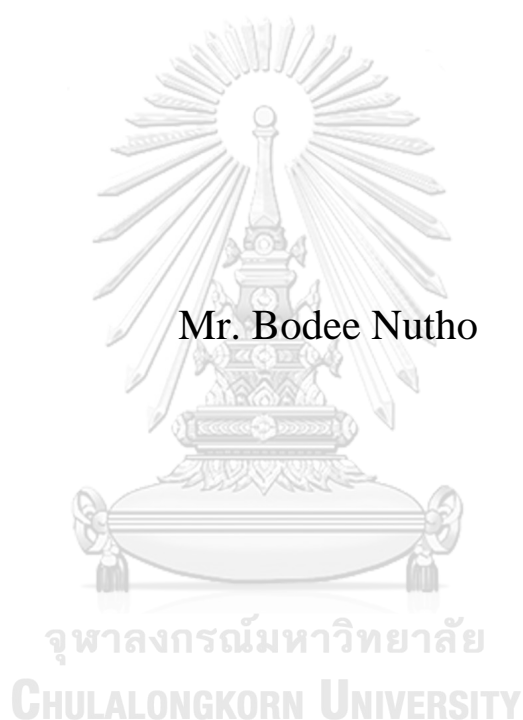


REACTION MECHANISMS OF SUBSTRATE AND
INHIBITOR WITH NS2B/NS3 SERINE PROTEASE OF ZIKA
VIRUS BY MM AND QM/MM SIMULATIONS



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Biotechnology
Common Course
Faculty of Science
Chulalongkorn University
Academic Year 2018
Copyright of Chulalongkorn University

กลไกการเกิดปฏิกิริยาของซัลเฟตและตัวยับยั้งกับเอ็นเอสทูปี/เอ็นเอสทีเซอร์อินโปรทีเอสของ
เชื้อไวรัสซิกาด้วยการจำลองแบบเอ็มเอ็มและคิวเอ็ม/เอ็มเอ็ม



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีชีวภาพ ไม่สังกัดภาควิชา/เทียบเท่า
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	REACTION MECHANISMS OF SUBSTRATE AND INHIBITOR WITH NS2B/NS3 SERINE PROTEASE OF ZIKA VIRUS BY MM AND QM/MM SIMULATIONS
By	Mr. Bodee Nutho
Field of Study	Biotechnology
Thesis Advisor	Assistant Professor THANYADA RUNGROTMONGKOL, Ph.D.
Thesis Co Advisor	Professor Adrian Mulholland, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy

..... Dean of the Faculty of Science
(Professor POLKIT SANGVANICH, Ph.D.)

DISSERTATION COMMITTEE

..... Chairman
(Assistant Professor RATH PICHYANGKURA, Ph.D.)

..... Thesis Advisor
(Assistant Professor THANYADA RUNGROTMONGKOL, Ph.D.)

..... Thesis Co-Advisor
(Professor Adrian Mulholland, Ph.D.)

..... Examiner
(Professor SUPOT HANNONGBUA, Dr.rer.nat.)

..... Examiner
(Associate Professor APHICHART KARNCHANATAT, Ph.D.)

..... External Examiner
(Associate Professor Panida Surawatanawong, Ph.D.)

บทคัดย่อ : กลไกการเกิดปฏิกิริยาของซับสเตรตและตัวยับยั้งกับเอ็นเอสทูบี/เอ็นเอสทีเรเซอร์อิน โพรทีเอสของเชื้อไวรัสซิกาด้วยการจำลองแบบเอ็มเอ็มและคิวเอ็ม/เอ็มเอ็ม. (REACTION MECHANISMS OF SUBSTRATE AND INHIBITOR WITH NS2B/NS3 SERINE PROTEASE OF ZIKA VIRUS BY MM AND QM/MM SIMULATIONS) อ.ที่ปรึกษาหลัก : ผศ. ดร.ธัญญา รุ่งโรจน์มงคล, อ.ที่ปรึกษาร่วม : ศ. ดร.เอเดรียน มูฮอลแลนด์

โรคติดเชื้อไวรัสซิกากลายเป็นปัญหาสาธารณสุขที่สำคัญทั่วโลก จากข้อมูลทางระบาดวิทยาแสดงให้เห็นถึงความเชื่อมโยงของการติดเชื้อไวรัสกับความผิดปกติของระบบประสาทที่เกิดขึ้นทั้งในเด็กแรกเกิดและผู้ใหญ่ อย่างไรก็ตามในปัจจุบันยังไม่มียาหรือวัคซีนที่ใช้ป้องกันหรือควบคุมการติดเชื้อไวรัสซิกา หนึ่งในโปรตีนเป้าหมายที่สำคัญของไวรัสชนิดนี้ คือ เอ็นไซม์เอ็นเอสทูบี/เอ็นเอสทีเรเซอร์อิน โพรทีเอสซึ่งมีบทบาทสำคัญในการตัดพันธะเปปไทด์ระหว่างการจำลองตัวของเชื้อไวรัส โดยในขั้นต้นได้อาศัยเทคนิคการจำลองพลวัตเชิงโมเลกุลแบบดั้งเดิมเพื่อศึกษารูปแบบการจดจำการเข้าจับและอันตรกิริยาที่เกิดขึ้นระหว่างเอ็นไซม์เอ็นเอสทูบี/เอ็นเอสทีเรเซอร์อิน โพรทีเอสกับซับสเตรต 4 ตัว จากผลการคำนวณพบว่าซับสเตรตที่ตำแหน่ง P1 และ P2 มีบทบาทสำคัญในการเข้าจับกับเอ็นไซม์ ขณะที่ตำแหน่ง P3 และ P4 มีความสำคัญที่น้อยกว่าในการเกิดอันตรกิริยาการเข้าจับ หลังจากนั้นทำการศึกษากลไกการเร่งปฏิกิริยาการตัดพันธะเปปไทด์ของเอ็นไซม์กับเปปไทด์ซับสเตรต (TGKRS) โดยอาศัยวิธีการจำลองที่ผสมผสานระหว่างกลศาสตร์ควอนตัมและกลศาสตร์โมเลกุล หรือการจำลองแบบคิวเอ็ม/เอ็มเอ็ม ผลการจำลองพลวัตเชิงโมเลกุลแบบคิวเอ็ม/เอ็มเอ็ม โดยใช้ระเบียบวิธี PM6/ff14SB พบว่าขั้นตอนการถ่ายโอนโปรตอนจากกรดอะมิโนเซอร์อิน 135 ไปสู่กรดอะมิโนฮิสติดีน 51 และปฏิกิริยาการชนนิวคลีโอไฟล์บนตัวซับสเตรตด้วยกรดอะมิโนเซอร์อิน 135 จัดเป็นปฏิกิริยาคำเนินไปแบบต่อเนื่องพร้อมกันในครั้งเดียว ซึ่งผลการคำนวณแบบคิวเอ็ม/เอ็มเอ็มในระดับทฤษฎีที่สูงขึ้นยังคงสนับสนุนรูปแบบการเกิดปฏิกิริยาเคมีดังกล่าว นอกจากนี้กลไกการยับยั้งปฏิกิริยาของเอ็นไซม์กับตัวยับยั้งไดเปปไทด์แอลดีไฮด์ (acyl-KR-aldehyde) ยังได้รับการศึกษาคู่ด้วยระเบียบวิธีคิวเอ็ม/เอ็มเอ็ม ผลการคำนวณพบว่าขั้นตอนการถ่ายโอนโปรตอนจากกรดอะมิโนเซอร์อิน 135 ไปสู่กรดอะมิโนฮิสติดีน 51 เกิดขึ้นพร้อมกับขั้นตอนการเดมิ นิวคลีโอไฟล์บนหมู่แอลดีไฮด์ด้วยกรดอะมิโนเซอร์อิน 135 โดยเกิดเป็นสารเชิงซ้อนโคเวเลนต์ที่มีประจุลบระหว่างไดเปปไทด์แอลดีไฮด์และเอ็นไซม์ที่จำลองโครงสร้างทางเคมีคล้ายคลึงกับสารมีซันดรีในการเกิดปฏิกิริยาไฮโดรไลซิสของซับสเตรต ดังนั้นวิธีทางคอมพิวเตอร์ที่ได้นำเสนอในวิทยานิพนธ์นี้จึงสามารถเป็นตัวช่วยในการออกแบบตัวยับยั้งเอ็นไซม์เอ็นเอสทูบี/เอ็นเอสทีเรเซอร์อิน โพรทีเอสที่มีประสิทธิภาพต่อไป

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา เทคโนโลยีชีวภาพ
ปีการศึกษา 2561

ลายมือชื่อนิสิต
ลายมือชื่อ อ.ที่ปรึกษาหลัก
ลายมือชื่อ อ.ที่ปรึกษาร่วม

5872853023 : MAJOR BIOTECHNOLOGY

KEYWORD Zika virus, NS2B/NS3 serine protease, MD simulation, QM/MM
D: simulation

Bodee Nutho : REACTION MECHANISMS OF SUBSTRATE AND INHIBITOR WITH NS2B/NS3 SERINE PROTEASE OF ZIKA VIRUS BY MM AND QM/MM SIMULATIONS. Advisor: Asst. Prof. THANYADA RUNGROTMONGKOL, Ph.D. Co-advisor: Prof. Adrian Mulholland, Ph.D.

Zika virus (ZIKV) infection has become a public health concern worldwide. The recent epidemiological data have revealed a possible association of ZIKV infection with neurological manifestations in both newborn children and adults. Currently, there is no anti-ZIKV drug or vaccine available for preventing or controlling ZIKV infection. An attractive drug target for ZIKV treatment is NS2B/NS3 serine protease that plays a crucial role to cleave the peptide bond during the viral replication process. Initially, the ZIKV NS2B/NS3 serine protease in complex with four peptide substrates were studied to investigate the binding recognition and protein-substrate interactions using conventional molecular dynamics (MD) simulations. The results indicate that the P1 and P2 positions of the substrate play a major role in binding with the enzyme, while the P3 and P4 positions show a less contribution in binding interaction. Afterwards, the cleavage reaction mechanism for the ZIKV protease with its peptide substrate (TGKRS) was studied by hybrid quantum mechanics/molecular mechanics (QM/MM) approach. QM/MM (PM6/ff14SB) free-energy simulations indicate that proton transfer from S135 to H51 and nucleophilic attack on the substrate by S135 are concerted. Importantly, higher-level QM/MM calculations also support a concerted reaction mechanism for this particular reaction. In addition, QM/MM calculations were performed to determine the inhibition mechanism of the ZIKV protease by a dipeptidyl aldehyde inhibitor (acyl-KR-aldehyde). The results show that the transfer of proton from the catalytic S135 to H51 take places in concert with nucleophilic addition on the aldehyde warhead by S135. The anionic covalent complex between the dipeptidyl aldehyde and the ZIKV protease resembles the tetrahedral intermediate for substrate hydrolysis. Therefore, the computational approaches presented here are helpful for further designing of potent NS2B/NS3 inhibitors.

Field of Study: Biotechnology

Student's Signature

Academic 2018

.....
Advisor's Signature

Year:

.....
Co-advisor's Signature

.....

ACKNOWLEDGEMENTS

This dissertation can be succeeded by the attentive support from many people who help and support me to overcome the difficulties during the years of my Ph.D. study at Program in Biotechnology, Faculty of Science, Chulalongkorn University.

Firstly, I would like to thank my advisor, Assistant Professor Dr. Thanyada Rungrotmongkol who gives a great opportunity to study in my Ph.D. program. Her understanding in protein knowledge not only guides but also inspires me to do and finish this dissertation. I would also particularly thank Professor Dr. Adrian J. Mulholland, my co-advisor, for useful suggestions and valuable experiences in his laboratory at the University of Bristol, UK. My grateful thanks are extended to Dr. Kara E. Ranaghan for her kindly advice about the QM/MM fundamental knowledge and its calculation procedures.

In addition, I would like to express my sincere gratitude to all of my thesis committees, Assistant Professor Dr. Rath Pichyangkura, Professor Dr. Supot Hannongbua, Associate Professor Dr. Aphichart Karnchanatat and Associate Professor Dr. Panida Surawatanawong for their essential advice and very useful suggestions in my examination. My thanks are also expressed to all members in the Biosim laboratory at Department of Biochemistry, Faculty of Science, Chulalongkorn University for their useful discussions in both academic and non-academic frameworks during a period of my study.

I would acknowledge the Royal Golden Jubilee Ph.D., Program and the 90th Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund) for financial support, the Centre for Computational Chemistry (CCC) at School of Chemistry, University of Bristol and the the Center of Excellence in Computational Chemistry (CECC) at Department of Chemistry, Chulalongkorn University for computational facilities and resources.

Lastly, I would like to give my deepest gratitude to my family for their carefulness, support, and everlasting love.

Bodee Nutho

TABLE OF CONTENTS

	Page
ABSTRACT (THAI)	iii
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS.....	xvi
CHAPTER I INTRODUCTION.....	1
1.1 Research rationality.....	1
1.2 Viral structure.....	2
1.3 Flavivirus Life Cycle.....	3
1.4 Pathogenesis and Clinical Presentations	6
1.5 Flaviviral NS2B/NS3 serine protease.....	7
1.6 Reaction mechanism of flaviviral NS2B/NS3 serine protease.....	9
1.7 Inhibition mechanism of flaviviral NS2B/NS3 protease by peptidyl aldehyde inhibitors.....	11
1.8 Objectives of the study	15
CHAPTER II THEORIES	16
2.1 Molecular mechanics.....	16
2.2 Molecular dynamics	18
2.2.1 Integration algorithm.....	21
2.2.2 Periodic boundary condition	22
2.2.3 Constraint dynamics	23
2.2.4 Water models.....	23
2.2.5 Continuum solvation	24
2.2.5.1 Poisson-Boltzmann (PB) model	25

2.2.5.2 Generalized Born (GB) model.....	26
2.3 Investigation of pK _a of a titratable amino acid.....	27
2.3.1 Constant pH molecular dynamics.....	28
2.3.2 Constant pH replica exchange molecular dynamics.....	28
2.4 Quantum mechanics	29
2.5 Combined quantum mechanics and molecular mechanics (QM/MM).....	30
2.5.1 QM/MM calculations	31
2.5.2 QM/MM boundary treatment	32
2.5.3 Calculations of energy profiles and barriers for reactions	33
2.5.3.1 The Weighted histogram analysis method (WHAM).....	35
CHAPTER III RESEARCH METHODOLOGY	37
3.1 Part I: Binding recognition of substrates in ZIKV NS2B/NS3 serine protease	37
3.1.1 Explicit-solvent CpHMD and CpH-REMD	37
3.1.2 System preparation for MD simulations	37
3.1.3 Molecular dynamics simulations.....	40
3.2 Part II: Reaction mechanism of the ZIKV protease with its substrate	40
3.2.1 System preparation	40
3.2.2 Molecular dynamics simulations.....	41
3.2.3 QM/MM MD free-energy calculations	42
3.2.4 High-level QM/MM calculations for the first step of the acylation process	43
3.3 Part III: Inhibition mechanism of the ZIKV protease with dipeptidyl aldehyde inhibitor	45
3.3.1 System preparation	45
3.3.2 Molecular dynamics simulation	45
3.3.3 QM/MM MD free-energy calculations	46
3.3.4 QM/MM potential energy calculations	47
CHAPTER IV RESULTS AND DISCUSSION.....	49
4.1 Part I: Binding recognition of substrates in ZIKV NS2B/NS3 serine protease	49
4.1.1 Protonation states of ionizable residues	49

4.1.2 Stability of the global structure	51
4.1.3 Intermolecular hydrogen bonds	52
4.1.4 Key residues for substrate binding	57
4.1.5 Binding free energy of protein-substrate complexes.....	61
4.2 Part II: Reaction mechanism of the ZIKV protease with its substrate	63
4.2.1 MM MD Simulations	63
4.2.2 Distances at the catalytic triad and oxyanion hole	65
4.2.3 PM6/MM free energy profiles.....	67
4.2.4 Potential energy surface for reaction and effect of basis set size.....	75
4.2.5 Conformational sampling and ab initio energy calculations	76
4.2.6 Structural analysis and hints for designing new inhibitors.....	85
4.3 Part III: Inhibition mechanism of the ZIKV protease with dipeptidyl aldehyde inhibitor	89
4.3.1 QM/MM free-energy reaction path	89
4.3.2 QM/MM potential energy reaction path.....	96
CHAPTER V CONCLUSIONS	107
5.1 Part I: Binding recognition of substrates in ZIKV NS2B/NS3 serine protease	107
5.2 Part II: Reaction mechanism of the ZIKV protease with its substrate	107
5.3 Part III: Inhibition mechanism of the ZIKV protease with dipeptidyl aldehyde inhibitor	109
REFERENCES	110
APPENDICES	131
APPENDIX 1: The results from MM/GBSA method giving the energy components and average binding free energies (in kcal mol ⁻¹) for the complexes of the four substrates with the ZIKV protease compared with the experimental K_m (in μ M).	132
APPENDIX 2: Free energy profiles for the TI formation for the reaction of the ZIKV protease with its substrate, calculated from umbrella sampling simulations at the PM3/ff14SB, PM6/ff14SB, AM1-d/ff14SB and SCC-DFTB/ff14SB levels of theory.	133

APPENDIX 3: Free energy profile for formation of the tetrahedral adduct relative to the noncovalent Michaelis complex for the reaction of the ZIKV protease with dipeptidyl aldehyde inhibitor, calculated from umbrella sampling simulations at the PDDG-PM3/ff14SB, PM3/ff14SB, PM6/ff14SB and AM1-d/ff14SB levels of theory.....	134
APPENDIX 4: Comparison of potential energy profiles for the reaction mechanism on substrate and dipeptidyl aldehyde inhibitor catalyzed by the ZIKV protease.	135
VITA.....	136



LIST OF TABLES

	Page
Table 1. The cleavage sites between P1 and P1' junction for each substrate.	10
Table 2. Force field in molecular mechanics	16
Table 3. Predicted pK _a values of the lysine residue of each peptide substrate obtained from cpHMD and cpH-REMD simulations	51
Table 4. The results from MM/PBSA method giving the energy components and average binding free energies (in kcal mol ⁻¹) for the complexes of the four substrates with the ZIKV protease compared with the experimental K _m (in μM)	62
Table 5. Free energies, relative to the MC, obtained at the PM6/ff14SB QM/MM level from two-dimensional umbrella sampling MD simulations	70
Table 6. Average Mulliken charges (PM6/ff14SB) in atomic units for atoms involving the formation of the TI, averaged from the QM/MM free energy profile ...	73
Table 7. Energy barriers (kcal mol ⁻¹) for the TI formation.....	78
Table 8. Average values of the interatomic distances involving active site interactions (in Å) for the noncovalent Michaelis complex, transition state and covalent complex structures	93
Table 9. Average Mulliken charges (PDDG-PM3/ff14SB) in atomic units (au) for atoms involving the formation of the tetrahedral adduct, averaged from the QM/MM free-energy profile	96
Table 10. Potential energy barriers (Δ [‡] V) and reaction energies (Δ _r V).....	99
Table 11. Reaction energetics (in kcal mol ⁻¹ , relative to the reactant state) and reaction coordinate (RC) values (in Å) from CI-NEB method for the reaction of the tetrahedral adduct.....	105

LIST OF FIGURES

	Page
Figure 1. The genome of Zika virus.....	3
Figure 2. Life cycle of Zika virus	5
Figure 3. Clinical manifestations and consequences of ZIKV infection	7
Figure 4. Three-dimensional structures of the ZIKV protease in complex with the tetrapeptide “TGKR” of the NS2B domain	9
Figure 5. Proposed reaction mechanism for the reaction of the acylation process catalyzed by the ZIKV protease.....	11
Figure 6. (A) Three-dimensional structure of the ZIKV NS2B/NS3 serine protease in complex with a dipeptide inhibitor (acyl-KR-aldehyde) (PDB ID: 5YOF [79]), where surface charge of the ZIKV protease structure was calculated using coulombic surface coloring with default setting implemented in UCSF Chimera [82], shading from red (high negative charge) to blue (high positive charge). (B) Two-dimensional chemical structure of a dipeptidyl aldehyde inhibitor used in this study	13
Figure 7. Proposed reaction mechanism of the ZIKV protease inhibition by peptidyl aldehyde	14
Figure 8. Range of time scales for dynamics in biological systems	19
Figure 9. Simplified schematic of MD algorithm.....	21
Figure 10. Schematic representation of periodic boundary condition	23
Figure 11. Type of water molecule models generated by 3-, 4-, 5- and 6-site of point charges	24
Figure 12. Illustration of the QM/MM concept	31
Figure 13. Illustration of subtractive QM/MM coupling scheme	32
Figure 14. Different approaches to treat the QM/MM boundary.....	33
Figure 15. Two different approaches used in QM/MM reaction modelling.....	34
Figure 16. (A) Three-dimensional structure of the tetrapeptide “TGKR” binding to the active site of the ZIKV protease (PDB accession code 5GJ4), where the surface charge with negative and positive charge accumulation is shaded from red to blue, respectively. (B) Comparison of the four substrate sequences used in this study	39

Figure 17. Chemical structures of (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC used in this study corresponding to substrates 1 to 4, respectively	39
Figure 18. Active site of ZIKV protease with the fragment of substrate.....	42
Figure 19. Titration curves for the lysine residue of each peptide substrate calculated with (Left) CpHMD and (Right) CpH-REMD	50
Figure 20. RMSD plots for complex all heavy atoms	52
Figure 21. Percentage of hydrogen bond occupation of the ZIKV protease residues formed with the four different substrates.....	54
Figure 22. Binding patterns of (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC peptides in the ZIKV protease active site extracted from the last MD snapshot	55
Figure 23. Per-residue decomposition free energy (in kcal mol ⁻¹) calculated with MM/PBSA method for the ZIKV protease in complex with (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC	58
Figure 24. (Left) Per-residue decomposition free energies are given as total energy (black bars), side chain (dark grey bars) and backbone contributions (light gray bars) for the four studied complexes of the ZIKV protease with (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC. (Right) The electrostatic and vdW contributions are shown in black and gray lines, respectively.	60
Figure 25. (Top panel) Time evolution of the RMSD (Å) of backbone atoms of all three replicate simulations of bound and apo systems.....	64
Figure 26. Time evolution of the radius of gyration (R _g) of residues within 4 Å of substrate for all three replicate simulations of bound (left) and apo (right) systems...	64
Figure 27. Superposition of the representative structures for each replica in bound state and apo state (obtained by cluster analysis) of the binding pockets.....	65
Figure 28. (A) Definition of the interatomic distances involved in the cleavage reaction (d1–d6). (B) Histograms of key interatomic distances, d1–d6, from MM MD simulations, sampling from 100 to 200 ns	66
Figure 29. (A) The geometric parameters describing the nucleophilic attack trajectory are depicted. θ_y is the angle defined by the enzyme serine oxygen, the carbonyl carbon and the carbonyl oxygen on the substrate. θ_x is the angle between (i) the plane defined by the enzyme serine oxygen, the substrate's carbonyl carbon, and the substrate's carbonyl oxygen, and (ii) the plane defined by the peptide bond. The dashed line represents the distance between the enzyme serine oxygen and the substrate's carbonyl	

carbon. (B) Histograms of these two geometric parameters, θ_x and θ_y , from three independent MD simulations, sampling from 100 to 200 ns	67
Figure 30. (A) 2D free energy surfaces for the TI formation and (B) the breakdown of the peptide bond for the reaction of the ZIKV protease with its substrate	68
Figure 31. 1D free energy profiles for the TI formation and the breakdown of the peptide bond relative to the MC for the reaction of the ZIKV NS2B/NS3 serine protease with its substrate	69
Figure 32. 1D free energy profiles for the tetrahedral intermediate formation using one reaction coordinate involving in the proton transfer from S135 oxygen to H51 ϵ nitrogen in the Michaelis complex.....	71
Figure 33. Average values of (A) the distances involving proton transfer and nucleophilic attack and (B) the hydrogen bond distances between QM and MM atoms along the reaction pathway	72
Figure 34. Average values of the distances along the reaction pathway involving the second step of acylation process	74
Figure 35. QM/MM energy profiles for the first step of the acylation process along the reaction pathway obtained at BH&HLYP-D3 QM/MM energy calculations using different basis sets for snapshot 1	76
Figure 36. Potential energy profiles for the first step of the acylation process calculated at the (A) BH&HLYP-D3/6-31G(d)/ff14SB, (B) LMP2/(aug)-cc-pVTZ/ff14SB, (C) SCS-LMP2/(aug)-cc-pVTZ/ff14SB, and (D) LCCSD(T)/(aug)-cc-pVTZ/ff14SB QM/MM level.....	77
Figure 37. Superposition of the QM/MM minimized structures of the five MD snapshots at the Michaelis complex (MC), transition state (TS1) and tetrahedral intermediate (TI)	78
Figure 38. Correlation between energy barriers and reaction energies found in each snapshot calculated at the (A) BH&HLYP-D3/6-31G(d)/ff14SB, (B) LMP2/(aug)-cc-pVTZ/ff14SB, (C) SCS-LMP2/(aug)-cc-pVTZ/ff14SB, and (D) LCCSD(T)/(aug)-cc-pVTZ/ff14SB QM/MM levels	81
Figure 39. Potential energy profiles for the first step of the acylation process calculated at the B3LYP-D3/6-31G(d)/ff14SB level of theory	82
Figure 40. Comparison of the average reaction profiles (relative to the Michaelis complex) for the formation of tetrahedral intermediate calculated with MP2 and SCS-MP2 (using the BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries) with and without local approximations.....	82

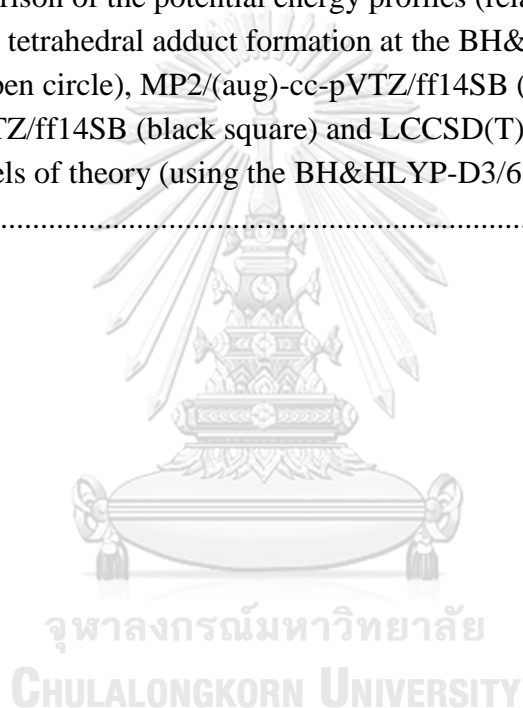
Figure 41. Comparison of the average reaction profiles (relative to the Michaelis complex) for the formation of tetrahedral intermediate calculated at SCS-LMP2/(aug)-cc-pVTZ/ff14SB level	84
Figure 42. (A) Representation of the extended QM region of the Michaelis complex, showing the QM atoms as sticks with magenta carbon atoms and the MM region with green carbon atoms. The QM region includes the side-chains of the catalytic triad (H51, D75 and S135) and the fragments of substrate and oxyanion hole. The hydrogen bonds are represented by dashed lines. (B) Potential energy profiles for the first step of the acylation process calculated at the BH&HLYP-D3/6-31G(d)/ff144SB level with two different QM subsystems	85
Figure 43. QM/MM (BH&HLYP-D3/6-31G(d)/ff14SB) optimized structures of the Michaelis complex (MC), transition state (TS1), and tetrahedral intermediate (TI) in the active site of ZIKV NS2B/NS3 protease	87
Figure 44. Time evolution of RMSD (Å) for the alpha carbon of the ZIKV protease in complex with the dipeptide inhibitor	90
Figure 45. DSSP plot of the secondary structure of the ZIKV protease with the dipeptide inhibitor bound along the simulation time	90
Figure 46. Free energy profile for the reaction of charged tetrahedral adduct relative to the noncovalent Michaelis complex using the combined reaction coordinates at the PM3/ff14SB level of QM/MM theory	91
Figure 47. Free-energy profile for formation of the charged tetrahedral adduct relative to the noncovalent Michaelis complex for the reaction between the ZIKV protease and dipeptidyl aldehyde inhibitor using the combined reaction coordinates at the PDDG-PM3/ff14SB level of QM/MM theory	92
Figure 48. Representative structures from the free-energy reaction path at the PDDG-PM3/ff14SB level of (I) noncovalent Michaelis complex, (II) transition state and (III) tetrahedral adduct in the active site of ZIKV protease	94
Figure 49. Distance between oxygen atom (O_{γ}) of the tetrahedral adduct and H_{2H51} atom during 1 ns QM/MM MD simulation of the ZIKV protease bound with the dipeptide inhibitor	96
Figure 50. Potential energy profiles for the reaction of the tetrahedral adduct formation calculated at the BH&HLYP-D3/6-31G(d)/ff14SB (black), BH&HLYP-D3/6-311+G(d)/ff14SB (red), MP2/(aug)-cc-pVTZ/ff14SB (blue), SCS-LMP2/(aug)-cc-pVTZ/ff14SB (magenta), and LCCSD(T)/(aug)-cc-pVTZ/ff14SB (green) QM/MM levels	99

Figure 51. Free energy profile for the reaction of protonated hemiacetal adduct (relative to the reactant state, tetrahedral adduct) using reaction coordinate involving the direct proton transfer between H51 and oxygen atom of the anionic hemiacetal 100

Figure 52. QM/MM (BH&HLYP-D3/6-31G(d)/ff14SB) optimized structures of (I) noncovalent Michaelis complex, (II) transition state, and (III) covalent tetrahedral adduct in the active site of ZIKV protease..... 102

Figure 53. Climbing image NEB profiles compared to adiabatic mapping (black circles, adiabatic mapping profile; grey triangles, CI-NEB profile from 7 starting images)..... 104

Figure 54. Comparison of the potential energy profiles (relative to the reactant) for the reaction of the tetrahedral adduct formation at the BH&HLYP-D3/6-31G(d)/ff14SB (open circle), MP2/(aug)-cc-pVTZ/ff14SB (white circle), SCS-MP2/(aug)-cc-pVTZ/ff14SB (black square) and LCCSD(T)/(aug)-cc-pVTZ/ff14SB (white square) levels of theory (using the BH&HLYP-D3/6-31G(d)/ff14SB CI-NEB geometries)..... 106



LIST OF ABBREVIATIONS

(-) ssRNA	negative-sense single-stranded RNA
(+) ssRNA	positive-sense single-stranded RNA
Ac-DKKOR-ACC	acyl- <i>D</i> Lysine-Lysine-Ornithine-Arginine-7-amino-4-carbamoylmethylcoumarin
Ac-DRKOR-ACC	acyl- <i>D</i> Arginine-Lysine-Ornithine-Arginine-7-amino-4-carbamoylmethylcoumarin
Ac-nKRR-ACC	acyl-Norleucine-Lysine-Lysine-Arginine-7-amino-4-carbamoylmethylcoumarin
AE	acyl enzyme
Bz-nKRR-AMC	benzyl-Norleucine-Lysine-Arginine-Arginine-aminomethylcoumarin
C	capsid
CG	conjugated gradient
CI-NEB	climbing image nudged-elastic band
CpHMD	constant pH molecular dynamics
CpH-REMD	constant pH replica exchange molecular dynamics
DENV	Dengue virus
DENV-2	Dengue virus serotype 2
DFT	density functional theory
dsRNA	double stranded RNA
E	envelope
ER	endoplasmic reticulum
ESP	electrostatic potential
GAFF	generalized AMBER force field
GB	generalized Born
GBS	Guillain-Barré syndrome
HCV	hepatitis C virus

HF	Hartree-Fock
LAs	link atoms
LYP	Lee-Yang-Parr
MC	Monte carlo or Michaelis complex
MD	molecular dynamics
MEP	minimum energy path
MM	molecular mechanics
MNDO	modified neglect of diatomic differential overlap
MP	Møller-Plesset
NMR	nuclear magnetic resonance
NPT	isothermal-isobaric ensemble
NS	nonstructural
NTPase	helicase-nucleoside triphosphatase
PB	Poisson-Boltzmann
PDB	Protein Data Bank
PDDG-PM3	pairwise distance directed Gaussian modification of PM3
PME	particle mesh Ewald
prM/M	premembrane/membrane
QM	quantum mechanics
QM/MM	quantum mechanics and molecular mechanics
RC	reaction coordinate
RdRp	RNA-dependent RNA polymerase
RESP	restrained electrostatic potential
RMSD	root-mean-square displacement
SASA	solvent accessible surface area
SCC-DFTB	self-consistent charge density functional tight binding
SCS-MP2	spin-component scaled second-order Møller-Plesset

SD	steepest descent
TI	tetrahedral intermediate
TS	transition state
WHAM	weighted histogram analysis method
WNV	West Nile virus
ZIKV	Zika virus



CHAPTER I

INTRODUCTION

1.1 Research rationality

Zika virus (ZIKV), a mosquito-borne flavivirus, was first isolated from sentinel rhesus monkey in the Zika Forest of Uganda in 1947. Originally, it was not considered a relevant human pathogen because the majority of its infections is asymptomatic until the large outbreak starting in Brazil in 2015. The virus is mainly transmitted to humans by *Aedes* species mosquitoes. In addition, the recent epidemiological data have revealed that this relatively harmless virus was now related to serious pathological disorders including Guillain–Barré syndrome (GBS) in adults and microcephaly in newborns. The incidences of GBS and microcephaly associated with ZIKV infection have led the World Health Organization to declare ZIKV infection as a global public health emergency since February of 2016. Unfortunately, there is no either currently available vaccine or antiviral drug for preventing or controlling ZIKV infection; thus, drug design and development of anti-ZIKV agents with effective and safe properties have become an imperative necessity. One of the attractive drug-targets for ZIKV treatment is the viral NS2B/NS3 serine protease, which plays a crucial role during viral replication process. Basically, one approach for the rational design of protease inhibitors begins with the optimization of the substrate sequence, which is possibly transformed into the new effective inhibitors. Therefore, the aims of this research are to investigate the protein-ligand interactions of natural substrates/inhibitors by means of molecular dynamics (MD) simulations and to understand the reaction mechanism of ZIKV NS2B/NS3 serine protease by hybrid Quantum Mechanics and Molecular Mechanics (QM/MM) approach. Molecular understanding from this observation would be very useful for the design of novel protease inhibitors with high specificity and efficiency toward the ZIKV protease.

1.2 Viral structure

As mentioned above, ZIKV mainly transmitted to humans by mosquitoes has emerged as a global public health concern. ZIKV infection is associated with serious neurological complications in adults, and infection in pregnant women is linked to congenital abnormalities, including fetal and newborn microcephaly. ZIKV was first isolated from rhesus monkey cells in 1947 in Uganda [1]. Moreover, it can be further classified into two genotypes, African and Asian, based on phylogenetic analysis [2]. ZIKV belongs to the Genus *Flavivirus* in the Family *Flaviviridae* and has the same genome organization as all flaviviruses, including Dengue virus (DENV), West Nile virus (WNV), yellow fever virus and Japanese encephalitis virus [3]. The ZIKV genome containing positive-sense single-stranded RNA ((+) ssRNA) of approximately 11 kb is translated into a single polyprotein precursor in the cytoplasm of infected host cells. After that, the polyprotein is cleaved and processed by both host and viral proteases into three structural proteins (capsid (C), premembrane/membrane (prM/M), and envelope (E)), and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) at a later stage (**Figure 1**). The structural proteins play a key role in forming the viral particle, while the NS proteins have important functions in viral genome replication and polyprotein processing [4]. It should be noted that the (+) ssRNA genome of ZIKV is enclosed by an icosahedral capsid [5], surrounded by the E protein [6] and prM/M proteins which are embedded to a lipid membrane [7]. In addition, the E protein with high thermostability is possibly related to the virulence of ZIKV by increasing the survival rate of the virus in the unfavorable environments such as saliva, urine and semen [8].

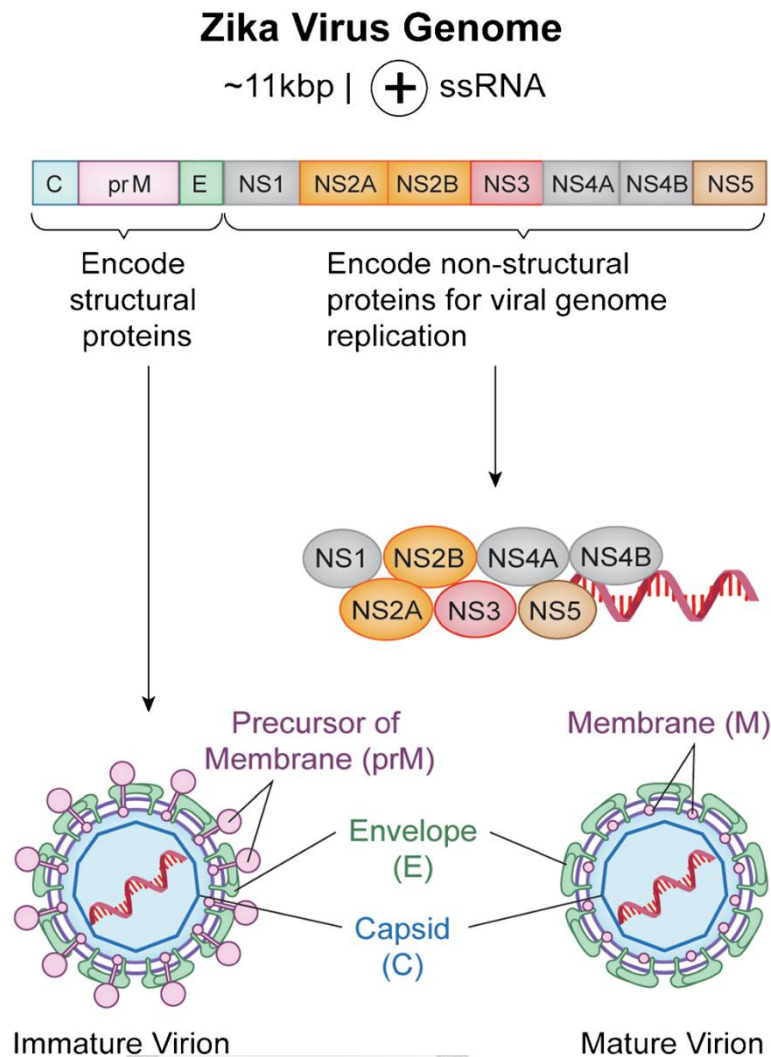


Figure 1. The genome of Zika virus. The single-stranded positive-sense RNA ((+) ssRNA) genome encodes three structural and seven nonstructural (NS) proteins. Structural proteins include capsid (C), precursor of membrane/membrane (prM/M) and envelope (E) proteins. The (+) ssRNA genome is encapsulated by the prM, E and C proteins, forming as an immature virion. Whereas mature virion is formed when the prM protein is cleaved into the M protein. The open reading frame of virus also encodes seven NS proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5), forming the replication complex necessary for viral genome replication [9].

1.3 Flavivirus Life Cycle

At the initial step of the ZIKV infection, the E protein on the mature ZIKV virion binds to the host cell receptors, leading to an uptake of the virus through endocytosis to form an endosome in the infected cell (**Figure 2**) [10]. Due to the acidic

condition inside the endosome, the E protein undergoes conformational change into the form which can fuse with the host endosome membrane and consequently release the RNA genome [11]. The viral (+) ssRNA genome is translated on the endoplasmic reticulum (ER) membrane into a single polyprotein complex, which is further cleaved by viral protease in the cytoplasm and host protease in the ER lumen [12].

As the polyprotein processing progresses, the original viral (+) ssRNA genome is replicated at the ER membrane [13]. Newly produced NS proteins form the replication complex to activate the subsequent replication course. The (+) ssRNA is firstly transcribed into double stranded RNA (dsRNA) through the activity of RNA-dependent RNA polymerase (RdRp) on the C-terminus of NS5 protein [14, 15]. Afterward, the dsRNA is separated and remodeled into (+) and negative-sense (-) ssRNA strands using the helicase on the C-terminus of NS3 protein as a catalyst [16]. Note that the original (+) ssRNA template is longer than the (-) ssRNA, as it contains a methylated group. The (-) ssRNA strand is further transcribed into dsRNA by RdRp enzyme [14]. After that, the dsRNA is unwound [15] and remodeled by NS3 helicase, producing the separated (-) and (+) ssRNA strands in the second round of the process [16]. It is worth mentioning that the resulting (-) ssRNA may be reused as a template to transcribe multiple (+) ssRNA strands. Finally, the newly synthesized (+) ssRNA is methylated by the methyltransferase on the N-terminus of NS5 protein, in preparation for viral particle assembly at the late stage of infection [14].

The immature virions consisting of the methylated (+) ssRNA genome, C, prM and E proteins are formed in the ER [17]. The immature virions move out from the ER to the Golgi apparatus, and then mature in the trans-Golgi network. Due to the acidic condition of the trans-Golgi network, it induces a conformational change of the spiky E-prM heterodimer on the immature virion into a herringbone-like array of E homodimers parallel to the viral surface [18, 19]. This structural reorganization induces cleavage of the prM into M by the host protease (i.e. furin), maturing the viral particle, reducing the viral diameter from ~60 nm to ~50 nm, and inducing exocytosis [20].

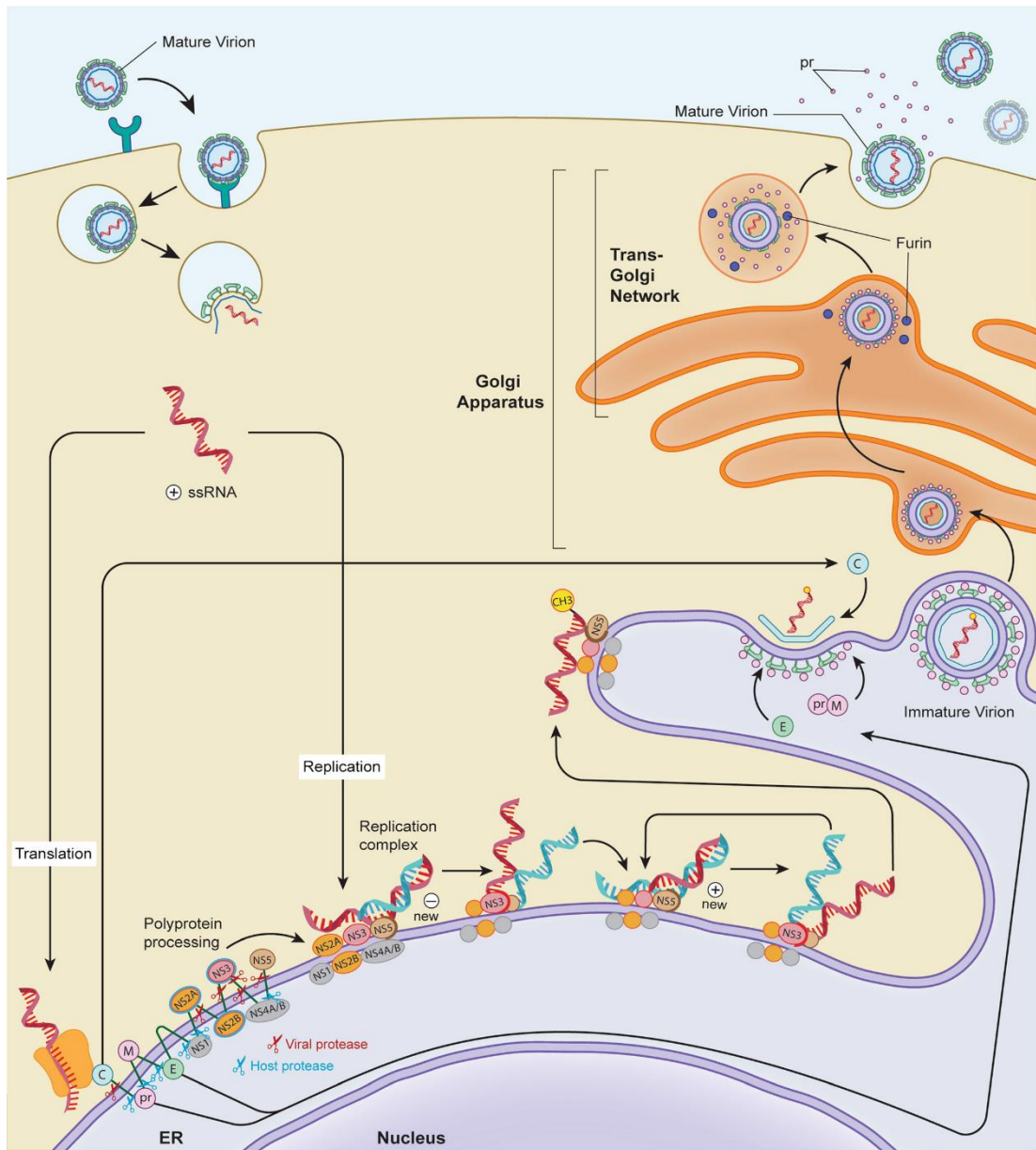


Figure 2. Life cycle of Zika virus. Initially, the mature virion binds to the host cell membrane receptor, and undergoes receptor-mediated endocytosis. Afterward, the viral E protein fuses with the endosomal membrane, resulting in the release of the viral genome into the host cytoplasm. The (+) ssRNA genome is then translated by host ribosomes at the endoplasmic reticulum (ER) membrane. A polyprotein is formed and cleaved by host and viral proteases. After polyprotein processing, seven NS proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) form the replication complex essential for viral genome replication at the ER membrane. RNA dependent RNA polymerase (RdRp) on the C-terminus of NS5 protein transcribes a complementary (–)

ssRNA strand from the original (+) ssRNA genome, forming the dsRNA. The dsRNA is further unwound and remodeled into two separated (-) and (+) ssRNA strands by NS3 helicase. The (-) ssRNA is used as a template to generate new (+) ssRNA by NS5 RdRp, producing another dsRNA strand. This new dsRNA is unwound and remodeled by NS3 helicase, forming separated (-) and (+) ssRNA strands used for another round of replication process. The resulting (-) ssRNA strand may be reused to transcribe several (+) ssRNA products. Alternately, the newly separated (+) ssRNA strand is capped by methyl group using NS5 methyltransferase. The methylated (+) ssRNA gathers at the ER membrane with the C, E and prM proteins to form an immature virion, which moves out of the ER into the Golgi apparatus. The immature virion matures in the trans-Golgi network, where prM protein is cleaved into M protein by furin. The M, E and C proteins enclose the methylated (+) ssRNA, forming a mature virion, then undergoing exocytosis and infecting other cells [9].

1.4 Pathogenesis and Clinical Presentations

ZIKV is mainly transmitted to human by the *Aedes* mosquito vector during blood feeding [21]. Besides blood, ZIKV has been isolated from human urine [22], saliva [23], breastmilk [24], female genital tract [25] and semen [26]. In addition, sexual transmission of ZIKV has been reported [27, 28]. More seriously, ZIKV transmission can also occur during pregnancy, in which ZIKV may cross the placenta and lead to symptoms in the developing fetus [29]. Nevertheless, ZIKV infection is asymptomatic in most cases (75% to 80%) [30]. Of the symptomatic cases, the typical clinical appearances are nonspecific, self-limiting flu-like symptoms together with myalgia (muscle pain), arthralgia (joint pain), fever, headache, rash, conjunctivitis (pink eye), and fatigue (tiredness) [31]. These symptoms are quite similar to other viral infectious symptoms, such as DENV and WNV fever.

There are now growing evidences, indicating that ZIKV infection may be associated with severe neurological problems, such as encephalitis [32], acute myelitis [33], macular atrophy [34], GBS [35], and microcephaly [36]. GBS related to ZIKV infection, first described in the French Polynesian outbreak in 2013, is an acute inflammatory demyelinating polyradiculoneuropathy with transient or permanent flaccid paralysis and loss of motor function [35]. Additionally, GBS can severely cause

death to human due to respiratory failure, pneumonia, sepsis and pulmonary embolism [37]. In the large outbreak in Brazil in 2015, there were many reports, showing that ZIKV infection during pregnancy was linked to microcephaly in the developing fetus. From these information, ZIKV has been confirmed as the causal agent. In addition to microcephaly, other neuronal abnormalities to fetal caused by ZIKV have been determined, and the overall phenotype associated with ZIKV infection can lead to the term of congenital Zika syndrome [38]. Although the severity of neurological complications related to ZIKV infection has been well studied, there are no effective vaccines or antiviral drugs currently available for clinical use [39, 40].

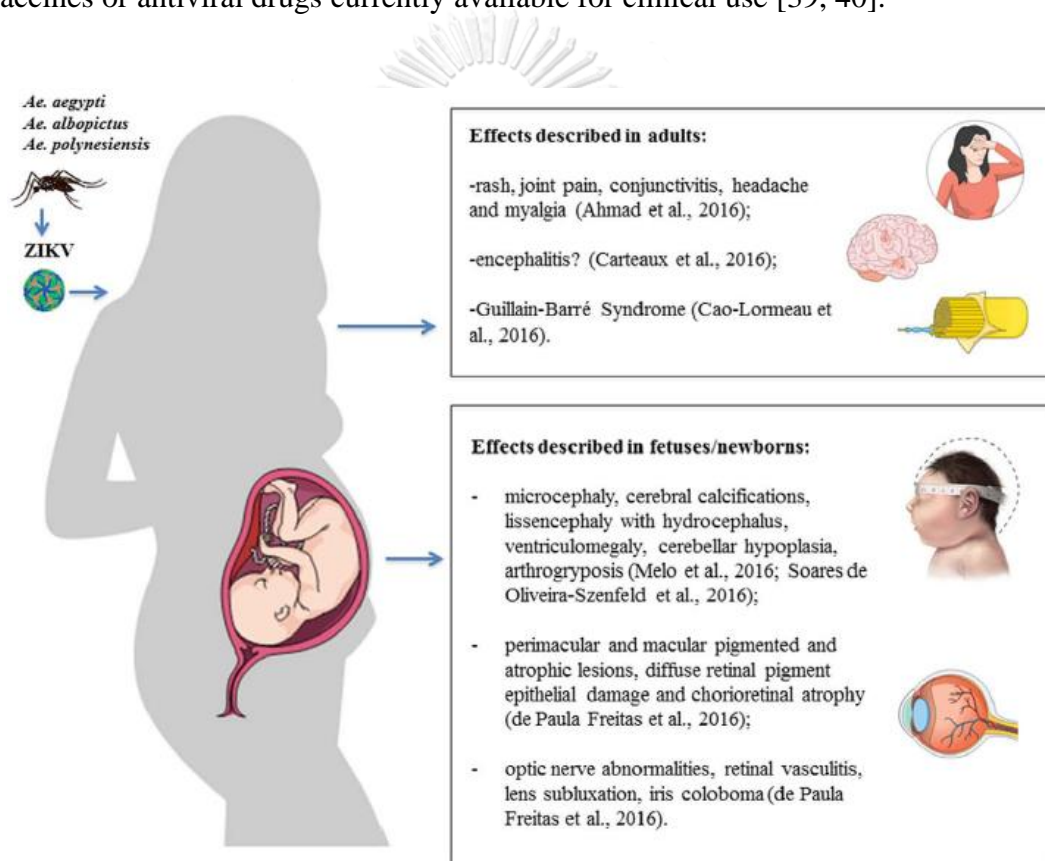


Figure 3. Clinical manifestations and consequences of ZIKV infection. Schematic illustration of the effects described in adults and fetuses/newborns after ZIKV infection [41].

1.5 Flaviviral NS2B/NS3 serine protease

Similar to other flaviviral NS3 proteins, the ZIKV NS3 protein (~70 kDa) consists of two functional domains consisting of a protease domain and a helicase

domain. The C-terminal region of the ZIKV NS3 protein encodes a helicase–nucleoside triphosphatase (NTPase), whereas the N-terminal region of the NS3 protein encodes a chymotrypsin-like serine protease that specifically recognizes substrates containing amino acid residues with basic property such as arginine at the P1 position and lysine at the P2 position [42, 43]. Based on the common structures of flavivirus proteases, the active form of the ZIKV protease composes of the N-terminal protease domain of NS3 protein, which contains the catalytic residues, known as catalytic triad (H51, D75 and S135), and the membrane-bound NS2B protein. Assembly of NS3 protease with the membrane-bound NS2B rearranges the NS3 active site to yield the optimal structure for enzyme catalysis, as has observed for related flavivirus proteases [44, 45]. The two-component NS2B/NS3 serine protease is necessary for viral polyprotein processing to produce mature structural and NS proteins; hence, this enzyme is a particularly promising flavivirus therapeutic target for the design of novel antiviral agents [46-48].

Recently, several research groups have reported the X-ray structures of the NS3 protease domain of ZIKV complexed with the membrane-bound NS2B (~40 amino acids), covalently bonded through a glycine-rich linker to the N-terminus of NS3, which shows strong peptidolysis [42, 49]. Interestingly, a peptidomimetic boronic acid inhibitor (cn-716) covalently linked to the ZIKV protease induces a conformational change in NS2B protein structure, showing a more compact “closed” form, with the NS2B polypeptide wrapped around NS3 [42]. In addition, a high-resolution of 1.8 Å crystal structure of ZIKV NS2B/NS3 protease bound with the last four amino acids T¹²⁷-G¹²⁸-K¹²⁹-R¹³⁰ (TGKR) of the NS2B protein has also been solved (**Figure 4**) [43]. This “TGKR” proteolytic product could bind to S1, S2 and only partially to S3 pockets of the ZIKV protease through mainly charge-charge and hydrogen bonding interactions. More recently, the unlinked ZIKV NS2B/NS3 protease construct in apo state and bound form with a reversely oriented peptide has been produced and investigated by X-ray crystallography [50]. The closed conformation of this construct suggests that NS2B is necessary for NS3 to generate a substrate-binding pocket. This is distinguished from the relaxed or open form, as found in the NS2B-(linker)-NS3 construct [49] and as has been reported for other flavivirus NS2B/NS3 proteases [45]. Furthermore, it was observed from the study reported in ref 50 that four amino acids K¹⁴-K¹⁵-G¹⁶-E¹⁷ (KKGE) from the neighboring NS3 could bind to the ZIKV protease

and resist proteolysis. Altogether, this information has provided a good starting point for further design of anti-ZIKV agents with high specificity and efficiency.

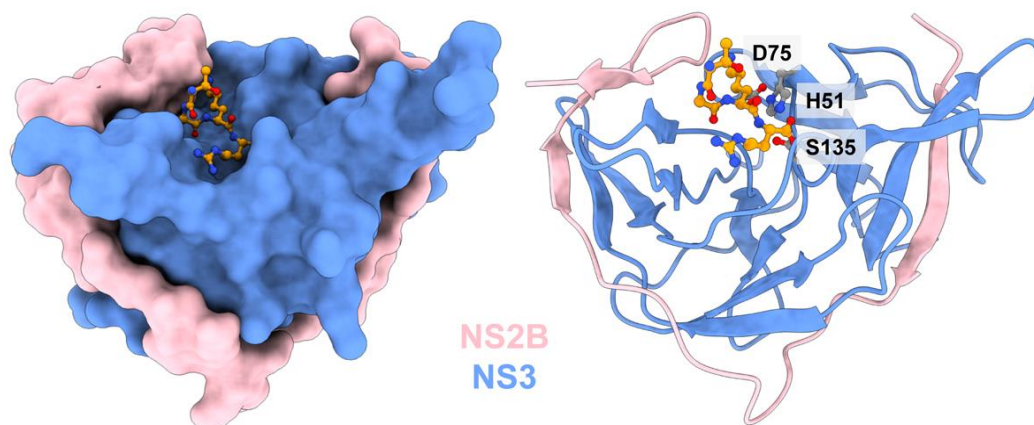


Figure 4. Three-dimensional structures of the ZIKV protease in complex with the tetrapeptide “TGKR” of the NS2B domain (PDB ID: 5GJ4). The catalytic triad (D75, H51 and S135) and the co-crystallized tetrapeptide (orange) are shown.

1.6 Reaction mechanism of flaviviral NS2B/NS3 serine protease

The NS2B/NS3 serine protease exhibits a favorable cleavage site of a peptide bond within a peptide substrate, which has an amino acid sequence generally represented by -P4-P3-P2-P1-P1'-P2'-P3'-P4'-. Note that the peptide bond is cleaved at the P1 and P1' boundary by the protease (**Table 1**). With the structural similarities of flaviviral proteases to other proteins in the same family [51], it has been proposed that these enzymes should present a similar catalytic mechanism. Theoretical and experimental evidence provide a common reaction mechanism for chymotrypsin-like serine proteases involved in two steps; acylation and deacylation processes [52-54]. In the acylation process, the imidazole ring of H51 is in the position to deprotonate the nucleophilic hydroxyl group of S135 in the Michaelis complex (MC), and the D75 side chain is in the orientation to stabilize the positively charged H51 after accepting the proton from S135. Subsequently, the peptide bond is cleaved by a nucleophilic attack on the substrate carbonyl carbon by the oxygen atom on the hydroxyl group of S135, producing a tetrahedral intermediate (TI) formation that is stabilized by hydrogen bonds with some residues located in the oxyanion hole (formed by the backbone amides of G133 and S135). The peptide bond is then cleaved, releasing the N-terminus of the

peptide substrate and acyl enzyme (AE) (**Figure 5**). The deacylation process involves the hydrolysis reaction on the AE, leading to a liberation of the C-terminus of the substrate and the catalytic residues of the protease. According to the reaction mechanism proposed for serine proteases, the acylation process is considered as the rate-limiting step, in which the reaction rate is determined by formation of a relatively stable TI [53, 55, 56]. In addition, for most serine proteases, the acylation process is known as the rate-limiting step for hydrolysis of amide bonds, while the deacylation process is rate-limiting step for hydrolysis of ester bonds [57-59].

Table 1. The cleavage sites between P1 and P1' junction for each substrate.

Substrate	Peptide positions							
	P4	P3	P2	P1	P1'	P2'	P3'	P4'
Capsid C	V	T	R	R	G	S	A	Y
NS2A/NS2B	S	G	K	R	S	W	P	P
NS2B/NS3	T	G	K	R	S	G	A	L
NS3/NS4A	A	G	K	R	G	A	A	F
NS4B/NS5	V	K	R	R	G	G	G	T

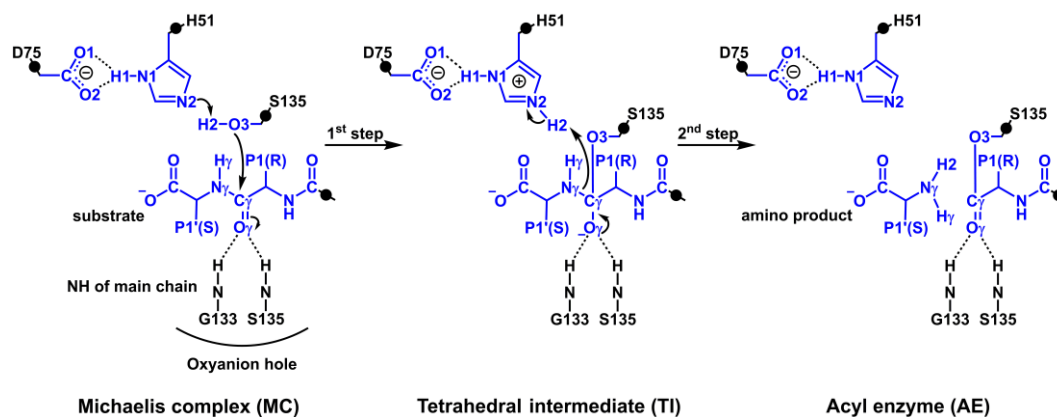


Figure 5. Proposed reaction mechanism for the reaction of the acylation process catalyzed by the ZIKV protease. This process is involved in proton transfer and nucleophilic attack, leading to the formation of a tetrahedral intermediate (TI), and thereafter the peptide bond breaking to produce acyl enzyme (AE). The QM region (labeled in blue) consists of a fragment of the catalytic residues (H51, D75 and S135) of the NS3 protease and its substrate. Four hydrogen link atoms are displayed in black circles. The atomic labels involved in the reaction mechanism are also given.

1.7 Inhibition mechanism of flaviviral NS2B/NS3 protease by peptidyl aldehyde inhibitors

Interestingly, protease inhibitors have been successful for the treatment of human immunodeficiency virus and hepatitis C virus (HCV) infections, for which several peptidomimetic inhibitors are currently in clinical use [60]. Highlighting this point is the fact that the substrate-binding site in HCV and flavivirus proteases is relatively shallow and thus not easy to bind to small-molecule compounds. Moreover, the flaviviral proteases is preferential to recognize dibasic or polybasic substrate sequences at enzyme recognition sites [61-63]. Of the highly anionic nature at the protease active site, the recognition motifs in inhibitors should present basic or polar characters, which could lead to the limitation of cell membrane permeability [64-67]. Even though it is challenging to develop peptidomimetic inhibitors into drug-like molecules targeting the protease active site with basic recognition preferences, the NS2B/NS3 protease is well conserved across the flaviviruses and no mutations have been reported in the ZIKV protease sequence, making it a promising viral target for the development of inhibitors with broad activity [68]. Fortunately, some successful

examples for development of inhibitors toward other proteases with basic recognition preferences (e.g. thrombin and factor Xa) have been introduced [69, 70], and thus may be valuable for the development of therapeutically effective inhibitors of flavivirus proteases.

Many studies have focused on the design of peptidomimetic inhibitors. The rationale designs include optimization of the P1 and P2 basic residues through natural or non-natural building blocks to enhance binding affinity [65, 71], incorporation of the C-terminal electrophilic group to react with catalytic residue in the active site [64, 66, 72, 73], and addition of a suitable cap to the N-terminal improve inhibitory activity [65, 74-76]. Peptidyl aldehydes are powerful reversible inhibitors that showed high inhibitory potency against DENV serotype 2 (DENV-2) [73] and WNV [64, 77] proteases. For instance, cationic tripeptidyl aldehyde (phenacetyl-KKR-aldehyde) derived from -P1-P2-P3- residue positions of substrate was demonstrated to be active toward WNV protease. In addition, it was stable in serum and cell permeable, as well as showed antiviral activity against WNV without detected cytotoxicity [64]. Recently, dipeptidyl aldehyde (acyl-KR-aldehyde) derived from P1 and P2 of ZIKV protease substrate has been shown to be a potent competitive inhibitor toward the ZIKV protease. Moreover, the X-ray crystal structure revealed that the aldehyde moiety on this inhibitor could form a covalent bond with the catalytic S135 of NS3, inducing the closed conformation of the ZIKV protease [78, 79] (**Figure 6**). Nuclear magnetic resonance (NMR) studies also demonstrated that the ZIKV protease in complex with this dipeptide inhibitor formed a stable complex in solution, which is in line with the X-ray crystal structure [78]. From this point of view, these information are very useful for further design of potent protease inhibitors. Nonetheless, the clinical potential of peptidyl aldehydes is still uncertain, but it is important for understanding the molecular recognition and inhibition mechanism toward the flavivirus proteases [66, 80, 81].

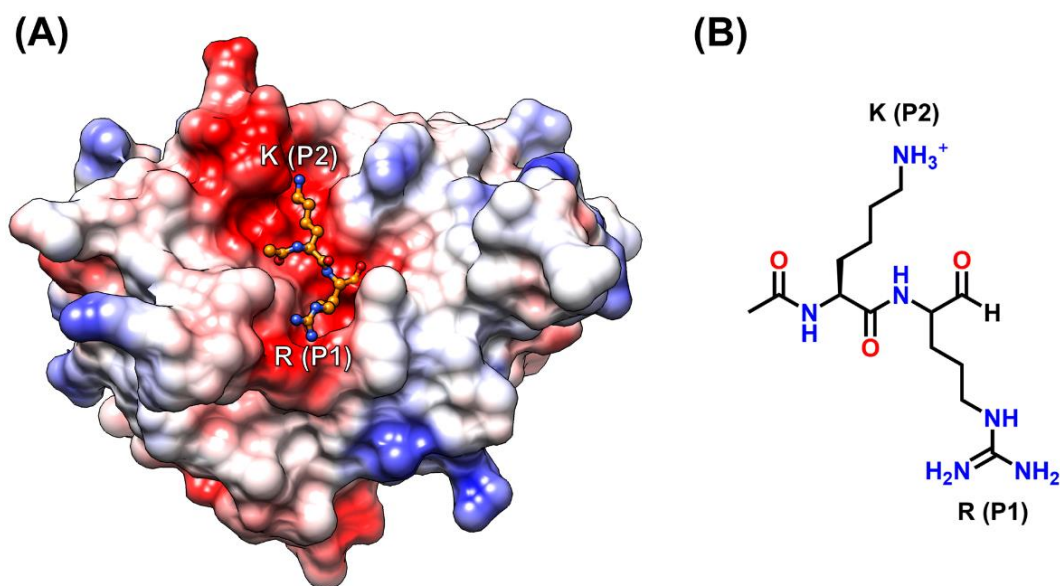


Figure 6. (A) Three-dimensional structure of the ZIKV NS2B/NS3 serine protease in complex with a dipeptide inhibitor (acyl-KR-aldehyde) (PDB ID: 5YOF [79]), where surface charge of the ZIKV protease structure was calculated using coulombic surface coloring with default setting implemented in UCSF Chimera [82], shading from red (high negative charge) to blue (high positive charge). (B) Two-dimensional chemical structure of a dipeptidyl aldehyde inhibitor used in this study.

As mentioned above (**section 1.6**), the general cleavage mechanism of substrate by serine proteases has been proposed comprising two steps; acylation and deacylation processes [53, 54, 83]. The acylation process is involved in the formation of a transient anionic tetrahedral intermediate and a covalent acyl-enzyme, while the deacylation step is involved in the hydrolysis reaction on acyl-enzyme to liberate the C-terminus of the substrate. Peptidyl aldehydes are reversible inhibitors covalently bonded to the catalytic S135 of flavivirus proteases [72, 80, 81], in which a covalent tetrahedral adduct or known as hemiacetal formation is formed (**Figure 7**). This covalent tetrahedral adduct mimics the tetrahedral transition state on the protease reaction pathway. The peptidyl aldehydes are thus considered as transition-state analogues [84]. Although the inhibitory potency of the peptidyl aldehydes has been assessed experimentally among several flaviviruses [72, 73, 77, 78], mechanistic details and energetics of the inhibition mechanism between ZIKV protease and the peptidyl aldehydes, that can be revealed computationally, remain unknown. Here, we explored the activation barriers, the

reaction energies and the key interactions corresponding to stationary points, including noncovalent Michaelis complex, transition state, and covalent complex, which would be helpful for the design of other transition state analogues with desirable properties.

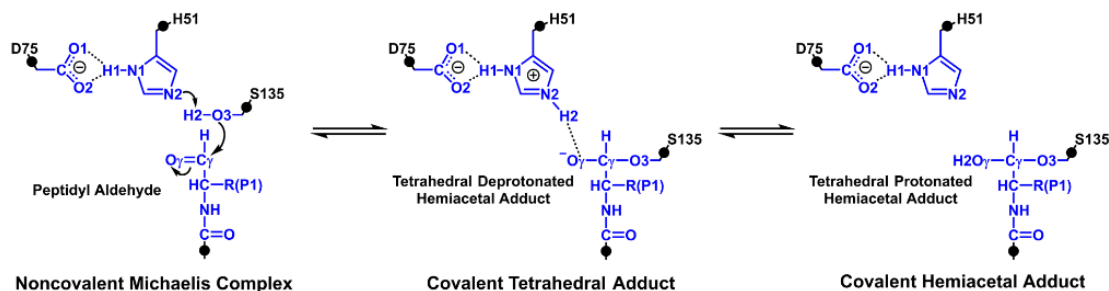
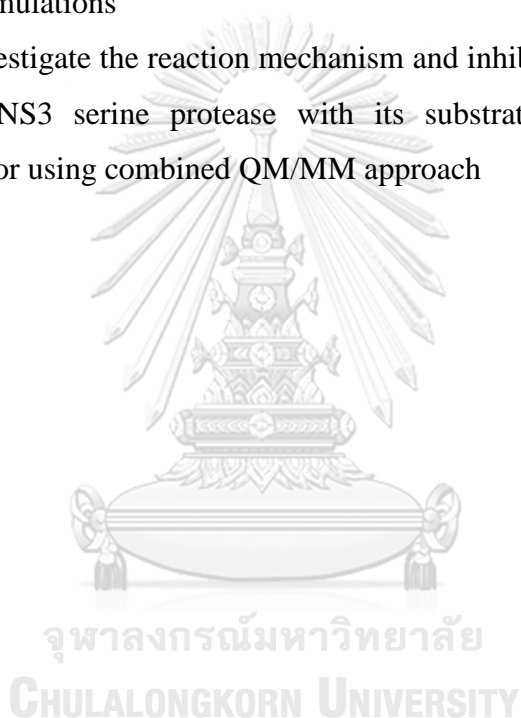


Figure 7. Proposed reaction mechanism of the ZIKV protease inhibition by peptidyl aldehyde. The process is involved in proton transfer and nucleophilic addition reaction to produce a covalent anionic hemiacetal adduct. The relevant atoms in QM region are labeled in blue comprising fragments of the catalytic triad (H51, D75 and S135) of the NS3 protease and the peptidyl aldehyde as well as four hydrogen link atoms (black circles) applied to saturate the QM/MM boundary. The atomic labels involved in the inhibition mechanism are also given for further discussion.

1.8 Objectives of the study

In this research, the accepted computational chemistry approaches such as molecular dynamics simulation (MD) and hybrid quantum mechanics/molecular mechanics (QM/MM) were selected to investigate the molecular interactions and catalytic mechanism of the ZIKV protease with its substrate and peptidyl aldehyde inhibitor. The aims of this work cover:

1. To determine the protein-ligand interactions, structure and dynamic properties of the ZIKV NS2B/NS3-substrate/inhibitor complexes using MD simulations
2. To investigate the reaction mechanism and inhibition mechanism of ZIKV NS2B/NS3 serine protease with its substrate and peptidyl aldehyde inhibitor using combined QM/MM approach



CHAPTER II

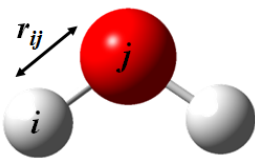
THEORIES

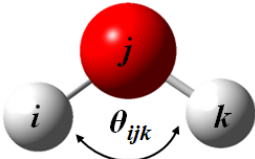
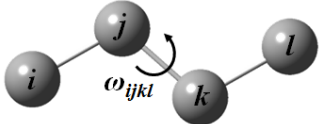
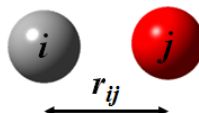
2.1 Molecular mechanics

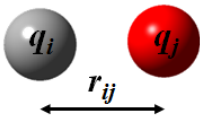
Molecular mechanics (MM) is a classical mechanics-based model, in which molecular structure is described by a series of atoms (point charges) connected to bonds (springs). The types of atoms and bonds are identified in terms of atomic size and bond length, respectively. MM framework is based on empirical approximations, called “force field”, which are used for calculating interactions and evaluating the potential energy of the systems as a function of atomic coordination. All common force fields can be classified potential energy functions into two classes: Bonded and non-bonded interactions. Bonded interactions compose of covalent bond stretching (E_{bond}), angle bending (E_{angle}), and torsional rotation (E_{dihedral}), all which are generally fixed throughout a simulation. Meanwhile, non-bonded interactions comprise van der Waals (E_{vdW}) and electrostatic (E_{ele}) interactions. The definition and corresponding equations of each term are summarized in **Table 2**. The total potential energy (E_{total}) of a molecule [85] is a summation of each interaction term as following equation:

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{vdW}} + E_{\text{ele}} \quad (1)$$

Table 2. Force field in molecular mechanics

Interaction	Definition and equations
Bonded terms:	
Bond stretching 	<p>The covalent bond stretching is a bonded interaction between two atoms i and j. The bond stretching energy on each pair of atoms is based on Hooke's law for a spring in a harmonic form as following equation:</p> $E_{\text{bond}} = \sum_{\text{bond}} \frac{1}{2} k_b (r - r_0)^2 \quad (2)$

Interaction	Definition and equations
	<p>where k_b is the force constant for bond stretching; r and r_0 are the actual and equilibrium bond length, respectively.</p>
<p>Bond bending</p> 	<p>The bond angle interaction is the interaction between three atoms i, j, and k. Similar to the bond length energy term, the bond angle energy is based on Hooke's law with respect to angle as following equation:</p> $E_{\text{angle}} = \sum_{\text{angle}} \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 \quad (3)$ <p>where k_{θ} is the force constant for bond bending; θ and θ_0 are the actual and equilibrium bond angle, respectively.</p>
<p>Dihedral</p> 	<p>The dihedral-angle interaction is the interaction between four atoms i, j, k, and l. The torsional angle energy can be written as following equation:</p> $E_{\text{dihedral}} = \sum_{\text{dihedral}} \frac{V_n}{2} (1 + \cos(n\omega - \delta)) \quad (4)$ <p>where V_n is the rotation barrier height; n is the rotate periodic; ω is the torsional angle; δ is the conformation of molecule in phase.</p>
<p>Non-bonded terms:</p>	
<p>van der Waal</p> 	<p>The van der Waal interactions are commonly used to describe steric interactions of molecule. Basically, the Lennard-Jones potential which represents the summation of attractive and repulsive terms is used to evaluate the van der Waal interactions as expressed below.</p> $E_{\text{vdw}} = \sum_{i < j}^{\text{atoms}} 4\epsilon \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right) \quad (5)$

Interaction	Definition and equations
	<p>where ε is a constant determining the depth of the potential well; σ is the diameter of the particle; The term $(\sigma/r_{ij})^{12}$ reflects a strong repelling force at very short distances, whereas the term $(\sigma/r_{ij})^6$ indicates an attracting force with a longer interaction range.</p>
<p>Electrostatics</p>  <p>The diagram shows two spheres representing atoms. The left sphere is grey and labeled q_i. The right sphere is red and labeled q_j. A double-headed arrow between them is labeled r_{ij}.</p>	<p>The electrostatic energy is related to the electrostatic interactions between two distributed charged atoms. The electrostatic interactions are described by Coulombic potential function as following equation:</p> $E_{\text{ele}} = \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (6)$ <p>where q_i and q_j are the fractional charges on atoms i and j, respectively; r_{ij} is the distance between atoms i and j; ε_0 is the dielectric constant.</p>

2.2 Molecular dynamics

Molecular dynamics (MD) is a form of computer simulation technique which is widely used to describe equilibrium and dynamics properties of complex system that cannot be calculated analytically. In this technique, atoms and molecules are allowed to interact for a period of time, providing a view of the particles movement. MD can be a powerful research tool for scientists to study the motion of individual atoms in a case which is impossible in laboratory experiments [86].

MD simulation has been considered as “statistical mechanics by numbers” and “Laplace’s vision of Newtonian mechanics” for prediction of future by animating nature’s forces. This provides detailed insight into molecular motion at atomic level. MD simulation represents an interface between theoretical information and laboratory experiments, known as a virtual experiment. It combines the relationship between molecular structure, motion and function. All classical MD methods depend on force

field for calculating interactions and evaluating the potential energy of the systems as a function of atomic coordination. A force field consists of both sets of equations used to calculate the potential energy and forces from particle coordinates, including a collection of parameters used in the equations [87] (see **Molecular mechanics section**). MD simulation is a very powerful technique in modern molecular modeling, and enables us to follow and understand structure and dynamics of biological molecules with extreme details. This method is based on the calculation of the time dependent behavior of a molecular system. MD simulations are able to provide theoretical information in terms of fluctuations and conformational changes of biological systems including proteins, carbohydrate, fatty acids, and nuclei acids. Moreover, this method is commonly used to determine the structures, dynamical behavior, and thermodynamics properties of biological molecules and their complexes taken from X-ray crystallography and from NMR experiments. In addition, biological systems exhibit a wide range of time scales for dynamics as shown in **Figure 8**

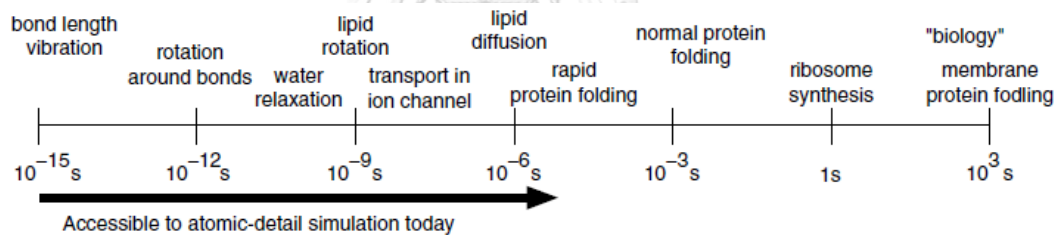


Figure 8. Range of time scales for dynamics in biological systems.

Classical MD simulation is an important method used to investigate the microscopic behaviors by integrating the motion of particles or molecules based on the Newton's second law equation of motion. Fundamentally, the acceleration of particles can be calculated by the first-order derivative of velocity (v_i) or second-order derivative of the atomic position (r_i) with respect to time t (Eq. 7)

$$F_i = m_i a_i = m_i \frac{dv_i}{dt} = m_i \frac{d^2 r_i}{dt^2} \quad (7)$$

where F_i is the total force exerted on the particle i , m_i and a_i are mass and acceleration of the particle i at given time t , respectively. In addition, the external force acting on the particle i can be obtained from the negative gradient of potential energy (U) as shown in Eq. 8

$$F_i = -\nabla_i U \quad (8)$$

Combination of all mentioned equations leads us to obtain (Eq. 9)

$$F_i = -\nabla_i U = m_i a_i = m_i \frac{dv_i}{dt} = m_i \frac{d^2 r_i}{dt^2} \quad (9)$$

The total potential energy is the summation between the bonded and non-bonded interactions, as shown in Eq. (1). In bonded term, the mechanical molecular model considers atoms as spheres and bonds as springs. The mathematics of spring deformation can be used to describe the ability of bonds to stretch, bend, and twist. In non-bonded term, intermolecular interactions consist of van der Waals attraction, steric repulsion and electrostatic attraction/repulsion.

In molecular system, the potential energy is the summation of MM energies (i.e., bonded and non-bonded interactions), as previously described. Again, bonded interactions compose of covalent bond-stretching, angle-bending, and dihedral angle potentials as described by harmonic oscillator function. Meanwhile, non-bonded interactions consist of electrostatic and van der Waals interactions, which are described by coulomb potential and Lennard-Jones potential, respectively. The sum of potential energy interactions [85] can be expressed as follows (Eqs. 10-11):

$$U = (E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}})_{\text{bonded}} + (E_{\text{ele}} + E_{\text{vdW}})_{\text{nonbonded}} \quad (10)$$

$$U = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_{eq})^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{1}{2} V_n (1 + \cos(n\phi - \gamma)) + \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{D r_{ij}} + \sum_{i < j}^{\text{atoms}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (11)$$

Currently, MD simulation can be achieved by several free or commercial software packages (e.g. AMBER, GROMACS, NAMD and CHARMM). A simplified scheme of the standard MDs algorithm is schematically illustrated in **Figure 9**.

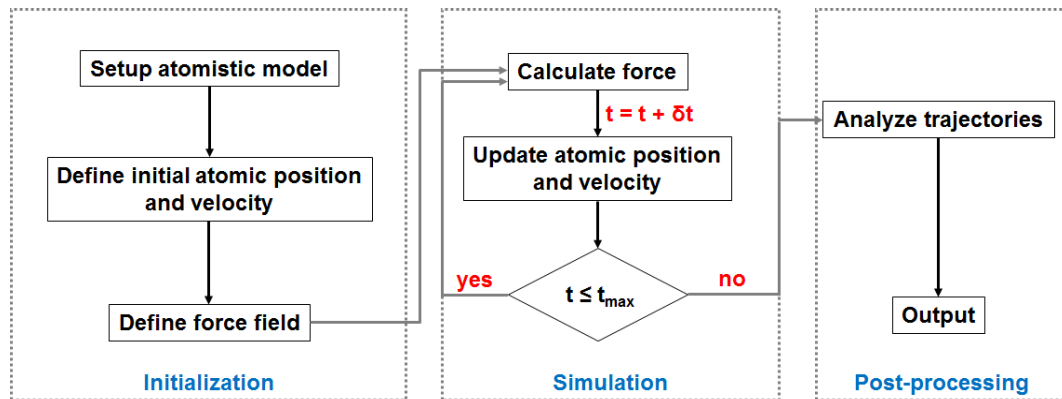


Figure 9. Simplified schematic of MD algorithm.

2.2.1 Integration algorithm

MD simulation generates the structures along the simulation time yielding trajectories of the system. The integration algorithm is needed to calculate the movement of atoms through a simulation time step. The integration algorithms include Verlet, Leap frog and velocity Verlet, and so on. The Leap frog algorithm [88] is widely used as simple and numerically stable of integration algorithm. Additionally, this integration algorithm implemented in AMBER program was used throughout this research as described below. The velocities are calculated at a half of time step ($\delta t/2$) to improve the accuracy of simulation (Eqs. 12-13).

$$v\left(t + \frac{\delta t}{2}\right) = v(t) + \left(\frac{dv}{dt}\right)\left(\frac{\delta t}{2}\right) + \frac{1}{2}\left(\frac{d^2v}{dt^2}\right)\left(\frac{\delta t}{2}\right)^2 \quad (12)$$

$$v\left(t - \frac{\delta t}{2}\right) = v(t) - \left(\frac{dv}{dt}\right)\left(\frac{\delta t}{2}\right) + \frac{1}{2}\left(\frac{d^2v}{dt^2}\right)\left(\frac{\delta t}{2}\right)^2 \quad (13)$$

After subtraction of the two above equations, the leap frog equation in terms of velocity and position are obtained in the similar way (Eqs. 14-15).

$$v\left(t + \frac{\delta t}{2}\right) = v\left(t - \frac{\delta t}{2}\right) + a(t)\delta t = v\left(t - \frac{\delta t}{2}\right) + \left(\frac{F}{m}\right)\delta t \quad (14)$$

$$r(t + \delta t) = r(t) + v\left(t + \frac{\delta t}{2}\right)\delta t \quad (15)$$

2.2.2 Periodic boundary condition

To avoid surface artifacts, periodic boundary condition is used for typical biomolecular simulations because a molecule that exits to the left of periodic box can reappear on the right (**Figure 10**). A molecule will not significantly interact with their periodic replicates if the box is sufficiently large. This is associated with the non-bonded interactions, which ideally should be summed over all neighbors in the resulting infinite periodic system. The simple cut-offs can effectively work for Lennard-Jones potential that decay very rapidly, while Coulomb interactions can lead to large errors if a sudden cut-off is applied. The use of particle mesh Ewald summation (PME) is a good option to calculate the infinite electrostatic interactions by splitting the summation into short-range and long-range parts. In PME summation method, the long-range part is treated by assigning charges to a grid that is solved in reciprocal space through Fourier transforms [89].

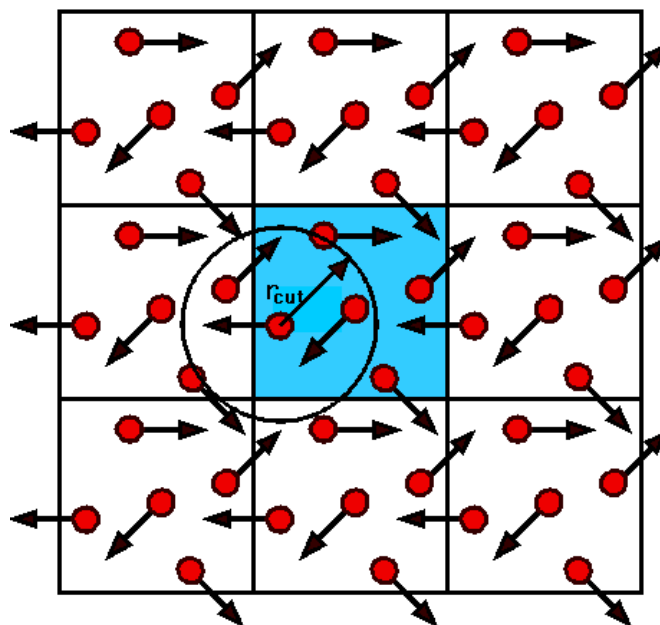


Figure 10. Schematic representation of periodic boundary condition.
 (<http://wiki.cs.umt.edu/classes/cs477/images/e/e3/Pbc.png>)

2.2.3 Constraint dynamics

One of the most demanding part of MD simulations is the computation of non-bonded interactions, because millions of pairs must be calculated for each time step. Extending the time step is an important way to improve simulation performance. Unfortunately, errors are occurred in bond vibrations already at 1 fs. However, the bond vibrations are not of interest in most simulations; thus, it can be removed entirely by introducing bond constraint algorithms such as the SHAKE or LINCS algorithms. Constraints allow to use a larger time steps of 2 fs, and fixed-length bonds are likely better approximations of the quantum mechanical grounds state than harmonic springs [90].

2.2.4 Water models

The most important solvent in nature is water. Nevertheless, it has many special properties due to the ability of water molecules to form hydrogen bonds with other water molecules or other molecules in three-dimensional networks. In addition, water molecules are necessary to include in computational simulations to get more accurate results and to compare with experimental data. Three-dimensional model of water

molecules is developed from the important parameters of the macroscopic properties (e.g. geometry, density, hydrogen bond, atomic charges and Lennard-Jones interaction). For the simple water model, water molecule is treated as rigid molecule and only involved in non-bonded interactions. The non-bonded interactions rely on the electrostatic interaction based on Coulomb's law and van der Waal interaction using the Lennard-Jones potential as expressed below:

$$E_{ab} = \sum_i \sum_j^{\text{on a on b}} \frac{k_c q_i q_j}{r_{ij}} + \frac{A}{r_{oo}^{12}} - \frac{B}{r_{oo}^6} \quad (16)$$

where k_c is the electrostatic constant; q_i and q_j are the partial atomic charges; r_{ij} is the distance between two atoms or dummy charged sites (e.g. lone pairs); A and B are the Lennard-Jones parameters.

It should be noted that the Lennard-Jones term is only treated with oxygen-oxygen interaction. Furthermore, the exact parameters of model geometries (i.e. O-H distance and H-O-H angle) are based on each water model. There are many types of water models such as 3-, 4-, 5- and 6-site depending on the number of interacting sites (**Figure 11**). For instance, the 3-site water models consisting of three interaction points over three atoms of a water molecule is widely used in MD simulations such as TIP3P, SPC and SPC/E water models, balancing accuracy and computational cost.

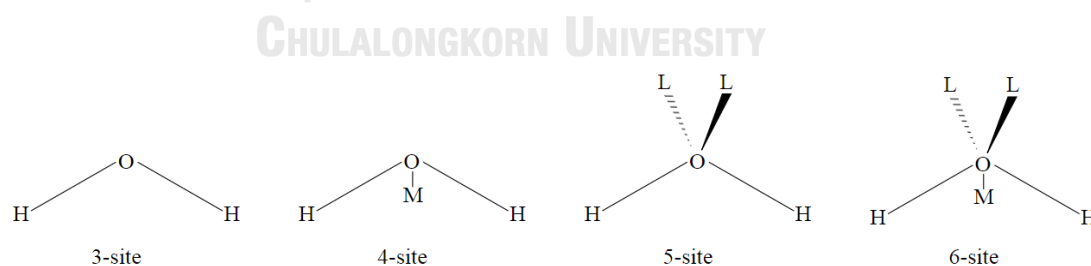


Figure 11. Type of water molecule models generated by 3-, 4-, 5- and 6-site of point charges. (https://en.wikipedia.org/wiki/Water_model)

2.2.5 Continuum solvation

To reduce computational expense relating all-atom MD simulations, continuum solvation or implicit solvent models are used instead of individual “explicit” solvent

molecules based on different water models. The continuum model is frequently used to roughly calculate free energy interaction between solute and solvent. This model treats the solvation as an electrostatic interaction of dielectric media and a solute molecule using a specified dielectric constant to represent the average properties of the real solvent, which has a specific value for each solvent. The total solvation free energy (ΔG_{solv}) is the summation of the nonpolar ($\Delta G_{\text{solv}}^{\text{nonpolar}}$) and electrostatic ($\Delta G_{\text{solv}}^{\text{ele}}$) terms (Eq. 17):

$$\Delta G_{\text{solv}} = \Delta G_{\text{solv}}^{\text{nonpolar}} + \Delta G_{\text{solv}}^{\text{ele}} \quad (17)$$

The implicit solvation states that the continuous medium can be disrupted by a cavity formation of a solute molecule. The favorable dispersion interaction and the unfavorable cavity formation between solute and solvent environment are typically assigned as the nonpolar solvation free energy. The $\Delta G_{\text{solv}}^{\text{nonpolar}}$ contribution can be approximated by the linear function of the solvent accessible surface area (SASA) approach (see Eq. 18). The SASA is evaluated by an exact radius of solvent as a probe rolling on the van der Waals surface of a solute molecule, whereas the $\Delta G_{\text{solv}}^{\text{ele}}$ can be calculated with the two well-known approaches: Poisson-Boltzmann (PB) and generalized Born (GB) models, which are described in more details below.

$$\Delta G_{\text{solv}}^{\text{nonpolar}} = \Delta G_{\text{cavity}} + \Delta G_{\text{dispersion}} \approx \gamma \text{SASA} + \beta \quad (18)$$

where γ and β are surface tension and offset values, respectively.

2.2.5.1 Poisson-Boltzmann (PB) model

For the Poisson's equation, the electrostatic potential, $\varphi(r)$, is computed as a function of charge distribution, $\rho(r)$, and position-dependent dielectric constant, $\varepsilon(r)$, as expressed in Eqs. 19-20.

$$\nabla \varepsilon(r) \nabla \varphi(r) = -4\pi \rho(r) \quad (19)$$

$$\nabla^2 \varphi(r) = \frac{-4\pi\rho(r)}{\varepsilon(r)} \quad (20)$$

However, this equation is valid when mobile ions are absent. For the existing of electrolyte in solvation, the Boltzmann distribution is employed to describe the distribution of mobile ions inside the potential field, leading to the so-called Poisson-Boltzmann (PB) equation. The linear PB equation is simplified for adopting in biomolecular simulations as shown in Eq.21, where q_i and n_i are the atomic charge of electrolyte and density of each ion, respectively. k_B is the Boltzmann constant and T is temperature.

$$\nabla \varepsilon(r) \nabla \varphi(r) = -4\pi \left[\rho(r) + \sum_i q_i n_i \exp\left(-\frac{q_i \varphi(r)}{k_B T}\right) \right] \quad (21)$$

For linear PB equation, the $\Delta G_{\text{solv}}^{\text{ele}}$ is the difference of the electrostatic potentials in solution (φ_{sol}) and gas (φ_{gas}) phases, as expressed below:

$$\Delta G_{\text{solv}}^{\text{ele}} = \frac{1}{2} \int \rho(r) (\varphi_{\text{sol}} - \varphi_{\text{gas}}) dr \quad (22)$$

2.2.5.2 Generalized Born (GB) model

The generalized Born (GB) model is a spherical cavity model approximation, which is more simplicity and computational efficiency compared with the PB model. Based on the GB model, each atom in a molecule is represented as a sphere of the Born radius (R_i) and a charge (q_i) which has the lower dielectric constant than environment. The interior atom uniformly defined by a dielectric constant of 1 in gas phase is surrounded by the high dielectric constant of solvent (e.g. $\varepsilon = 80$ of water). The GB model is used to calculate the $\Delta G_{\text{solv}}^{\text{ele}}$, as shown in Eq. 23 below:

$$\Delta G_{\text{solv}}^{\text{ele}} = -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)} \left(1 - \frac{\exp(-\kappa f_{GB})}{\epsilon} \right) \quad (23)$$

where r_{ij} is the distance between atoms i and j , R_i is the effective Born radii, f_{GB} is a smooth function interpolating the atomic radii and distance between pair of atoms (Eq. 24).

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp\left(\frac{-r_{ij}^2}{4R_i R_j}\right) \right]^{1/2} \quad (24)$$

Note that the effect of the salt concentration can be achieved by the Debye-Hückel screening parameter (κ). For a single ion, the atom radius is equal to its van der Waals radius, ρ_i . In pure water ($\kappa = 0$), the $\Delta G_{\text{solv}}^{\text{ele}}$ of a single ion is written below:

$$\Delta G_{\text{solv}}^{\text{ele}} = -\frac{q_i^2}{2\rho_i} \left(1 - \frac{1}{\epsilon} \right) \quad (25)$$

The effective Born radii are approximated as equal to the distance from atom to the molecular surface. The conformational change of molecule can directly affect the effective Born radii, leading to an inaccuracy of the calculation. Moreover, in the case of atom is surrounded by the other atoms, its electrostatics will be screened according to the low dielectric constant of environment, instead of high dielectric of solvent, called as de-screening. The effective radii of each atom can be evaluated from the degree of de-screening. Thus, the different atomic de-screening results in many GB models.

2.3 Investigation of pK_a of a titratable amino acid

The instantaneous pK_a of a titratable group is determined by its electrostatic environment, which is affected by the given conformation of protein and protonation states of other titratable residues. Changes in the conformation of the protein can affect

the electrostatics, which may, in turn, induce a shift in the pK_a of titratable groups. Furthermore, complex formation between protein and small molecules or other proteins can also induce changes in the pK_a values of titratable groups on either binding partner. The most widely used techniques to investigate the pK_a values of titratable amino acids are constant pH molecular dynamics (CpHMD) and constant pH replica exchange molecular dynamics (CpH-REMD) simulations.

2.3.1 Constant pH molecular dynamics

The CpHMD method uses the GB implicit solvent model with periodic Monte Carlo (MC) sampling of protonation states. Electrostatic interactions are modeled through GB potential, which are used to calculate the protonation state energies. At each MC step, a titratable residue and a new protonation state are randomly assigned. The free energy of transition (ΔG) for protonation or deprotonation is calculated as:

$$\Delta G = k_B T (\text{pH} - \text{pK}_{a,\text{ref}}) \ln 10 + \Delta G_{\text{elec}} - \Delta G_{\text{elec,ref}} \quad (26)$$

where k_B is the Boltzmann constant, T is the temperature, pH denotes the specified solvent pH, $\text{pK}_{a,\text{ref}}$ is the pK_a of the reference compound, ΔG_{elec} is the electrostatic energy component of the titratable residue, and $\Delta G_{\text{elec,ref}}$ is the electrostatic component of the transition energy of the chosen reference peptide. The free energy calculated from this equation forms the basis of the application of the Metropolis criterion to determine whether the transition of the protonation state is accepted. If the MC move is accepted, MD is continued with the changed protonation state of the residue; otherwise, the simulation is carried on with the residue in the unchanged protonation state.

2.3.2 Constant pH replica exchange molecular dynamics

Based on this approach, each replica is simulated at a separated pH. In CpH-REMD, adjacent replicas in the pH ladder swap pH with the MC exchange probability for replicas i and j :

$$P_{i \rightarrow j} = \min \{1, \exp[\ln 10(N_i - N_j)(\text{pH}_i - \text{pH}_j)]\} \quad (27)$$

where N_i is the number of titratable protons present in replica i and pH_i is the pH of replica i prior to the exchange attempt.

2.4 Quantum mechanics

One of the main limitations of standard MM potential functions is that it cannot model the breaking and making of chemical bonds and electronic reorganization involved in a chemical reaction. Hence, modelling enzyme-catalyzed reactions need to use a method capable of describing the bond breaking and bond making. Quantum mechanical (QM) (also known as electronic structure or quantum chemical) methods can solve these problems. Such methods model the distributions of electrons in molecules explicitly. They are widely used to study the geometries, electronic structure and reactions of small molecules. A number of different QM methods exist, and many (so called ‘wavefunction-based’ methods) involve finding the solution to the Schrödinger equation. The Schrödinger equation cannot be solved exactly for molecular systems containing more than one electron and hence approximations are required for many-body systems. Different QM methods differ depending on the approximations made, and can be divided roughly into three types: semi-empirical density functional theory (DFT) and ab initio approaches.

Semi-empirical methods are generally used and considered as the least computationally intensive of the QM methods presented here, but they typically exhibit the least accuracy, unless specifically parameterized for a particular property. Such methods can be applied to larger systems than DFT or correlated ab initio methods. They can also be combined with MD simulations. Examples of semi-empirical QM (molecular orbital) methods include those based on the Modified Neglect of Diatomic Differential Overlap (MNDO) approximation, such as AM1, PM3 and PM6. The self-consistent charge density functional tight binding (SCC-DFTB) semi-empirical method is based on DFT and has been shown in many cases to provide geometries and relative energies comparable to DFT and ab initio calculations.

Density functional theory methods are superior to semi-empirical methods and can provide accuracy approaching that of the correlated ab initio methods with the lower computational cost. The basis of DFT is that the ground-state energy of a molecule can be calculated just from a knowledge of the electron density distribution. The density is a function of only three variables and is thereby much simpler than the ab initio wavefunction, a function of $3N$ variables, where N is the number of electrons. However, the exact form of the functional relating the density to the energy remain unknown. Numerous approximate functionals have been developed based on a mixture of trial and error and known limiting features of the exact functional, but there is (as yet) no systematic way to improve them. One popular density functional is B3LYP, termed a ‘hybrid’ functional, in which a degree of Hartree–Fock (HF) exact exchange is mixed with contributions from other functionals including the Becke88 exchange functional and the Lee–Yang–Parr (LYP) correlation functional.

The simplest useful ab initio QM methods apply HF theory, in which it is approximated that each electron’s spatial distribution is not dependent on the instantaneous motion of the other electrons. This approximation turns out to be the main flaw of HF theory: it ignores electron correlation, the tendency of electrons to avoid each other. The neglect of this effect in the calculation of the total energy has significant implications for chemistry: HF calculations on reactions often give large errors. Many ‘correlated’ ab initio methods, including those based on Møller–Plesset perturbation theory (e.g. MP2), configuration interaction, or coupled cluster theory, use HF wavefunctions as a starting point. These methods offer a significant improvement in accuracy over HF calculations, but also have much higher computational cost, which currently makes their application for systems with many tens of atoms difficult.

2.5 Combined quantum mechanics and molecular mechanics (QM/MM)

Combined or hybrid QM/MM approaches have become the method of choice for modeling reactions in biomolecular systems. In QM/MM methods, the region of the system in which the chemical process takes place is treated at an appropriate level of quantum chemistry (QM) theory, while the remainder is described by force-field-based molecular mechanics (MM) methods. In general, the decision for dividing a system into regions that are described at different levels of theory is the local characteristic of most

chemical reactions in condensed phases. For instance, to model large biomolecular system, the logical approach is to combine the two techniques and to use a QM method for the chemically active region (e.g., substrates and active site in an enzymatic reaction) and an MM treatment for the surrounding environment (e.g., the rest of protein and solvent). The common concept of hybrid QM/MM strategy is illustrated in **Figure 12**.

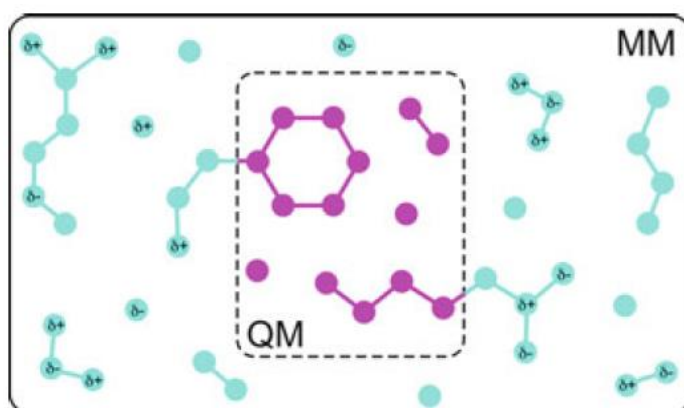


Figure 12. Illustration of the QM/MM concept. The QM region (labeled in violet), in which a chemical reaction occurs, is treated at a sufficiently high level of QM theory. The rest of the system (labeled in light blue) is described at the MM level [91].

2.5.1 QM/MM calculations

After partition of QM and MM regions, the energy calculation on both regions can be carried out using the commonly used two coupling schemes: additive and subtractive coupling schemes. In the additive scheme, the total QM/MM energy of the system is obtained from the summation of the energy of the QM region, MM region, and between QM and MM region as follows:

$$E_{\text{TOTAL}} = E_{\text{QM}}(\text{QM}) + E_{\text{MM}}(\text{MM}) + E_{\text{QM-MM}}(\text{QM} + \text{MM}) \quad (28)$$

The terms QM and MM in the parenthesis indicate the atoms in the QM and MM subsystems, respectively. The subscripts denote the level of theory at which the potential energies are calculated.

In the subtractive scheme (**Figure 13**), the total QM/MM energy of the system is obtained from the MM energy of the QM and MM regions, plus the QM energy of the isolated QM subsystem, minus the MM energy of the QM subsystem. The subtraction term is used to correct for double counting of the interactions within the QM subsystem:

$$E_{\text{TOTAL}} = E_{\text{MM}}(\text{QM} + \text{MM}) + E_{\text{QM}}(\text{QM}) - E_{\text{MM}}(\text{QM}) \quad (29)$$

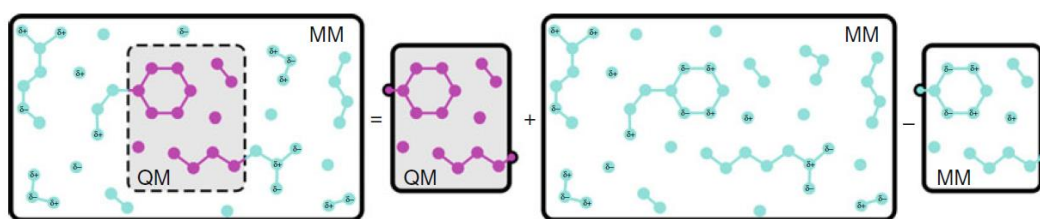


Figure 13. Illustration of subtractive QM/MM coupling scheme [91].

2.5.2 QM/MM boundary treatment

The QM/MM boundary treatment is important for the QM/MM method. The selection of the QM region can be performed by two different ways: with and without the QM/MM bond cutting. In many cases, a straightforward cut through the covalent bonds that cross the QM/MM boundary is commonly used, resulting in the creation of one or more unpaired electrons in the QM subsystem. The approaches which can solve the artefact of such open valences are needed. One of the most commonly used methods is the “link atoms” approach, which can place an additional atom at a suitable position along the cut bonds between the QM and MM atoms (**Figure 14A**). Hydrogen like atom is most general used due to its simplicity; however, any type of the link atoms can be used to treat the QM/MM boundary.

Alternatively, a double occupied molecular orbital is another popular approach to replace a chemical bond that passes through the QM and MM regions. The two most commonly used schemes are the localized hybrid orbital method, or known as the localized self-consistent field method [92], which introduces orbitals at the QM atom (**Figure 14B**), and the generalized hybrid orbital approach [93], which places additional orbitals on the MM atom (**Figure 14C**).

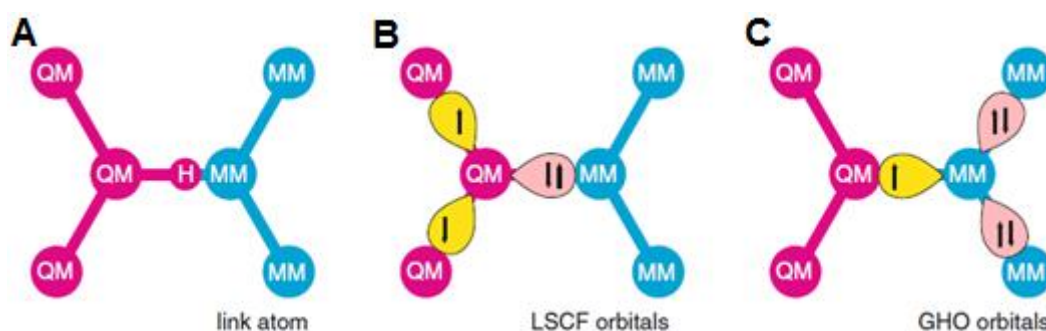


Figure 14. Different approaches to treat the QM/MM boundary. (A) link atoms and (B and C) frozen orbitals [91].

2.5.3 Calculations of energy profiles and barriers for reactions

Structural and energetic information involving the reactant, transition state and product along the reaction path are the main points in modeling enzyme-catalyzed reactions. Such information can be obtained either from static methods or dynamical methods (**Figure 15**). The former methods are involved in the optimization of structures along the reaction path, while the latter methods are involved in the generation of structure ensembles. For the static methods, one approach commonly used in locating approximate transition states and reaction paths on QM/MM potential energy profiles is the so-called “adiabatic mapping” method. In this approach, a reaction coordinate is defined (e.g. group of distances, angles, or torsions). The energy of the system is further minimized while constraining this coordinate to a set of slowly changed values using a force constant, giving a potential energy surface. The difference in electronic energy between the transition state and the reactants provides the potential energy barrier for the reaction. By addition of zero-point energy correction (for the atoms in the QM region) and calculation of vibrational frequencies, it is possible to obtain the activation energy, enthalpy and free energy. Thus, this free energy can be directly compared to experimental activation free energy.

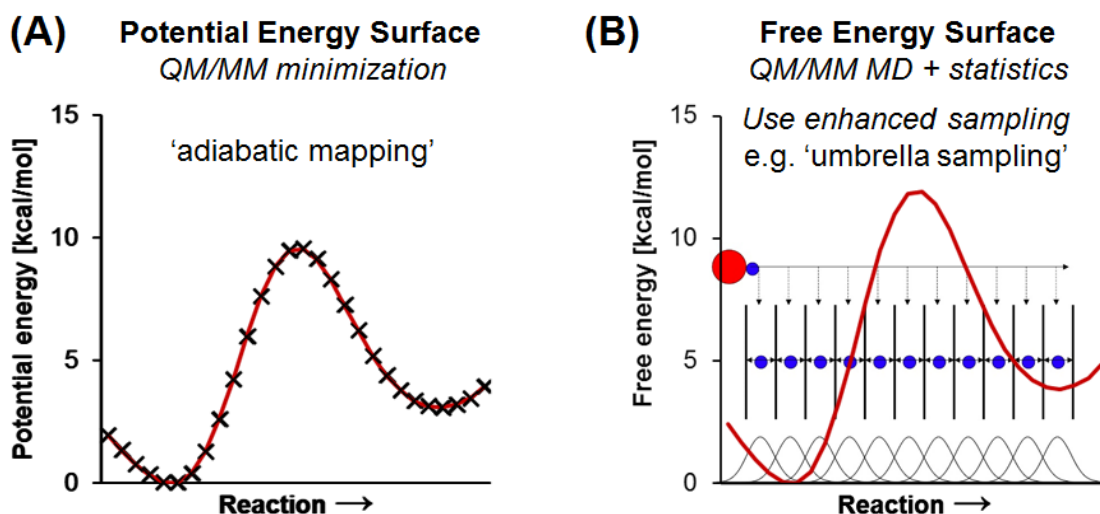


Figure 15. Two different approaches used in QM/MM reaction modelling (A) static methods and (B) dynamical methods.

Nevertheless, the reaction potential energy profile may not be obtained using the adiabatic mapping approach (e.g. large structural or charge changes occur along the reaction path). In these cases, it can be overcome by using dynamical methods to generate an ensemble of reacting structures. One of the commonly used methods for doing this is the umbrella sampling approach. Similar to adiabatic mapping, a reaction coordinate is chosen, then a set of MD simulations are performed, in each of which a harmonic or other restraining “umbrella” potential is applied to maintain the reaction coordinate close to a desired value. The set of selected values is designed to cover the range from the reactants to the products at a given interval. This results in an ensemble of structures representative of reaction at a particular temperature. The bias introduced by the umbrella potential can be removed using the weighted histogram analysis method (WHAM), to construct a free energy profile along the reaction path.

Sampling reaction paths with umbrella sampling (or using related approaches) is a challenge, as the potential energy is calculated with QM/MM methods. This is because the underlying simulations require a very large number of computations of the potential energy and its gradient (10^5 or more). Therefore, this technique is currently practical for low-level QM methods, such as AM1, PM3, PM6 and SCC-DFTB. However, such low-level QM methods are likely to yield quite inaccurate energy barriers. Another difficulty with such methods is that the computed free energy profile

may not be converged with the length of time scale along the reaction path, as one may not sample a truly equilibrium ensemble in the given time. The accuracy of QM/MM umbrella sampling free energy barriers based on low-level QM methods is possible to improve by applying a correction derived from adiabatic mapping studies at both the low level used for umbrella sampling and at a higher level.

2.5.3.1 The Weighted histogram analysis method (WHAM)

The weighted histogram analysis method (WHAM) is the most commonly used technique to calculate the probability distribution along the reaction coordinate. The WHAM algorithm is applied to estimate the uncertainty of the unbiased probability distribution given by the umbrella simulations and then compute the potential of mean force (PMF). The WHAM equations can be written below (Eqs. 30-31):

$$P(\xi) = \frac{\sum_{i=1}^{N_w} g_i^{-1} h_i(\xi)}{\sum_{j=1}^{N_w} n_j g_j^{-1} \exp[-\beta(w_j(\xi) - F_j)]} \quad (30)$$

and

$$\exp(-\beta F_i) = \int d\xi \exp[-\beta w_j(\xi)] P(\xi) \quad (31)$$

where g_i is the statistical inefficiency as expressed following (Eq. 32):

$$g_i = 1 + 2\tau_i \quad (32)$$

N_w is the number of umbrella simulations or umbrella windows, $h_i(\xi)$ is the umbrella histogram. τ_i is the integrated autocorrelation time of umbrella window i in the units of the simulation frame time. F_i is the free energy constant. β is equal to $1/k_B T$, where k_B is the Boltzmann constant, and T is the temperature. n_j is the total number of the

data points in histogram h_j . $P(\xi)$ is the unbiased probability distribution which is associated with the PMF as expressed below (Eq. 33):

$$W(\xi) = -\beta^{-1} \ln \left(\frac{P(\xi)}{P(\xi_0)} \right) \quad (33)$$

ξ_0 is the reference point where the PMF is defined as zero.



CHAPTER III

RESEARCH METHODOLOGY

3.1 Part I: Binding recognition of substrates in ZIKV NS2B/NS3 serine protease

3.1.1 Explicit-solvent CpHMD and CpH-REMD

Constant pH molecular dynamics (CpHMD) and constant pH replica exchange molecular dynamics (CpH-REMD) in explicit solvent were performed by AMBER16 coupled with the AMBER ff10 forcefield [94]. Starting from the X-ray crystal structure of the ZIKV NS2B/NS3-tetrapeptide (TGKR) complex (PDB ID: 5GJ4) [43], the amino acid residues sequence of the original tetrapeptide from P4 to P1 positions were changed to the corresponding peptide substrate (**Table 1**) for each position. Ionizable Lys (K) residue of each peptide substrate was made titratable. Each complex was solvated into truncated octahedron box of TIP3P water molecules, and sodium ions added to keep the whole system neutral. Stepwise minimizations with heating to 300 K and equilibration procedure were carried out to give starting points for the simulations of 50 ns covering the pH range 0-14 for CpHMD simulations and 7-14 for CpH-REMD simulations with 8 replicas. Protonation state statistics were calculated using the cphstats module of AMBER16.

3.1.2 System preparation for MD simulations

The starting coordinates of the ZIKV NS2B/NS3-tetrapeptide (TGKR) complex were obtained from the RSCB Protein Data Bank (PDB ID: 5GJ4) [43]. A tetrapeptide “TGKR” is associated with the P4 to P1 positions of the substrate, respectively (**Figure 16A**). To build the models of the ZIKV protease in complex with different substrates, the amino acid residues of the original tetrapeptide were changed to the corresponding substrate for each position. This modification was performed based on the similar orientation with the template peptide using the small molecules tool implemented in Accelrys Discovery Studio 2.5^{Accelrys Inc.}. Note that four fluorogenic substrates used in this study are selected from the literatures [43, 95] on the basis of being the recently designed substrates compared with the most commonly used substrate in the assays. All

determined substrates were consisted of: (i) substrate 1, acyl-Norleucine-Lysine-Lysine-Arginine-7-amino-4-carbamoylmethylcoumarin (Ac-nKKR-ACC); (ii) substrate 2, acyl-DArginine-Lysine-Ornithine-Arginine-7-amino-4-carbamoylmethylcoumarin (Ac-DRKOR-ACC); (iii) substrate 3, acyl-DLysine-Lysine-Ornithine-Arginine-7-amino-4-carbamoylmethylcoumarin (Ac-DKKOR-ACC); (iv) substrate 4, benzol-Norleucine-Lysine-Arginine-Arginine-aminomethylcoumarin (Bz-nKRR-AMC), as shown in **Figures 16B** and **17**. Each simulated system was assessed the protonation states of all ionizable amino acid side chains at pH 8.5 using PROPKA 3.1 [96], except for the catalytic residue H51 that was assigned as the neutral form with protonation at the delta position (HID type) due to mechanistic consideration. The electrostatic potential (ESP) charges of the substrates were calculated with the HF/6-31(d) level of theory using the Gaussian09 program [97]. Afterward, the restrained ESP charges and missing parameters of the substrates were achieved by the antechamber and parmchk modules of AMBER16, respectively, [98]. The AMBER ff14SB force field [99] and generalized AMBER force field (GAFF) [100] were used to assign bonded and non-bonded parameters for the protein and substrates, accordingly. The complexes were solvated in the TIP3P [101] water box with the minimum distance of 10 Å between the protein surface and the solvation box edge that the dimensions were roughly set to 81 Å × 85 Å × 87 Å. After that, sodium ions were randomly added to neutralize the simulated systems (2, 1, 1 and 2 ions for the substrates 1 to 4 systems, respectively). Prior to performing the MD simulations, the added hydrogen atoms and water molecules were energy-minimized with 500 iterations of steepest descent (SD) and 1500 iterations of conjugated gradient (CG) methods, while the remaining molecules were held fixed. The protein and substrates were then energy-minimized by SD (500 steps) and CG (1500 steps) methods with constrained solvent molecules. Finally, the entire system was fully energy-minimized with the same minimization procedure.

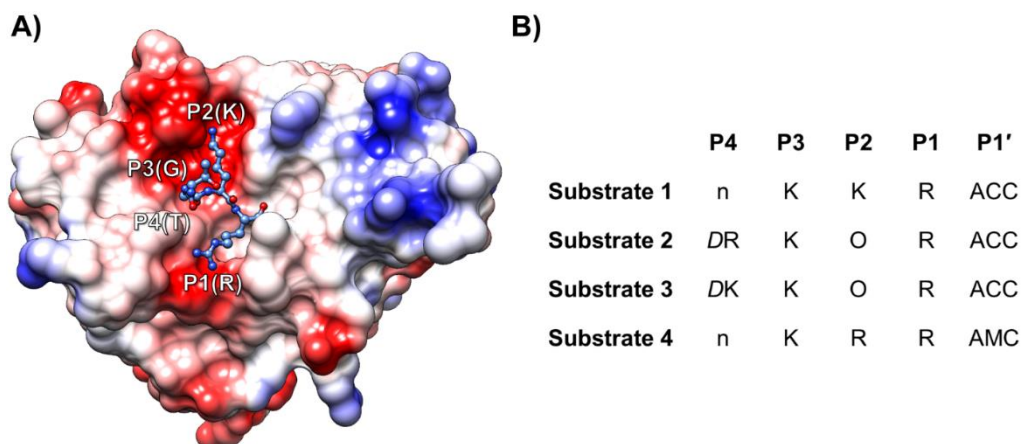


Figure 16. (A) Three-dimensional structure of the tetrapeptide “TGKR” binding to the active site of the ZIKV protease (PDB accession code 5GJ4), where the surface charge with negative and positive charge accumulation is shaded from red to blue, respectively. (B) Comparison of the four substrate sequences used in this study.

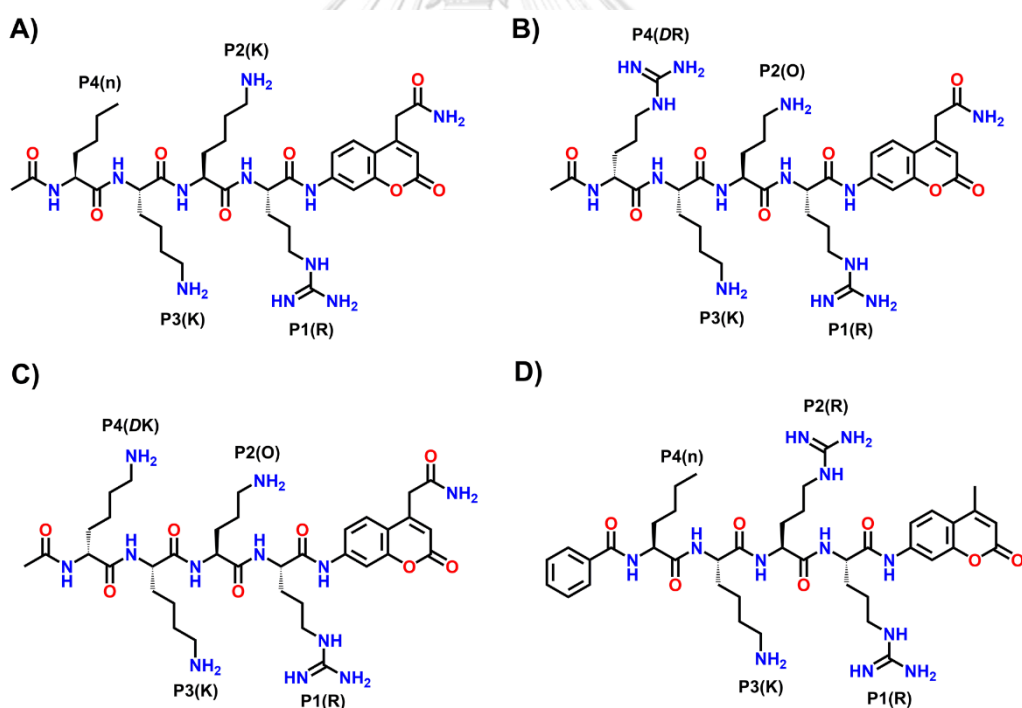


Figure 17. Chemical structures of (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC used in this study corresponding to substrates 1 to 4, respectively.

3.1.3 Molecular dynamics simulations

Each complex was simulated under the periodic boundary condition with the isothermal–isobaric (NPT) ensemble, as previously described [102-104]. In brief, a residue-based cutoff of 10 Å was employed for non-bonded interactions, and the particle mesh Ewald summation method [105] was used to treat the electrostatic interactions. The SHAKE algorithm [90] was applied to constrain all covalent bonds involving hydrogen atoms. A simulation time step of 2 fs was used along the MD simulation. The Langevin thermostat [106] with a collision frequency of 2 ps⁻¹ was employed for temperature control, while the Berendsen barostat [107] with a pressure-relaxation time of 1 ps was used to maintain the standard pressure of the system. Each simulated system was gradually heated from 10 to 310 K for 100 ps using positional restraints of 50.0 kcal mol⁻¹ Å⁻² to the C α atoms of protein. Afterward, the system was submitted to five stages of restrained MD simulations at 310 K with positional restraints of 20, 15, 10, 5 and 2.5 kcal mol⁻¹ Å⁻² for 200 ps of each stage, and another 200 ps without any restraint to equilibrate the systems. Consequently, the whole system was simulated under the NPT scheme (310 K, 1 bar) for 50 ns. The MD trajectories were collected every 2 ps. The global root-mean-square displacement were used to determine the system stability. The protein-substrate interactions and the binding free energies were evaluated using the cpptraj [108] and MMPBSA.py [109] modules of AMBER16, respectively.

3.2 Part II: Reaction mechanism of the ZIKV protease with its substrate

3.2.1 System preparation

The coordinates of the ZIKV protease enzyme in complex with TGKR peptide from the X-ray structure (Protein Data Bank (PDB), code 5GJ4) [43] were employed. Note that a tetrapeptide ‘TGKR’ corresponds to the -P4-P3-P2-P1- positions of the NS2B/NS3 cleavage site. The carboxylate group presented on the peptide C-terminus was replaced by the P1’ serine (S) residue, corresponding to the NS2B/NS3-mediated cleavage site [110]. The addition of missing hydrogen atoms and protonation states of ionizable amino acid residues were assigned at pH = 8.5, using the H++ server (<http://biophysics.cs.vt.edu/H++>) [111], except for H51 which was set as the neutral form with protonated δ -NH (HID type) due to mechanistic consideration. In addition,

their environments were considered based on the possibility of hydrogen bond formation with the surrounding residues. The ff14SB [99] AMBER force field was used to assign bonded and non-bonded parameters to the protein and peptide substrate. The system was solvated in an octahedron box of TIP3P [101] water model and neutralized with sodium ions [112], consisting of ~6000 water molecules and 3 Na⁺ ions.

3.2.2 Molecular dynamics simulations

MD simulations of the ZIKV NS2B/NS3–TGKRS complex were performed with the AMBER16 software package coupled with the *sander* and *pmemd* modules [113]. The system was simulated under periodic boundary conditions and the electrostatic interactions were treated using the Particle Mesh Ewald method [105]. Temperature was maintained by the Langevin dynamics technique with a collision frequency of 2 ps⁻¹ [106], and pressure was controlled by the simple Berendsen barostat [107], with a pressure-relaxation time of 1 ps. A cut-off distance of 10 Å was set for non-bonded interactions, while all covalent bonds involving hydrogen atoms were constrained by the SHAKE algorithm [114], allowing a simulation time step of 2 fs. Prior to performing MD simulations, the water molecules and counter ions were energy-minimized by 1000 iterations of steepest descent (SD), followed by 2000 iterations of conjugate gradient (CG) methods, while the rest of the molecules were restrained. Afterward, the protein was energy-minimized by SD (1000 steps) and CG (2000 steps) methods together with restrained solvent. Finally, the entire system was fully energy-minimized using the same minimization procedure. After an initial energy minimization step, the complex was slowly heated up to 300 K in 100 ps using positional restraints of 50.0 kcal mol⁻¹ Å⁻² for the alpha carbon (C α) atoms of protein. The system was further equilibrated for 1000 ps with positional restraints, which were consequently decreased by 5.0 kcal mol⁻¹ Å⁻² every 100 ps, and another 100 ps without any positional restraints. Subsequently, the whole system was simulated under an isothermal-isobaric (*NPT*) ensemble for 200 ns (300 K, 1 bar) in three independent simulations, leading to 600 ns in total of simulation time. In addition, a single MD simulation was performed for the apo form for 1 μ s.

3.2.3 QM/MM MD free-energy calculations

QM/MM MD simulations were carried out using the *sander* module implemented in the AMBER16 package. One suitable representative conformation from the previous MM MD simulations that fulfilled the following criteria: (i) $d(\text{N2-H2}) < 2.0 \text{ \AA}$ and $d(\text{O3-C}\gamma) < 3.0 \text{ \AA}$, and (ii) the nucleophilic attack angle, $\theta_y(\text{O3-C}\gamma\text{-O}\gamma)$, close to 90° , was selected as the starting geometry for the QM/MM simulations. The QM region consisting of 60 atoms included the side-chains of the catalytic triad (H51, D75 and S135) and the substrate amino acids (P1 and P1') close to the scissile bond (see **Figure 18**). Moreover, a negative total charge ($-1e$) was set for the QM region and 4 hydrogen link atoms (LAs) [115, 116] were applied to saturate the QM/MM boundary. The SHAKE algorithm was not calculated in the QM region. The QM region was treated with the PM6 semiempirical method [117], while the remaining region was described at the MM level using the ff14SB force field.

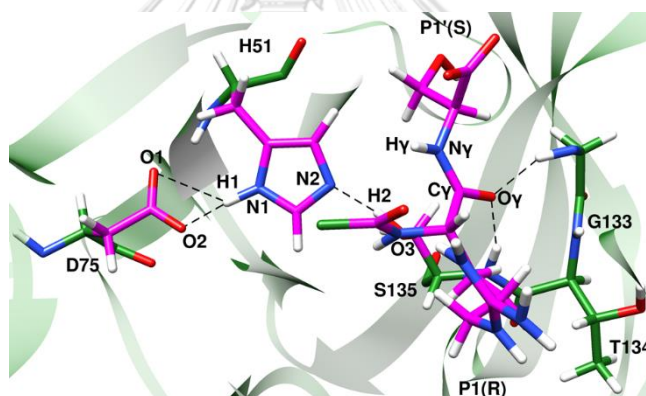


Figure 18. Active site of ZIKV protease with the fragment of substrate. The QM atoms and the MM regions are displayed as magenta and green carbon atoms, respectively. Four hydrogen link atoms were used to saturate the QM/MM boundary, where the color is switched from magenta to green. The hydrogen bonds are represented by dashed lines. The atomic labels involved in the reaction mechanism are also given for further discussion.

QM/MM umbrella sampling MD simulations were performed for each step of the acylation process, harmonically restraining the reaction coordinate (RC) with a force constant of $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. Each window was composed of 50 ps of restrained QM/MM MD. The RC was scanned at 0.1 \AA intervals between adjacent windows, using

the previous geometry as the starting point for the subsequent window. The free energy profiles for each step were calculated by the weighted histogram analysis method (WHAM) [118-120]. The acylation process was categorized into two steps, in which the first and second steps were defined by TI formation and the breaking of peptide bond, respectively. The following reaction coordinates (RCs) were used: (i) *the first step of the acylation* – a 2D umbrella sampling simulation involving the proton transfer from S135 oxygen to H51 ϵ nitrogen, defined as $d(\text{O3-H2})-d(\text{N2-H2})$, ranging from -1.0 to 1.0 Å, and the nucleophilic attack of S135 oxygen to the carbonyl carbon of substrate, defined as $d(\text{O3-C}\gamma)$, varying from 3.0 to 1.5 Å. In order to calculate a less constrained pathway, a 1D umbrella sampling for the reaction was run using a RC described by the difference of the breaking and forming bonds ($d(\text{O3-H2})-d(\text{N2-H2})-d(\text{O3-C}\gamma)$, from -3.8 to -0.6 Å). For (ii) *the second step of the acylation* – 2D umbrella sampling with two RCs accounting for the breaking of $\text{C}\gamma\text{-N}\gamma$ peptide bond, defined as $d(\text{C}\gamma\text{-N}\gamma)$, with values ranging from 1.5 to 1.9 Å, and the transfer of proton from the ϵ nitrogen of the H51 to the peptide bond nitrogen of the substrate, defined as $d(\text{N}\gamma\text{-H2})-d(\text{N2-H2})$, varying from 1.6 to -1.0 Å, was simulated. Similar to the first step of the acylation process, a single RC described by the antisymmetric combination of $d(\text{C}\gamma\text{-N}\gamma)$ and $d(\text{N}\gamma\text{-H2})-d(\text{N2-H2})$, ($d(\text{C}\gamma\text{-N}\gamma)-d(\text{N}\gamma\text{-H2})+d(\text{N2-H2})$, from -0.2 to 2.8 Å) was used.

3.2.4 High-level QM/MM calculations for the first step of the acylation process

High-level QM/MM calculations (up to the LCCSD(T) level, which can provide “chemically accurate” barriers for reactions) [121-123] were performed to obtain accurate potential energy profiles for the reaction. TS-like conformations obtained from PM6/ff14SB 1D umbrella sampling MD simulations were used as the starting geometries for an adiabatic mapping approach along the RC of $d(\text{O3-H2})-d(\text{N2-H2})-d(\text{O3-C}\gamma)$ for the first step of the acylation process. Five initial geometries of TS structure were selected from 30–50 ps simulation at every 5 ps of the TS-sampling window at the PM6/ff14SB of stationary point ($d(\text{O3-H2})-d(\text{N2-H2})-d(\text{O3-C}\gamma)$ at -1.8 Å). To reduce the size of QM/MM calculations, all water molecules above the sphere of the protein were removed, and so

the resulting protein was in a 10 Å layer of water surrounding protein. The QM region consisted of the atoms that were used in QM/MM umbrella sampling MD simulations as mentioned above. All protein residues plus water molecules within a 5-Å sphere of the substrate were included in the active region of the optimization process and allowed to move freely, while the rest of atoms were kept frozen. All QM/MM calculations were performed using ChemShell [124, 125], by combining the Orca package [126] for the QM part and DL_POLY [127] for the MM part. A full electrostatic embedding scheme was employed to calculate the polarizing influence of the enzyme environment on the QM region. Hydrogen LAs with the charge-shift model [105, 128] were applied to treat the QM/MM boundary. During the QM/MM geometry optimizations, all atoms in QM region were described at the BH&HLYP-D3/6-31G(d) level of theory. The D3 in the acronym indicates that an empirical dispersion correction was applied to the DFT calculations [129]. The inclusion of dispersion correction plays an important in DFT QM and QM/MM calculations, particularly in those reactions catalyzed by enzymes [130, 131]. The MM region was described by the ff14SB AMBER force field implemented in the DL_POLY code. A RC restraint was used to drive the reaction from the TS toward the reactant and product states at 0.1 Å intervals during the geometry optimizations. Note that the BH&HLYP [132-134] is known to provide better results than B3LYP [132, 135, 136] for some proton and charge transfers [123, 137] and also a better description of hydrogen bonding [138-140]. In addition, the BH&HLYP method gave the good results in line with *ab initio* methods [141]. Herein, single-point energy QM/MM calculations on the BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries were performed at the LMP2/(aug)-cc-pVTZ/ff14SB, SCS-LMP2/(aug)-cc-pVTZ/ff14SB, and LCCSD(T)/(aug)-cc-pVTZ/ff14SB levels of theory using ChemShell together with MOLPRO for the QM region [142]. SCS denotes the spin component scaled method [143] applied for the MP2 [144] calculations. It is worth noting that the combination of SCS and MP2 calculations provided results in good agreement with the coupled cluster methods for the reaction mechanisms catalyzed by other enzymes [123, 145, 146]. The (aug) in the notation of the (aug)-ccpVTZ [147] basis set means that a basis set augmented with diffuse functions were only used for the oxygen atoms. Meanwhile, the L in these acronyms for the correlated *ab initio* methods

indicates that local approximations [148-150] were adopted in the QM/MM calculations.

3.3 Part III: Inhibition mechanism of the ZIKV protease with dipeptidyl aldehyde inhibitor

3.3.1 System preparation

The structure of a covalent complex between the ZIKV protease and dipeptide inhibitor (acyl-KR-aldehyde) was taken from a high-resolution (1.5 Å) X-ray crystal structure (PDB accession code 5YOF) [79]. The ionizable residues were assessed by H++ web-prediction of protonation [151] (<http://biophysics.cs.vt.edu/H++>) at pH 8.5. However, the catalytic H51 was protonated at both δ - and ϵ -nitrogen positions (HIP type) to resemble the tetrahedral intermediate adduct of the bound aldehyde. The partial atomic charges and the electrostatic potential (ESP) charges of dipeptidyl aldehyde were obtained by the restrained electrostatic potential (RESP) method [152, 153] using HF/6-31G(d) calculations. The system was immersed in a cubic box of TIP3P water molecules and neutralized with 4 Na⁺ ions using the LEaP module implemented in AMBER16. The final system was composed of 25965 atoms (box size of 90 × 90 × 90 Å³).

3.3.2 Molecular dynamics simulation

MD simulation of the ZIKV protease-inhibitor complex was carried out using the AMBER16 software package [113] with the AMBER ff14SB force field [99] for the protein and the generalized Amber force field version 2 (GAFF2) [100] for the inhibitor. The MD protocol used in this study was performed as previously described [102, 154, 155]. In brief, the system was energy-minimized and heated up from 10 to 300 K and equilibrated under *NPT* scheme ($P = 1$ bar). Afterward, system was submitted to a MD simulation in the *NPT* ensemble for 200 ns to produce the starting structure for QM/MM simulations. A cut-off distance of 10 Å was employed for full electrostatic and van der Waals interactions. The long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method [105], whereas all covalent bonds involving hydrogen atoms were constrained with the SHAKE algorithm [90], allowing to use a larger time step of 2 fs. The Langevin thermostat [156] with a collision

frequency of 2 ps⁻¹, and the Berendsen barostat [107], with a pressure relaxation time of 1 ps, were used for temperature control and pressure control, respectively.

3.3.3 QM/MM MD free-energy calculations

QM/MM MD simulations were performed using the *sander* module implemented in the AMBER16. A representative structure extracted from the most populated structures within the clustering of the MD trajectories was energy-minimized prior to performing the QM/MM simulations. The QM region composing of 49 atoms included the fragments of the inhibitor, and the side chains of the catalytic triad (H51, D75 and S135), which participate in the catalytic mechanism (blue label in **Figure 7**). Moreover, an overall charge of zero was set for the QM region and four hydrogen link atoms (LAs) [115] were applied to saturate the covalent bonds that cross the QM/MM boundary. The QM atoms were treated with the semi-empirical Hamiltonian Pairwise Distance Directed Gaussian modification of PM3 (PDDG-PM3) [157], whereas the MM atoms in the system were calculated using the ff14SB all-atom AMBER force field.

QM/MM umbrella sampling MD simulations were performed for studying the inhibition mechanism of dipeptidyl aldehyde inhibitor toward the ZIKV protease and used to compute the free-energy reaction path. 50 ps of restrained QM/MM MD per umbrella was carried out with a harmonic potential force constant of 100 kcal mol⁻¹ Å⁻², in which the first 10 ps was excluded to ensure the equilibration of the system. The reaction coordinate (RC) was scanned at 0.1 Å intervals between neighboring windows, using the previous geometry as the starting structure for the next umbrella window. The approximate free-energy profiles from the QM/MM umbrella sampling MD simulations were estimated by the weighted histogram analysis method (WHAM) [158]. The transition state was approximately investigated, as the highest energy point along the reaction pathway on the free-energy profile. In this work, the RC used for describing the formation of the intermediate adduct (hemiacetal adduct) was defined as a linear combination of the following interatomic distances: (i) the distance involved in the transfer of proton from S135 hydroxyl group to H51 ε nitrogen, and (ii) the distance involved in the nucleophilic attack of S135 oxygen to the carbonyl carbon on aldehyde moiety, $RC = [d(O3-H2) - d(N2-H2) - d(O3-C\gamma)] \text{ \AA}$. This representation of the RC was

selected, as it has been previously shown to be suitable for describing the first step of the acylation process catalyzed by serine proteases as well as the ZIKV protease [56, 159-162].

3.3.4 QM/MM potential energy calculations

Potential energy reaction path calculations were performed with the QM/MM minimization or adiabatic mapping procedure [163, 164] at the BH&HLYP-D3/6-31G(d)/ff14SB level using the ChemShell [165] interface. The setup of QM/MM calculations was performed as our previous report [162]. Briefly, all water molecules beyond a 10 Å sphere of the protein were removed to reduce the computational expenses for QM/MM calculations. The QM region consisted of the atoms that were used in QM/MM free-energy simulations as described above. All atoms situated outside a 5 Å sphere of the inhibitor were held fixed during the QM/MM geometry optimizations. The QM calculations were performed using the Orca package [126] and the rest of the system was treated with the ff14SB AMBER force field using the DL_POLY code [166]. To account for the polarizing effect between QM and MM regions, a full electrostatic embedding scheme [167] was employed, whereas the QM/MM boundary was saturated with hydrogen LAs using the charge-shift model. Note that the D3 dispersion correction was applied for the BH&HLYP energy and gradient calculations. QM/MM optimizations were conducted with the RC restraint using a harmonic potential force constant of 500 kcal mol⁻¹ Å⁻² applied to drive the backward reaction from the product complex (covalent complex) to the reactant state (noncovalent Michaelis complex) at 0.1 Å intervals. To further evaluate the energetics of the reaction, single-point energy QM/MM calculations on the structures optimized by BH&HLYP-D3/6-31G(d)/ff14SB level were carried out at the MP2/(aug)-cc-pVTZ/ff14SB, SCS-MP2/(aug)-cc-pVTZ/ff14SB, and LCCSD(T)/(aug)-cc-pVTZ/ff14SB levels using ChemShell to interface with MOLPRO [142]. The (aug) in the notation of the (aug)-cc-pVTZ basis set denotes that the augmented basis set is used to treat oxygen atoms only. In addition, the acronym SCS in the SCS-MP2 method stands for the spin-component scaled method developed by Grimme [143] for the MP2 [144] calculations. Meanwhile, the acronym LCCSD(T) denotes that local

approximations [148, 150] were adopted in the calculations within the coupled-cluster framework.

Additionally, the resulting potential energy surfaces were refined using the improved tangent climbing image nudged-elastic band (CI-NEB) [168, 169] implemented in ChemShell to characterize and optimize the reaction pathway by removing the RC constraint. Indeed, they involve the simultaneous optimization of a sequence of structures along the minimum-energy reaction path. In this study, a NEB path was generated at the BH&HLYP-D3/6-31G(d)/ff14SB level by optimizing seven structures along the path linking to the reactant, transition state and product. Note that intermediate structures along the NEB path were generated by linear interpolation between the structures of reactant and transition state or transition state and product. Furthermore, single-point energy QM/MM calculations on the structures generated by the NEB techniques were carried out at the MP2, SCS-MP2, and LCCSD(T) levels of theory to investigate the consistency of the results.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Part I: Binding recognition of substrates in ZIKV NS2B/NS3 serine protease

4.1.1 Protonation states of ionizable residues

Since the preferential substrates of the ZIKV protease contain basic amino acids (R or K) at the P1 and P2 positions, the pK_a values and protonation states of ionizable residues were taken into account. In this present work, the pK_a values of lysine (K) residues in peptide substrates which are probably changed in the negatively charged microenvironments of the ZIKV protease's active site were investigated. Constant pH molecular dynamics (CpHMD) and constant pH replica exchange MD (CpH-REMD) methods were performed on the ZIKV protease-substrate complexes in order to calculate the pK_a values of titrable residues of each peptide substrate for 50 ns. The titration curves of each peptide substrate and the corresponding pK_a values are shown and summarized in **Figure 19** and **Table 3**, respectively. The results suggest that the predicted pK_a values of the K residue in peptide substrates are in the range of 8.7-10.2, which are also consistent with the pK_a values calculated with PropKa program. This reflects that the K residue should be presented in the cationic form due to the protein microenvironment. Therefore, the K residue of the peptide substrates was assigned as the positively charged form throughout this research.

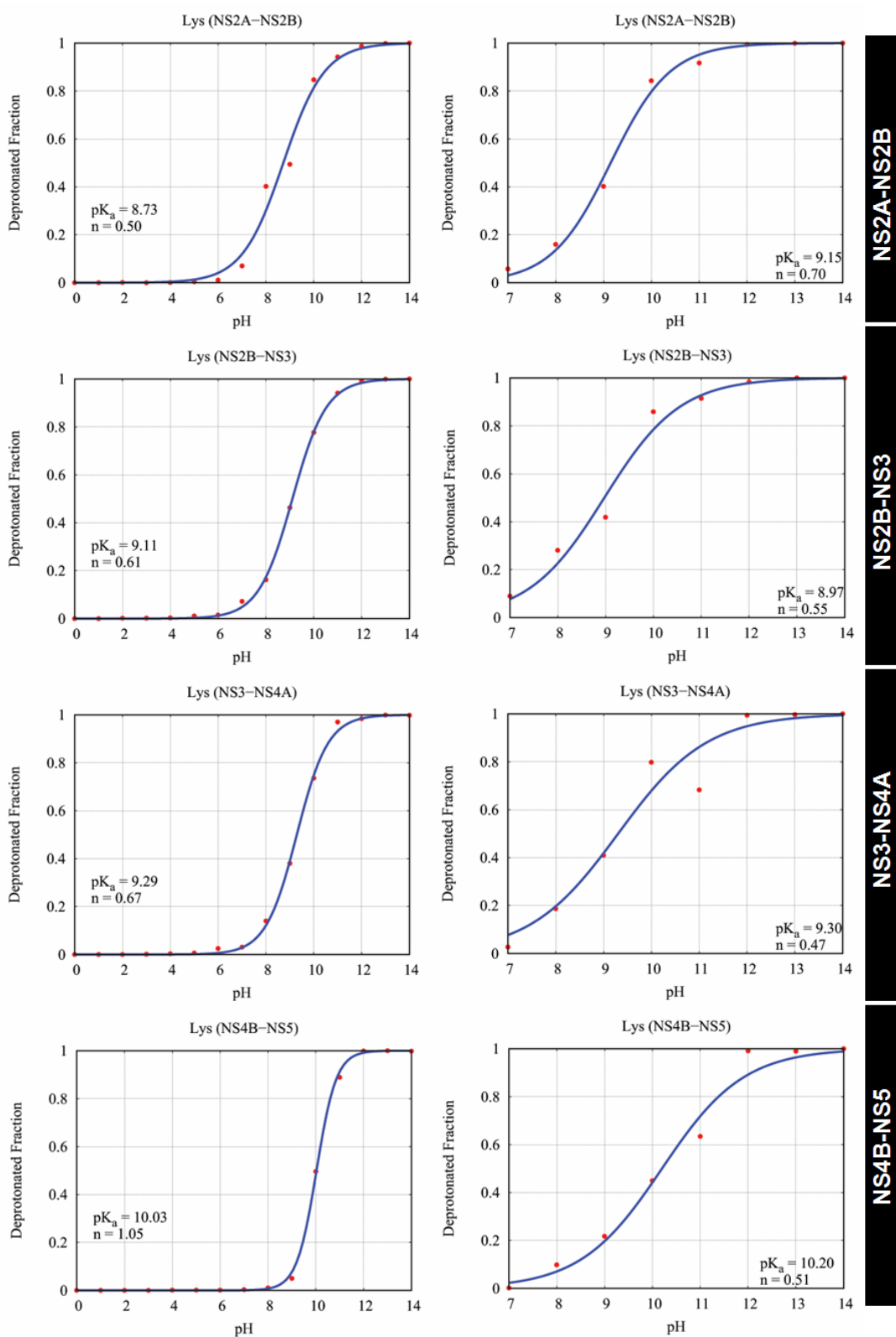


Figure 19. Titration curves for the lysine residue of each peptide substrate calculated with (Left) CpHMD and (Right) CpH-REMD.

Table 3. Predicted pK_a values of the lysine residue of each peptide substrate obtained from cpHMD and cpH-REMD simulations.

Peptide substrate	Predicted pK _a values		
	cpHMD	cpH-REMD	PropKa
NS2A/NS2B: S(P4)-G(P3)- K(P2) -R(P1)-S(P1')	8.73 ± 0.09	9.15 ± 0.06	10.65
NS2B/NS3: T(P4)-G(P3)- K(P2) -R(P1)-S(P1')	9.11 ± 0.02	8.97 ± 0.11	11.38
NS3/NS4A: A(P4)-G(P3)- K(P2) -R(P1)-G(P1')	9.29 ± 0.03	9.30 ± 0.22	11.16
NS4B/NS5: V(P4)- K(P3) -R(P2)-R(P1)-G(P1')	10.03 ± 0.02	10.20 ± 0.13	10.95

4.1.2 Stability of the global structure

To monitor the structural stability of the four simulated systems, the time evolution of the root-mean-square displacement (RMSD) relative to the starting structure for all heavy atoms in the complex was assessed and plotted in **Figure 20**. The result shows that the RMSD values of all systems obviously increase in the first 5 ns and then fluctuated at around 2.0-2.2 Å till the end of simulation. All complexes were likely to be stable after 20 ns. In this study, the last 20 ns of the corresponding coordinates were extracted as the production phase for further analyses in terms of (i) intermolecular hydrogen bonds, (ii) key residues for substrate binding and (iii) binding free energy of protein-substrate complexes.

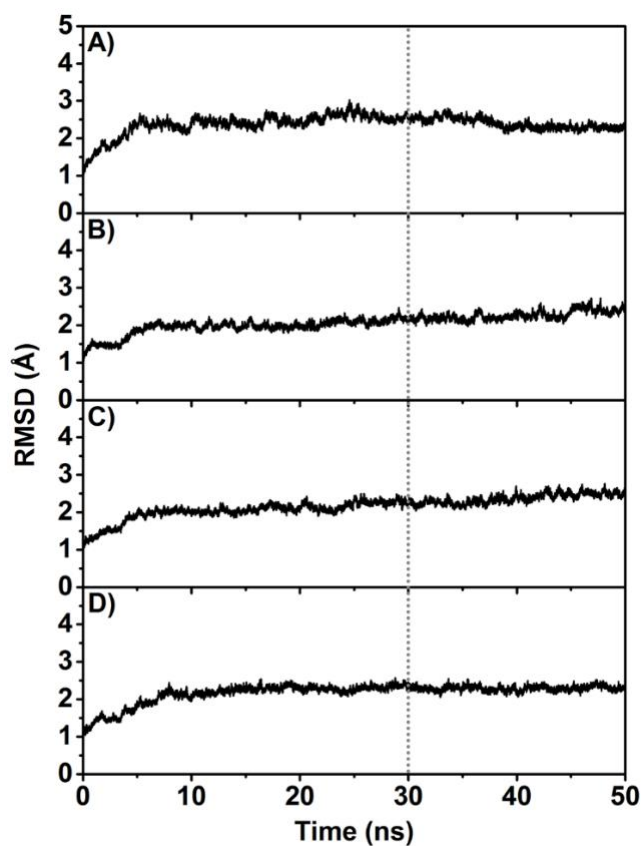


Figure 20. RMSD plots for complex all heavy atoms of the ZIKV protease with (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC bound.

4.1.3 Intermolecular hydrogen bonds

The intermolecular hydrogen bonds formed between substrate and surrounding amino acid residues at the enzyme active site play an important role in binding recognition and ZIKV NS2B/NS3-mediated cleavage of individual substrate. To evaluate the hydrogen bond strength of the protein-substrate complexes, the number and percentage of hydrogen bond formations with the ZIKV protease at each position of the substrate were monitored according to the two geometric criteria of (i) a proton donor (D) and acceptor (A) distance ≤ 3.5 Å and (ii) a D-H \cdots A bond angle $\geq 120^\circ$. Hydrogen bond occupations of $\geq 80\%$, 50-79% and $< 50\%$ are considered as strong, moderate and weak hydrogen bond interactions, respectively. The results of hydrogen bond interactions are shown in **Figure 21**, while the representative structures determined from the last MD snapshot of the ZIKV protease bound to each studied

substrate are depicted in **Figure 22**. Note that no hydrogen bond interactions between the P4 position and the residues located at the S4 pocket of the enzyme are observed. Among three positions of the substrate (P1, P2 and P3), the highest number of hydrogen bond formations was found at the P1(R) position in all systems, which is the center for cleavage reaction. It should be mentioned that the hydrogen bond interactions formed at the S1 subsite of the enzyme are only contributed from the NS3 residues with different levels of hydrogen bond strength. This center P1 residue forms strong and/or moderate hydrogen bonds with G133, T134, S135 and G151 of NS3 protease, which are conserved among the four substrates to some extent. However, the hydrogen bond interactions of the P1(R) position of the substrates 1 (Ac-nKRR-ACC) and 4 (Bz-nKRR-AMC) with D129 and Y130 of NS3 domain are remarkably decreased or almost absent when their P2 position is changed to the longer amino acid side chain, compared with the P2(O) residue of the substrates 2 (Ac-DRKOR-ACC) and 3 (Ac-DKKOR-ACC). This phenomenon suggests the indirect effect of the weaker binding affinity on the substrates 1 and 4 toward the ZIKV protease.

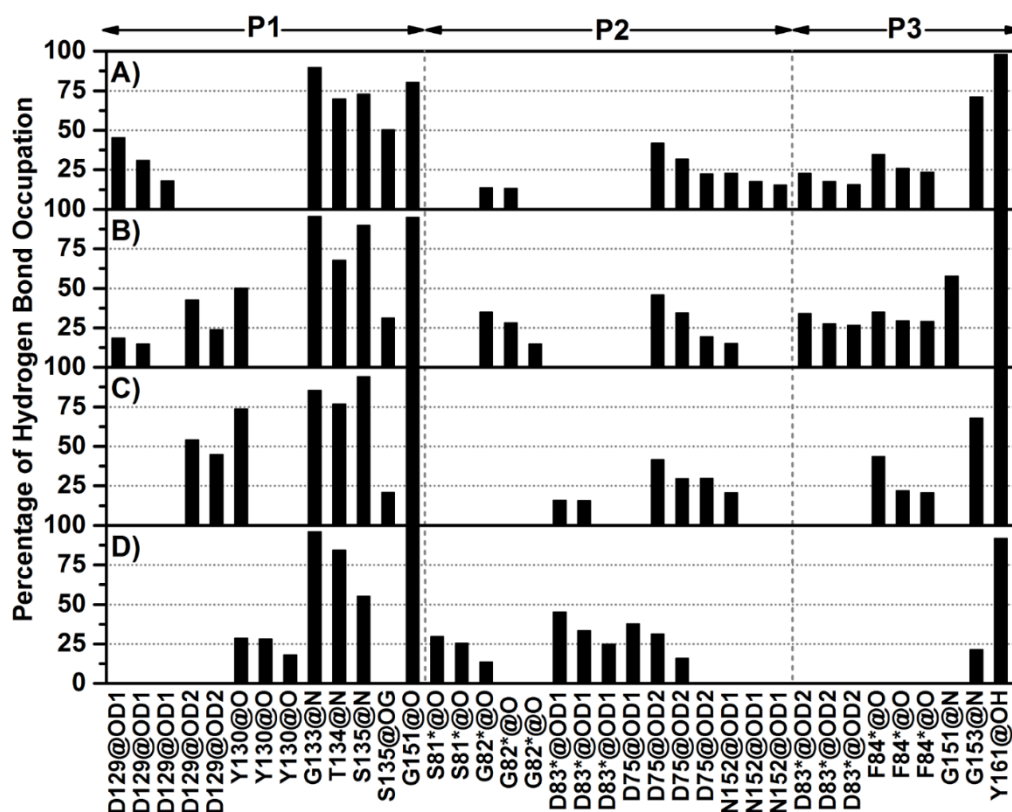


Figure 21. Percentage of hydrogen bond occupation of the ZIKV protease residues formed with the four different substrates: (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC. The residues associated with the NS2B protein are indicated by asterisks.

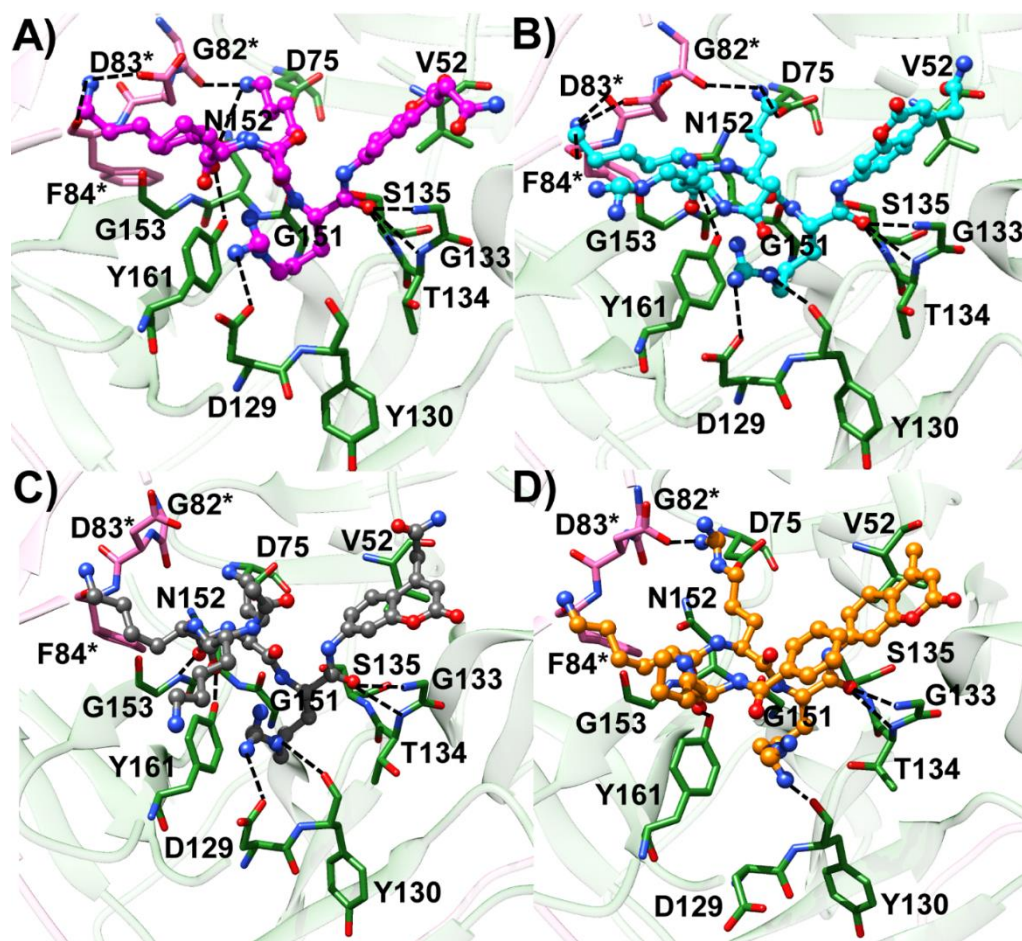


Figure 22. Binding patterns of (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC peptides in the ZIKV protease active site extracted from the last MD snapshot. Hydrogen bond interactions are represented by dashed lines. The residues related to the NS2B protein are given by asterisks.

The second highest number of hydrogen bond formations between the ZIKV protease and each substrate was observed at the P2 position. The positively charged P2 residue of all substrates is stabilized by the negatively charged residue D75 of NS3. In addition, it is notably seen that the interactions with the D83 of NS2B are significantly reduced and totally vanished in the substrates 1 (Ac-nKKR-ACC), 2 (Ac-DRKOR-ACC) and 3 (Ac-DKKOR-ACC) systems due to the substitution of the positively charged short side chain of lysine (K) or ornithine (O) at this position in comparison with the substrate 4 (Bz-nKRR-AMC) comprising the longer amino acid side chain of arginine (R). However, this event probably culminates in the bad contacts between the

P2(R) residue of the substrate 4 and the surrounding residues at the S2 subsite of the enzyme. It can also indirectly induce the conformational change of a guanidinium group of the P1(R) residue, which moves away from the D129 of NS3 protease. Consequently, the P2 position of the substrate 4 is able to form rather weak hydrogen bond interactions with additional residues S81 and G82 of NS2B domain instead.

At the P3 position, one strong (Y161 of NS3) hydrogen bond is found in all complexes. One moderate (G153 of NS3) hydrogen bond is determined in the substrates 1 and 3, while the interaction with this residue is dramatically decreased and totally disappeared in the substrates 4 and 2, respectively. Although the P3 residue of all studied substrates is R residue, the hydrogen bond patterns for substrate binding at this site are different in each system. The P3 residue of substrates 1 and 2 forms weak hydrogen bonds with the two residues of NS2B (*i.e.* D83 and F84). For the remaining systems, the interactions with both residues are reduced or completely absent particularly for the protein-substrate 4 complex.

According to the total number of hydrogen bonds and interaction strength of the four focused complexes, the ZIKV NS2B/NS3-substrate 2 complex shows the highest number of hydrogen bond interactions in total, which might be correlated with the strongest binding efficiency of this substrate toward the ZIKV protease compared with the other substrates. A high number of strong and moderate hydrogen bond interactions are likely formed at the P1 and P2 positions. This finding is associated with the common binding recognition preferences of flavivirus proteases, which are favorable to bind with basic amino acids at the S1 and S2 subsites [43, 50, 62, 63]. The preferential interactions of substrates at the P1 residue with the residues G133, T134 and S135 of NS3 located in the oxyanion hole, which stabilize the carbonyl oxygen of the P1 reacting residue in the cleavage reaction, are likely conserved in all complexes. These interactions are quite similar to those of our previous MD simulations of DENV type 2 NS2B/NS3 protease [170] and the related WNV NS2B/NS3 X-ray structures [80, 171]. Taken together, the overall substrate conformation in the studied systems is in the active conformation and can undergo the cleavage reaction in the next step.

4.1.4 Key residues for substrate binding

To explore the key residues essential for substrate recognition to the ZIKV protease, the per-residue free energy decomposition ($\Delta G_{\text{bind}}^{\text{residue}}$) based on MM/PBSA method was calculated over the 100 structures extracted from the last 20 ns of MD simulations. With an energy stabilization of $< -1.0 \text{ kcal mol}^{-1}$ in **Figure 23**, the computed $\Delta G_{\text{bind}}^{\text{residue}}$ for all systems displays that the residues D83 of NS2B as well as D75, A132, G133, T134, G151, N152 and V155 of NS3 play a key role for substrate binding. The interactions of P1(R) residue of the substrates 1 (Ac-nKKR-ACC), 2 (Ac-DRKOR-ACC) and 3 (Ac-DKKOR-ACC) with residues D129, A132, G133 and S135 of NS3 protease located at the S1 subsite are much stronger than that with the P1 of the substrate 4. Likewise, the P2 residue of the substrates 1 to 3 is strongly stabilized by the residue D75 located at the S2 subsite, while this interaction is dramatically reduced in the substrate 4 system. This is probably due to a suitable side chain of the P2 residue on substrates 1 to 3 (K or O) interacting with the D75 of NS3 better than a longer side chain with a bulky group of the P2(R) on substrate 4, which can lead to the steric effect at the S2 subsite. In addition, the introduction of R residue at the P2 position of the substrate 4 results in the conformational changes of the remaining positions, causing the loss of several interactions with the amino acid residues especially at the S1 pocket (**Figure 23D**), as mentioned in **section 4.1.3**. Some residues of NS2B such as G82 and D83 also form hydrogen bonds with the P2 residue of the substrate. Note that the binding pattern of the P3 residue at the S3 subsite is similar in all complexes without significant energy contributions predicted, while there is no obviously attractive interaction observed between the P4 residue and the S4 pocket. This result can also imply that the S1 and S2 pockets of the ZIKV NS2B/NS3 protease play an important role in recognition of basic amino acids more than the other subsites, as previous reports [43, 78]. The obtained results are in line with the experimental data that the K_m values (**Table 4**) of substrate 4 are higher than those of substrates 1 to 3 [43, 95], indicating a lower binding affinity between that substrate and the ZIKV protease. Furthermore, our computationally derived results agree well with the reported experimental data [95, 172], showing that the preferential residues of each subsite of the ZIKV protease are $R > K, O > K > R, K > R$ and $V > L > K/R$ for P1, P2, P3 and P4, respectively.

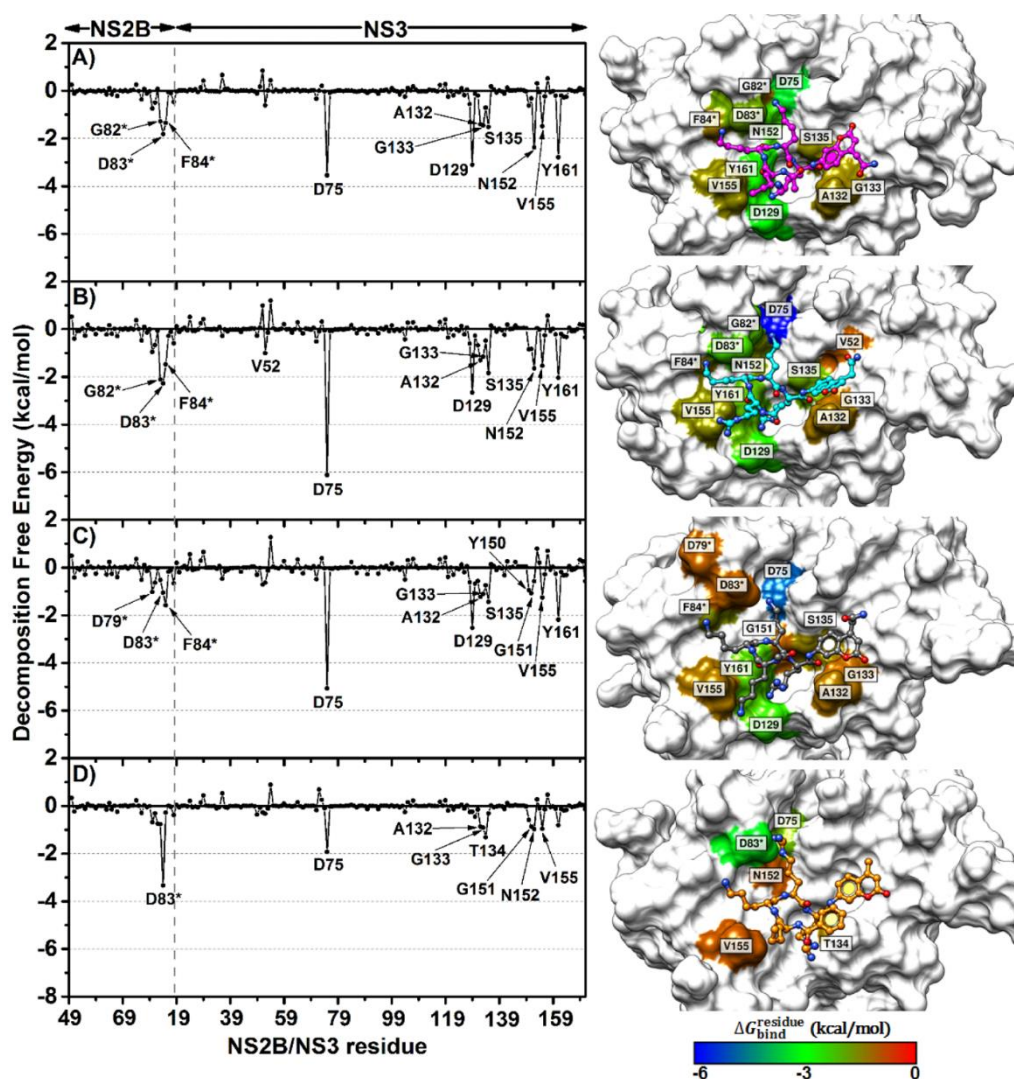


Figure 23. Per-residue decomposition free energy (in kcal mol⁻¹) calculated with MM/PBSA method for the ZIKV protease in complex with (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC, where the amino acids involved in substrate binding are shaded based on their $\Delta G_{\text{bind}}^{\text{residue}}$ values. The lowest and highest energies are ranged from blue to red, respectively. Note that other unimportant residues are illustrated in the white surface. The residues corresponding to the NS2B protein are indicated by asterisks.

To further evaluate the energetic contributions, the degrees of stabilization and/or destabilization from the twenty-four residues of NS2B/NS3 protease involved in substrate binding were separately calculated in terms of the energy contributions

from their backbone and side chain atoms, as shown in **Figure 24 (left)**. Besides, the polar ($\Delta E_{\text{ele}} + \Delta G_{\text{polar}}$) and nonpolar ($\Delta E_{\text{vdW}} + \Delta G_{\text{nonpolar}}$) interactions were given in **Figure 24 (right)**. The negative and positive values are associated with the substrate stabilization and destabilization, respectively. The results suggest that these residues tend to stabilize all substrates mostly through their side chains, as can be seen from the free energy highly contributed from the side chain atoms in all complexes (dark grey bars in **Figure 24, left**). The major energy contribution for binding of substrates with each residue comes from the nonpolar term. However, this is except for polar residues (*e.g.* D83 of NS2B and D75 of NS3), where the polar contribution plays a significant role for substrate binding. This is because these polar residues can form hydrogen bonds with each position of substrates via the attractive electrostatic interaction (discussed in **section 4.1.3**).



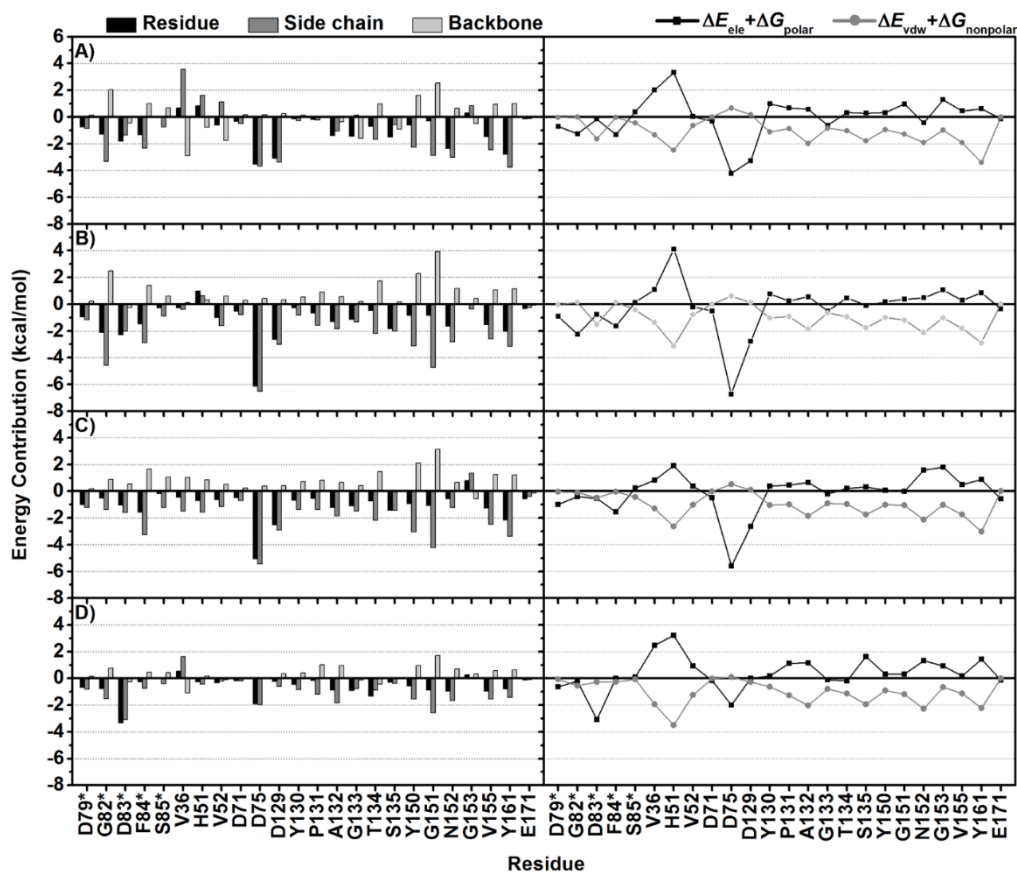


Figure 24. (Left) Per-residue decomposition free energies are given as total energy (black bars), side chain (dark grey bars) and backbone contributions (light gray bars) for the four studied complexes of the ZIKV protease with (A) Ac-nKKR-ACC, (B) Ac-DRKOR-ACC, (C) Ac-DKKOR-ACC and (D) Bz-nKRR-AMC. (Right) The electrostatic and vdW contributions are shown in black and gray lines, respectively. The residues corresponding to the NS2B protein are indicated by the asterisk.

For further design of the antiviral agents against the ZIKV protease based on our calculations, it suggests here that the P1 and P2 positions of the substrate are the most important parts for interacting the ZIKV protease via favorable electrostatic interactions with the residues at the S1 and S2 subsites. Consequently, peptidomimetic inhibitors can be designed based on the P1 and P2 sites and should present a positively charged group to resemble the side chain of basic amino acids, for optimal interaction at the S1 pocket particularly with D129 of NS3 and also with D83 of NS2B and D75 of NS3 at the S2 pocket. Nevertheless, it should be noted that the addition of the long side

chain with bulky group at the P2 position might cause the loss of interaction with D75 due to the steric effect like in the case of substrate 4. In addition, the P3 and P4 residues do not significantly interact with the NS2B/NS3 residues. Therefore, the side chain of the P3 and P4 sites can be substituted by another short side chain, such as G or A to reduce the molecular weight. These strategies recently lead to the design of dipeptide inhibitor (acyl-KR-aldehyde) derived from the P2 and P1 positions of ZIKV's substrate, indicating a potent competitive inhibitor toward the ZIKV protease [78]. In summary, the effective inhibitors for the ZIKV protease should contain a cationic group at the P1 and P2 positions. The side chain of P3 and P4 can be changed into a shorter side chain. Moreover, the choice of the side chain length at the P2 position should be selected with caution.

4.1.5 Binding free energy of protein-substrate complexes

To estimate the binding free energies of the four different substrates in complex with the ZIKV protease, the MM/PBSA method implemented in AMBER16 was performed. Based on this approach, it combines interaction energies in gas phase and the solvation free energy calculations. The electrostatic contribution to the free energy of solvation is calculated with the Poisson-Boltzmann (PB) model, whereas the nonpolar solvation term is assessed by the solvent accessible surface area (SASA). In general, the binding free energy of the complex can be estimated from difference of free energy between the complex, protein and substrate. The energetic contributions comprise the interaction energies in gas phase (ΔE_{MM}) calculated from the summation of the electrostatic (ΔE_{ele}) as well as van der Waals interactions (ΔE_{vdW}), free energy of solvation (ΔG_{solv}) and entropic contribution ($T\Delta S$). The computed binding free energies (ΔG_{bind}) and their corresponding energy contributions of all complexes averaged over 100 frames from the last 20 ns are shown in **Table 4**.

Table 4. The results from MM/PBSA method giving the energy components and average binding free energies (in kcal mol⁻¹) for the complexes of the four substrates with the ZIKV protease compared with the experimental K_m (in μ M).

	Ac-nKKR-ACC	Ac-DRKOR-ACC	Ac-DKKOR-ACC	Bz-nKRR-AMC
ΔE_{ele}	-657.83 ± 40.68	-774.83 ± 34.80	-744.52 ± 33.93	-511.84 ± 33.79
ΔE_{vdW}	-58.25 ± 5.22	-57.68 ± 4.84	-56.76 ± 3.83	-57.39 ± 3.86
ΔE_{MM}	-716.08 ± 39.87	-832.52 ± 35.00	-801.29 ± 34.14	-569.23 ± 33.41
$\Delta G_{\text{solv}}^{\text{ele}}$	657.17 ± 36.63	765.24 ± 31.12	742.24 ± 31.89	517.94 ± 31.00
$\Delta G_{\text{solv}}^{\text{nonpolar}}$	-5.91 ± 0.27	-5.97 ± 0.20	-6.01 ± 0.16	-5.54 ± 0.26
ΔG_{solv}	651.26 ± 36.62	759.26 ± 31.11	736.23 ± 31.85	512.39 ± 30.88
ΔH	-64.82 ± 6.70	-73.25 ± 7.20	-65.05 ± 6.16	-56.84 ± 5.40
$-T\Delta S$	32.32 ± 4.79	30.29 ± 7.96	35.10 ± 5.11	30.83 ± 5.06
ΔG_{bind}	-32.51 ± 8.24	-42.96 ± 10.73	-29.95 ± 7.88	-26.01 ± 7.40
K_m (μ M)	13.85 ± 0.30 [95]	4.20 ± 0.40 [95]	7.99 ± 0.73 [95]	20.42 ± 5.26 [43]

By considering the ΔE_{MM} term, the main contribution to the binding of protein-substrate complexes is the attractive electrostatic interaction (ΔE_{ele} in **Table 4**). Meanwhile, the van der Waal term plays a minor role in binding of the substrates. The calculated ΔG_{bind} values of substrates 1 (Ac-nKKR-ACC), 2 (Ac-DRKOR-ACC), 3 (Ac-DKKOR-ACC) and 4 (Bz-nKRR-AMC) are -32.51 , -42.96 , -29.95 and -26.01 , respectively. These predictions suggest a strong binding of the studied substrates with the ZIKV protease in all complexes. However, the ZIKV NS2B/NS3-substrate 2 complex shows the most preferential binding, while the substrate 4 has the lowest binding affinity, in reasonable agreement with the previously reported experimentally derived K_m values [43, 95]. The reduction in the ΔG_{bind} value of substrate 4 may be due to the occurrence of bad contacts between P2(R) and the residues in the S2 pocket, as previously mentioned in the **sections 4.1.3 and 4.1.4**. For substrates 1 to 3, the substitution of P2 with a shorter amino acid (substrate 1: R2 \rightarrow K2, substrates 2 and 3: R2 \rightarrow O2) results in a decreased ΔG_{bind} by c.a. 4-15 kcal mol⁻¹, compared with the

substrate 4 complex. It is worth noting that the binding free energies calculated with the MM/PBSA method do not provide an absolute binding free energy value compared to the experimental ones, but it is still useful for giving the trend of a relative binding efficiency of individual substrate toward the ZIKV protease. In addition, the positive values of solvation free energy (ΔG_{solv}) calculated with PB model are varied for each system and play a role in penalization of the total interaction energies of protein-substrate complexes. It should be noted that ΔG_{bind} and its energetic components of all systems calculated with the MM/GBSA method are also calculated and given in **appendix 1**. However, the results from this method do not give the significant difference in binding affinity of each peptide substrate based on the comparison with the MM/PBSA method as mentioned above.

4.2 Part II: Reaction mechanism of the ZIKV protease with its substrate

4.2.1 MM MD Simulations

To monitor the structural stability of the three independent MD simulations, the root-mean-square deviation (RMSD) relative to the initial minimized structure for the backbone atoms of protein was evaluated and plotted along the simulation time (see **Figure 25**). The results show that the protein is stable and shows similar behavior in the three independent replicas (average RMSD of 1.5 Å). Upon substrate binding, the active site (residues within 4 Å of substrate) does not show significant structural changes caused by ligand binding, in comparison with the apo form (average RMSD value of 1.0 Å for all heavy atoms of residues in the active site). In addition, the structure of the apo state stays in a closed conformation over the course of the simulation (1 μs), and exhibits a stable substrate-binding pocket, which is very similar to the X-ray crystal structure of ZIKV protease as a free enzyme [50]. To evaluate the structural compactness at the protein active site, radius of gyration (R_g) of amino acids within 4 Å of substrate was calculated. The result shows relatively constant R_g values (average value of 8.7 Å) for both forms, reflecting a closed conformation of ZIKV protease (see **Figure 26**). Representative structures for each replica in the bound state and apo state (obtained by cluster analysis) of the binding pocket are superimposed to illustrate the closed conformation of enzyme with and without substrate bound are

shown in **Figure 27**. Our MD simulation results support the closed conformation of ZIKV protease in both apo and bound forms that evidently found in the experiments based on the X-ray structures with and without peptide/inhibitor bound [50, 78].

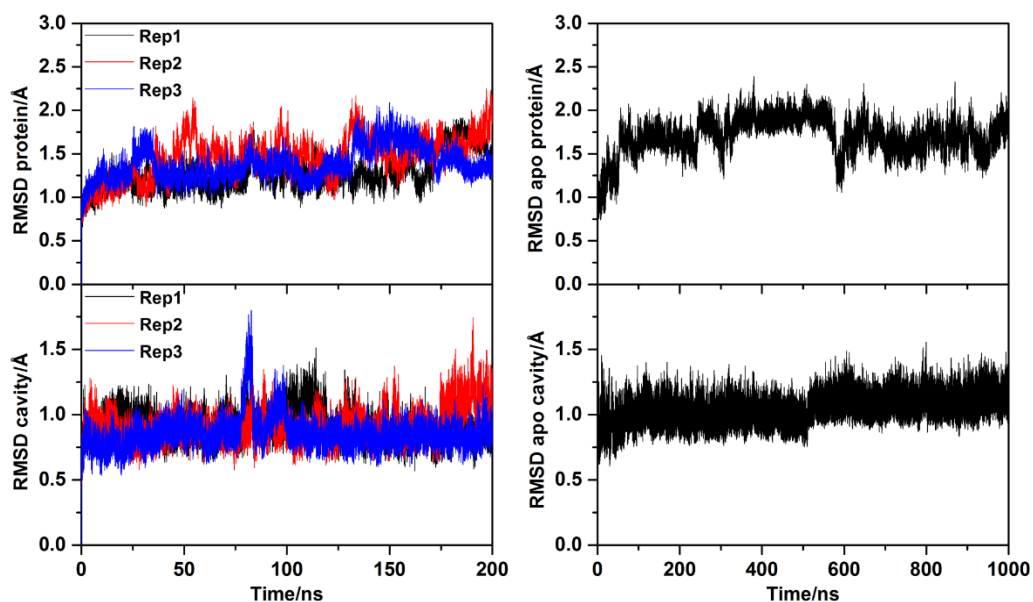


Figure 25. (Top panel) Time evolution of the RMSD (\AA) of backbone atoms of all three replicate simulations of bound and apo systems. (Bottom panel) RMSD evolution (\AA) of the heavy atoms of residues in the active site (defined as residues within 4\AA of substrate) in bound state (left) and apo state (right).

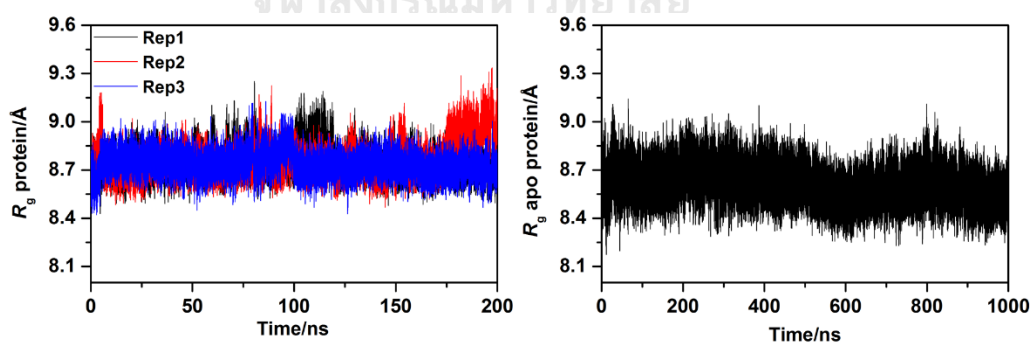


Figure 26. Time evolution of the radius of gyration (R_g) of residues within 4\AA of substrate for all three replicate simulations of bound (left) and apo (right) systems.

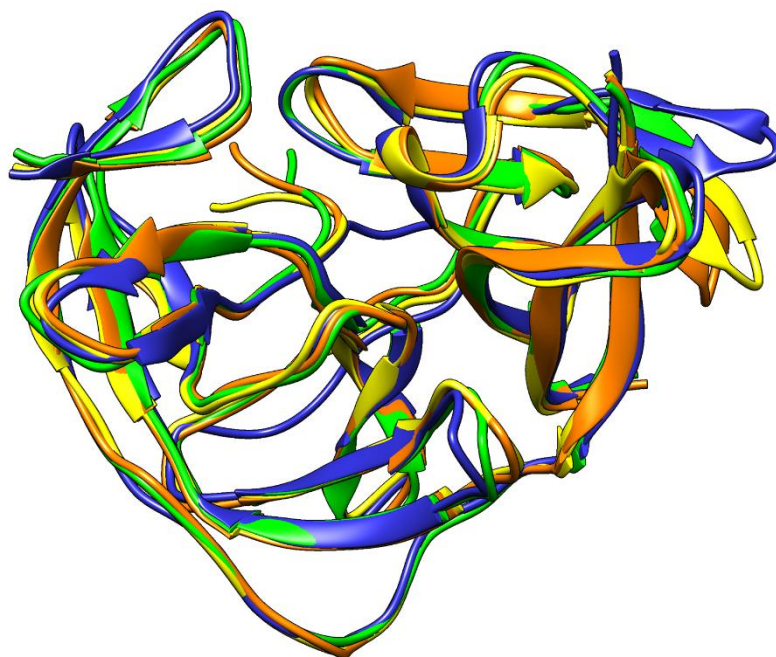


Figure 27. Superposition of the representative structures for each replica in bound state and apo state (obtained by cluster analysis) of the binding pockets: replica 1 (yellow), replica 2 (green), replica 3 (blue) and apo form (orange).

4.2.2 Distances at the catalytic triad and oxyanion hole

The distances of the catalytic residues H51, D75 and S135 ($d1-d3$), nucleophilic attack ($d4$) and oxyanion hole constructed by the backbone amides of G133 and S135 ($d5$ and $d6$), are illustrated in **Figure 28A**. Histograms of these important distances ($d1-d6$) are also plotted in **Figure 28B**, where the data presented here is a combination of all three independent MD simulations collected from the last 100 ns of each replica. The histograms for the distances of $d1$ and $d2$ show a peak in the population at ~ 3.0 Å for $d1$ and ~ 2.7 Å for $d2$ (**Figure 28B**), as monitored from the carboxylate oxygen atoms (O1 and O2) of D75 to the N1-imidazole nitrogen of H51, indicating the presence of two hydrogen bonds. Similar to the X-Ray structures of WNV NS2B/NS3 protease in complex with peptide inhibitors, the O1 atom was located closer to the N1 atom than the O2 atom ($d1$ of ~ 3.3 Å and $d2$ of ~ 2.8 Å) [45, 171]. The histogram of the distance between the N2-imidazole nitrogen of H51 and the O3-hydroxyl oxygen of S135 ($d3$) displays a peak at ~ 2.7 Å, involving the forming of a strong hydrogen bond and feasibly

facilitating the proton transfer from S135 to H51. Besides, the nucleophilic attack involving the first step of the acylation process is defined as the distance between the hydroxyl oxygen atom (O3) of S135 and the carbonyl carbon (C_{γ}) of the scissile peptide P1(R) position ($d4$). Based on the histogram for this distance, there is a maximum in the population at ~ 3.0 Å. This finding demonstrates that all three catalytic triad residues and the substrate are in an appropriate configuration for initiating the cleavage reaction.

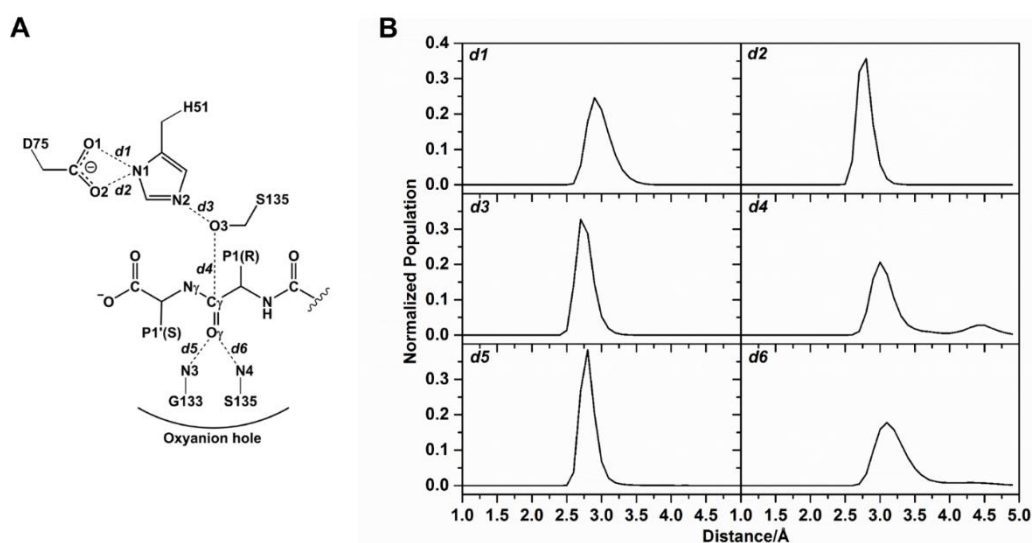


Figure 28. (A) Definition of the interatomic distances involved in the cleavage reaction ($d1$ – $d6$). (B) Histograms of key interatomic distances, $d1$ – $d6$, from MM MD simulations, sampling from 100 to 200 ns.

The oxyanion hole of the flavivirus NS2B/NS3 protease is formed by the backbone amides of residues G133 and S135, which interact with the O $_{\gamma}$ -carbonyl oxygen of the P1(R) reacting residue in the cleavage reaction ($d5$ and $d6$). The histograms of these two distances indicate a maximum in the population at ~ 3.1 Å for $d5$ and ~ 2.8 Å for $d6$, showing stable interactions. Altogether, the overall substrate conformation and interatomic interactions present in the active site of ZIKV NS2B/NS3 protease over the course of MD simulations are relatively similar to those of the WNV NS2B/NS3 X-ray structures [45, 171] and our previous MD simulations of DENV type 2 NS2B/NS3 [170]. This demonstrates that the simulation models are stable.

In addition to these six important distances, the angles related to nucleophilic attack, defined as θ_x and θ_y , are measured. The histograms of both angles show a peak

with the highest population at about 98° and 82° for θ_x and θ_y , respectively (**Figure 29**). These two calculated parameters that describe the nucleophilic attack had been introduced from the average value of superimposed 79 serine protease complexes with inhibitors, consistently close to 90° [173]. Additionally, the molecular geometry found in these protease-inhibitor complexes might be considered as a universal hypothesis of good models for the reactive MC, resulting in rapid progression to the AE. Therefore, both parameters, $d4$ and θ_y , would be helpful to guide in selecting an initial structure for the QM/MM simulations in the next step (see **Chapter III Research Methodology**).

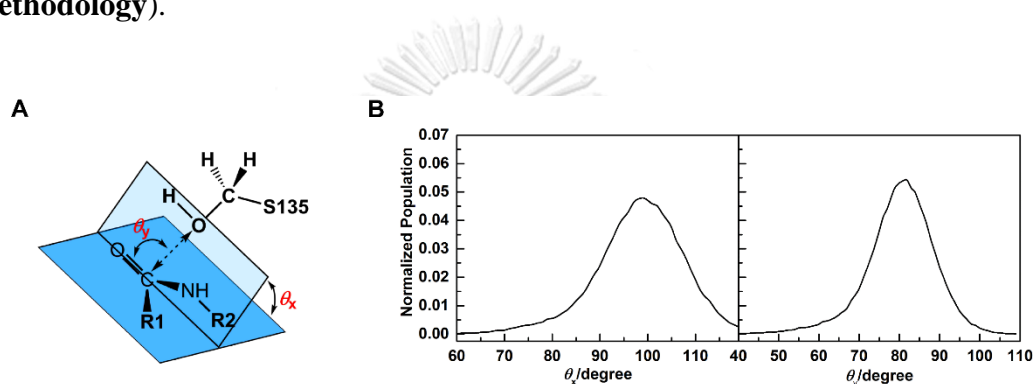


Figure 29. (A) The geometric parameters describing the nucleophilic attack trajectory are depicted. θ_y is the angle defined by the enzyme serine oxygen, the carbonyl carbon and the carbonyl oxygen on the substrate. θ_x is the angle between (i) the plane defined by the enzyme serine oxygen, the substrate's carbonyl carbon, and the substrate's carbonyl oxygen, and (ii) the plane defined by the peptide bond. The dashed line represents the distance between the enzyme serine oxygen and the substrate's carbonyl carbon. (B) Histograms of these two geometric parameters, θ_x and θ_y , from three independent MD simulations, sampling from 100 to 200 ns.

4.2.3 PM6/MM free energy profiles

QM/MM approaches can be used with enhanced sampling methods, e.g. the commonly used umbrella sampling technique, to calculate activation free energies in good agreement with experimental data [174-177]. Herein, we applied QM/MM (PM6/ff14SB) MD umbrella sampling simulations to model the reaction involving the acylation step catalyzed by the ZIKV protease and to map the corresponding free energy surfaces (potential of mean force, PMF). The free energy profiles consist of the TI

formation and the breakdown of peptide bond, which are calculated by iteratively changing the value of a RC (typically defined as a combination of relevant interatomic distances; see **Chapter III Research Methodology**). The computational time for extensive sampling is very costly, the use of semiempirical level is more practical. In this study, we selected the PM6 semiempirical method to treat the QM region. PM6 has been successfully applied to study reaction mechanisms involved in enzyme-catalyzed reactions [178-180]. Moreover, our preliminary results show that the energy barrier obtained at the PM6/ff14SB level of theory is in more reasonable agreement with the experimental activation free energy compared to PM3, AM1-d and SCC-DFTB results. Note that the free energy profiles obtained at the PM3/ff14SB, PM6/ff14SB, AM1-d/ff14SB and SCC-DFTB/ff14SB levels of theory are compared and given in **appendix 2**. The resulting free energy profiles at PM6/ff14SB level are depicted in **Figures 30 and 31**.

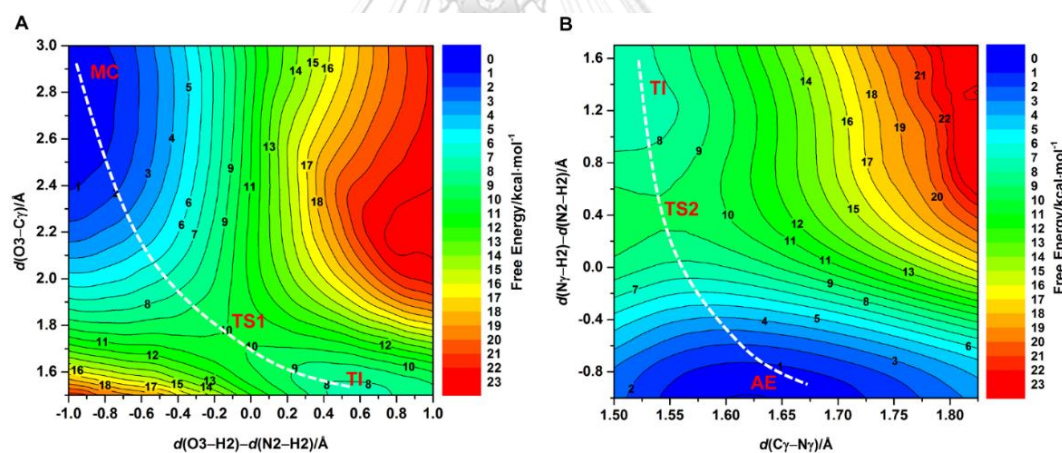


Figure 30. (A) 2D free energy surfaces for the TI formation and (B) the breakdown of the peptide bond for the reaction of the ZIKV protease with its substrate, calculated from umbrella sampling simulations at the PM6/ff14SB level of theory.

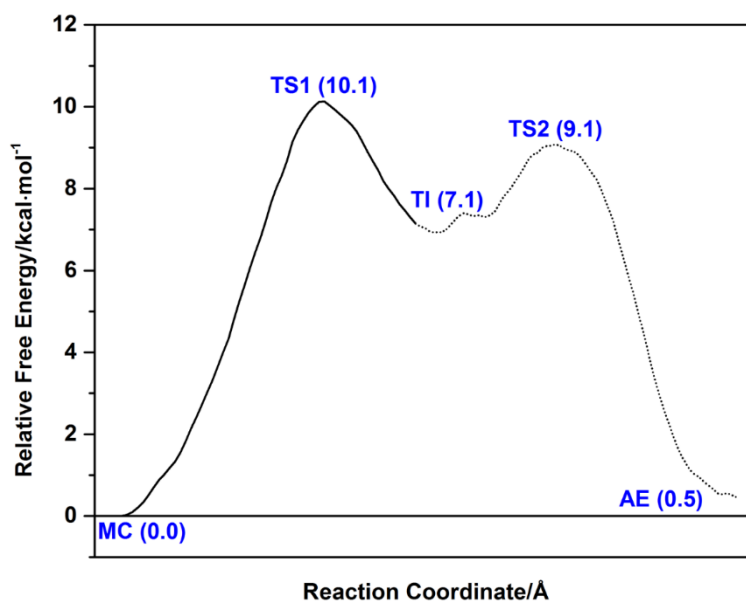


Figure 31. 1D free energy profiles for the TI formation and the breakdown of the peptide bond relative to the MC for the reaction of the ZIKV NS2B/NS3 serine protease with its substrate using the combined RCs at the PM6/ff14SB level of QM/MM theory.

The 2D free energy surfaces for the corresponding reaction are calculated along the RCs. The minimum energy path (MEP) [181-183] on the surface for the first step of acylation (**Figure 30**) shows an activation free energy (ΔG^\ddagger) of 10.1 kcal mol⁻¹ relative to the MC (TS1; **Table 5**). The apparent experimental activation free energy is ~ 18.2 kcal mol⁻¹ (k_{cat} of 0.95 s⁻¹ at 37 °C) [184]; hence, the PM6/ff14SB reaction barrier is too low compared to the experimental data. This is due to limitations of the PM6 semiempirical QM method for this reaction, as shown by higher level QM/MM calculations (see below) [185]. Afterward, this TS further evolves to the TI formation with a lower barrier of 7.1 kcal mol⁻¹ relative to the MC. Only one TS is found in the first step of the acylation, i.e. the reaction occurs in a concerted manner [186], in which the transfer of proton from S135 to H51 and the nucleophilic addition of S135 oxygen to the carbonyl carbon on P1(R) of the substrate occur together. This is similar to the *ab initio* QM/MM results from MD-FEP calculations obtained by Kato and Ishida on trypsin [186]. Moreover, the concerted reaction mechanism of the TI formation has been studied extensively by theoretical investigations on other serine proteases [55, 186-190]. However, this is in contrast to the QM/MM (PDDG-PM3/ff99SB) study suggesting a stepwise mechanism of the TI formation catalyzed by DENV type 2

NS2B/NS3 serine protease, in which the nucleophilic attack takes place only after the proton transfer [191]. Thus, we additionally performed QM/MM MD umbrella sampling simulations using only one RC involving the proton transfer from S135 oxygen to H51 ϵ nitrogen ($d(\text{O3-H2})-d(\text{N2-H2})$) in the MC. During the proton transfer from S135 to H51, the nucleophilic attack on the substrate happens spontaneously. Therefore, this is a further evidence to support a concerted mechanism. The resulting free energy profile at the PM6/ff14SB level has an activation barrier of 9.9 kcal mol⁻¹ (See **Figure 32**), which is slightly lower than that of the energy barrier obtained by using combination of the interatomic distances at the same level of theory (**Figure 31**). Nevertheless, Warshel and colleagues suggested a stepwise reaction with prior proton transfer in the MC [192]. Indeed, the question “How do serine proteases really work?” [193], is still open, and the exact sequence of steps is under debate.

Table 5. Free energies, relative to the MC, obtained at the PM6/ff14SB QM/MM level from two-dimensional umbrella sampling MD simulations (i.e. from the free energy surfaces shown in **Figure 30**).

	Free energy (kcal mol ⁻¹)
MC	0.0
TS1	10.1
TI	7.1
TS2	9.1
AE	0.5

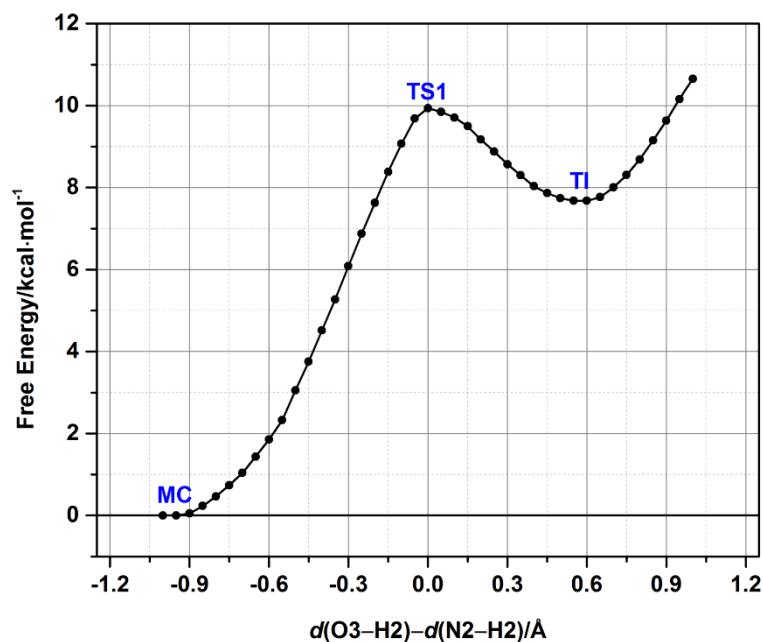


Figure 32. 1D free energy profiles for the tetrahedral intermediate formation using one reaction coordinate involving in the proton transfer from S135 oxygen to H51 ϵ nitrogen in the Michaelis complex.

Using the combination of the interatomic distances, **Figure 31** shows 1D free energy profiles obtained from QM/MM MD 1D umbrella sampling simulations for both steps of the acylation. All of the intermediates and reaction energy barriers observed in 2D free energy surfaces are apparent. The free energy profile for the first step of the acylation has the same energy barrier of 10.1 kcal mol⁻¹ with 2D free energy surfaces at the RC of ~ 1.8 Å. The distance calculations also support that the first step of the acylation process takes place in a concerted manner, as can be seen from a simultaneous decrease in $d(\text{N2-H2})$ and $d(\text{O3-C}\gamma)$ involving the distances of the proton transfer and nucleophilic attack, respectively (**Figure 33A**). Moreover, the alterations of the two hydrogen bond distances between H51 and D75, defined as $d(\text{O1-H1})$ and $d(\text{O2-H1})$, are monitored along the RC. The calculated distances show that $d(\text{O1-H1})$ increases from ~ 2.0 up to ~ 2.4 Å at the RC accounting for TS1, while $d(\text{O2-H1})$ shortens from ~ 2.0 to ~ 1.6 Å with the progress of the reaction (**Figure 33B**). This finding is described by the conformational change of the imidazole ring of H51 that prepares its geometry to accept the proton from S135. The rotation of the H51 imidazole ring also results in the formation of two strong hydrogen bonds with the carboxylate group of D75. Apart

from these two hydrogen bonds, other hydrogen bonds are also detected in the oxyanion hole. In the MC complex, two hydrogen bonds are formed between the carbonyl oxygen on the scissile substrate and the NH group of G133 and S135, defined as $d(\text{NH}_{\text{G133}}-\text{O}_\gamma)$ and $d(\text{NH}_{\text{S135}}-\text{O}_\gamma)$ (**Figure 33B**). As the reaction proceeds, both distances continuously decrease along the RC. They remain approximately constant at ~ 1.9 Å to stabilize negative charge centered on the carbonyl oxygen when the TI is formed (for comparison: the Mulliken charges of the carbonyl oxygen are -0.69 and -0.93 a.u. in the MC and the TI formation, respectively, **Table 6**). In addition, as expected, the Mulliken charge analysis revealed that the electronic properties of the $\text{N}_{2\text{H51}}$ atom change from basic to acidic after receiving the proton from S135. On the other hand, as the $\text{N}_{2\text{H51}}$ becomes more acidic (more positive charge), the Mulliken charge on the substrate's N_γ atom decreases. This event plays an important role for the peptide bond breaking in the next step, where the H2 hydrogen is transferred from $\text{N}_{2\text{H51}}$ to the N_γ nitrogen on the TI. However, the atomic charge of the $\text{O}_{3\text{S135}}$ atom remains approximately constant (around -0.6 a.u.) suggesting that the reaction proceeds by a concerted pathway rather than by a stepwise manner as mentioned previously. Also, this information is in good agreement with the single TS found in the 1D and 2D free energy profiles.

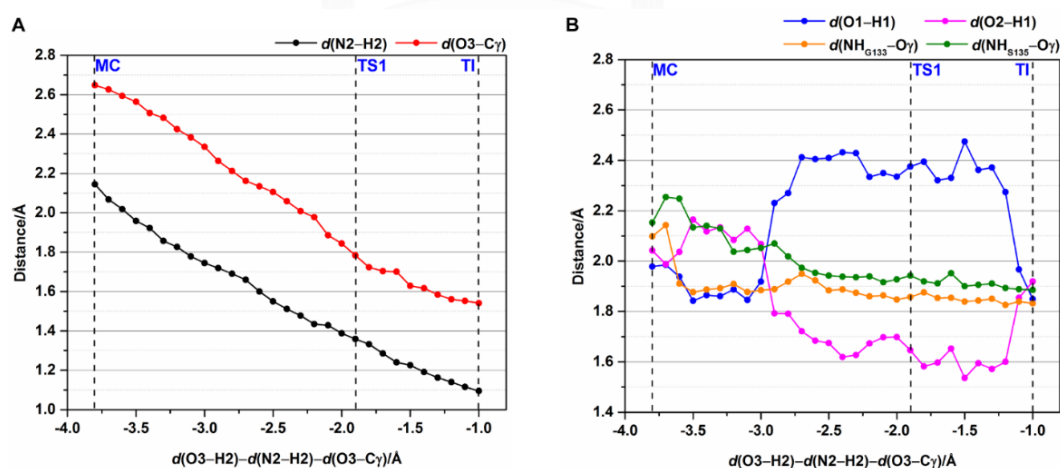


Figure 33. Average values of (A) the distances involving proton transfer and nucleophilic attack and (B) the hydrogen bond distances between QM and MM atoms along the reaction pathway.

Table 6. Average Mulliken charges (PM6/ff14SB) in atomic units for atoms involving the formation of the TI, averaged from the QM/MM free energy profile.

Relevant atom	Mulliken charge		
	MC	TS1	TI
N γ	-0.45	-0.56	-0.61
O γ	-0.69	-0.82	-0.93
C γ	0.61	0.72	0.74
N2	-0.40	-0.34	-0.18
H2	0.35	0.41	0.38
O3	-0.60	-0.64	-0.63

The free energy profile for the second step of the acylation process connecting the TI to the AE product is shown in **Figures 30B** and **31**. The simulations show that the H2 atom is transferred from N2_{H51} to the substrate's N γ , and thereafter the peptide bond breaks. The calculated distances along the reaction pathway evidently show that $d(\text{N}\gamma\text{-H}2)$ continuously decreases indicating the progression of the proton transfer to N γ , while $d(\text{C}\gamma\text{-N}\gamma)$ involving peptide bond breakdown slightly increases up to ~ 1.6 Å when the AE product is formed (**Figure 34**). The energy difference between TI and TS2 is considered as the energy barrier for this step and is only 2.0 kcal mol⁻¹. This value is lower than that of the barrier height observed in the TI formation, such that the TI formation is the rate-limiting step for the acylation process.

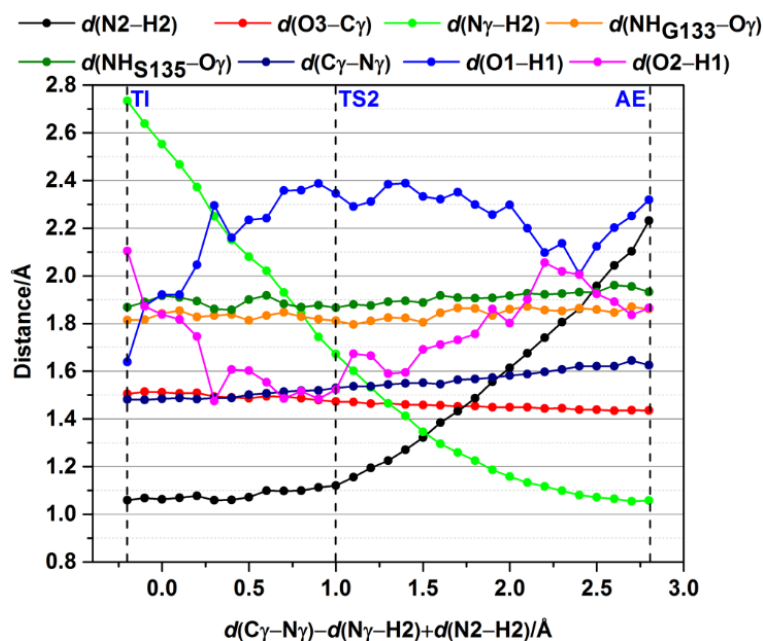


Figure 34. Average values of the distances along the reaction pathway involving the second step of acylation process. Interatomic distances are shown in different colors, as indicated.

During the peptide bond breaking process, the bond between $O3_{H51}$ and $C\gamma$ of the substrate becomes stronger, which is shown by the acyl-enzyme bond, $d(O3-C\gamma)$, approaching its lowest value of 1.41 Å upon reaching the AE product (**Figure 34**). Furthermore, as the reaction progresses, hydrogen bonds between $H51$ and $D75$ show the similar situation as found in the first step of acylation involving the rotation of the imidazole ring of $H51$ to facilitate the proton transfer from $N2_{H51}$ to the $N\gamma$ nitrogen on the substrate. The calculated distances demonstrate that $d(O1-H1)$ increases up to ~2.4 Å, whereas $d(O2-H1)$ decreases from ~2.1 to ~1.5 Å. In this event, the position of $H2$ attached to $N2_{H51}$ atom is in close proximity to the $N\gamma$ nitrogen of substrate, allowing easy transfer the proton. The other two hydrogen bonds formed in the oxyanion hole are maintained along the reaction pathway, shown by the $d(NHG133-O\gamma)$ and $d(NHS135-O\gamma)$ distances of around 1.9 Å. This indicates that the AE formation is still stabilized by both hydrogen bonds. It is worth noting that structural changes are also found in the carbonyl group on the acyl-portion. As the reaction progresses, the carbonyl bond length becomes shorter in AE with the similar value to that found in MC complex (~1.24 Å), reflecting the conversion of $C\gamma$ hybridization from a sp^3 (the

carbonyl bond length of ~ 1.29 Å at the TI stationary point) to a sp^2 hybridization upon AE product.

4.2.4 Potential energy surface for reaction and effect of basis set size

To examine the influence of the basis set on the energy profiles for the first step of the acylation process, single point QM/MM energy calculations were performed on the BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries (see below) using a variety of polarization and diffuse functions, as well as a larger basis set for the DFT method (6-311++G(d,p)) (**Figure 35**). The results show that the energy profiles are very similar in terms of shape, reaction energy barriers and energies of the TI formation, except for 6-31G. The BH&HLYP-D3 energy profiles calculated with different basis sets, including the 6-31G(d), 6-31G(d,p), 6-31+G(d), 6-31++G(d,p) and 6-311++G(d,p) are all similar, with energy differences within 1.5 kcal mol⁻¹ at transition states (TSs). The addition of polarizable functions (i.e., 6-31G(d) and 6-31G(d,p)) is clearly important for correct description of reaction energetics, changing the energies of the TS and TI by ~ 9 kcal mol⁻¹, compared to the calculation with 6-31G basis set. However, additional inclusion of diffuse functions (i.e., 6-31+G(d) and 6-31++G(d,p)) does not significantly change the calculated energies. Thus, the inclusion of only polarization functions is sufficient for giving the reasonable BH&HLYP-D3 energy profiles in this case; hence, the 6-31G(d) basis set provides a good choice for geometry optimization, balancing accuracy and computational cost.

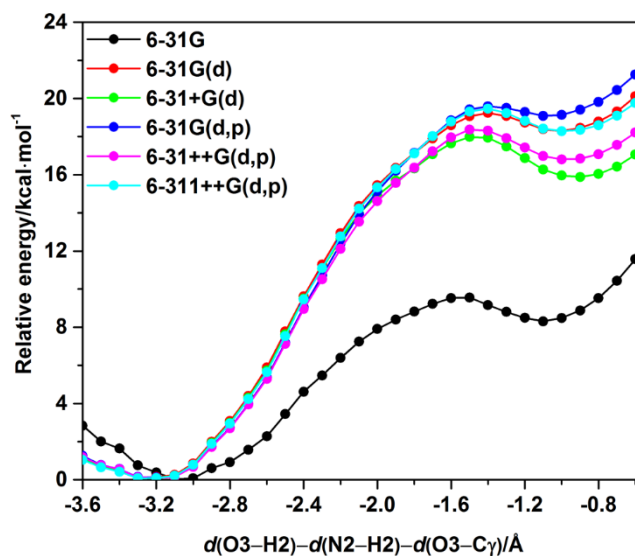


Figure 35. QM/MM energy profiles for the first step of the acylation process along the reaction pathway obtained at BH&HLYP-D3 QM/MM energy calculations using different basis sets for snapshot 1.

4.2.5 Conformational sampling and *ab initio* energy calculations

Using TS-like conformations from the 1D PMF profile as starting points, potential energy profiles were calculated with an adiabatic mapping procedure along the RC at the BH&HLYP-D3/6-31G(d)/ff14SB level of theory. Afterward, *ab initio* QM/MM (LMP2/(aug)-cc-pVTZ/ff14SB, SCS-LMP2/(aug)-cc-pVTZ/ff14SB and LCCSD(T)/(aug)-cc-pVTZ/ff14SB) single point energy calculations were used to calculate higher-level potential energy surfaces (PESs).

As can be seen from the PESs, there are two local energy minima corresponding to the MC complex and the TI with a single approximate TS along the MEP at all levels of QM/MM theory (**Figure 36**). This indicates that this step of the acylation occurs in a concerted manner, as also found in the PM6/MM free energy profiles. To account for conformational sampling, the results were averaged over multiple pathways using five different starting structures. Based on the QM/MM optimization at the BH&HLYP-D3/6-31G(d)/ff14SB level (**Figure 36A**), the average activation energy barrier is 19.3 kcal mol⁻¹ (**Table 7**). The location of energy barrier is shifted to the product side at the RC of ~ -1.5 Å at higher level QM/MM calculations in comparison with the location of barrier obtained at the PM6/ff14SB level (~ -1.8 Å). All structures for each stationary point (i.e., MC, TS and TI structures) were superimposed on top of the QM/MM

structures calculated at the BH&HLYP-D3/6-31G(d)/ff14SB level to verify the consistency of the five calculated pathways (**Figure 37**). All five of the optimized MC structures show similar configurations of the substrate and catalytic and important residues in the active site of enzyme. As the reaction proceeds, the transfer of proton from S135 to H51 and the nucleophilic attack of S135 oxygen to the carbonyl carbon on substrate occur simultaneously (see the detail in structural analysis section below). The TS and TI geometries are very similar in structure (**Figure 37**).

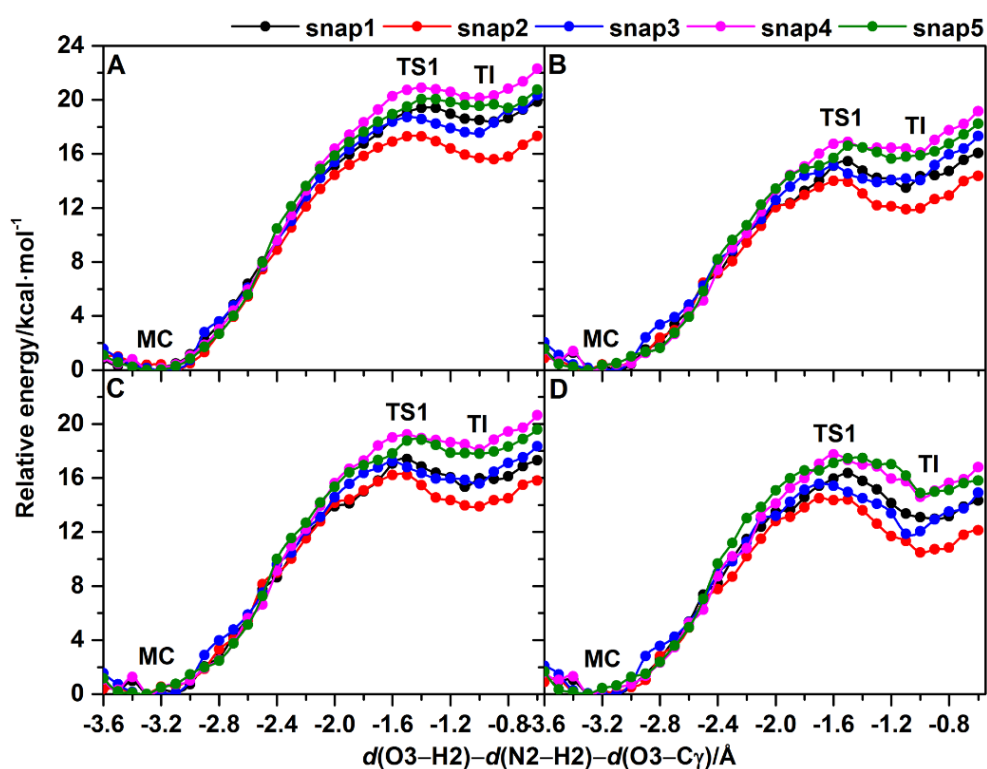


Figure 36. Potential energy profiles for the first step of the acylation process calculated at the (A) BH&HLYP-D3/6-31G(d)/ff14SB, (B) LMP2/(aug)-cc-pVTZ/ff14SB, (C) SCS-LMP2/(aug)-cc-pVTZ/ff14SB, and (D) LCCSD(T)/(aug)-cc-pVTZ/ff14SB QM/MM level. All geometries were optimized at the BH&HLYP-D3/6-31G(d)/ff144SB level.

Table 7. Energy barriers (kcal mol⁻¹) for the TI formation calculated with the BH&HLYP-D3/6-31G(d), LMP2/(aug)-cc-pVTZ, SCS-LMP2/(aug)-cc-pVTZ, and LCCSD(T)/(aug)-cc-pVTZ QM/MM methods^a

calculation method	activation energy					average ^c
	snap 1 ^b	snap 2	snap 3	snap 4	snap 5	
BH&HLYP-D3	19.4	17.3	18.7	20.9	20.0	19.3 ± 1.4
LMP2	15.5	13.4	15.1	16.9	16.6	15.6 ± 1.2
SCS-LMP2	17.4	16.2	17.2	19.2	18.8	17.8 ± 1.2
LCCSD(T)	16.4	14.5	15.6	17.7	17.5	16.3 ± 1.3

^a The L in these acronyms indicates that the calculation of local approximations was included in the ab initio methods and (aug) indicates that a basis set augmented with diffuse functions were only applied for the oxygen atoms. ^bThe conformational details of MC, TS and TI structures are presented in Figure 10. ^cThe average value is calculated through the simple arithmetic mean of 5 data points.

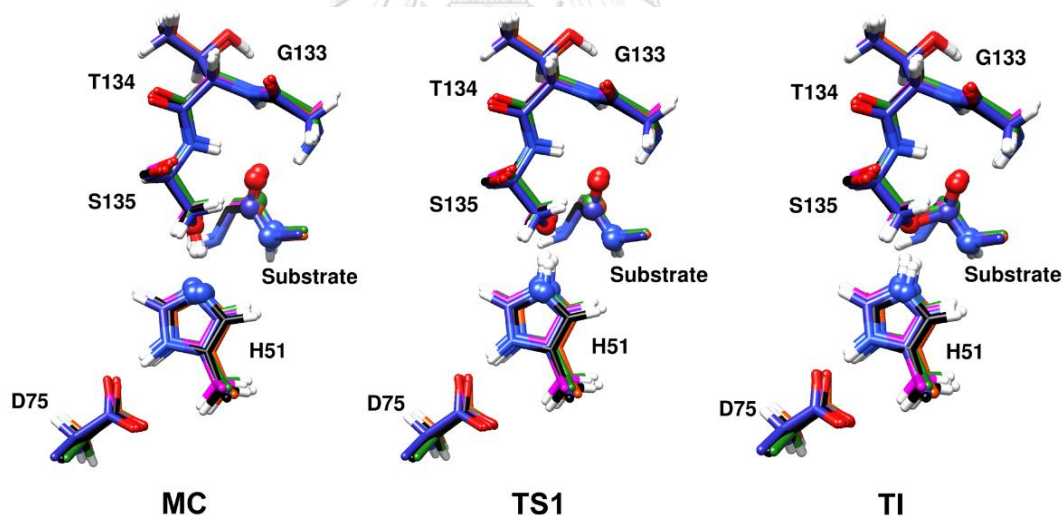


Figure 37. Superposition of the QM/MM minimized structures of the five MD snapshots at the Michaelis complex (MC), transition state (TS1) and tetrahedral intermediate (TI).

Single-point energy calculations were performed on the geometries optimized with BH&HLYP-D3/6-31G(d)/ff14SB level of theory along the reaction path using LMP2, SCS-LMP2, and LCCSD(T) methods with the Dunning correlation consistent (aug)-cc-pVTZ basis set for the QM region. The results obtained for the five different

initial structures with the correlated *ab initio* methods are shown in **Figure 37B-D** and **Table 7**. From conventional transition state theory (TST) [194], the apparent activation free energy barrier from the experiment is ~ 18.2 kcal mol⁻¹ (converted from the value of $k_{\text{cat}} = 0.95$ s⁻¹ at 37 °C) [184]. Note that the barriers here are activation potential energy that cannot be compared directly to a free energy barrier; this would require the calculation of the entropic, tunneling and zero-point energy contributions [195]. The most accurate calculations here, from first principles, are at the LCCSD(T)/(aug)-cc-pVTZ level (barrier height of ~ 16.3 kcal mol⁻¹), which is also in good agreement with experiment (from a rough estimate of these additional contributions) and we use it as a reference for the comparisons. The average barriers from LMP2/(aug)cc-pVTZ and SCS-LMP2/(aug)cc-pVTZ are close (within 1.5 kcal mol⁻¹) to the LCCSD(T)/(aug)-cc-pVTZ/ff14SB results. The LMP2 result underestimates the average activation energy barrier compared to LCCSD(T)/(aug)-cc-pVTZ level by only 0.7 kcal mol⁻¹, while SCS-LMP2 and BH&HLYP-D3 overestimate the barrier by 1.5 and 3.0 kcal mol⁻¹, respectively. Meanwhile, the relative energy of the TI shows an obviously higher reaction energy than the MC for all levels of QM/MM calculations, indicating that formation of the TI is a relatively unstable species compared to the MC, and the reaction is endothermic. The TI has average energies relative to the MC of 18.2 ± 1.8 kcal mol⁻¹, 14.2 ± 1.7 kcal mol⁻¹ and 16.1 ± 1.8 kcal mol⁻¹ with BH&HLYP-D3/6-31G(d), LMP2/(aug)cc-pVTZ and SCS-LMP2/(aug)cc-pVTZ, respectively. However, these reaction energies are larger than that of the energy obtained at the LCCSD(T)/(aug)-cc-pVTZ level (reaction energy of 13.0 ± 1.7 kcal mol⁻¹). In addition, the energy barriers found in each snapshot linearly correlate with the reaction energies (i.e. the energy of the TI relative to the MC) at all levels of QM/MM theory presented here (see **Figure 38**). We suggest here that higher-level QM calculations provide significantly less deviation of the observed activation energy barrier and reaction energy from the LCCSD(T)/(aug)-cc-pVTZ energy description of the reaction. Even though the value for the BH&HLYP-D3/6-31G(d) average reaction barrier is higher than those of the LMP2 and SCS-LMP2 methods, it is somewhat close to the SCS-LMP2 value indicating that the BH&HLYP functional in combination with dispersion correction performs well for this particular reaction. It should be noted that dispersion-corrected density functional theory (DFT-D3) is very efficient to treat general intermolecular

interactions at a reasonably accurate level and has emerged as a suitable approach for modeling the electronic structure of large molecular systems [196]. Moreover, BH&HLYP has been shown previously to give better results than the widely used B3LYP for some proton and charge transfers [137] and hydrogen bonding [138, 140]. Nevertheless, the calculation at the B3LYP-D3/6-31G(d)/ff14SB level of theory should be tested, as we do here. We tested by selecting three from five different starting structures (snapshots 1-3) to perform adiabatic mapping at this level of theory. The computed potential energy profiles show that the energy barrier and the reaction energy are not predictable (see **Figure 39**), suggesting that the calculation at the B3LYP-D3/6-31G(d)/ff14SB level fails to describe and locate the TI minimum. Additionally, the QM/MM calculations at the MP2 and SCS-MP2 levels with and without local approximations were tested to examine the effect of local approximations on their accuracy and so clarify their use at the coupled cluster level. Calculations at the (L)MP2 and SCS-(L)MP2 levels of theory with and without local approximations show that the local approximations do not significantly affect the results, with the largest difference around only ~ 0.5 kcal mol⁻¹ (see **Figure 40**). It can be inferred from this that local approximations introduce only very small errors at the LCCSD(T) level, indicating that the use of local approximations is an effective way to reduce computational cost to achieve coupled cluster accuracy.

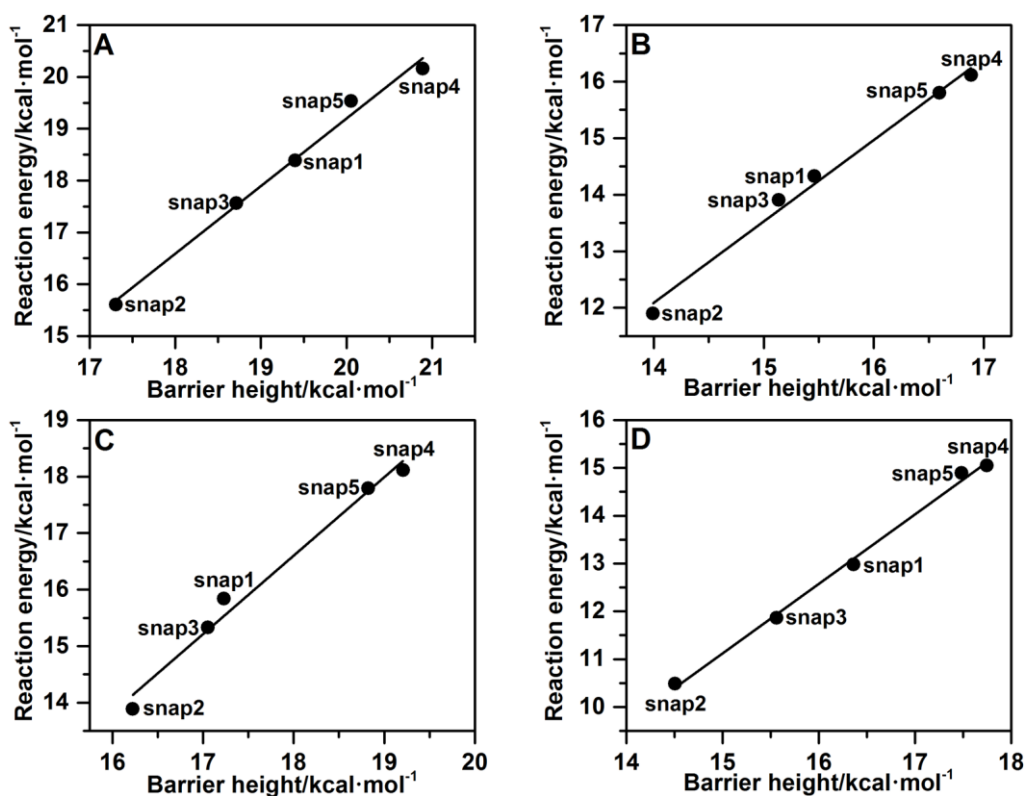


Figure 38. Correlation between energy barriers and reaction energies found in each snapshot calculated at the (A) BH&HLYP-D3/6-31G(d)/ff14SB, (B) LMP2/(aug)-cc-pVTZ/ff14SB, (C) SCS-LMP2/(aug)-cc-pVTZ/ff14SB, and (D) LCCSD(T)/(aug)-cc-pVTZ/ff14SB QM/MM levels.

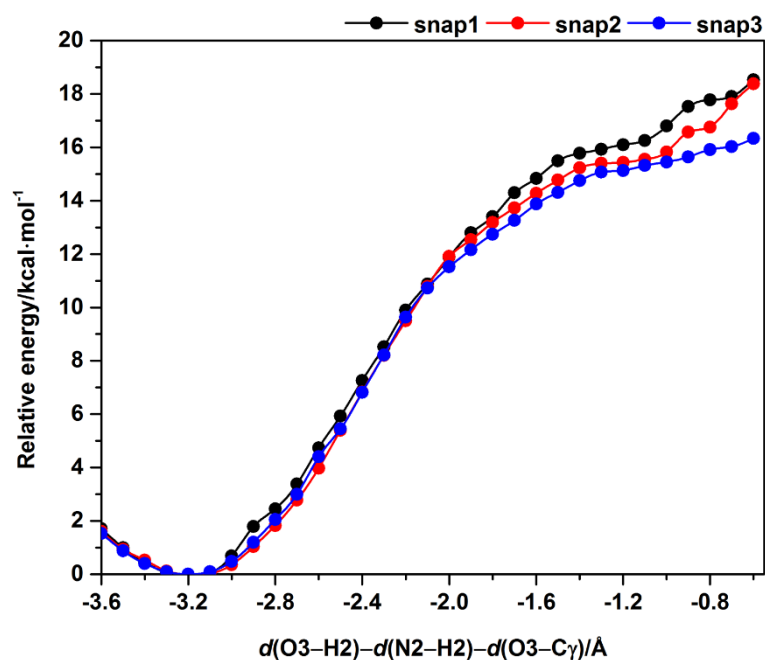


Figure 39. Potential energy profiles for the first step of the acylation process calculated at the B3LYP-D3/6-31G(d)/ff14SB level of theory.

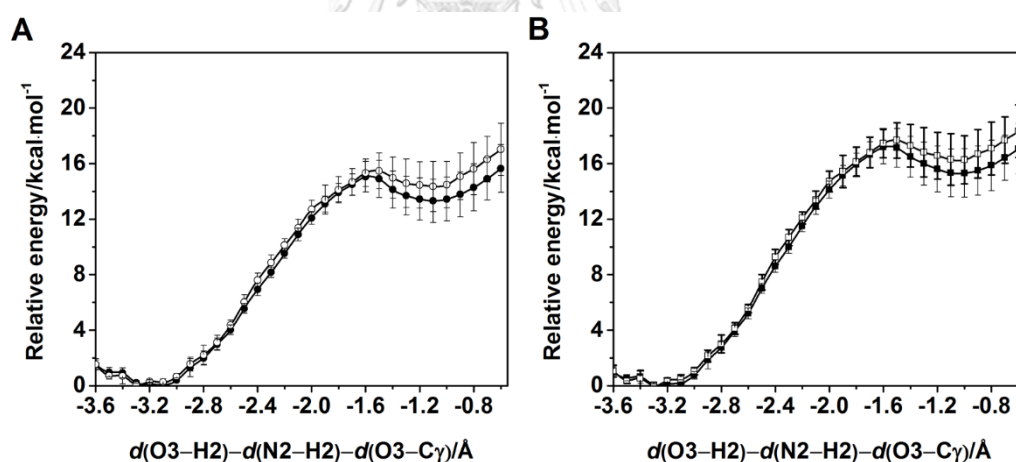


Figure 40. Comparison of the average reaction profiles (relative to the Michaelis complex) for the formation of tetrahedral intermediate calculated with MP2 and SCS-MP2 (using the BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries) with and without local approximations. (A) black circles: MP2/(aug)-cc-pVTZ/ff14SB; white circles: LMP2/(aug)-cc-pVTZ/ff14SB; (B) black squares: SCS-MP2/(aug)-cc-pVTZ/ff14SB; white squares: SCS-LMP2/(aug)-cc-pVTZ/ff14SB. Error bars indicate the standard deviation of the average energy.

The negative charge on the TI formation is produced from the substrate carbonyl oxygen atom. This high-energy species is stabilized by proton donating groups of amino acid backbone nearby, termed as the oxyanion hole which, in the case of flavivirus NS2B/NS3 protease, is formed by the backbone amide groups of G133 and S135 of the protease. It has been known that the most important effect of stability on the TI is the electrostatic interactions between the oxyanion hole region and the negatively charged carbonyl oxygen of the scissile substrate. The contributions of the oxyanion hole to electrostatic stabilization were evaluated at the SCS-LMP2/(aug)cc-pVTZ/ff1SB level. The energy contribution of the oxyanion hole region was calculated by subtracting the energy of the full QM region from the QM region without the residues G133, T134 and main chain NH of S135. The oxyanion hole region stabilizes both the TS1 and TI (on average by ~ 1.5 kcal mol⁻¹ and ~ 7.0 kcal mol⁻¹ for TS1 and TI, respectively), as shown in **Figure 41**. The residues of the oxyanion hole stabilize the TI significantly more than TS1, because charge is more localized in the TI than in the TS1, and stabilizes both relative to the neutral components in the MC complex. Besides, preliminary tests of the influence of enlarging the minimal QM region (see **Chapter III Research Methodology**) by an additional inclusion of the oxyanion hole region, QM/MM geometry optimizations on the snapshot 1 at the BH&HLYP-D3/6-31G(d)/ff14SB level were carried out by including fragments of G133, T134 and S135 in the QM region, leading to a total of 79 atoms. In the energy profiles calculated with two different sizes of the QM region (with and without residue fragments of oxyanion hole in the QM region), the TS1 and TI formation are found at almost the same values of the RC, as shown in **Figure 42**. Furthermore, as expected, the computed potential energy barrier and the reaction energy were lowered by 1.8 kcal mol⁻¹ and 2.5 kcal mol⁻¹, respectively for the larger QM subsystem, compared with the minimal QM subsystem. This is possibly due to the fact that the hydrogen bonds between the backbone amides of residues G133 and S135 and the carbonyl oxygen of the scissile substrate are stronger when these residues are treated quantum mechanically. However, it can be seen that addition of the residues located in the oxyanion hole that stabilize the TS1 and TI formation in the QM region does not significantly affect the activation and reaction energies, compared to the minimal QM subsystem. In particular, the difference in activation energy computed with both choices of the size of the QM region is relatively

small ($1.8 \text{ kcal mol}^{-1}$), indicating that an MM treatment of the oxyanion hole region is still acceptable in terms of giving the similar barrier shapes and reducing computational expenses.

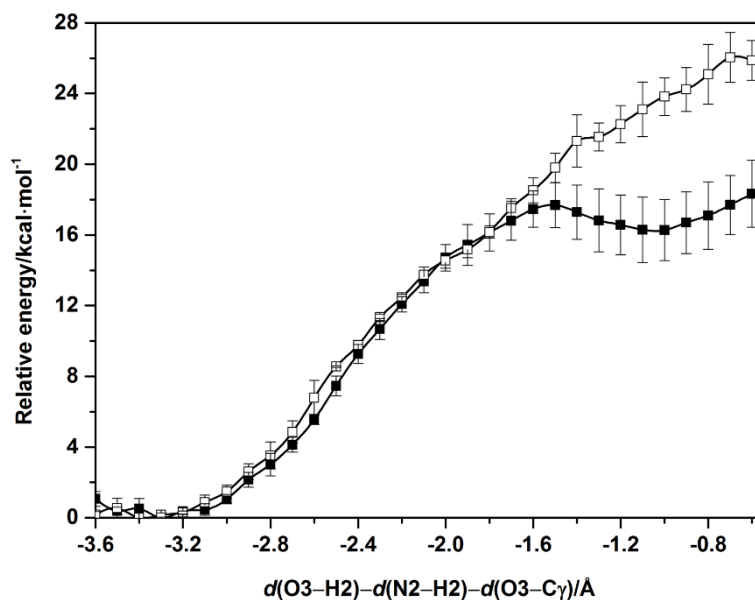


Figure 41. Comparison of the average reaction profiles (relative to the Michaelis complex) for the formation of tetrahedral intermediate calculated at SCS-LMP2/(aug)-cc-pVTZ/ff14SB level (using the BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries). Black squares: full QM region; white squares: QM region without the oxyanion hole region including residues G133, T134 and main chain NH of S135. Error bars indicate the standard deviation of the average energy.

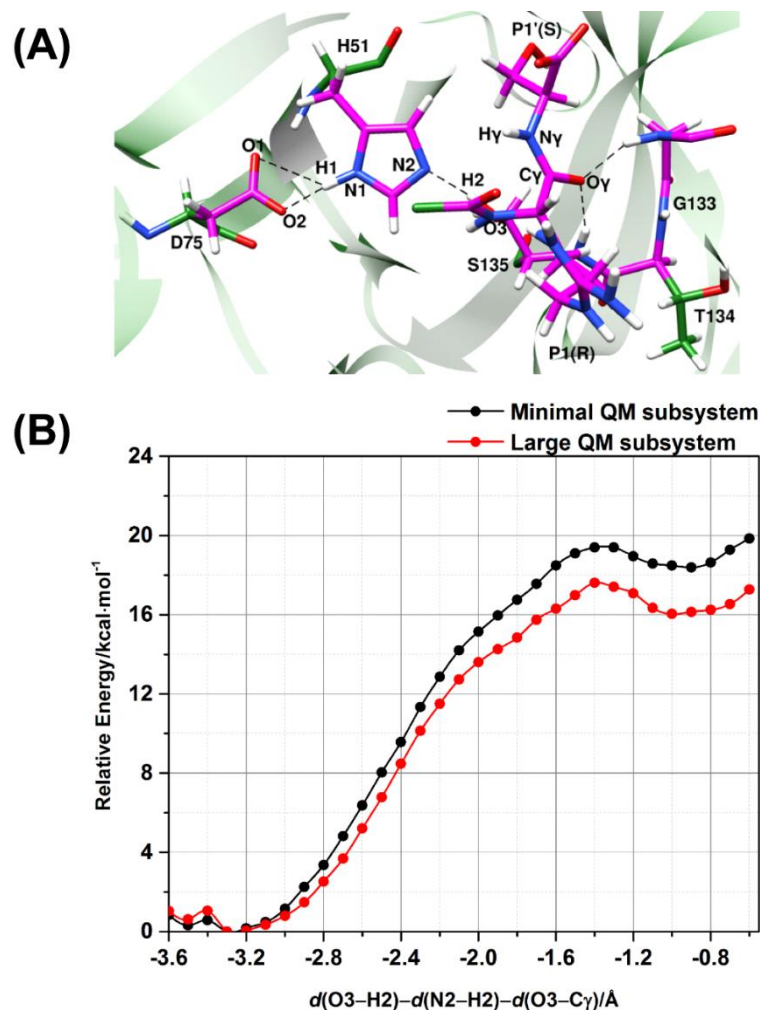


Figure 42. (A) Representation of the extended QM region of the Michaelis complex, showing the QM atoms as sticks with magenta carbon atoms and the MM region with green carbon atoms. The QM region includes the side-chains of the catalytic triad (H51, D75 and S135) and the fragments of substrate and oxyanion hole. The hydrogen bonds are represented by dashed lines. (B) Potential energy profiles for the first step of the acylation process calculated at the BH&HLYP-D3/6-31G(d)/ff144SB level with two different QM subsystems. Black circles: minimal QM subsystem; red circles: large QM subsystem (minimal QM subsystem plus residue fragments of oxyanion hole).

4.2.6 Structural analysis and hints for designing new inhibitors

The geometries of the MC, TS and TI in the first step of the acylation process obtained from the representative snapshot 1 of the combined interatomic distances are depicted in **Figure 43**. Again, it is noticeable that the two RC distances involving the

formation of TI simultaneously decrease ($d(\text{O3-H2})-d(\text{N2-H2})$: -0.6 to 0.6 Å; and $d(\text{O3-C}\gamma)$: 2.5 to 1.5 Å) during the reaction, reflecting a concerted reaction mechanism. Moreover, the main chain NH groups of G133 and S135 located in the oxyanion hole retain strong hydrogen bonding interactions with the carbonyl oxygen of the scissile bond and tend to exhibit the stronger electrostatic stabilization when the TI is formed (**Figure 41**). Interestingly, the step of proton transfer from S135 to H51 shows a complete proton transfer between the nitrogen and oxygen atoms at the transition state (TS1), as indicated by the decrease in $d(\text{N2-H2})$ from 1.6 Å (MC) to 1.0 Å (TS1). At the same time, the oxygen nucleophilic attack on the carbon of the substrate's carbonyl group occurs in concert with the proton transfer step, in which the nucleophilic attack distance, $d(\text{O3-C}\gamma)$, shortens in the order of 2.5 Å (MC), 2.0 Å (TS1) and 1.5 Å (TI). Thus, the reaction mechanism for the TI formation observed in common serine proteases (e.g. trypsin and chymotrypsin) and here in ZIKV NS2B/NS3, is consistent with the generally accepted reaction mechanism. Nevertheless, the acyl enzyme hydrolysis or deacylation process is typically rate-limiting for the hydrolysis of certain ester substrates by several serine proteases such as chymotrypsin [197], trypsin [198], subtilisin [199], and elastase [200]. Likewise, a kinetic study on DENV type 4 protease with its substrates revealed that the deacylation is rate-limiting for ester bond hydrolysis, whereas the acylation is the rate-determining step for amide bond hydrolysis [62]. From this point of view, it can be postulated that the acylation reaction is the rate-limiting step for substrate amide bond hydrolysis by ZIKV protease and thus the acyl enzyme hydrolysis has not been considered in this study.

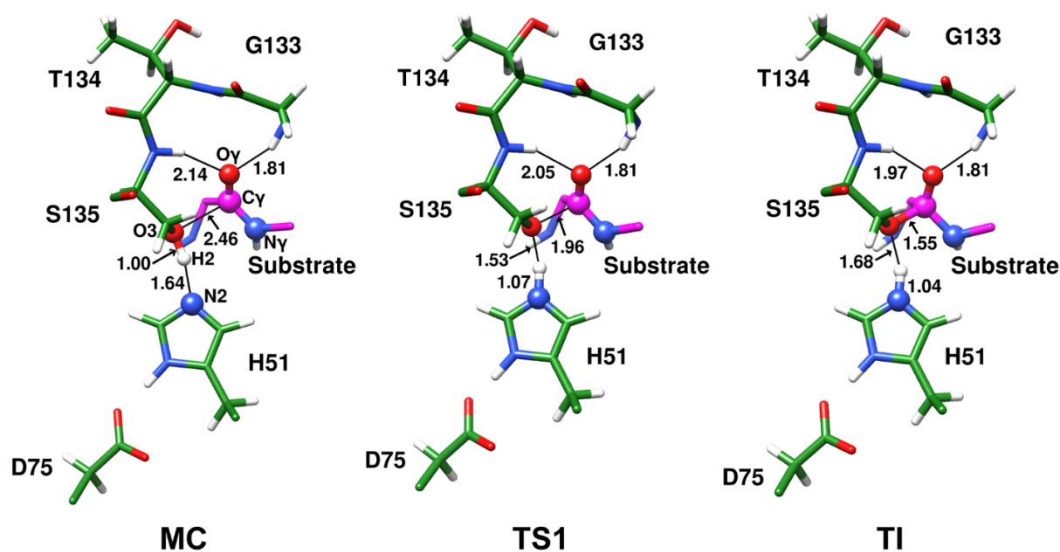


Figure 43. QM/MM (BH&HLYP-D3/6-31G(d)/ff14SB) optimized structures of the Michaelis complex (MC), transition state (TS1), and tetrahedral intermediate (TI) in the active site of ZIKV NS2B/NS3 protease obtained from the combined interatomic distances (in Å) for snapshot 1.

In this present study, we have succeeded in detecting the TI, which has a very short lifetime; a very hard task experimentally. The determination of the reaction mechanism of ZIKV protease here could help to design new inhibitors based on TS analogues [201]. Moreover, it seems that the stability of the TI can be a crucial factor in biocatalysis of several enzymes. Therefore, compounds which effectively mimic such an intermediate are considered as good candidates to become potential inhibitors of the targeted enzymes. It is worth noting that tetrahedral intermediates occur in many enzymes, particularly proteases and metallo-enzyme catalyzed reactions involved in hydrolysis of ester, amide or other acyl bonds. For instance, to resemble the tetrahedral TS of serine proteases, reversible covalent inhibitors able to form hemiketal formation have been introduced [202]. This attempt has led to two FDA approved drugs, telaprevir and boceprevir, which are covalent inhibitors against hepatitis C virus (HCV) protease through an α -ketoamide warhead [203]. Another example is proteasome inhibitor bortezomib that imitates the tetrahedral TS via boronic acid warhead and binds covalently to a threonine residue in the proteasome catalytic site. This inhibitor has been approved by FDA and used for treating multiple myeloma [204]. One more

example is from the discovery and development of β -lactamase inhibitors. These inhibitors mimic tetrahedral TS catalyzed by β -lactamase and inactivate enzyme activity [205]. Similar to other serine proteases, β -lactamases catalyze an amide bond hydrolysis, and their catalytic mechanisms involve the occurrence of tetrahedral intermediates along the reaction pathway [206]. Currently clinically approved β -lactamase inhibitors are clavulanate, tazobactam, sulbactam, avibactam and vaborbactam [207]. In the case of ZIKV protease, potential inhibitors could form a covalent bond through their warheads with the catalytic residue S135 and mimic the TI structure. Development of peptidomimetic inhibitors has been successful, and some of them derived from flavivirus protease substrates have been shown to be active toward both WNV and DENV proteases [72, 77, 208]. In addition, some of the available inhibitors against WNV and DENV proteases have also been shown to be active toward ZIKV protease [42, 78]. This is due to the high structural similarities among DENV, WNV, and ZIKV proteases [45, 49] [209]. Structures of ZIKV NS2B/NS3 serine protease complexed with several inhibitors are available, which assist antiviral inhibitor design [42, 43, 50]. For example, the dipeptidic inhibitor acyl-KR-aldehyde derived from P1 and P2 residues of the substrate is a potent competitive inhibitor, in which the aldehyde moiety forms a covalent bond with the catalytic S135 of NS3 [78]. Although it shows high ligand efficiency, it is a big challenge in developing them into drug-like molecules and for use in clinical applications because of their limitations such as cell penetration and stability caused by their positive charge [78], which has stimulated medicinal chemists to further find new inhibitors [210, 211]. Small-molecule inhibitors such as pyrazole ester derivatives identified from high-throughput screening of a small molecule library from the National Institutes of Health are much more interesting in terms of drug-likeness with small-molecular weight molecules and being potent inhibitors against both WNV and DENV proteases [212, 213]. Recently, one of the pyrazole ester derivatives, namely 5-amino-1-((4-methoxyphenyl)sulfonyl)-1*H*-pyrazol-3-yl benzoate, in complex with ZIKV protease has been revealed by crystallographic study, in which the benzoyl moiety of this compound could form a covalent bond with the catalytic residue S135 [79]. These examples provide structural information for covalent inhibitor design. Simulations such those presented here can

help guide design [214] of efficient protease inhibitors against ZIKV and other flaviviruses.

4.3 Part III: Inhibition mechanism of the ZIKV protease with dipeptidyl aldehyde inhibitor

4.3.1 QM/MM free-energy reaction path

Prior to performing QM/MM umbrella sampling MD simulations, the covalent complex between the ZIKV protease and the dipeptidyl aldehyde inhibitor was submitted to energy minimization, and followed by 200 ns of unbiased MD simulation to produce the initial structure for the subsequent QM/MM calculations. The quality of the MD simulation was evaluated by calculating the root mean-square deviation (RMSD) of the snapshots with respect to the initial structure of the simulation and by monitoring the stability of defined secondary structure of the protein (DSSP) [215]. Detailed plots of RMSD values and DSSP secondary structure analysis *versus* the simulation time are shown in **Figures 44** and **45**. After the first 20 ns of the MD, the structural features of the complex are stable and show an average RMSD value for the C_α atoms of 1.3 Å over the course of simulation time (**Figure 44**). Moreover, analysis of the secondary structure indicates that the structural elements of the protein are retained over the entire simulation time (**Figure 45**). These results display that the simulation model is stable.

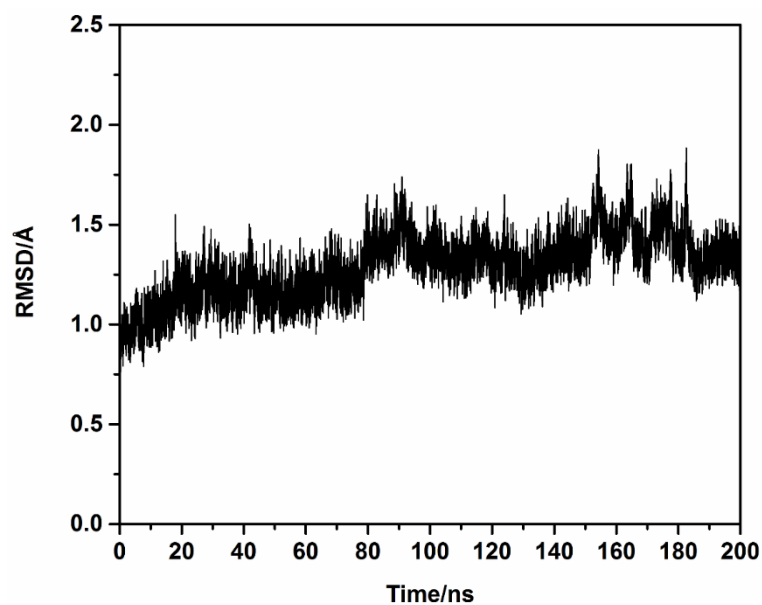


Figure 44. Time evolution of RMSD (Å) for the alpha carbon of the ZIKV protease in complex with the dipeptide inhibitor.

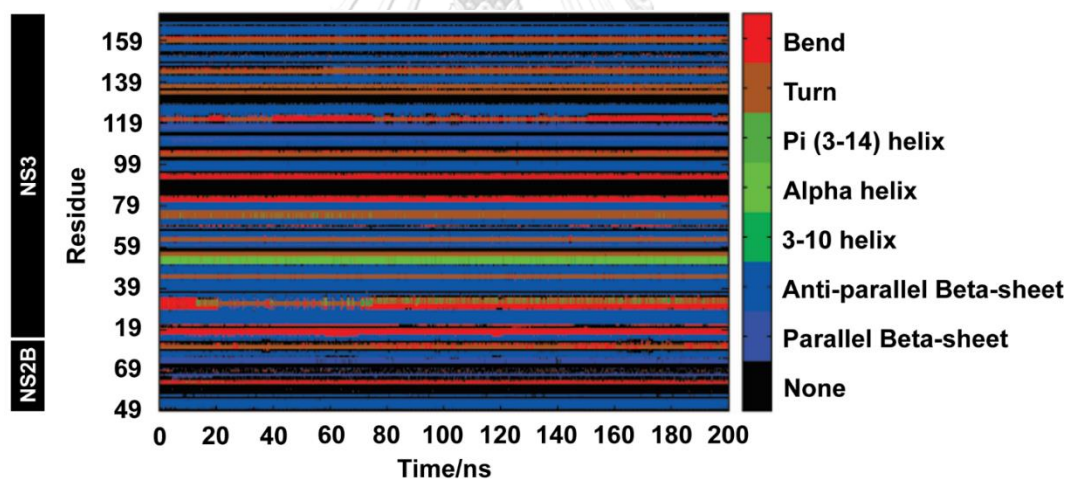


Figure 45. DSSP plot of the secondary structure of the ZIKV protease with the dipeptide inhibitor bound along the simulation time.

To choose a representative structure with statistically more rigorous selection, average linkage hierarchical clustering [216] of the MD trajectories sampling from 100 to 200 ns was used to generate a suitable representative structure averaged from the most populated structures. Afterward, the initial structure obtained from the clustering technique was submitted to the QM/MM simulations. Here, we performed QM/MM (PDDG-PM3/ff14SB) umbrella sampling MD simulations to explore the inhibition

mechanism of the ZIKV protease by the dipeptidyl aldehyde inhibitor, and to calculate free-energy profile (potential of mean force). The free-energy profile was calculated along the RC, defined as a linear combination of relevant interatomic distances (see **Chapter III Research Methodology**). To treat the QM region, we selected the PDDG-PM3 semi-empirical method, which is modified the pairwise core repulsion function to introduce functional group information via pairwise atomic interactions [157]. This semi-empirical QM method has been successful to give the relative activation energies for many reactions in line with experimental data [217]. In addition, our preliminary results show that the activation free-energy barrier calculated with the PDDG-PM3/ff14SB level provides a much more reasonable energy barrier than the original PM3 result, where the free-energy barrier obtained at the PM3 level is found to be too high (**Figure 46**). It should be noted that the free energy profiles obtained at the PDDG-PM3/ff14SB, PM3/ff14SB, PM6/ff14SB and AM1-d/ff14SB levels of theory are compared and given in **appendix 3**. The calculated free-energy profile at the PDDG-PM3/ff14SB level for the reaction of the tetrahedral adduct is shown in **Figure 47**.

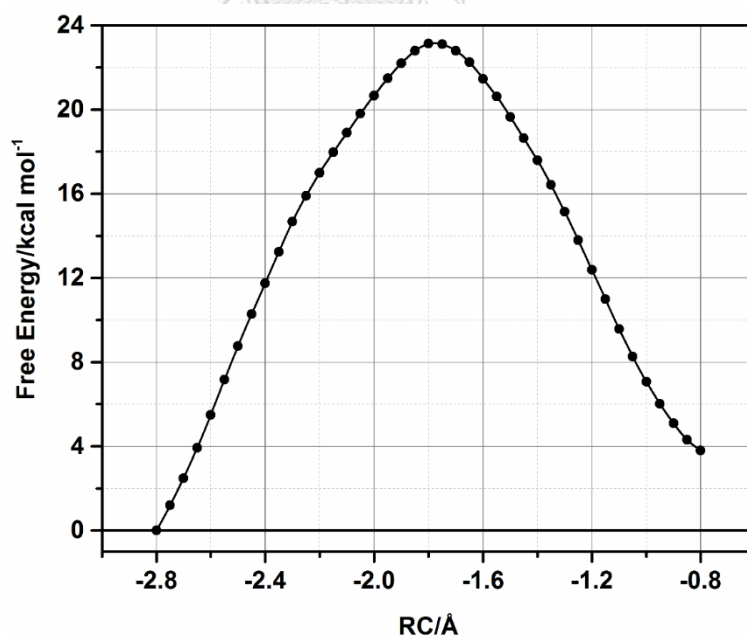


Figure 46. Free energy profile for the reaction of charged tetrahedral adduct relative to the noncovalent Michaelis complex using the combined reaction coordinates at the PM3/ff14SB level of QM/MM theory.

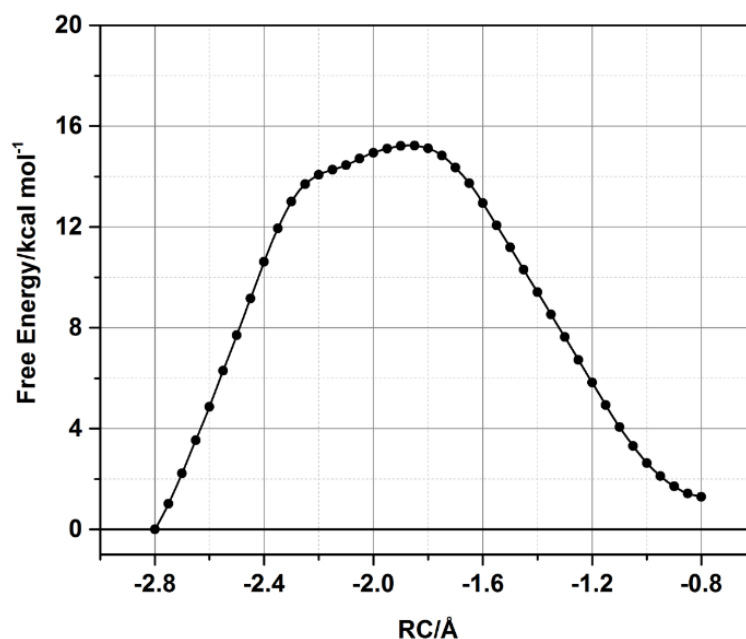


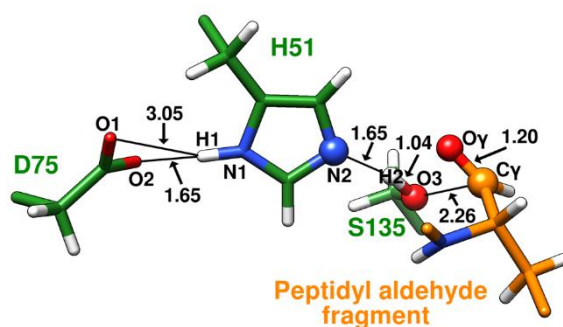
Figure 47. Free-energy profile for formation of the charged tetrahedral adduct relative to the noncovalent Michaelis complex for the reaction between the ZIKV protease and dipeptidyl aldehyde inhibitor using the combined reaction coordinates at the PDDG-PM3/ff14SB level of QM/MM theory.

Figure 47 highlights that the calculated free-energy barrier for the reaction is $15.2 \text{ kcal mol}^{-1}$, and the reaction is slightly endothermic with the reaction free-energy of $2.0 \text{ kcal mol}^{-1}$. In addition, a single transition state is explored on the free-energy profile, as indicative of a concerted reaction [186], in which the proton transfer from S135 to H51 and the nucleophilic attack on the carbonyl group of the dipeptidyl aldehyde inhibitor occur simultaneously. This is also supported by the calculation of relevant interatomic distances, where the distances involved in the proton transfer [$d(\text{N2-H2})$] and nucleophilic attack [$d(\text{O3-C}\gamma)$] decrease together (**Table 8, Figure 48**). The concerted mechanism was also revealed by the ab initio QM/MM calculations combined with molecular dynamics free-energy perturbation (MD-FEP) approach on trypsin and its substrate, in which proton transfer occurred in concert with nucleophilic attack on the substrate for the first step of the acylation process [186]. Moreover, there were numerous theoretical studies on other serine proteases [159-161, 188] and our recent high-level QM/MM calculations on the ZIKV protease with its substrate [162], suggesting that the concerted reaction was observed for the first step of the acylation

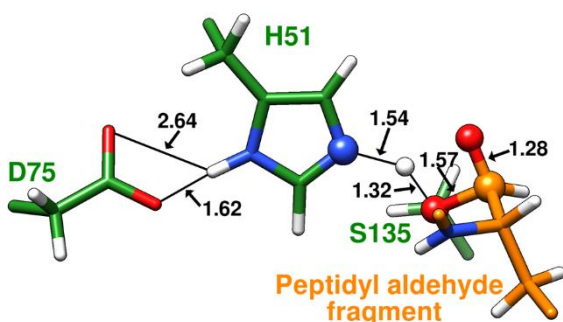
process. Thus, it is clearly seen from our finding that peptidyl aldehyde mimics the transition state/tetrahedral intermediate of the cleavage reaction on a substrate via the concerted mechanism. Additionally, it is noticeable that the structural changes are found in the aldehyde carbonyl group of the dipeptidyl aldehyde. As the progress of the reaction, the aldehyde carbonyl bond length [$d(\text{C}\gamma\text{--O}\gamma)$] becomes longer in the product state (~ 1.32 Å), reflecting the interconversion of $\text{C}\gamma$ hybridization from an sp^2 [$d(\text{C}\gamma\text{--O}\gamma) \sim 1.20$ Å in the reactant state] to an sp^3 hybridization upon reaching the product state. Furthermore, when the covalent complex is formed, the protonated-imidazole ring of H51 is stabilized by the carboxylate group of D75 with a presence of two hydrogen bonds [$d(\text{O1--H1}) = 2.60$ Å and $d(\text{O2--H1}) = 1.59$ Å]. Besides these two hydrogen bonds, the negatively charged oxygen on the tetrahedral adduct also directly interacts with the H51 through a strong hydrogen bond [$d(\text{O}\gamma\text{--H2}) = 1.78$ Å] rather than the amino acid residues in the oxyanion hole region (the main chain NH groups of G133 and S135), as generally stabilize the negative charge on the carbonyl oxygen of the substrate scissile bond [218, 219]. This is due to a resemblance of the hemiacetal adduct to the native catalytic reaction mechanism, in which the hemiacetal oxygen is analogous to the nitrogen leaving group on natural substrates in the acylation process.

Table 8. Average values of the interatomic distances involving active site interactions (in Å) for the noncovalent Michaelis complex, transition state and covalent complex structures, from the free-energy reaction path obtained at the PDDG-PM3/ff14SB level of theory.

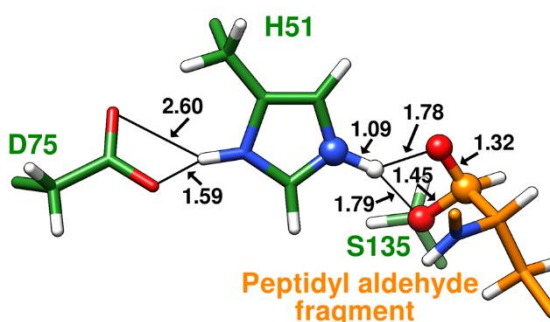
	Noncovalent Michaelis complex	Transition state	Covalent complex
$d(\text{N2--H2})$	1.65 ± 0.05	1.54 ± 0.05	1.09 ± 0.03
$d(\text{O3--C}\gamma)$	2.26 ± 0.06	1.57 ± 0.04	1.45 ± 0.03
$d(\text{O1--H1})$	3.05 ± 0.26	2.64 ± 0.16	2.60 ± 0.14
$d(\text{O2--H1})$	1.65 ± 0.06	1.62 ± 0.05	1.59 ± 0.05



(I) Noncovalent Michaelis Complex



(II) Transition State



(III) Covalent Tetrahedral Adduct

Figure 48. Representative structures from the free-energy reaction path at the PDDG-PM3/ff14SB level of (I) noncovalent Michaelis complex, (II) transition state and (III) tetrahedral adduct in the active site of ZIKV protease.

To monitor the charge redistribution/reorganization of the relevant atoms in the QM region for each stationary point, the Mulliken charge distribution was assessed and shown in **Table 9**. As the reaction progresses, the atomic charge on the N2_{H51} significantly changes from the negative charge to the positive charge due to charge transfer from the S135 hydroxyl proton. In contrast, as the N2 atom acts as a general

base in abstracting a proton to activate S135 as a nucleophile, the atomic charge on the carbonyl oxygen ($O\gamma$) of the aldehyde warhead becomes more negative charge, reaching its lowest value of -0.86 au at the product state. As previously described, this phenomenon resembles the peptide bond breaking process of the natural substrate, in which the protonation of the nitrogen leaving group by $H2_{H51}$ proton would facilitate the cleavage of the peptide bond in the acylation step. On the other hand, protonation of the anionic aldehyde carbonyl oxygen by $H2_{H51}$ proton could lead to a nonproductive stabilization of the tetrahedral intermediate, making it a stable covalent complex (product would not be cleaved). To highlight this point, we also performed unbiased QM/MM MD simulation at the PDDG-PM3/ff14SB level of theory for 1 ns using the covalent anionic tetrahedral adduct as the initial structure to elucidate the possibility of this transformation. The results found that the direct proton transfer from $N2_{H51}$ atom to $O\gamma$ atom of the tetrahedral adduct occurs spontaneously from the beginning to the end of simulation time and forms the protonated neutral form of the hemiacetal. This observation can be obviously seen from the distance between the $O\gamma$ atom and $H2_{H51}$ atom [$d(O\gamma-H2)$] of ~ 0.98 Å over the course of simulation time (see **Figure 49**). Our simulation also supports the previous study of aldehyde association to trypsin [220], suggesting that the hemiacetal adduct in the neutral form is likely to be the final product of the inhibition reaction. Along the reaction pathway, it should be noted that the atomic charge on the $O3_{S135}$ atom rises at the transition state (-0.26 au). This is the result of the partially formed covalent bond between $O3_{S135}$ atom and the carbonyl carbon ($C\gamma$) of the peptidyl aldehyde at this stationary point [$d(O3-C\gamma) = 1.57$ Å, **Table 7** and **Figure 48**, label II], resulting in the charge redistribution to $C\gamma$ and $O\gamma$ atoms.

Table 9. Average Mulliken charges (PDDG-PM3/ff14SB) in atomic units (au) for atoms involving the formation of the tetrahedral adduct, averaged from the QM/MM free-energy profile.

Relevant atom	Noncovalent Michaelis complex	Transition state	Covalent complex
O γ	-0.31 ± 0.03	-0.72 ± 0.03	-0.86 ± 0.03
C γ	0.24 ± 0.02	0.27 ± 0.03	0.24 ± 0.02
N2	-0.13 ± 0.03	-0.11 ± 0.03	0.24 ± 0.04
H2	0.30 ± 0.01	0.37 ± 0.01	0.28 ± 0.01
O3	-0.46 ± 0.02	-0.26 ± 0.04	-0.31 ± 0.02

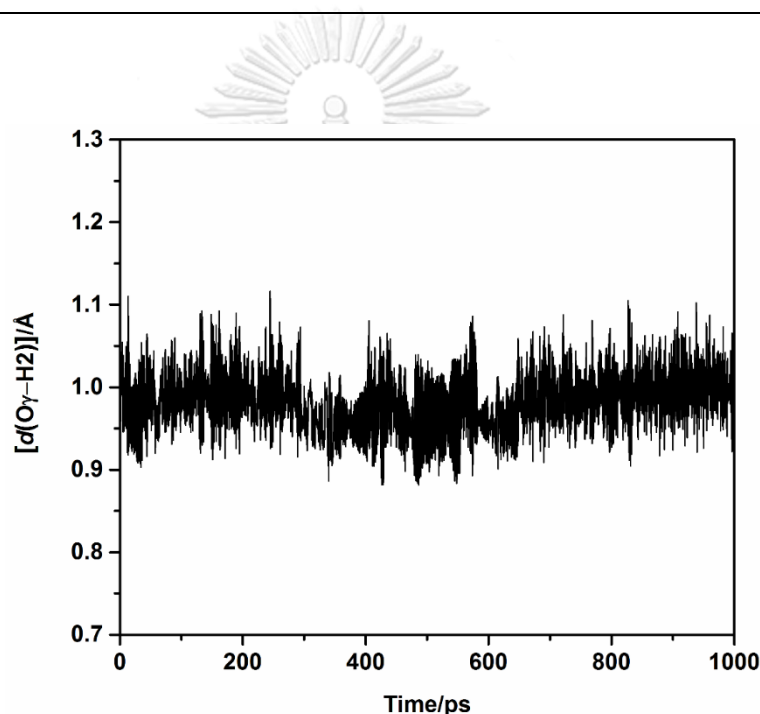


Figure 49. Distance between oxygen atom (O γ) of the tetrahedral adduct and H2_{H51} atom during 1 ns QM/MM MD simulation of the ZIKV protease bound with the dipeptide inhibitor.

4.3.2 QM/MM potential energy reaction path

The starting geometry for QM/MM potential energy adiabatic mapping calculations is the covalent anionic tetrahedral adduct, minimized at the BH&HLYP-D3/6-31G(d)/ff14SB level of theory. The potential energy barrier and the reaction energy (relative to the noncovalent Michaelis complex) for the reaction of the intermediate adduct formation are summarized in **Table 10**. The DFT method shows

that the potential energy barrier for the reaction is $6.2 \text{ kcal mol}^{-1}$ and the reaction is slightly endothermic with the reaction energy of $0.5 \text{ kcal mol}^{-1}$. Single-point energy calculations were carried out on the BH&HLYP/6-31G(d)/ff14SB geometries using a larger basis set for the DFT method (i.e. 6-311+G(d)) to examine the effect of basis set size on the potential energy profile for the reaction (**Table 10**). The results reveal that an increase in the basis set size from 6-31G(d) to 6-311+G(d) does not show a significant effect on the DFT energetics. The reaction barrier is increased by only $0.3 \text{ kcal mol}^{-1}$ when the larger basis set was used (**Figure 50**). Therefore, the minimal basis set (6-31G(d)) provides a good choice for geometry optimization, compromising between accuracy and computational cost. Afterward, single-point energies calculated at the MP2, SCS-MP2, and LCCSD(T) levels of theory were conducted on the same set of structures optimized by the BH&HLYP-D3/6-31G(d)/ff14SB level. The corresponding potential energy profiles for the reaction indicate that the reaction proceeds in a concerted manner of proton transfer and nucleophilic addition reaction at all levels of QM/MM theory tested here, as previously also found in the QM/MM (PDDG-PM6/ff14SB) free-energy simulations. This process is involved in only one approximate transition state along the minimum-energy path connecting to two local energy minima between the noncovalent Michaelis complex and the covalent hemiacetal adduct (**Figure 50**). The LCCSD(T)/(aug)-cc-pVTZ level (energy barrier of $4.3 \text{ kcal mol}^{-1}$) is the most accurate result presented here and used as a reference method for the comparisons with the two ab initio methods (i.e. MP2 and SCS-MP2). The potential energy barrier of $3.2 \text{ kcal mol}^{-1}$ calculated with the MP2 method is predicted to be lower than that of the LCCSD(T) result by $1.1 \text{ kcal mol}^{-1}$, whereas SCS-MP2, BH&HLYP-D3, and BH&HLYP-D3 with the larger basis set show the higher barriers by $0.7 \text{ kcal mol}^{-1}$, $1.9 \text{ kcal mol}^{-1}$, $2.2 \text{ kcal mol}^{-1}$, respectively. Note that coupled cluster theory can be used as an accurate descriptor for the reactions and is considered as the “gold standard” of ab initio methods [150]. Based on our calculation, it indicates that the SCS-MP2/(aug)-cc-pVTZ/ff14SB result is in very good agreement with the local coupled cluster result. Our finding supports previous QM/MM calculations for the reactions catalyzed by other enzymes [123, 145, 221], suggesting that SCS-MP2 was a good choice for calculations on enzyme-catalyzed reactions. In addition, it evidently shows that the spin component scaled method developed for MP2 calculations is

necessary for giving more accurate results. Meanwhile, the reaction energies (i.e. the energy of the tetrahedral adduct relative to the noncovalent Michaelis complex) is slightly different for each level of theory. Both BH&HLYP-D3 with 6-31G(d) basis set and SCS-MP2 results display that the reaction is slightly endothermic, while MP2 and LCCSD(T) results indicate that the reaction is weakly exothermic. In fact, peptidyl aldehyde inhibitors undergo a covalently reversible inhibition with flavivirus NS2B/NS3 protease [73, 78, 80, 81] and thus the predicted reaction energy for the reaction on such inhibitors is expected to be only weakly exothermic [222, 223]. However, in this case the anionic hemiacetal is not thought to be the final product for the inhibition reaction, as should be the protonated neutral form of the hemiacetal adduct due to the proton transfer from H51. To clarify this point, we additionally performed QM/MM (BH&HLYP-D3/6-31G(d)/ff14SB) adiabatic mapping using the RC involved in the direct proton transfer between H51 and oxygen atom of the anionic hemiacetal, [$d(\text{N2-H2})-d(\text{O}\gamma\text{-H2})$]. The resulting potential energy profile for this reaction suggests that the protonated neutral form is predicted to be $\sim 12 \text{ kcal mol}^{-1}$ much more stable than the deprotonated anionic form without the activation energy predicted (**Figure 51**). This also previously supports the result from unbiased QM/MM MD simulations (see **QM/MM Free-Energy Reaction Path Section**), in which the proton on ϵ nitrogen of the H51 transfers spontaneously onto the anionic oxygen atom of hemiacetal, producing the neutral form of the tetrahedral adduct in the final state. Thus, the results present here are still comparable to each other.

Table 10. Potential energy barriers ($\Delta^\ddagger V$) and reaction energies ($\Delta_r V$) calculated with BH&HLYP-D3/6-31G(d), BH&HLYP-D3/6-311+G(d), MP2/(aug)-cc-pVTZ, SCS-MP2/(aug)-cc-pVTZ, and LCCSD(T)/(aug)-cc-pVTZ QM/MM methods on BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries^a

Calculation method	$\Delta^\ddagger V$	$\Delta_r V$
	(RC = -1.80 Å)	(RC = -1.00 Å)
BH&HLYP-D3/6-31G(d)/ff14SB	6.2	0.5
BH&HLYP-D3/6-311+G(d)/ff14SB	6.5	-0.4
MP2/(aug)-cc-pVTZ/ff14SB	3.2	-1.4
SCS-MP2/(aug)-cc-pVTZ/ff14SB	5.0	0.1
LCCSD(T)/(aug)-cc-pVTZ/ff14SB	4.3	-2.8

^aEnergies (relative to the reactant) are reported in kcal mol⁻¹. The (aug) indicates that augmented functions were treated with oxygen atoms only.

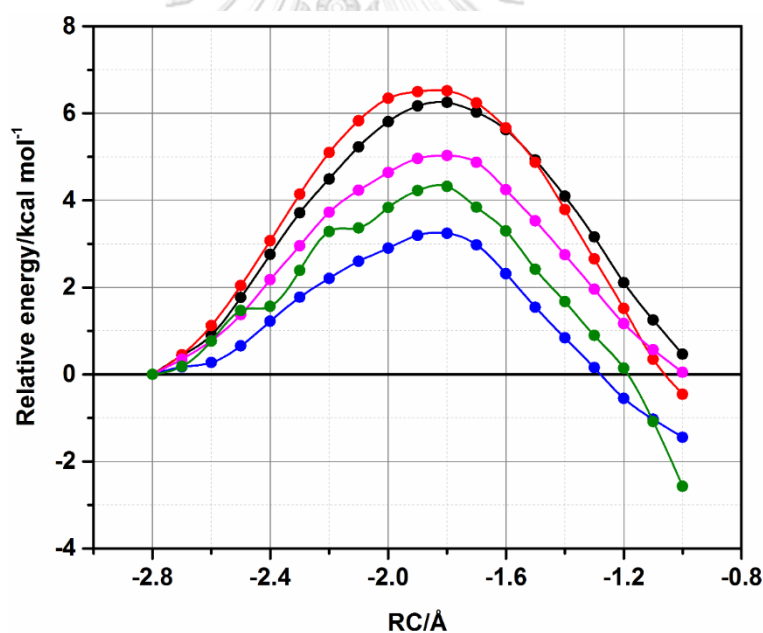


Figure 50. Potential energy profiles for the reaction of the tetrahedral adduct formation calculated at the BH&HLYP-D3/6-31G(d)/ff14SB (black), BH&HLYP-D3/6-311+G(d)/ff14SB (red), MP2/(aug)-cc-pVTZ/ff14SB (blue), SCS-LMP2/(aug)-cc-pVTZ/ff14SB (magenta), and LCCSD(T)/(aug)-cc-pVTZ/ff14SB (green) QM/MM levels. All geometries were optimized at the BH&HLYP-D3/6-31G(d)/ff14SB level.

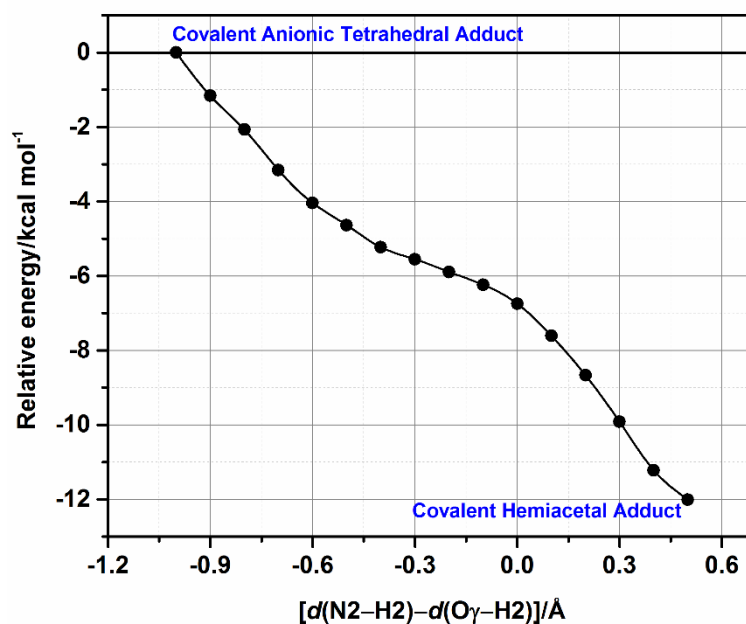


Figure 51. Free energy profile for the reaction of protonated hemiacetal adduct (relative to the reactant state, tetrahedral adduct) using reaction coordinate involving the direct proton transfer between H51 and oxygen atom of the anionic hemiacetal.

Although the MP2 method gives a good predictive description for the reaction energies, as is close to coupled cluster result to some extent, this method underestimates the energy barrier, as is somewhat found for some enzyme-catalyzed reactions [195, 221]. For the BH&HLYP-D3 method, it has been also shown to be a good description in terms of barrier shape, but the ab initio methods particularly SCS-MP2 method give the results much more similar to the coupled cluster theory. Therefore, we suggest here that the correlated ab initio method (i.e. SCS-MP2) and the more precious coupled cluster theory should be carried out to obtain accurate energetics for this particular reaction. Nevertheless, all levels of QM/MM theory presented here predict the energy barriers lower than that of the reaction mechanism on the ZIKV protease and its substrate by $\sim 12\text{--}13$ kcal mol⁻¹, as we performed previously [162]. Comparison of potential energy profiles for the reaction mechanism on substrate and dipeptidyl aldehyde inhibitor catalyzed by the ZIKV protease is summarized in **appendix 4**. This finding reflects that the peptidyl aldehyde inhibitor is a good transition state analogue toward the ZIKV protease, which mimics the native catalytic reaction transition state

and reduces the energy barriers in much more efficient way compared with the natural substrate.

Figure 52 illustrates the representative structures for each stationary point of the noncovalent Michaelis complex, transition state, and tetrahedral adduct extracted from adiabatic mapping. The reaction mechanism is involved in the transfer of proton from the O3 position of S135 to the N2 position of H51, and nucleophilic attack of the O3_{S135} on the carbonyl carbon (C γ) of the peptidyl aldehyde (see **Figure 52**, label I, corresponding to reactant state in **Figure 7**). In the reactant state, the structural analysis shows that the proton (H2) to acceptor (N2_{H51}) distance is 1.59 Å, and the distance between the N2 nitrogen and H2 proton is reduced to 1.19 Å, and the distance between these two atoms become shorter of 1.05 Å upon reaching the covalent complex (**Figure 52**, label III), as a result of the complete proton transfer. Likewise, the distance involved in nucleophilic addition reaction, [$d(\text{O3}-\text{C}\gamma)$], continuously decreases until the covalent bond is formed (2.23 Å in the reactant state, 1.90 Å in the transition state and 1.55 Å in the covalent complex). Again, it is clearly seen from these stationary points that the proton transfer and the nucleophilic attack occur in a concerted manner. In addition, the negatively charged hemiacetal oxygen (O γ) directly interacts with protonated H51 through a strong hydrogen bond interaction [$d(\text{O}\gamma-\text{H2}) = 2.29$ Å in the covalent complex]. As already mentioned, this is in contrast to the negative charge generated on the carbonyl oxygen of substrate, which was stabilized by the residues located in the oxyanion hole. The reason is that the anionic hemiacetal adduct produced by peptidyl aldehyde resembles the interactions of amine leaving group on the substrate with protonated H51 in the acylation reaction.

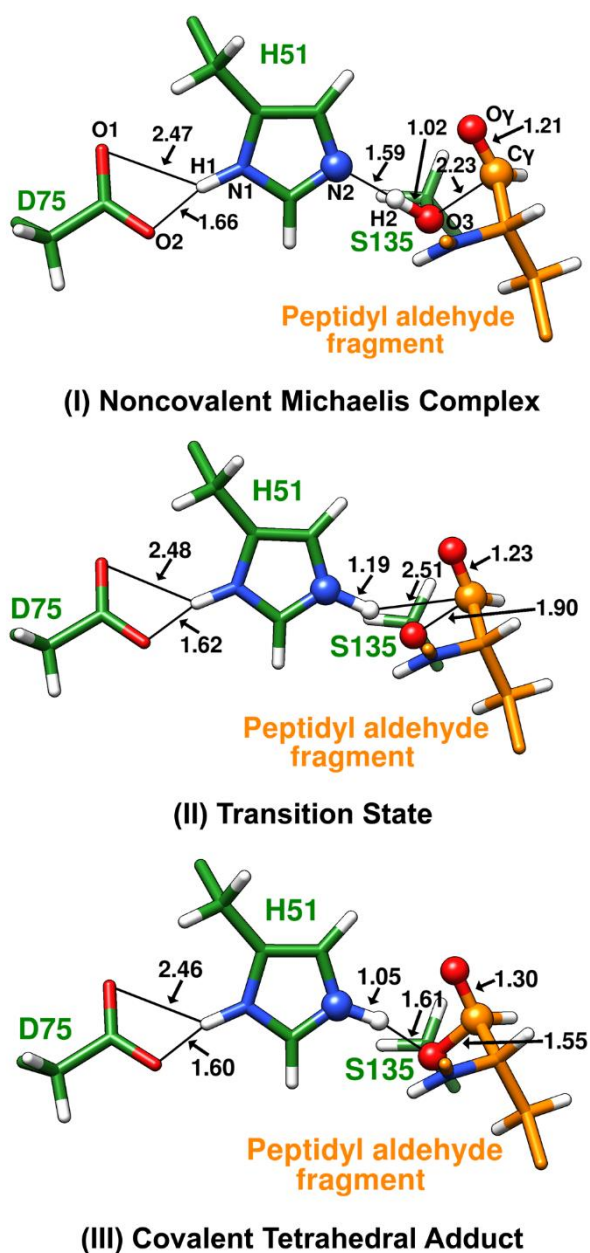


Figure 52. QM/MM (BH&HLYP-D3/6-31G(d)/ff14SB) optimized structures of (I) noncovalent Michaelis complex, (II) transition state, and (III) covalent tetrahedral adduct in the active site of ZIKV protease.

Fundamentally, in this study structures were generated along the adiabatic mapping path by moving backward to the corresponding noncovalent Michaelis complex at 0.1 Å intervals along the defined RC. To validate the RC used in the adiabatic mapping, the refinement of the pathway using CI-NEB techniques was performed. The results show that the potential energy profiles from the CI-NEB method

and adiabatic mapping are very similar in terms of barrier height and transition state location (**Figure 53**). The potential energy barrier for the reaction is $6.1 \text{ kcal mol}^{-1}$ obtained after refinement with the CI-NEB method at the BH&HLYP/6-31G(d)/ff14SB level of theory (**Table 11**). The energy barrier differs from adiabatic mapping by only $0.1 \text{ kcal mol}^{-1}$, confirming that the RC used for adiabatic mapping approach here provides a reasonable description of the reaction. The transition state from CI-NEB calculations is located at $\text{RC} = -1.86 \text{ \AA}$, [$d(\text{O3-H2}) = 1.25 \text{ \AA}$, $d(\text{N2-H2}) = 1.21 \text{ \AA}$ and $d(\text{O3-C}\gamma) = 1.90 \text{ \AA}$], which is in good agreement with the approximate transition state generated by adiabatic mapping of $\text{RC} = -1.80 \text{ \AA}$. However, it should be mentioned that there is no RC used in the generation of the CI-NEB path, but the RC value is a useful geometric descriptor to compare the pathways. In addition, single-point energies calculated with the correlated ab initio methods (MP2, SCS-MP2 and LCCSD(T)) exhibit the similar trend in terms of giving the similar barrier shapes compared with the DFT result (see **Figure 54**), and are much correlated well with the energy barriers calculated with the adiabatic mapping. Our calculations show that the predictions made by the BH&HLYP method are qualitatively correct of mechanistic conclusions, associated with how the pathway was generated in comparison with the ab initio methods.

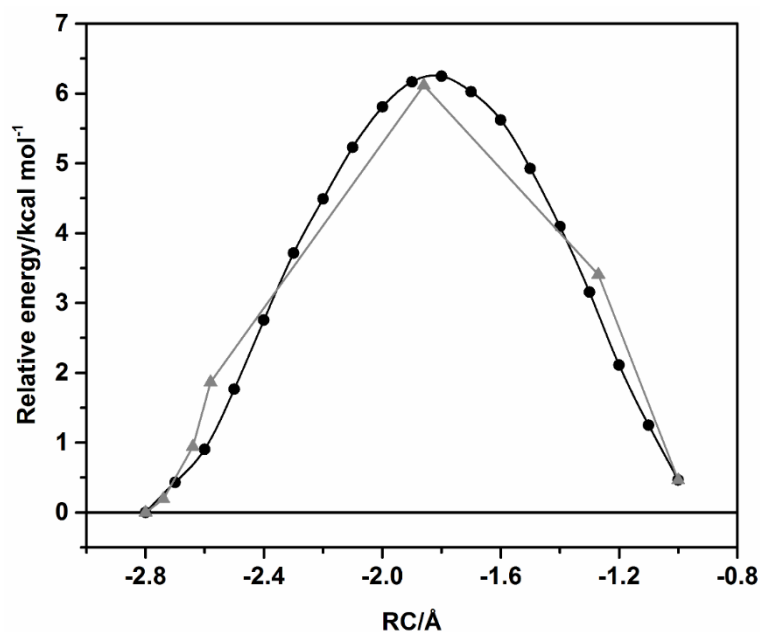


Figure 53. Climbing image NEB profiles compared to adiabatic mapping (black circles, adiabatic mapping profile; grey triangles, CI-NEB profile from 7 starting images). Note that there is no reaction coordinate defined in the CI-NEB calculations; the data is plotted relative to the energy of the reactant and against the reaction coordinate to aid the comparison with the adiabatic mapping profile. All profiles are calculated at the BH&HLYP-D3/6-31G(d)/ff14SB level of theory and correspond to the reaction of the formation of tetrahedral addct.

Table 11. Reaction energetics (in kcal mol⁻¹, relative to the reactant state) and reaction coordinate (RC) values (in Å) from CI-NEB method for the reaction of the tetrahedral adduct calculated with BH&HLYP-D3/6-31G(d), MP2/(aug)-cc-pVTZ, SCS-MP2/(aug)-cc-pVTZ, and LCCSD(T)/(aug)-cc-pVTZ QM/MM methods.

Calculation method	Noncovalent Michaelis complex (RC = -2.80 Å)	Transition state (RC = -1.86 Å)	Covalent complex (RC = -1.00 Å)
BH&HLYP-D3/6-31G(d)/ff14SB	0.0	6.1	0.5
MP2/(aug)-cc-pVTZ/ff14SB ^a	0.0	3.2	-1.4
SCS-MP2/(aug)-cc-pVTZ/ff14SB ^a	0.0	5.0	0.1
LCCSD(T)/(aug)-cc-pVTZ/ff14SB ^a	0.0	4.3	-2.8

^aSingle-point energies based on BH&HLYP-D3/6-31G(d)/ff14SB geometries generated by the NEB techniques with MM point charges included in the QM Hamiltonian.

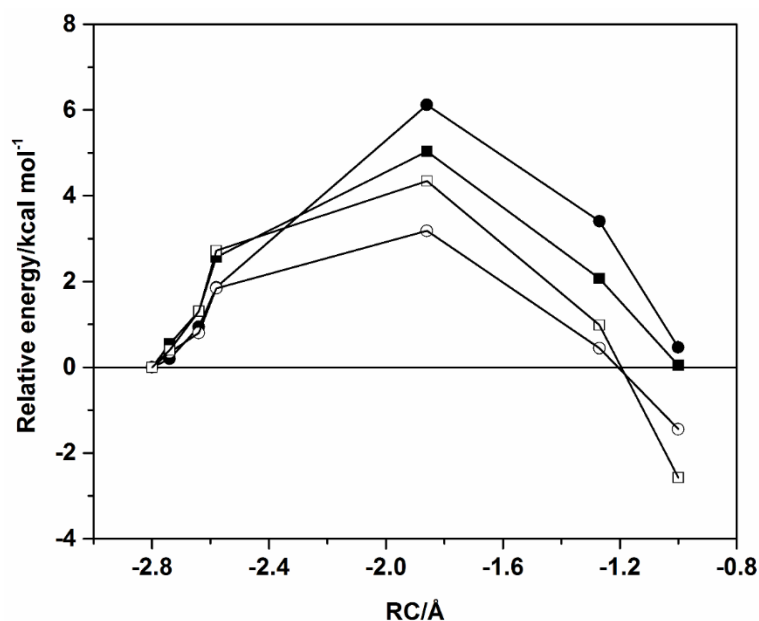


Figure 54. Comparison of the potential energy profiles (relative to the reactant) for the reaction of the tetrahedral adduct formation at the BH&HLYP-D3/6-31G(d)/ff14SB (open circle), MP2/(aug)-cc-pVTZ/ff14SB (white circle), SCS-MP2/(aug)-cc-pVTZ/ff14SB (black square) and LCCSD(T)/(aug)-cc-pVTZ/ff14SB (white square) levels of theory (using the BH&HLYP-D3/6-31G(d)/ff14SB CI-NEB geometries). There is no reaction coordinate defined in the CI-NEB calculations; however, energies are plotted as a function of reaction coordinate for comparison.

Altogether, we applied two different computational QM/MM methodologies to explore the feasibility of the proposed reaction mechanism of the ZIKV protease inhibition by dipeptidyl aldehyde inhibitor, to characterize structures of the stationary points (noncovalent Michaelis complex, transition state, and covalent complex), and to determine the energetics of the reaction. Note that the structures of the stationary points and the observed interactions characterized in the two QM/MM strategies are similar. The information derived from the potential energy reaction path was complemented by the dynamical information which was taken from the QM/MM free-energy simulations. Therefore, the detailed insights in the inhibition mechanism of the ZIKV protease by the peptidyl aldehyde would be a theoretical guidance [214] in designing efficient protease inhibitors with desired property based on transition state analogues against ZIKV and other flaviviruses.

CHAPTER V

CONCLUSIONS

5.1 Part I: Binding recognition of substrates in ZIKV NS2B/NS3 serine protease

In this present work, binding recognition and specificity of the four peptide substrates in the ZIKV protease were revealed by the conventional MD simulations. In all studied complexes, the results show that the basic residues at the P1 and P2 positions of the substrate have an important role in binding with the ZIKV protease, while the P3 and P4 sites play a less contribution. At the S1 subsite of the enzyme, the interactions occurred with the P1 position of the substrate are only contributed from the residues of NS3 protease. In contrast, the P2 residue of the substrate favorably interact with amino acid residues at the S2 subsite from both the C-terminal domain of NS2B (e.g. D83) and NS3 domain (e.g. D75). Whereas the energy contributions of the P3 and P4 residues to the amino acids at the S3 and S4 pockets are significantly decreased and almost disappeared, respectively. It can suggest here that the P1 and P2 sites should be kept for the further design of peptidomimetic inhibitors. The nonpolar interaction plays a predominant role in maintaining the interaction of the substrate at each subsite of enzyme. Altogether, we hope that this computational study provides the fundamental knowledge to guide for the design of efficient protease inhibitors.

5.2 Part II: Reaction mechanism of the ZIKV protease with its substrate

In this present study, QM/MM methods have elucidated the reaction mechanism of the acylation process of ZIKV protease and its substrate cleavage reaction. Although the reaction mechanism of common serine proteases has been widely studied, understanding the mechanistic details of the reaction catalyzed by ZIKV NS2B/NS3 serine protease is still necessary to provide fundamental knowledge of catalysis and may also help in the design and development of potential antiviral agents using TS analogues.

By combining MM MD simulations, QM/MM MD free-energy simulations and QM/MM adiabatic mapping, we investigated ZIKV protease in terms of both dynamical

and mechanistic properties. Our results show that the mechanism of the acylation reaction catalyzed by ZIKV protease behaves according to the proposed reaction mechanism for general serine proteases. Only one TS is determined along the reaction pathway suggesting a concerted reaction mechanism, in which the transfer of proton from S135 to H51 takes place in synchrony with the nucleophilic attack on the substrate. The formation of the TI is the rate-determining step of the acylation process. The calculations reveal that the PM6/ff14SB level underestimates the activation free energy in comparison with experimentally determined apparent catalytic rate, but even so the barrier shapes have been found to be similar to those of other serine proteases [55, 159, 186-190]. Thus, we have performed multiple fully optimized potential energy profiles for the first step of the acylation, using the correlated *ab initio* QM/MM methods to obtain accurate energetics. We have optimized geometries with BH&HLYP-D3/6-31G(d)/ff14SB level of theory. Similar to the free energy profiles, the formation of TI is concluded to occur through a concerted manner that involves a single transition state (TS1), where a stable species of TI has been found at all levels of theory. The potential energy barriers are $16.3 \text{ kcal mol}^{-1}$ at the LCCSD(T)/(aug)-cc-pVTZ/ff14SB//BH&HLYP-D3/6-31G(d)/ff14SB level of theory. DFT and *ab initio* results presented here are reasonably close to the coupled cluster energies. The LMP2 result slightly underestimates the barrier compared to the BH&HLYP-D3 and SCS-LMP2 results. However, these results are in good agreement with the apparent experimental free energy of the activation of $\sim 18.2 \text{ kcal mol}^{-1}$ (at 310 K). It is indeed of significance that potential energy barriers here cannot directly be compared to activation free energies derived from experimental kinetics data. This is due to the fact that the activation free energies include additional effects such as entropy and quantum tunneling [122].

By comparisons with the most accurate (LCCSD(T)/(aug)-cc-pVTZ/ff14SB) results, although the BH&HLYP-D3/6-31G(d) result gives a good description, the *ab initio* LMP2/(aug)-cc-pVTZ and SCS-LMP2/(aug)-cc-pVTZ methods give results much more similar to the LCCSD(T) result. We suggest that it is necessary to use higher-level QM methods for obtaining accurate energetics (precisely at the coupled cluster level, or the more practicable, SCS-LMP2) for ZIKV protease and for other related enzymes having the similar reaction mechanism. We hope that this

computational information can be a good starting point to design and develop novel inhibitors toward ZIKV protease and other flavivirus proteases based on TS analogues.

5.3 Part III: Inhibition mechanism of the ZIKV protease with dipeptidyl aldehyde inhibitor

In the present study, the mechanistic study on enzymatic reaction of ZIKV inhibition by dipeptidyl aldehyde was elucidated for the first time using two different methodologies, QM/MM free-energy and QM/MM potential energy path calculations. The results show that the inhibition mechanism of the ZIKV protease by dipeptidyl aldehyde behaves similarly to the proposed reaction mechanism for common serine proteases with their substrates. The anionic hemiacetal formation mimics the transition state/tetrahedral intermediate of the native catalytic reaction. The free-energy profile displays a single transition state observed in the reaction path, indicating a concerted mechanism. In addition, the adiabatic mapping approach was used to generate QM/MM potential energy surfaces of the reaction. DFT and ab initio methods provide the reasonable energetics, as compared with the local coupled cluster theory. In particular, the potential energy barrier obtained at the SCS-MP2 level is in good agreement with the local coupled cluster result, indicating that more reliable energetics can be obtained by addition of the spin component scaling calculation to the MP2 method. Comparison of the DFT results with those from ab initio methods shows that the BH&HLYP functional gives a good description of the reaction in this system. It can suggest here that the use of DFT method is qualitatively sufficient to describe this kind of reaction mechanism and useful for the comparative analysis. Moreover, the CI-NEB method reveals that the RC used in this study is a good choice to describe the reaction. The transition state structures from adiabatic mapping are very similar to the transition state from refinement with the CI-NEB techniques. Consequently, it is our hope that this computational study sheds light on the mechanistic details of the ZIKV protease inhibition by peptidyl aldehyde and can be a valuable starting point in design and development of potential protease inhibitors against the ZIKV protease and other flavivirus proteases with desired properties based on transition state analogues.

REFERENCES

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1. Dick, G.W.A., *Zika Virus (I). Isolations and serological specificity*. Transactions of the Royal Society of Tropical Medicine and Hygiene, 1952. **46**(5): p. 509-520.
2. Faye, O., et al., *Molecular Evolution of Zika Virus during Its Emergence in the 20th Century*. PLoS Neglected Tropical Diseases, 2014. **8**(1): p. 36.
3. Ledermann, J.P., et al., *Aedes hensilli as a Potential Vector of Chikungunya and Zika Viruses*. PLoS Neglected Tropical Diseases, 2014. **8**(10).
4. Fernandez-Garcia, M.D., et al., *Pathogenesis of Flavivirus Infections: Using and Abusing the Host Cell*. Cell Host and Microbe, 2009. **5**(4): p. 318-328.
5. Turrini, F., et al., *Zika virus: A re-emerging pathogen with rapidly evolving public health implications*. New Microbiologica, 2016. **39**(2): p. 86-90.
6. Dai, L., et al., *Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody*. Cell Host and Microbe, 2016. **19**(5): p. 696-704.
7. Sirohi, D., et al., *The 3.8 Å resolution cryo-EM structure of Zika virus*. Science, 2016. **352**(6284): p. 467-470.
8. Kostyuchenko, V.A., et al., *Structure of the thermally stable Zika virus*. Nature, 2016. **533**: p. 425-428.
9. Abrams, R.P.M., J. Solis, and A. Nath, *Therapeutic Approaches for Zika Virus Infection of the Nervous System*. Neurotherapeutics, 2017. **14**(4): p. 1027-1048.
10. Perera-Lecoin, M., et al., *Flavivirus entry receptors: An update*. Viruses, 2013. **6**(1): p. 69-88.
11. Smit, J.M., et al., *Flavivirus cell entry and membrane fusion*. Viruses, 2011. **3**(2): p. 160-171.
12. Murray, C.L., C.T. Jones, and C.M. Rice, *Architects of assembly: Roles of Flaviviridae non-structural proteins in virion morphogenesis*. Nature Reviews Microbiology, 2008. **6**(9): p. 699-708.
13. Gillespie, L.K., et al., *The endoplasmic reticulum provides the membrane platform for biogenesis of the flavivirus replication complex*. Journal of Virology, 2010. **84**(20): p. 10438-10447.

14. Wu, J., et al., *Perturbation in the conserved methyltransferase-polymerase interface of flavivirus NS5 differentially affects polymerase initiation and elongation*. Journal of Virology, 2015. **89**(1): p. 249-261.
15. Welsch, S., et al., *Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites*. Cell Host and Microbe, 2009. **5**(4): p. 365-375.
16. Luo, D., et al., *Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein*. EMBO Journal, 2008. **27**(23): p. 3209-3219.
17. Yu, I.M., et al., *Structure of the immature dengue virus at low pH primes proteolytic maturation*. Science, 2008. **319**(5871): p. 1834-1837.
18. Kuhn, R.J., et al., *Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion*. Cell, 2002. **108**(5): p. 717-725.
19. Kampmann, T., et al., *In silico screening of small molecule libraries using the dengue virus envelope E protein has identified compounds with antiviral activity against multiple flaviviruses*. Antiviral Research, 2009. **84**(3): p. 234-241.
20. Stadler, K., et al., *Proteolytic activation of tick-borne encephalitis virus by furin*. Journal of Virology, 1997. **71**(11): p. 8475-8481.
21. Marchette, N.J., R. Garcia, and A. Rudnick, *Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia*. The American journal of tropical medicine and hygiene, 1969. **18**(3): p. 411-415.
22. Gourinat, A.C., et al., *Detection of zika virus in urine*. Emerging Infectious Diseases, 2015. **21**(1): p. 84-86.
23. Musso, D., et al., *Detection of Zika virus in saliva*. Journal of Clinical Virology, 2015. **68**: p. 53-55.
24. Dupont-Rouzeyrol, M., et al., *Infectious Zika viral particles in breastmilk*. The Lancet, 2016. **387**(10023): p. 1051.
25. Prisant, N., et al., *Zika virus in the female genital tract*. The Lancet Infectious Diseases, 2016. **16**(9): p. 1000-1001.
26. Mansuy, J.M., et al., *Zika virus: High infectious viral load in semen, a new sexually transmitted pathogen?* The Lancet Infectious Diseases, 2016. **16**(4): p. 405.

27. Fréour, T., et al., *Sexual transmission of Zika virus in an entirely asymptomatic couple returning from a Zika epidemic area, France, April 2016*. *Eurosurveillance*, 2016. **21**(23).
28. Davidson, A., et al., *Suspected female-to-male sexual transmission of zika virus — New York city, 2016*. *Morbidity and Mortality Weekly Report*, 2016. **65**(28): p. 716-717.
29. Besnard, M., et al., *Evidence of perinatal transmission of zika virus, French Polynesia, December 2013 and February 2014*. *Eurosurveillance*, 2014. **19**(13).
30. Costa, F., et al., *Emergence of congenital Zika syndrome: Viewpoint from the front lines*. *Annals of Internal Medicine*, 2016. **164**(10): p. 689-691.
31. Zammarchi, L., et al., *Zika virus infections imported to Italy: Clinical, immunological and virological findings, and public health implications*. *Journal of Clinical Virology*, 2015. **63**: p. 32-35.
32. Rozé, B., et al., *Zika virus detection in cerebrospinal fluid from two patients with encephalopathy, Martinique, February 2016*. *Eurosurveillance*, 2016. **21**(16).
33. Mécharles, S., et al., *Acute myelitis due to Zika virus infection*. *The Lancet*, 2016. **387**(10026): p. 1481.
34. Ventura, C.V., et al., *Zika virus in Brazil and macular atrophy in a child with microcephaly*. *The Lancet*, 2016. **387**(10015): p. 228.
35. Cao-Lormeau, V.M., et al., *Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: A case-control study*. *The Lancet*, 2016. **387**(10027): p. 1531-1539.
36. Mlakar, J., et al., *Zika virus associated with microcephaly*. *New England Journal of Medicine*, 2016. **374**(10): p. 951-958.
37. Lawn, N.D. and E.F.M. Wijdicks, *Fatal guillain-barre syndrome*. *Neurology*, 1999. **52**(3): p. 635-638.
38. Martines, R.B., et al., *Pathology of congenital Zika syndrome in Brazil: a case series*. *The Lancet*, 2016. **388**(10047): p. 898-904.
39. Barouch, D.H., S.J. Thomas, and N.L. Michael, *Prospects for a Zika Virus Vaccine*. *Immunity*, 2017. **46**(2): p. 176-182.

40. Bullard-Feibelman, K.M., et al., *The FDA-approved drug sofosbuvir inhibits Zika virus infection*. Antiviral Research, 2017. **137**: p. 134-140.
41. Saiz, J.C., et al., *Zika Virus: What have we learnt since the start of the recent epidemic?* Frontiers in Microbiology, 2017. **8**(AUG).
42. Lei, J., et al., *Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor*. Science, 2016. **353**(6298): p. 503-505.
43. Phoo, W.W., et al., *Structure of the NS2B-NS3 protease from Zika virus after self-cleavage*. Nature Communications, 2016. **7**.
44. Leung, D., et al., *Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors*. J Biol Chem, 2001. **276**(49): p. 45762-71.
45. Erbel, P., et al., *Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus*. Nat Struct Mol Biol, 2006. **13**(4): p. 372-373.
46. Chappell, K.J., et al., *Mutagenesis of the West Nile virus NS2B cofactor domain reveals two regions essential for protease activity*. Journal of General Virology, 2008. **89**(4): p. 1010-1014.
47. Skoreński, M., R. Grzywa, and M. Sieńczyk, *Why should we target viral serine proteases when developing antiviral agents?* Future Virology, 2016. **11**(12): p. 745-748.
48. Patick, A.K. and K.E. Potts, *Protease Inhibitors as Antiviral Agents*. Clinical Microbiology Reviews, 1998. **11**(4): p. 614-627.
49. Chen, X., et al., *Mechanisms of activation and inhibition of Zika virus NS2B-NS3 protease*. Cell Res, 2016. **26**(11): p. 1260-1263.
50. Zhang, Z., et al., *Crystal structure of unlinked NS2B-NS3 protease from Zika virus*. Science, 2016. **354**(6319): p. 1597-1600.
51. Aleshin, A.E., et al., *Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold*. Protein Science, 2007. **16**(5): p. 795-806.
52. Blow, D.M., J.J. Birktoft, and B.S. Hartley, *Role of a Buried Acid Group in the Mechanism of Action of Chymotrypsin*. Nature, 1969. **221**: p. 337.
53. Blow, D.M., *Structure and mechanism of chymotrypsin*. Accounts of Chemical Research, 1976. **9**(4): p. 145-152.

54. Warshel, A. and S. Russell, *Theoretical correlation of structure and energetics in the catalytic reaction of trypsin*. Journal of the American Chemical Society, 1986. **108**(21): p. 6569-6579.
55. Rodríguez, A., et al., *Comparison of Different Quantum Mechanical/Molecular Mechanics Boundary Treatments in the Reaction of the Hepatitis C Virus NS3 Protease with the NS5A/5B Substrate*. The Journal of Physical Chemistry B, 2007. **111**(44): p. 12909-12915.
56. Zhou, Y. and Y. Zhang, *Serine protease acylation proceeds with a subtle re-orientation of the histidine ring at the tetrahedral intermediate*. Chemical Communications, 2011. **47**(5): p. 1577-1579.
57. Zerner, B. and M.L. Bender, *The Kinetic Consequences of the Acyl-Enzyme Mechanism for the Reactions of Specific Substrates with Chymotrypsin*. Journal of the American Chemical Society, 1964. **86**(18): p. 3669-3674.
58. Hedstrom, L., L. Szilagyi, and W. Rutter, *Converting trypsin to chymotrypsin: the role of surface loops*. Science, 1992. **255**(5049): p. 1249-1253.
59. Harris, J.L., et al., *Definition and Redesign of the Extended Substrate Specificity of Granzyme B*. Journal of Biological Chemistry, 1998. **273**(42): p. 27364-27373.
60. De Clercq, E. and G. Li, *Approved antiviral drugs over the past 50 years*. Clinical Microbiology Reviews, 2016. **29**(3): p. 695-747.
61. Chambers, T.J., A. Nestorowicz, and C.M. Rice, *Mutagenesis of the yellow fever virus NS2B/3 cleavage site: Determinants of cleavage site specificity and effects on polyprotein processing and viral replication*. Journal of Virology, 1995. **69**(3): p. 1600-1605.
62. Li, J., et al., *Functional profiling of recombinant NS3 proteases from all four serotypes of dengue virus using tetrapeptide and octapeptide substrate libraries*. Journal of Biological Chemistry, 2005. **280**(31): p. 28766-28774.
63. Chappell, K.J., et al., *Insights to substrate binding and processing by West Nile Virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies*. Journal of Biological Chemistry, 2006. **281**(50): p. 38448-38458.

64. Stoermer, M.J., et al., *Potent cationic inhibitors of west Nile virus NS2B/NS3 protease with serum stability, cell permeability and antiviral activity*. Journal of Medicinal Chemistry, 2008. **51**(18): p. 5714-5721.
65. Behnam, M.A.M., et al., *Discovery of Nanomolar Dengue and West Nile Virus Protease Inhibitors Containing a 4-Benzoyloxyphenylglycine Residue*. Journal of Medicinal Chemistry, 2015. **58**(23): p. 9354-9370.
66. Nitsche, C., et al., *Peptide-Boronic Acid Inhibitors of Flaviviral Proteases: Medicinal Chemistry and Structural Biology*. Journal of Medicinal Chemistry, 2017. **60**(1): p. 511-516.
67. Chu, J.J.H., et al., *Antiviral activities of 15 dengue NS2B-NS3 protease inhibitors using a human cell-based viral quantification assay*. Antiviral Research, 2015. **118**: p. 68-74.
68. Yun, S.I., et al., *Complete genome sequences of three historically important, spatiotemporally distinct, and genetically divergent strains of Zika virus: MR-766, P6-740, and PRVABC-59*. Genome Announcements, 2016. **4**(4).
69. Peterlin-Mašič, L. and D. Kikelj, *Arginine mimetics*. Tetrahedron, 2001. **57**(33): p. 7073-7105.
70. Gustafsson, D., et al., *The direct thrombin inhibitor melagatran and its oral prodrug H 376/95: Intestinal absorption properties, biochemical and pharmacodynamic effects*. Thrombosis Research, 2001. **101**(3): p. 171-181.
71. Weigel, L.F., et al., *Phenylalanine and Phenylglycine Analogues as Arginine Mimetics in Dengue Protease Inhibitors*. Journal of Medicinal Chemistry, 2015. **58**(19): p. 7719-7733.
72. Yin, Z., et al., *Peptide inhibitors of dengue virus NS3 protease. Part 1: Warhead*. Bioorganic and Medicinal Chemistry Letters, 2006. **16**(1): p. 36-39.
73. Yin, Z., et al., *Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors*. Bioorganic and Medicinal Chemistry Letters, 2006. **16**(1): p. 40-43.
74. Nitsche, C., et al., *Thiazolidinone-peptide hybrids as dengue virus protease inhibitors with antiviral activity in cell culture*. Journal of Medicinal Chemistry, 2013. **56**(21): p. 8389-8403.

75. Behnam, M.A.M., et al., *C-terminal residue optimization and fragment merging: Discovery of a potent peptide-hybrid inhibitor of dengue protease*. ACS Medicinal Chemistry Letters, 2014. **5**(9): p. 1037-1042.
76. Bastos Lima, A., et al., *Dual inhibitors of the dengue and West Nile virus NS2B-NS3 proteases: Synthesis, biological evaluation and docking studies of novel peptide-hybrids*. Bioorganic and Medicinal Chemistry, 2015. **23**(17): p. 5748-5755.
77. Knox, J.E., et al., *Peptide inhibitors of West Nile NS3 protease: SAR study of tetrapeptide aldehyde inhibitors*. Journal of Medicinal Chemistry, 2006. **49**(22): p. 6585-6590.
78. Li, Y., et al., *Structural Dynamics of Zika Virus NS2B-NS3 Protease Binding to Dipeptide Inhibitors*. Structure, 2017. **25**(8): p. 1242-1250.e3.
79. Li, Y., et al., *Structural Insights into the Inhibition of Zika Virus NS2B-NS3 Protease by a Small-Molecule Inhibitor*. Structure, 2018. **26**(4): p. 555-564.e3.
80. Erbel, P., et al., *Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus*. Nature Structural and Molecular Biology, 2006. **13**(4): p. 372-373.
81. Noble, C.G., et al., *Ligand-bound structures of the dengue virus protease reveal the active conformation*. Journal of Virology, 2012. **86**(1): p. 438-446.
82. Pettersen, E.F., et al., *UCSF Chimera - A visualization system for exploratory research and analysis*. Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.
83. Blow, D.M., J.J. Birktoft, and B.S. Hartley, *Role of a buried acid group in the mechanism of action of chymotrypsin*. Nature, 1969. **221**(5178): p. 337-340.
84. Hedstrom, L., *Serine protease mechanism and specificity*. Chemical Reviews, 2002. **102**(12): p. 4501-4523.
85. Duan, Y., et al., *A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations*. Journal of Computational Chemistry, 2003. **24**(16): p. 1999-2012.
86. Hospital, A., et al., *Molecular dynamics simulations: Advances and applications*. Advances and Applications in Bioinformatics and Chemistry, 2015. **8**(1): p. 37-47.

87. Ponder, J.W. and D.A. Case, *Force fields for protein simulations*, in *Advances in Protein Chemistry*. 2003, Academic Press Inc. p. 27-85.
88. Cuendet, M.A. and W.F. Van Gunsteren, *On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm*. *Journal of Chemical Physics*, 2007. **127**(18).
89. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems*. *The Journal of Chemical Physics*, 1993. **98**(12): p. 10089-10092.
90. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. *Journal of Computational Physics*, 1977. **23**(3): p. 327-341.
91. Groenhof, G., *Introduction to QM/MM simulations*, in *Methods in Molecular Biology*. 2013, Humana Press Inc. p. 43-66.
92. Assfeld, X. and J.L. Rivail, *Quantum chemical computations on parts of large molecules: The ab initio local self consistent field method*. *Chemical Physics Letters*, 1996. **263**(1-2): p. 100-106.
93. Gao, J., et al., *A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations*. *Journal of Physical Chemistry A*, 1998. **102**(24): p. 4714-4721.
94. Hornak, V., et al., *Comparison of multiple amber force fields and development of improved protein backbone parameters*. *Proteins: Structure, Function and Genetics*, 2006. **65**(3): p. 712-725.
95. Rut, W., et al., *Extended substrate specificity and first potent irreversible inhibitor/activity-based probe design for Zika virus NS2B-NS3 protease*. *Antiviral Research*, 2017. **139**: p. 88-94.
96. Olsson, M.H.M., et al., *PROPKA3: Consistent treatment of internal and surface residues in empirical pK_a predictions*. *Journal of Chemical Theory and Computation*, 2011. **7**(2): p. 525-537.
97. Frisch, M.J., et al., *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.
98. Wang, J., et al., *Automatic atom type and bond type perception in molecular mechanical calculations*. *Journal of Molecular Graphics and Modelling*, 2006. **25**(2): p. 247-260.

99. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB*. Journal of Chemical Theory and Computation, 2015. **11**(8): p. 3696-3713.
100. Wang, J., et al., *Development and testing of a general Amber force field*. Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.
101. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
102. Meeprasert, A., S. Hannongbua, and T. Rungrotmongkol, *Key binding and susceptibility of NS3/4A serine protease inhibitors against hepatitis C virus*. Journal of Chemical Information and Modeling, 2014. **54**(4): p. 1208-1217.
103. Kongkaew, S., et al., *Interactions of HLA-DR and Topoisomerase I Epitope Modulated Genetic Risk for Systemic Sclerosis*. Scientific Reports, 2019. **9**(1).
104. Phanich, J., et al., *Role of R292K mutation in influenza H7N9 neuraminidase toward oseltamivir susceptibility: MD and MM/PB(GB)SA study*. Journal of Computer-Aided Molecular Design, 2016. **30**(10): p. 917-926.
105. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems*. The Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
106. Uberuaga, B.P., M. Anghel, and A.F. Voter, *Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation*. Journal of Chemical Physics, 2004. **120**(14): p. 6363-6374.
107. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath*. The Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
108. Roe, D.R. and T.E. Cheatham, *PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data*. Journal of Chemical Theory and Computation, 2013. **9**(7): p. 3084-3095.
109. Miller Iii, B.R., et al., *MMPBSA.py: An efficient program for end-state free energy calculations*. Journal of Chemical Theory and Computation, 2012. **8**(9): p. 3314-3321.
110. Shiryaev, S.A., et al., *Characterization of the Zika virus two-component NS2B-NS3 protease and structure-assisted identification of allosteric small-molecule antagonists*. Antiviral Research, 2017. **143**: p. 218-229.

111. Gordon, J.C., et al., *H++: a server for estimating pKa's and adding missing hydrogens to macromolecules*. *Nucleic Acids Research*, 2005. **33**(suppl_2): p. W368-W371.
112. Åqvist, J., *Ion-water interaction potentials derived from free energy perturbation simulations*. *Journal of Physical Chemistry*, 1990. **94**(21): p. 8021-8024.
113. Case, D.A., et al., *AMBER 16*. 2016, University of California, San Francisco.
114. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. *Journal of Computational Physics*, 1977. **23**(3): p. 327-341.
115. Reuter, N., et al., *Frontier bonds in QM/MM methods: A comparison of different approaches*. *Journal of Physical Chemistry A*, 2000. **104**(8): p. 1720-1735.
116. Field, M.J., P.A. Bash, and M. Karplus, *A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations*. *Journal of Computational Chemistry*, 1990. **11**(6): p. 700-733.
117. Stewart, J.J.P., *Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements*. *Journal of Molecular Modeling*, 2007. **13**(12): p. 1173-1213.
118. Kumar, S., et al., *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method*. *Journal of Computational Chemistry*, 1992. **13**(8): p. 1011-1021.
119. Roux, B., *The calculation of the potential of mean force using computer simulations*. *Computer Physics Communications*, 1995. **91**(1-3): p. 275-282.
120. Kumar, S., et al., *Multidimensional free-energy calculations using the weighted histogram analysis method*. *Journal of Computational Chemistry*, 1995. **16**(11): p. 1339-1350.
121. Helgaker, T., et al., *A priori calculation of molecular properties to chemical accuracy*. *Journal of Physical Organic Chemistry*, 2004. **17**(11): p. 913-933.
122. Claeysens, F., et al., *High-accuracy computation of reaction barriers in enzymes*. *Angewandte Chemie - International Edition*, 2006. **45**(41): p. 6856-6859.

123. Van Der Kamp, M.W., et al., *Testing high-level QM/MM methods for modeling enzyme reactions: Acetyl-CoA deprotonation in citrate synthase*. Journal of Physical Chemistry B, 2010. **114**(34): p. 11303-11314.
124. Sherwood, P., et al., *QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis*. Journal of Molecular Structure: THEOCHEM, 2003. **632**(1): p. 1-28.
125. Metz, S., et al., *ChemShell—a modular software package for QM/MM simulations*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2014. **4**(2): p. 101-110.
126. Neese, F., *The ORCA program system*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012. **2**(1): p. 73-78.
127. Smith, W. and T.R. Forester, *DL_POLY_2.0: A general-purpose parallel molecular dynamics simulation package*. Journal of Molecular Graphics, 1996. **14**(3): p. 136-141.
128. Senn, H.M. and W. Thiel, *QM/MM studies of enzymes*. Current Opinion in Chemical Biology, 2007. **11**(2): p. 182-187.
129. Grimme, S., S. Ehrlich, and L. Goerigk, *Effect of the damping function in dispersion corrected density functional theory*. Journal of Computational Chemistry, 2011. **32**(7): p. 1456-1465.
130. Lonsdale, R., J.N. Harvey, and A.J. Mulholland, *Inclusion of dispersion effects significantly improves accuracy of calculated reaction barriers for cytochrome P450 catalyzed reactions*. Journal of Physical Chemistry Letters, 2010. **1**(21): p. 3232-3237.
131. Lonsdale, R., J.N. Harvey, and A.J. Mulholland, *Effects of dispersion in density functional based quantum mechanical/molecular mechanical calculations on cytochrome P450 catalyzed reactions*. Journal of Chemical Theory and Computation, 2012. **8**(11): p. 4637-4645.
132. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*. Physical Review B, 1988. **37**(2): p. 785-789.
133. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior*. Physical Review A, 1988. **38**(6): p. 3098-3100.

134. Becke, A.D., *A new mixing of Hartree-Fock and local density-functional theories*. The Journal of Chemical Physics, 1993. **98**(2): p. 1372-1377.
135. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange*. The Journal of Chemical Physics, 1993. **98**(7): p. 5648-5652.
136. Kohn, W., A.D. Becke, and R.G. Parr, *Density functional theory of electronic structure*. Journal of Physical Chemistry, 1996. **100**(31): p. 12974-12980.
137. Zhao, Y. and D.G. Truhlar, *Benchmark databases for nonbonded interactions and their use to test density functional theory*. Journal of Chemical Theory and Computation, 2005. **1**(3): p. 415-432.
138. Zhao, Y., N. González-Garda, and D.G. Truhlar, *Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods*. Journal of Physical Chemistry A, 2005. **109**(9): p. 2012-2018.
139. Dilabio, G.A., E.R. Johnson, and A. Otero-De-La-Roza, *Performance of conventional and dispersion-corrected density-functional theory methods for hydrogen bonding interaction energies*. Physical Chemistry Chemical Physics, 2013. **15**(31): p. 12821-12828.
140. Gkionis, K., et al., *Performance of Becke's half-and-half functional for non-covalent interactions: Energetics, geometries and electron densities*. Journal of Molecular Modeling, 2009. **15**(9): p. 1051-1060.
141. Yu, W., et al., *Comparison of some representative density functional theory and wave function theory methods for the studies of amino acids*. Journal of Computational Chemistry, 2009. **30**(4): p. 589-600.
142. Werner, H.J., et al., *Molpro: A general-purpose quantum chemistry program package*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012. **2**(2): p. 242-253.
143. Grimme, S., *Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies*. Journal of Chemical Physics, 2003. **118**(20): p. 9095-9102.
144. Møller, C. and M.S. Plesset, *Note on an approximation treatment for many-electron systems*. Physical Review, 1934. **46**(7): p. 618-622.

145. Lawan, N., et al., *Comparison of DFT and ab initio QM/MM methods for modelling reaction in chorismate synthase*. Chemical Physics Letters, 2014. **608**: p. 380-385.
146. Antony, J. and S. Grimme, *Is spin-component scaled second-order møller-plesset perturbation theory an appropriate method for the study of noncovalent interactions in molecules?* Journal of Physical Chemistry A, 2007. **111**(22): p. 4862-4868.
147. Dunning Jr, T.H., *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen*. The Journal of Chemical Physics, 1989. **90**(2): p. 1007-1023.
148. Pulay, P., *Localizability of dynamic electron correlation*. Chemical Physics Letters, 1983. **100**(2): p. 151-154.
149. Pipek, J. and P.G. Mezey, *A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions*. The Journal of Chemical Physics, 1989. **90**(9): p. 4916-4926.
150. Werner, H.J. and M. Schütz, *An efficient local coupled cluster method for accurate thermochemistry of large systems*. Journal of Chemical Physics, 2011. **135**(14).
151. Anandakrishnan, R., B. Aguilar, and A.V. Onufriev, *H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations*. Nucleic Acids Research, 2012. **40**(W1): p. W537-W541.
152. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model*. Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
153. Cornell, W.D., et al., *Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation*. Journal of the American Chemical Society, 1993. **115**(21): p. 9620-9631.
154. Kaiyawet, N., T. Rungrotmongkol, and S. Hannongbua, *Effect of halogen substitutions on dUMP to stability of thymidylate synthase/dUMP/mTHF*

- ternary complex using molecular dynamics simulation.* Journal of Chemical Information and Modeling, 2013. **53**(6): p. 1315-1323.
155. Kongkaew, S., et al., *Molecular dynamics simulation reveals the selective binding of human leukocyte antigen alleles associated with behçet's disease.* PLoS ONE, 2015. **10**(9).
156. Andersen, H.C., *Molecular dynamics simulations at constant pressure and/or temperature.* The Journal of Chemical Physics, 1980. **72**(4): p. 2384-2393.
157. Repasky, M.P., J. Chandrasekhar, and W.L. Jorgensen, *PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods.* Journal of Computational Chemistry, 2002. **23**(16): p. 1601-1622.
158. Kumar, S., et al., *Multidimensional free-energy calculations using the weighted histogram analysis method.* Journal of Computational Chemistry, 1995. **16**(11): p. 1339-1350.
159. Rungrotmongkol, T., et al., *Combined QM/MM mechanistic study of the acylation process in furin complexed with the H5N1 avian influenza virus hemagglutinin's cleavage site.* Proteins: Structure, Function and Bioinformatics, 2009. **76**(1): p. 62-71.
160. Martínez-González, J.Á., et al., *Theoretical Study of the Free Energy Surface and Kinetics of the Hepatitis C Virus NS3/NS4A Serine Protease Reaction with the NS5A/5B Substrate. Does the Generally Accepted Tetrahedral Intermediate Really Exist?* ACS Catalysis, 2015. **5**(1): p. 246-255.
161. Rodríguez, A., et al., *Comparison of different quantum mechanical/molecular mechanics boundary treatments in the reaction of the hepatitis C virus NS3 protease with the NS5A/5B substrate.* Journal of Physical Chemistry B, 2007. **111**(44): p. 12909-12915.
162. Nutho, B., A.J. Mulholland, and T. Rungrotmongkol, *Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations Support a Concerted Reaction Mechanism for the Zika Virus NS2B/NS3 Serine Protease with Its Substrate.* The Journal of Physical Chemistry B, 2019.
163. Bolhuis, P.G., et al., *Transition path sampling: Throwing ropes over rough mountain passes, in the dark,* in *Annual Review of Physical Chemistry.* 2002. p. 291-318.

164. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: Survey of methods for simulating the activity of proteins*. Chemical Reviews, 2006. **106**(5): p. 1589-1615.
165. Metz, S., et al., *ChemShell-a modular software package for QM/MM simulations*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2014. **4**(2): p. 101-110.
166. Smith, W. and T.R. Forester, *DL-POLY-2.0: A general-purpose parallel molecular dynamics simulation package*. Journal of Molecular Graphics, 1996. **14**(3): p. 136-141.
167. Fox, S.J., et al., *Electrostatic embedding in large-scale first principles quantum mechanical calculations on biomolecules*. Journal of Chemical Physics, 2011. **135**(22).
168. Henkelman, G., B.P. Uberuaga, and H. Jónsson, *Climbing image nudged elastic band method for finding saddle points and minimum energy paths*. Journal of Chemical Physics, 2000. **113**(22): p. 9901-9904.
169. Goumans, T.P.M., et al., *An embedded cluster study of the formation of water on interstellar dust grains*. Physical Chemistry Chemical Physics, 2009. **11**(26): p. 5431-5436.
170. Yotmanee, P., et al., *Binding specificity of polypeptide substrates in NS2B/NS3pro serine protease of dengue virus type 2: A molecular dynamics Study*. Journal of Molecular Graphics and Modelling, 2015. **60**: p. 24-33.
171. Robin, G., et al., *Structure of West Nile Virus NS3 Protease: Ligand Stabilization of the Catalytic Conformation*. Journal of Molecular Biology, 2009. **385**(5): p. 1568-1577.
172. Gruba, N., et al., *Substrate profiling of Zika virus NS2B-NS3 protease*. FEBS Letters, 2016. **590**(20): p. 3459-3468.
173. Radisky, E.S. and D.E. Koshland Jr, *A clogged gutter mechanism for protease inhibitors*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(16): p. 10316-10321.
174. Garcia-Viloca, M., et al., *How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations*. Science, 2004. **303**(5655): p. 186-195.

175. Warshel, A., *Computer simulations of enzyme catalysis: Methods, progress, and insights*, in *Annual Review of Biophysics and Biomolecular Structure*. 2003. p. 425-443.
176. Van Der Kamp, M.W. and A.J. Mulholland, *Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology*. *Biochemistry*, 2013. **52**(16): p. 2708-2728.
177. Senn, H.M. and W. Thiel, *QM/MM methods for biomolecular systems*. *Angewandte Chemie - International Edition*, 2009. **48**(7): p. 1198-1229.
178. Jindal, G. and A. Warshel, *Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region*. *Journal of Physical Chemistry B*, 2016. **120**(37): p. 9913-9921.
179. Dhoke, G.V., et al., *QM/MM calculations revealing the resting and catalytic states in zinc-dependent medium-chain dehydrogenases/reductases*. *ACS Catalysis*, 2015. **5**(6): p. 3207-3215.
180. Korth, M. and W. Thiel, *Benchmarking semiempirical methods for thermochemistry, kinetics, and noncovalent interactions: OMx methods are almost as accurate and robust as DFT-GGA methods for organic molecules*. *Journal of Chemical Theory and Computation*, 2011. **7**(9): p. 2929-2936.
181. Chudyk, E.I., et al., *QM/MM simulations as an assay for carbapenemase activity in class A β -lactamases*. *Chemical Communications*, 2014. **50**(94): p. 14736-14739.
182. Callegari, D., et al., *L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib*. *Chemical Science*, 2018. **9**(10): p. 2740-2749.
183. Fritz, R.A., et al., *Multiscale Simulations of Clavulanate Inhibition Identify the Reactive Complex in Class A β -Lactamases and Predict the Efficiency of Inhibition*. *Biochemistry*, 2018. **57**(26): p. 3560-3563.
184. Kuiper, B.D., et al., *Increased activity of unlinked Zika virus NS2B/NS3 protease compared to linked Zika virus protease*. *Biochemical and Biophysical Research Communications*, 2017. **492**(4): p. 668-673.

185. Korth, M., *Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields*. Journal of Chemical Theory and Computation, 2010. **6**(12): p. 3808-3816.
186. Ishida, T. and S. Kato, *Theoretical perspectives on the reaction mechanism of serine proteases: The reaction free energy profiles of the acylation process*. Journal of the American Chemical Society, 2003. **125**(39): p. 12035-12048.
187. Daggett, V., S. Schröder, and P. Kollman, *Catalytic Pathway of Serine Proteases: Classical and Quantum Mechanical Calculations*. Journal of the American Chemical Society, 1991. **113**(23): p. 8926-8935.
188. Nemukhin, A.V., et al., *Modeling of serine protease prototype reactions with the flexible effective fragment potential quantum mechanical/molecular mechanical method*. Theoretical Chemistry Accounts, 2004. **111**(1): p. 36-48.
189. Guo, H., A. Wlodawer, and H. Guo, *A general acid-base mechanism for the stabilization of a tetrahedral adduct in a serine carboxyl peptidase: A computational study*. Journal of the American Chemical Society, 2005. **127**(45): p. 15662-15663.
190. Rodríguez, A., C. Oliva, and M. González, *A comparative QM/MM study of the reaction mechanism of the Hepatitis C virus NS3/NS4A protease with the three main natural substrates NS5A/5B, NS4B/5A and NS4A/4B*. Physical Chemistry Chemical Physics, 2010. **12**(28): p. 8001-8015.
191. Lima, M.C. and G.M. Seabra, *Reaction mechanism of the dengue virus serine protease: a QM/MM study*. Physical chemistry chemical physics : PCCP, 2016. **18**(44): p. 30288-30296.
192. Bentzien, J., et al., *Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: The nucleophilic attack in subtilisin*. Journal of Physical Chemistry B, 1998. **102**(12): p. 2293-2301.
193. Warshel, A., et al., *How do serine proteases really work?* Biochemistry, 1989. **28**(9): p. 3629-3637.
194. Laidler, K.J. and M.C. King, *The development of transition-state theory*. Journal of Physical Chemistry, 1983. **87**(15): p. 2657-2664.

195. Ranaghan, K.E., et al., *Ab Initio QM/MM Modeling of the Rate-Limiting Proton Transfer Step in the Deamination of Tryptamine by Aromatic Amine Dehydrogenase*. Journal of Physical Chemistry B, 2017. **121**(42): p. 9785-9798.
196. Antony, J., et al., *Protein-ligand interaction energies with dispersion corrected density functional theory and high-level wave function based methods*. Journal of Physical Chemistry A, 2011. **115**(41): p. 11210-11220.
197. Bender, M.L., F.J. Kezdy, and C.R. Gunter, *The Anatomy of an Enzymatic Catalysis. α -Chymotrypsin*. Journal of the American Chemical Society, 1964. **86**(18): p. 3714-3721.
198. Christensen, U. and H.H. Ipsen, *Steady-state kinetics of plasmin- and trypsin-catalysed hydrolysis of a number of tripeptide-p-nitroanilides*. BBA - Enzymology, 1979. **569**(2): p. 177-183.
199. Philipp, M. and M.L. Bender, *Kinetics of subtilisin and thiolsubtilisin*. Molecular and Cellular Biochemistry, 1983. **51**(1): p. 5-32.
200. Stein, R.L., *Catalysis by Human Leukocyte Elastase. 4.1 Role of Secondary-Subsite Interactions*. Journal of the American Chemical Society, 1985. **107**(20): p. 5767-5775.
201. Schramm, V.L., *Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes*, in *Annual Review of Biochemistry*. 2011. p. 703-732.
202. Bandyopadhyay, A. and J. Gao, *Targeting biomolecules with reversible covalent chemistry*. Current Opinion in Chemical Biology, 2016. **34**: p. 110-116.
203. Pan, Q., et al., *Telaprevir/boceprevir era: From bench to bed and back*. World Journal of Gastroenterology, 2012. **18**(43): p. 6183-6188.
204. Goldberg, A.L., *Development of proteasome inhibitors as research tools and cancer drugs*. Journal of Cell Biology, 2012. **199**(4): p. 583-588.
205. Kuzin, A.P., et al., *Inhibition of the SHV-1 β -lactamase by Sulfones: Crystallographic observation of two reaction intermediates with Tazobactam*. Biochemistry, 2001. **40**(6): p. 1861-1866.

206. Beadle, B.M., et al., *Structural milestones in the reaction pathway of an amide hydrolase: Substrate, acyl, and product complexes of cephalothin with AmpC β -lactamase*. *Structure*, 2002. **10**(3): p. 413-424.
207. Docquier, J.D. and S. Mangani, *An update on β -lactamase inhibitor discovery and development*. *Drug Resistance Updates*, 2018. **36**: p. 13-29.
208. Kang, C., et al., *Exploring the binding of peptidic West Nile virus NS2B-NS3 protease inhibitors by NMR*. *Antiviral Research*, 2013. **97**(2): p. 137-144.
209. Kang, C., T.H. Keller, and D. Luo, *Zika Virus Protease: An Antiviral Drug Target*. *Trends in Microbiology*, 2017. **25**(10): p. 797-808.
210. Nitsche, C., et al., *Biochemistry and medicinal chemistry of the dengue virus protease*. *Chemical Reviews*, 2014. **114**(22): p. 11348-11381.
211. Luo, D., S.G. Vasudevan, and J. Lescar, *The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development*. *Antiviral Research*, 2015. **118**: p. 148-158.
212. Johnston, P.A., et al., *HTS identifies novel and specific uncompetitive inhibitors of the two-component NS2B-NS3 proteinase of West Nile virus*. *Assay and Drug Development Technologies*, 2007. **5**(6): p. 737-750.
213. Koh-Stenta, X., et al., *Identification of covalent active site inhibitors of dengue virus protease*. *Drug Design, Development and Therapy*, 2015. **9**: p. 6389-6399.
214. Amaro, R.E. and A.J. Mulholland, *Multiscale methods in drug design bridge chemical and biological complexity in the search for cures*. *Nature Reviews Chemistry*, 2018. **2**: p. 0148.
215. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features*. *Biopolymers*, 1983. **22**(12): p. 2577-2637.
216. Zhao, Y. and G. Karypis, *Hierarchical clustering algorithms for document datasets*. *Data Mining and Knowledge Discovery*, 2005. **10**(2): p. 141-168.
217. Orlando, A. and W.L. Jorgensen, *Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions*. *Accounts of Chemical Research*, 2010. **43**(1): p. 142-151.

218. Asbóth, B. and L. Polgár, *Transition-State Stabilization at the Oxyanion Binding Sites of Serine and Thiol Proteinases: Hydrolyses of Thiono and Oxygen Esters*. *Biochemistry*, 1983. **22**(1): p. 117-122.
219. Bryan, P., et al., *Site-directed mutagenesis and the role of the oxyanion hole in subtilisin*. *Proceedings of the National Academy of Sciences of the United States of America*, 1986. **83**(11): p. 3743-3745.
220. Kuramochi, H., H. Nakata, and S.I. Ishij, *Mechanism of association of a specific aldehyde inhibitor, leupeptin, with bovine trypsin*. *Journal of Biochemistry*, 1979. **86**(5): p. 1403-1410.
221. Kaiyawet, N., et al., *High-level QM/MM calculations support the concerted mechanism for Michael addition and covalent complex formation in thymidylate synthase*. *Journal of Chemical Theory and Computation*, 2015. **11**(2): p. 713-722.
222. Schmidt, T.C., et al., *Protocol for Rational Design of Covalently Interacting Inhibitors*. *ChemPhysChem*, 2014. **15**(15): p. 3226-3235.
223. Schirmeister, T., et al., *Quantum Chemical-Based Protocol for the Rational Design of Covalent Inhibitors*. *Journal of the American Chemical Society*, 2016. **138**(27): p. 8332-8335.



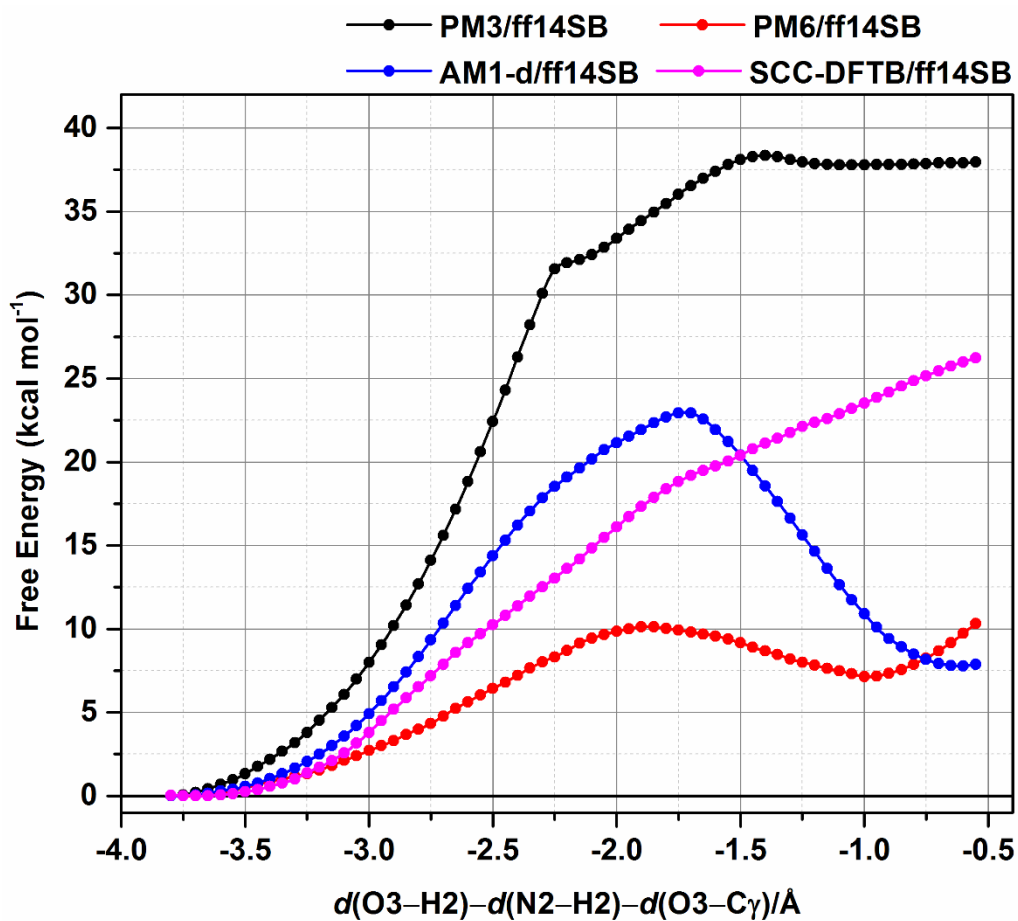
APPENDICES

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

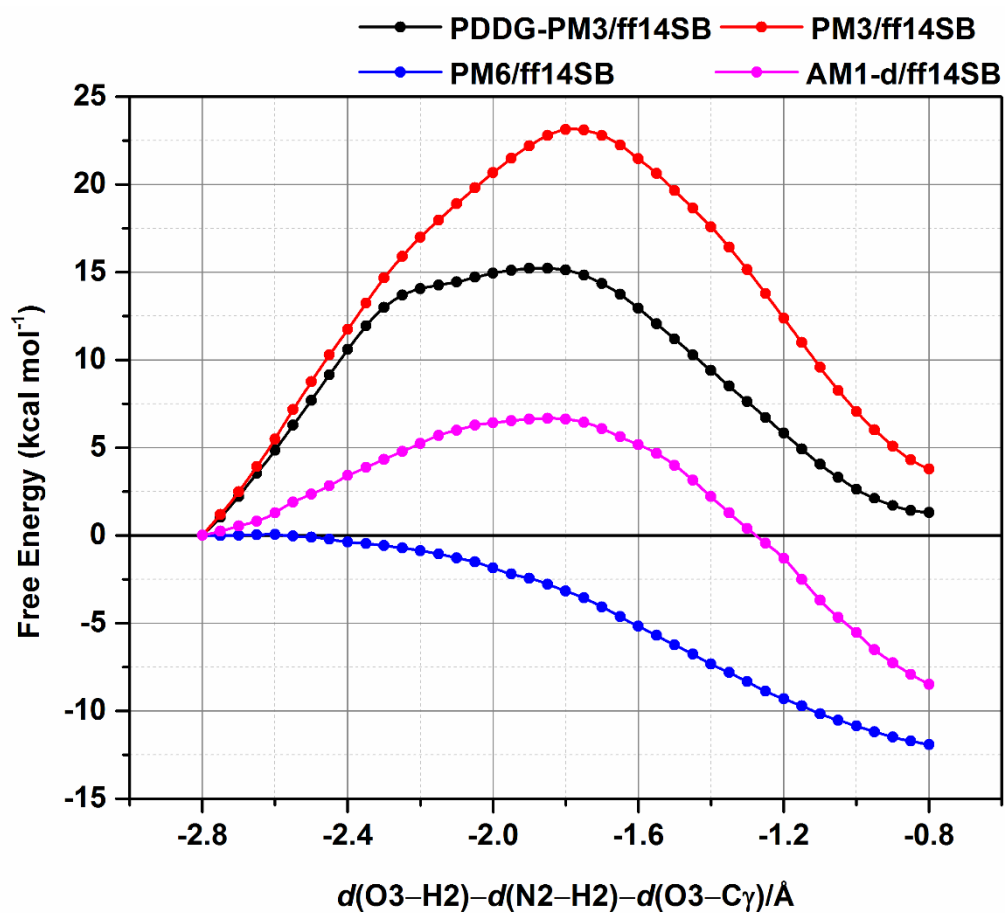
APPENDIX 1: The results from MM/GBSA method giving the energy components and average binding free energies (in kcal mol⁻¹) for the complexes of the four substrates with the ZIKV protease compared with the experimental K_m (in μM).

	Ac-nKKR-ACC	Ac-DRKOR-ACC	Ac-DKKOR-ACC	Bz-nKRR-AMC
ΔE_{ele}	-657.83 ± 40.68	-774.83 ± 34.80	-744.52 ± 33.93	-511.84 ± 33.79
ΔE_{vdW}	-58.25 ± 5.22	-57.68 ± 4.84	-56.76 ± 3.83	-57.39 ± 3.86
ΔE_{MM}	-716.08 ± 39.87	-832.52 ± 35.00	-801.29 ± 34.14	-569.23 ± 33.41
$\Delta G_{\text{solv}}^{\text{ele}}$	653.70 ± 36.51	765.52 ± 31.94	737.31 ± 31.46	511.45 ± 30.61
$\Delta G_{\text{solv}}^{\text{nonpolar}}$	-8.32 ± 0.39	-8.36 ± 0.36	-8.07 ± 0.25	-7.41 ± 0.36
ΔG_{solv}	645.38 ± 36.43	757.16 ± 31.87	729.24 ± 31.40	504.03 ± 30.44
ΔH	-70.70 ± 6.12	-75.35 ± 6.57	-72.04 ± 5.43	-65.20 ± 5.30
$-T\Delta S$	32.32 ± 4.79	30.29 ± 7.96	35.10 ± 5.11	30.83 ± 5.06
ΔG_{bind}	-38.38 ± 7.77	-45.06 ± 10.32	-36.94 ± 7.46	-34.36 ± 7.33
K_m (μM)	13.85 ± 0.30 [95]	4.20 ± 0.40 [95]	7.99 ± 0.73 [95]	20.42 ± 5.26 [43]

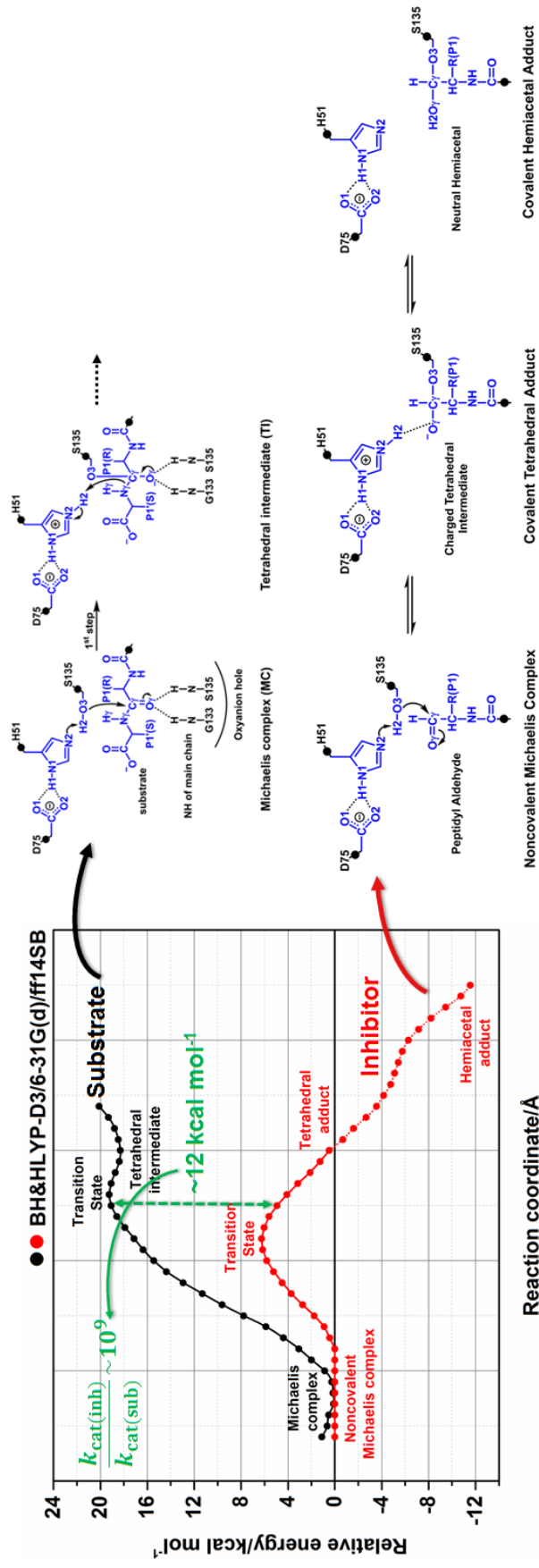
APPENDIX 2: Free energy profiles for the TI formation for the reaction of the ZIKV protease with its substrate, calculated from umbrella sampling simulations at the PM3/ff14SB, PM6/ff14SB, AM1-d/ff14SB and SCC-DFTB/ff14SB levels of theory.



APPENDIX 3: Free energy profile for formation of the tetrahedral adduct relative to the noncovalent Michaelis complex for the reaction of the ZIKV protease with dipeptidyl aldehyde inhibitor, calculated from umbrella sampling simulations at the PDDG-PM3/ff14SB, PM3/ff14SB, PM6/ff14SB and AM1-d/ff14SB levels of theory.



APPENDIX 4: Comparison of potential energy profiles for the reaction mechanism on substrate and dipeptidyl aldehyde inhibitor catalyzed by the ZIKV protease.



VITA

NAME	Mr. Bodee Nutho
DATE OF BIRTH	28 August 1991
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Chulalongkorn University
HOME ADDRESS	8 Ram Inthra Road, Ram Inthra, Khan Na Yao, Bangkok 10230
PUBLICATION	<ol style="list-style-type: none"> 1. B. Nutho, A. Mulholland, T. Rungrotmongkol, Reaction mechanism of Zika virus NS2B/NS3 serine protease inhibition by dipeptidyl aldehyde: A QM/MM study, <i>Physical Chemistry Chemical Physics</i>, Revised. 2. B. Nutho, T. Rungrotmongkol, Binding recognition of substrates in NS2B/NS3 serine protease of Zika virus revealed by molecular dynamics simulations, <i>Journal of Molecular Graphics and Modelling</i>, Revised. 3. B. Nutho, A. Mulholland, T. Rungrotmongkol, Quantum Mechanics/Molecular Mechanics (QM/MM) calculations support a concerted reaction mechanism for the Zika virus NS2B/NS3 serine protease with its substrate, <i>The Journal of Physical Chemistry B</i>, 2019, 123, 2889-2903. 4. J. Kammarabutr, P. Mahalapbutr, B. Nutho, N. Kungwan, T. Rungrotmongkol, Low susceptibility of asunaprevir towards R155K and D168A point mutations in HCV NS3/4A protease: A molecular dynamics simulation, <i>Journal of Molecular Graphics and Modeling</i>, 2019, 89, 122-130. 5. P. Mahalapbutr, B. Nutho, P. Wolschann, N. Kungwan, T. Rungrotmongkol, Molecular insights into inclusion complexes of mannonone E and H enantiomers with various β-cyclodextrins, <i>Journal of Molecular Graphics and Modeling</i>, 2018, 79, 72-80. 6. W. Panman, B. Nutho, S. Chamni, S. Dokmaisrijan, N. Kungwan, T. Rungrotmongkol, Computational screening of fatty acid synthase inhibitors against thioesterase domain, <i>Journal of Biomolecular Structure & Dynamics</i>, 2018, 36, 4114-4125. 7. B. Nutho, N. Nunthaboot, P. Wolschann, N. Kungwan, T. Rungrotmongkol, Metadynamics supports molecular dynamics simulation-based binding affinities of eucalyptol and beta-cyclodextrins inclusion complexes, <i>RSC Advances</i>, 2017, 7, 50899-50911.

8. P. Wongpituk, B. Nutho, W. Panman, N. Kungwan, P. Wolschann, T. Rungrotmongkol, N. Nunthaboot, Structural dynamics and binding free energy of neral-cyclodextrins inclusion complexes: molecular dynamics simulation, *Molecular Simulation*, 2017, 43, 1356-1363.
9. B. Nutho, M. Chulapa, A. Meeprasert, N. Kungwan, T. Rungrotmongkol, Screening of Hepatitis C NS5B polymerase inhibitors containing benzothiadiazine core: A steered molecular dynamics, *Journal of biomolecular structure and dynamics*, 2017, 8, 1743-1757.
10. W. Jetsadawisut, B. Nutho, A. Meeprasert, T. Rungrotmongkol, N. Kungwan, P. Wolschann, S. Hannongbua, Susceptibility of inhibitors against 3C protease of Coxsackievirus A16 and Enterovirus A71 causing hand, foot and mouth disease: A molecular dynamics study, *Biophysical Chemistry*, 2016, 219, 9-16.
11. B. Nutho, W. Khuntawee, C. Rungnim, P. Pongsawasdi, P. Wolschann, A. Karpfen, T. Rungrotmongkol, Binding mode and free energy prediction of fisetin/ β -cyclodextrin inclusion complex, *Beilstein Journal of Organic Chemistry*, 2014, 10, 2789-2799.

AWARD RECEIVED

1. B. Nutho, T. Rungrotmongkol. Susceptibility of potent inhibitors against NS2B/NS3 serine protease of Zika virus: A molecular dynamics study, The 21st International Annual Symposium on Computational Science and Engineering (ANSCSE 21). Thailand Science Park, Pathum Thani, Thailand, August 2-4, 2017. (3rd award for poster presentation)
2. B. Nutho, A. Mulholland, T. Rungrotmongkol. Reaction mechanism of the Zika virus NS2B/NS3 serine protease with its substrate: A QM/MM study, The 22nd International Annual Symposium on Computational Science and Engineering (ANSCSE 22). Faculty of Science, Chulalongkorn University, Bangkok, Thailand, August 2-3, 2018. (Outstanding award for oral presentation)