *Synthesis and conformational studies of constrained Oligosaccharides*. (Doctor of Philosophy), University of Alberta, Edmonton, Alberta, Canada.
- Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein engineering design & selection*, 8(2), 127-134.
- Wallach, I., Jaitly, N., & Lilien, R. (2010). A structure-based approach for mapping adverse drug reactions to the perturbation of underlying biological pathways. *PLOS ONE*, 5(8). doi:<https://doi.org/10.1371/journal.pone.0012063>
- Wallner, B., & Elofsson, A. (2006). Identification of correct regions in protein models using structural, alignment and consensus information. *Protein Science*, 15, 900-913.
- Wang, C., Nguyen, P. H., Pham, K., Huynh, D., Le, T.-B. N., Wang, H., Luo, R. (2016). Calculating protein-ligand binding affinities with MMPBSA: method and error Analysis. *Journal of Computational Chemistry*, 37(27), 2436-2446.
- Wang, J., Morin, P., Wang, W., & Kollman, P. A. (2001). Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of Efavirenz by docking and MM-PBSA. *Journal of the American Chemical Society*, 123(22), 5221-5230.
- Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*, 37, 623-633.

- Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wong, J., & Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*, *37*, 623-633.
- Watterson, D., Kobe, B., & Young, P. R. (2012). Residues in domain III of the dengue virus envelope glycoprotein involved in cell-surface glycosaminoglycan binding. *Journal of General Virology*, *93*, 72-82.
- Wei, B. Q., Weaver, L. H., Ferrari, A. M., Matthews, B. W., & Shoichet, B. K. (2004). Testing a flexible-receptor docking algorithm in a model binding site. *Journal of Molecular Biology*, *337*, 1161-1182.
- Wichapong, K., Nueangaudom, A., Pianwanit, S., Tanaka, F., & Kokpol, S. (2014). Molecular dynamic simulation, binding free energy calculation and molecular docking of human D-amino acid oxidase (DAAO) with its inhibitors. *Molecular Simulations*, *40*(14), 1167-1189.
- Wichmann, C., Becker, Y., Chen-Wichmann, L., Vogel, V., Vojtkova, A., Herglotz, J., Grez, M. (2010). Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity. *BLOOD*, *116*(4), 603-613.
- Wolf, A., & Kirschner, K. N. (2013). Principal component and clustering analysis on molecular dynamics data of the ribosomal L11.23S subdomain. *Journal of Molecular Modeling*, *19*(2), 539-549.
- Wong, C. F. (2008). Flexible ligand-flexible protein docking in protein kinase systems. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1784*(1), 244-251.
- Woods, C. J., Malaisree, M., Michel, J., Long, B., McIntosh-Smith, S., & Mulholland, A. J. (2014). Rapid decomposition & visualization of protein ligand binding free energies by residue and by water. *Faraday Discussions*, *169*, 477-499.
- Xu, C., Lam, H. Y., Zhang, Y., & Li, X. (2013). Convergent synthesis of MUC1 Glycopeptides via Serine ligation. *Chemical Communications*, *49*(55), 6200-6202.
- Xu, Y., Rahman, N. A., Othman, R., Hu, P., & Hu, M. (2012). Computational identification of self-inhibitory peptides from envelope protein. *Protein*, *80*(9), 2154-2168

- Yan, C., & Zou, X. (2014). Predicting peptide binding sites on protein surfaces by clustering chemical interactions. *Journal of Computational Chemistry*, 36(1), 49-61.
- Yang, C., Zhang, S., He, P., Wang, C., Huang, J., & Zhou, P. (2015). Self-binding peptides: folding or binding. *Journal of Chemical Information and Modeling*, 329-342.
- Yang, Y., Shen, Y., Liu, H., & Yao, X. (2011). Molecular dynamics simulation and free energy calculation studies of the binding mechanism of allosteric inhibitors with p38-alpha MAP Kinase. *Journal of Chemical Information and Modeling*, 51(12), 3235-3246.
- Yennamalli, R., Subbarao, N., Kampmann, T., McGeary, R. P., Young, P. R., & Kobe, B. (2009). Identification of novel target site & an inhibitor of the dengue virus E protein. *Journal of Computer-Aided Molecular Design*, 23(6), 333-341.
- Yi, X., Zhang, Y., Wang, P., Qi, J., Hu, M., & Zhong, G. (2015). Ligands binding and molecular simulation: the potential investigation of a biosensor based on an insect odorant binding protein. *International Journal of Biological Sciences*, 11(1), 75-87.
- Young, D. (2004). *Computational chemistry: A practical guide for applying techniques to real world problems*: John Wiley & Sons.
- Zaki, R., Roffeei, S. N., Hii, Y. L., Yahya, A., Appannan, M., Said, M. A., Rocklov, J. (2019). Public perception & attitude towards dengue prevention activity & response to dengue early warning in Malaysia. *PLOS ONE*, 14(2). doi:10.1371/journal.pone.0212497
- Zamri, A., Teruna, H. Y., Rahmawati, E. N., Frimayanti, N., & Ikhtiarudin, I. (2019). Synthesis & *In Silico* studies of a Benzenesulfonyl Curcumin Analogue as a new anti dengue virus type 2 NS2B/NS3. *Indonesian Journal of Pharmacy*, 30(2), 84-90.
- Zhang, X., Sun, L., & Rossmann, M. G. (2015). Temperature dependent conformational change of dengue virus. *Current Opinion in Virology*, 12, 109-112.
- Zhang, Y., Zhang, W., Ogata, S., Clements, D., Strauss, J. H., Baker, T. S., Rossmann, M. G. (2004). Conformational changes of the flavivirus E-glycoprotein. *Structure*, 12, 1607-1618.

- Zhang, Z., Witham, S., & Alexoy, E. (2011). On the role of electrostatic on protein-protein interactions. *Physical Biology*, 8(3). doi:10.1088/1478-3975/8/3/035001
- Zhao, S., Kumar, R., Sakai, A., Vetting, M. W., Wood, B. M., Brown, S., Jacobson, M. P. (2013). Discovery of new enzymes and metabolic pathways using structure and genome context. *Nature*, 502(7473), 698-702.
- Zhou, H.-X., & Bates, P. A. (2013). Modeling protein association mechanisms and kinetics. *Current Opinion in Structural Biology*, 23(6), 1-13.
- Zhu, Y.-L., Beroza, P., & Artis, D. R. (2014). Including explicit water molecules as part of the protein structure in MMPBSA calculations. *Journal of Chemical Information and Modeling*, 54, 462-469.

Universiti Malaya

LIST OF PUBLICATIONS AND PAPERS PRESENTED

1. **Asfarina Amir Hassan**, Vannajan Sanghiran Lee, Aida Baharuddin, Shatrah Othman, Yongtao Xu, Meilan Huang, Rohana Yusof, Noorsaadah Abd. Rahman, Rozana Othman, Conformational and Energy Evaluations of Novel Peptides Binding to Dengue Virus Envelope Protein. *Journal of Molecular Graphics & Modelling*, 2017,74, 273-287.
2. Aida Baharuddin, **Asfarina Amir Hassan**, Rozana Othman, Yongtao Xu, Meilan Huang, Bimo Ario Tejo, Rohana Yusuf, Noorsaadah Abd Rahman & Shatrah Othman, Dengue Envelope Domain III-Peptide Binding Analysis via Tryptophan Fluorescence Quenching Assay, *Chemical and Pharmaceutical Bulletin*, 2014, 62(10), 947-955.
3. Aida Baharuddin, **Asfarina Amir Hassan**, Gan Chye Sheng, Shah Bakhtiar Nasir, Shatrah Othman, Rohana Yusof, Noorsaadah Abd Rahman, Current Approaches in Antiviral Drug Discovery Against the Flaviviridae Family, *Current Pharmaceutical Design*, 2014, 20, 3428-3444.
4. **Asfarina Amir Hassan**, Vannajan Sanghiran Lee, Rohana Yusof, Noorsaadah Abd. Rahman, Rozana Othman, Characterizing Ligand-Binding Site for Dengue Virus Type 2 (DENV2) Envelope Protein. *Journal of Biomolecular Structure & Dynamics* (Manuscript in preparation).

List of Oral Presentations

1. **Asfarina Amir Hassan**, Vannajan Sanghiran Lee, Aida Baharuddin, Shatrah Othman, Yongtao Xu, Meilan Huang, Rohana Yusof, Noorsaadah Abd. Rahman & Rozana Othman. (2015, November). Docking & Molecular Dynamics Simulation Study of Dengue Virus Type 2 (DENV2) Envelope Protein with Peptides. The University of Malaya-Academia Sinica Research Symposium. High Impact Research Building, University Malaya.
2. **Asfarina Amir Hassan**, Rozana Othman, Vannajan Sanghiran Lee, Aida Baharuddin, Yongtao Xu, Meilan Huang, Rohana Yusof, Shatrah Othman, Noorsaadah Abd Rahman. (2014, February). Designing inhibitors against envelope of dengue virus: An *In Silico* Study, Drug Development & Design Research Group (DDDRG) colloquium. Eastin Hotel, Petaling Jaya.
3. **Asfarina Amir Hassan**, Rozana Othman, Vannajan Sanghiran Lee, Meilan Huang, Shatrah Othman, Rohana Yusof and Noorsaadah Abd. Rahman. (2014, January). Docking & Molecular Dynamics Simulation Study of Dengue Virus Type 2 (DENV2) Envelope Protein with Peptides. The 9th Mathematics & Physical Science Graduate Congress. Komplek Dewan Kuliah, Fakulti Sains, University Malaya.

List of Poster Presentations

1. **Asfarina Amir Hassan**, Rozana Othman, Vannajan Sanghiran Lee, Aida Baharuddin, Meilan Huang, Rohana Yusof, Shatrah Othman, Noorsaadah Abd Rahman. (2015, August). Peptide Binding to Dengue Virus Envelope Protein: Docking and MD Simulations Studies. CENAR, Annual Colloquium on Drug Development from Natural Products. Universiti Malaya.
2. **Asfarina Amir Hassan**, Rozana Othman, Vannajan Sanghiran Lee, Aida Baharuddin, Meilan Huang, Shatrah Othman, Rohana Yusof and Noorsaaadah Abd Rahman. (2014, February). Molecular Dynamics Simulation of Dengue Virus Type 2 (DENV2) Envelope Protein with Peptides as Potential Inhibitors. The 5th Trilateral Seminar UM-CU-NUS. Department of Chemistry, Faculty of Science, University of Malaya.
3. **Asfarina Amir Hassan**, Rozana Othman, Shatrah Othman, Rohana Yusof, Noorsaadah Abd. Rahman. (2013, June). Peptide Binding to Dengue Virus Type 2 Envelope Protein: A Docking Study. LRGS Dengue Colloquium & Workshop. Research Management & Innovation Complex (RMIC), University of Malaya.
4. **Asfarina Amir Hassan**, Rozana Othman, Shatrah Othman, Rohana Yusof, Noorsaadah Abd. Rahman. Docking Study of Lacto-N-Neotetraose as a Potential Inhibitor against Envelope of Dengue Virus Type 2. (2013, March). The Fifth HOPE Meeting with Nobel Laureates. Grand Prince Hotel New Takanawa, Tokyo, Japan.
5. **Asfarina Amir Hassan**, Rozana Othman, Rohana Yusof & Noorsaadah Abd Rahman. Molecular Dynamics Simulation & Docking of Dengue Virus Type 2 (DEN2). (2012, April). University of Malaya Researcher's Conference. University of Malaya.