

วิธีทางคอมพิวเตอร์สำหรับการวิเคราะห์สารยับยั้งไวรัสไข้เลือดออกเอ็นเอสทูปี/เอ็นเอสทีโปรทีเอสและวีวัน
ไคเนส

นายกนิณ วิชาผง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาเคมี ภาควิชาเคมี

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2552

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

COMPUTER-BASED METHODS FOR ANALYZING INHIBITORS OF DENGUE
VIRUS NS2B/NS3 PROTEASE AND OF WEE1 KINASE

Mr. Kanin Wichapong

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Chemistry
Department of Chemistry
Faculty of Science
Chulalongkorn University
Academic year 2009
Copyright of Chulalongkorn University

กนิณ วิชาผง: วิธีทางคอมพิวเตอร์สำหรับการวิเคราะห์สารยับยั้งไวรัสไข้เลือดออกเอ็นเอสทูปี/เอ็นเอสทีโปรทีเอสและวีวันไคเนส. (COMPUTER-BASED METHODS FOR ANALYZING INHIBITORS OF DENGUE VIRUS NS2B/NS3 PROTEASE AND OF WEE1 KINASE) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. ศิริรัตน์ กักผล, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร. สมศักดิ์ เพ็ญวณิช, ศ. ดร. โวลฟ์กัง ซิลลป์, 172 หน้า.

ได้สร้างโครงสร้างสามมิติของเอ็นเอสทูปี/เอ็นเอสทีโปรทีเอสของไวรัสไข้เลือดออกที่เป็นสารเชิงซ้อนกับตัวยับยั้งเทระเปปไทด์ แล้วนำโครงสร้างที่ได้ไปทำการคำนวณโมเลกุลาร์ไดนามิกส์ซิมูเลชัน จากนั้นนำโครงสร้างที่เหมาะสมไปทำการคำนวณโมเลกุลาร์ดอกกิ้งกับสารยับยั้งที่มีขนาดเล็กรวมทั้งโมเลกุลาร์ไดนามิกส์ซิมูเลชัน ผลการคำนวณโมเลกุลาร์ไดนามิกส์ซิมูเลชันของสารเชิงซ้อนระหว่างเอ็นไซม์ชนิดนี้กับตัวยับยั้งเทระเปปไทด์และสารยับยั้งที่มีขนาดเล็กพบว่าเรซิดิวส์บริเวณปลายด้านซีของสายเอ็นเอสทูปี (แอสปาติกลำดับที่ 81 ถึง เซอร์รีนลำดับที่ 85), บริเวณเอสวัน (ลิวซีนลำดับที่ 128 ถึง ทีโอนีนลำดับที่ 134 ของสายเอ็นเอสที), ฮีสทีดีนลำดับที่ 51, แอสปาติกลำดับที่ 75, เซอร์รีนลำดับที่ 135, โกลซีนลำดับที่ 151, แอสปาราจีนลำดับที่ 152, โกลซีนลำดับที่ 153 และ ไทโรซีนลำดับที่ 161 ของสายเอ็นเอสทีมีความสำคัญต่อการเกิดอันตรกิริยากับสารยับยั้ง นอกจากนี้ผลการคำนวณยังแสดงให้เห็นว่า เอ็นเอสทูปีมีความสำคัญต่อการทำให้บริเวณการเข้าจับเกิดความเสถียรและการเกิดอันตรกิริยากับสารยับยั้ง จากนั้นได้ทำการคัดกรองเสมือนตามลำดับขั้นเพื่อหาสารยับยั้งเอ็นเอสทูปี/เอ็นเอสทีโปรทีเอสของไวรัสไข้เลือดออกโดยใช้เทคนิคการค้นหาฟามาโคฟอร์ร่วมกับโมเลกุลาร์ดอกกิ้งและตามด้วยการคำนวณพลังงานเสรีการยึดจับ สุดท้ายได้เสนอโครงสร้างของสารจากฐานข้อมูลทางการค้าที่คาดว่าจะมีประสิทธิภาพในการยับยั้งเอ็นไซม์ชนิดนี้เพื่อนำไปทดสอบฤทธิ์ทางชีวภาพ

ในส่วนที่สองได้ใช้เทคนิคคอมพิวเตอร์ดีควเอสเออาร์โมเดลสำหรับสารกลุ่ม ไพโรโรคาร์บาโซล ซึ่งเป็นสารยับยั้งวีวันไคเนส โมเดลที่ได้มีประสิทธิภาพในการทำนายที่ดี ซึ่งแสดงด้วยค่าทางสถิติต่างๆ ($r^2 = 0.870$, $q^2_{LOO} = 0.764$ และ $r^2_{pred.} = 0.790$) คอมพิวเตอร์ที่ได้ให้ข้อมูลที่สำคัญเกี่ยวกับอันตรกิริยาที่สำคัญของสารยับยั้งและบ่งบอกโครงสร้างที่สำคัญที่จะส่งผลกระทบต่อฤทธิ์ของสารยับยั้ง จากนั้นได้พัฒนาวิธีที่ใช้ในการทำนายพลังงานเสรีการยึดจับของสารยับยั้งกับเอ็นไซม์วีวันไคเนส โมเดลสำหรับสารกลุ่ม ไพโรโรคาร์บาโซล และสารกลุ่มพิริโดพรีมีดีน ได้ถูกสร้างขึ้น ซึ่งโมเดลที่ได้ให้ผลการคำนวณพลังงานเสรีการยึดจับคล่องกับผลการทดลอง นอกจากนี้การศึกษาเอนริชเม้นท์พบว่าเมื่อทำการคัดกรองฐานข้อมูลโดยใช้โมเดลที่สร้างขึ้นนี้ สารยับยั้งที่มีฤทธิ์เกือบจะทั้งหมดอยู่ใน 10% แรกของผลการคัดกรอง สุดท้ายได้ทำการคัดกรองเสมือนสำหรับสารยับยั้งวีวันไคเนสและได้เสนอสารที่คาดว่าจะมีฤทธิ์ในการยับยั้งเอ็นไซม์นี้เพื่อนำไปทดสอบฤทธิ์ทางชีวภาพต่อไป

ภาควิชา.....เคมี.....ลายมือชื่อ.....
 สาขาวิชา.....เคมี.....ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....
 ปีการศึกษา.....2552.....ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....
 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

4873801223 : MAJOR CHEMISTRY

KEYWORDS : Dengue Virus / NS2B/NS3 Protease / Wee1 Kinase

KANIN WICHAPONG: COMPUTER-BASED METHODS FOR ANALYZING INHIBITORS OF DENGUE VIRUS NS2B/NS3 PROTEASE AND OF WEE1 KINASE. THESIS ADVISOR: ASSOC. PROF. SIRIRAT KOKPOL, Ph. D., THESIS CO-ADVISOR: ASST. PROF. SOMSAK PIANWANIT, Ph. D., PROF. WOLFGANG SIPPL, Dr. habil., 172 pp.

Homology models of the Dengue virus (DV) NS2B/NS3 protease complexed with a tetra-peptidic inhibitor were constructed. Molecular dynamics (MD) simulations of these complexes were carried out to rationalize the ligand interaction. The validated model was then used for molecular docking studies of small-molecule inhibitors. The results derived from MD simulations of the complex between DV NS2B/NS3 protease and the tetra-peptidic inhibitor as well as with small-molecule inhibitors revealed that residues at the C-terminus of NS2B (Asp81-Ser85), at the S1 pocket (Leu128-Thr134 of NS3), His51, Asp75, Ser135, Gly151, Asn152, Gly153 and Tyr161 of NS3 are important for inhibitor interaction. Results also demonstrated that NS2B is important for stabilizing the binding pocket of NS3 as well as for stabilizing the binding of the tetra-peptidic inhibitor. A stepwise virtual screening (VS) for DV NS2B/NS3 protease inhibitors was carried out by combining pharmacophore and molecular docking-based screening with subsequent binding free energy calculation. Hit compounds were selected from commercial compound libraries and proposed for biological testing using the DV NS2B/NS3 protease.

In the second part, a 3D-QSAR model using the CoMFA approach was constructed for pyrrolocarbazole derivatives reported as Wee1 kinase inhibitors. The derived model was found to be robust and predictive, indicated by good statistical values ($r^2 = 0.870$, $q^2_{LOO} = 0.764$ and $r^2_{pred.} = 0.790$). The analysis of the graphical CoMFA contour plot provided insight into the relevant interactions of the inhibitors and the essential features of potent Wee1 kinase inhibitors. Subsequently, a structure-based approach was developed to predict the binding activities of Wee1 kinase inhibitors. Linear interaction energy models for pyrrolocarbazole and pyridopyrimidine derivative were established. The obtained models yielded a good correlation between the experimental binding affinities and the calculated binding free energies. A carried out enrichment study showed that most of the true active compounds could be obtained by screening only the first 10% of compound databases containing actives and decoys. Based on the validated linear interaction energy models and VS carried out on different compound collections, several hits were selected for biological testing against Wee1 kinase.

Department : Chemistry Student's Signature

Field of Study : Chemistry Advisor's Signature

Academic Year : 2009 Co-Advisor's Signature

Co-Advisor's Signature

ACKNOWLEDGMENTS

First of all, I would like to express my deep gratitude to my beloved parents and my family for their considerable and unconditional understanding, encouragement and support throughout my entire life.

I am truly grateful to Assoc. Prof. Sirirat Kokpol, Assist. Prof. Somsak Pianwanit and Prof. Dr. Wolfgang Sippl for giving me an opportunity and freedom while always pointing me in the right direction to work on this project. I very much appreciate their useful guidance, helpful comments and discussion and valuable suggestion throughout my study. I would like to additionally sincerely thank Prof. Dr. Wolfgang Sippl for his kind hospitality during my stay in Halle(Saale), Germany.

I am grateful to my thesis committee: Prof. Dr. Supot Hannongbua, Assoc. Prof. Vudhichai Parasuk, Assoc. Prof. Jiraporn Ungwitajatorn and Assoc. Prof. Anan Tongraar, for their time to read my dissertation, their advice and their contribution in taking part in my examination.

I would like to give special thanks to all present and former members of the Computation Chemistry Unit Cell (CCUC), Chulalongkorn University for their assistance, support and for providing a friendly and nice working atmosphere. I also would like to extend my additional thanks to all of my colleagues and friends in Halle(Saale) and Leipzig, Germany for assisting me to survive in Germany, sharing new ideas at work and supporting me in various ways.

Finally, this dissertation cannot be completely finished without financial and facilities support. Therefore, I gratefully acknowledge the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program Scholarship for the full financial support for my study and the expense for my first time stay in Germany. I also would like to acknowledge the DAAD (German Academic Exchange Service) scholarship for the expense for my second stay in Germany. The grant from Center of Petroleum, Petrochemical and Advanced Material, Chulalongkorn University is appreciated for supporting me to participate in an international conference. The CCUC, Department of Chemistry, Faculty of Science, Chulalongkorn University, and the Medicinal Research Group, Department of Pharmaceutical Chemistry, Martin-Luther University Halle-Wittenberg are acknowledged for all their computational support.

CONTENTS

	Page
ABSTRACT IN THAI.....	iv
ABSTRACT IN ENGLISH.....	v
ACKNOWLEDGMENTS.....	vi
CONTENTS.....	vii
LIST OF TABLES.....	xii
LIST OF FIGURES.....	xiv
LIST OF CHARTS.....	xxii
LIST OF ABBREVIATION.....	xxiii
CHAPTER I - INTRODUCTION.....	1
1.1. NS2B/NS3 protease of Dengue virus (DV).....	1
1.1.1 The Disease and History.....	1
1.1.2 Current Status in the World and in Thailand.....	2
1.1.3 Structure of Dengue virus.....	4
1.1.4 Replication Cycle of Dengue Virus.....	5
1.1.5 Enzyme Target (NS2B/NS3 Protease).....	6
1.1.6 Literature Reviews on DV NS2B/NS3 protease.....	9
1.2 Wee1 Kinase.....	18
1.2.1 Cancer and Cell Cycle.....	18
1.2.2 Literature Reviews on Wee1 Kinase.....	22
1.3 Scope of this Research Work.....	25
CHAPTER II - THEORY AND COMPUTATIONAL METHODS.....	28
2.1 Quantitative Structure-Activity Relationship (QSAR).....	28
2.1.1 Comparative Molecular Field Analysis (CoMFA).....	29
2.2 Homology Model.....	32
2.2.1 Template Selection.....	33
2.2.2 Sequence Alignment.....	34

2.2.3 Model Generation.....	36
2.2.4 Model Evaluation and Refinement.....	36
2.3 Molecular Dynamics (MD) Simulation.....	38
2.3.1 Theoretical Background.....	38
2.4 Binding Free Energy Calculation.....	42
2.4.1 MM/PB(GB)SA.....	43
2.4.2 Linear Interaction Energy with Continuum Electrostatics (LIECE).....	45
2.5 Virtual Screening.....	48
2.5.1 Pharmacophore Model and Pharmacophore Search.....	48
2.5.2 GRID – Molecular Field Calculation.....	49
2.5.3 Molecular Docking and Scoring Function.....	50
2.5.4 Similarity Search.....	52
CHAPTER III - NS2B/NS3 PROTEASE OF DENGUE VIRUS.....	53
3.1 Homology modeling and MD simulations of DV NS2B/NS3 Protease.....	53
3.1.1 Materials and Computational Methods.....	53
3.1.1.1 Sequence Alignment and Homology Modeling.....	53
3.1.1.2 Stereochemical quality.....	55
3.1.1.3 Protein and Inhibitor Preparation.....	56
3.1.1.4 Molecular Dynamics Simulations.....	57
3.1.1.5 GRID calculations.....	58
3.1.2 Results and Discussion.....	59
3.1.2.1 Model Building and Evaluation.....	59
3.1.2.2 Molecular Dynamics Simulations.....	60
3.1.2.2.1 Model Stability and Overall Structure.....	60
3.1.2.2.2 NS2B-NS3 Interaction.....	61
3.1.2.2.3 Protease-Inhibitor Interaction.....	64
3.1.2.3 Selectivity profile of the inhibitors Bz-Nle-Lys-Arg- Ala-H compared to Bz-Nle-Lys-Arg-Arg-H.....	69

3.2 MD simulation of DV NS2B/NS3 protease complex with small-molecule inhibitors and Binding Free Energy Calculation.....	73
3.2.1 Material and Computational Methods.....	74
3.2.1.1 Molecular docking of known small-molecule inhibitors.....	74
3.2.1.2 Molecular Dynamics (MD) Simulation.....	75
3.2.1.3 Binding Free Energy Calculation (MM/PBSA).....	76
3.2.2 Results and Discussion.....	77
3.2.2.1 Binding Free Energy of known inhibitors.....	77
3.2.2.2 Decomposition (DC) Energy per residue Analysis.....	79
3.3 Virtual Screening for inhibitors of Dengue Virus NS2B/NS3 protease.....	81
3.3.1 Material and Computational Methods.....	82
3.3.1.1 Databases.....	82
3.3.1.2 Pharmacophore Model and Pharmacophore Search.....	83
3.3.1.3 Database Filtration.....	83
3.3.1.4 Molecular Docking and Energy Minimization of docking solution.....	83
3.3.1.5 MD simulation and Binding Free Energy Calculation....	84
3.3.2 Results and Discussion.....	85
3.3.2.1 Structure-based Pharmacophore Model.....	85
3.3.2.1.1 Virtual Screening and Binding Free Energy Calculation.....	87
3.3.2.2 Comparison between Static and Dynamic Pharmacophore model.....	92
3.3.2.2.1 Pharmacophore Models (static and dynamic model) Generation.....	92
3.3.2.2.2 Virtual Screening.....	94
CHAPTER IV- Wee1 Kinase.....	102
4.1 3D-QSAR (CoMFA) model of checkpoint Wee1 kinase inhibitors.....	102
4.1.1 Material and Computational Methods.....	102
4.1.1.1 Inhibitor Data Set.....	102

4.1.1.2 Ligand Preparation.....	107
4.1.1.3 Molecular Docking.....	107
4.1.1.4 Ligand Alignment Method.....	108
4.1.1.5 CoMFA Model.....	109
4.1.1.6 Partial Least Square (PLS) Analysis.....	109
4.1.2 Results and Discussion.....	110
4.1.2.1 Ligand Docking.....	110
4.1.2.2 CoMFA Models.....	112
4.1.2.3 CoMFA Graphical Contour Plot.....	120
4.2 Structure-based prediction of activities of Wee1 Kinase inhibitors using the linear response MM/PBSA approach.....	123
4.2.1 Material and Computational methods.....	124
4.2.1.1 Inhibitor data set.....	124
4.2.1.2 Ligand preparation and docking.....	126
4.2.1.3 Energy minimization of docking poses.....	126
4.2.1.4 Interaction energy calculation.....	127
4.2.1.5 Binding free energy calculation using LR-MM/PBSA and LIECE models.....	127
4.2.2 Results and discussion.....	128
4.2.2.1 Binding mode of pyridopyrimidine derivatives.....	128
4.2.2.2 LR-MM/PBSA based prediction.....	131
4.2.2.3 Linear interaction energy with continuum electrostatics (LIECE) models.....	138
4.2.2.4 Enrichment Study.....	140
4.3 Virtual Screening for novel Wee1 Kinase inhibitors.....	144
4.3.1 Materials and Computational Methods.....	144
4.3.2 Results and Discussion.....	145
4.3.2.1 Pharmacophore Models Generation.....	145
4.3.2.2 Virtual Screening.....	147

CHAPTER V - CONCLUSION	149
5.1 DV NS2B/NS3 protease.....	149
5.2 Wee1 Kinase.....	151
REFERENCES	154
APPENDIX	168
VITAE	172

LIST OF TABLES

	Page
Table 1.1. RMSD values (Å) of NS2B, NS3, and NS2B/NS3 protease, indicated as the first, the second, and the third values, respectively, between each pair of different X-ray structures.....	12
Table 1.2. Chemical Structures and their K_i values of WNV and DV NS2B/NS3 protease inhibitors.....	16
Table 3.1. Comparison of the H-bond occupancy between residues from NS2B and NS3. The number of H-bond between each residue pair are given in the parenthesis, behind the percentage of H-bond occupancy.....	61
Table 3.2. Chemical Structures of the known small-molecule inhibitors and their K_i values against NS2B/NS3 protease of DV and WNV.....	74
Table 3.3. Components of MM/PBSA and binding free energy of known inhibitors.....	77
Table 3.4. Number of Hits derived from each steps of virtual screening using the structure-based pharmacophore model.....	87
Table 3.5. $\Delta G_{\text{binding}}$ of the hit compounds derived from virtual screening using the structure-based pharmacophore model.....	89
Table 3.6. Number of hits derived from each steps of Virtual Screening using PH-1 and PH-2 model.....	94
Table 3.7. $\Delta G_{\text{binding}}$ of hit compounds derived from PH-1.....	98
Table 3.8. $\Delta G_{\text{binding}}$ of hit compounds derived from PH-2.....	98
Table 3.9. Tanimoto Coefficient of hits compared to known small-molecule inhibitors. K_i values of these compounds against the DV NS2B/NS3 are also reported.....	100
Table 4.1: RMSD values between crystal structure and the top-ranked docking pose using different docking settings.....	110
Table 4.2. Statistics of generated CoMFA models using different alignment methods.....	114
Table 4.3. Statistics of the resulting CoMFA models.....	114

Table 4.4 List of actual pIC_{50} , predicted pIC_{50} (derived from leave-one-out cross-validation), estimated pIC_{50} values and residual for the 139 training set compounds as derived from the CoMFA model 10.....	115
Table 4.5. Actual pIC_{50} , predicted pIC_{50} and residual values of the 30 test set compounds.....	119
Table 4.6. Statistical Values of training set of LR-MM/PBSA models.....	131
Table 4.7. Comparison between $\Delta G_{exp.}$ and $\Delta G_{pred.}$, derived from model 1, of compounds in the test set of pyrrolocarbazole data set.....	135
Table 4.8. Comparison between $\Delta G_{exp.}$ and $\Delta G_{pred.}$, calculated from model 4, of compounds in the test set of pyridopyrimidine data set.....	136
Table 4.9. Comparison between $\Delta G_{exp.}$ and predicted $\Delta G_{pred.}$, calculated from model 8, of compounds in the test set of the general model.....	136
Table 4.10. Statistical values of results derived from applying LIECE models.	138
Table 4.11. Enrichment Factor (EF) of each dataset, containing decoys from Chembridge KINAsset with 6 different sets of active Wee1 kinase inhibitors and also decoys from DUD combining with the same datasets from active Wee1 kinase inhibitors, at 5% and 10% of screened database using LR-MM/PBSA model 6 and LIECE model 20.....	142
Table 4.12 Number of hit compounds derived from each pharmacophore model.....	146

LIST OF FIGURES

	Page
Figure 1.1 <i>Aedes aegypti</i> distribution in the Americas in 1970, at the end of the mosquito control program, and in 2002.....	2
Figure 1.2 High-risk area for the dengue transmission indicated in yellow color.....	2
Figure 1.3 Average annual number of DF/DHF cases reported to WHO and average annual number of countries reporting dengue.....	3
Figure 1.4 Schematic representation of DV genome organization and polyprotein, and sites where host protease (signalase) and DV protease (NS2B/NS3 protease) cleave the polyprotein.....	4
Figure 1.5 The replication cycle of Dengue Virus.....	5
Figure 1.6 Diagram represents the organization of DV polyprotein and NS2B/NS3 protease. An essential cofactor domain of NS2B (residues 49-95) for activating activity of NS3 protease is displayed in the white box. The NS3 protease domain is from residue 1 to 179, whereas NS3 Helicase domain locates at residue 180 to 618 of NS3. Amino acids diagram of the X-ray structure (PDB code 2VBC), a full length NS3 complexed with 18 amino acid residue of NS2B (residue 49-66), linked with G4-S-G4, is also displayed.....	7
Figure 1.7 (A) Interaction of the tetrapeptidic inhibitor with the WNV NS2B/NS3 protease (taken from the crystal structure 2FP7). (B) Schematic representation of the interaction of the inhibitor with the substrate pockets of WNV NS2B/NS3.....	10
Figure 1.8 An example of patented DV NS3 protease inhibitor.....	14
Figure 1.9 Example of some small-molecule inhibitors of DV NS2B/NS3 protease.....	14
Figure 1.10 Natural compounds as DV NS2B/NS3 protease inhibitor.....	15
Figure 1.11 Inhibitors of DV NS2B/NS3 protease and their EC ₅₀ and CC ₅₀ values.....	17

- Figure 1.12** A schematic representation shows cell cycle and CDK-cyclin complexes which are important in each checkpoint..... 18
- Figure 1.13** Checkpoint at each stage in the cell cycle..... 19
- Figure 1.14** A representation of cell cycle at the G2/M checkpoint; **(A)** CDK1(cdc2)-cyclin B complex is phosphorylated by Wee1 and Myt1 resulting in an inactivated state of this complex, giving time for repairing DNA damage. This process, a phosphorylation of CDK1(cdc2)-cyclin B complex, is important for preventing the DNA damage cell to go into the mitosis phase as displayed in **(B)** .. 20
- Figure 1.15** Chemical structure of PD0166285..... 21
- Figure 1.16** Overall structure of the Wee1 kinase (X-Ray Structure of the Wee1/PD0407824 complex, pdb code 1X8B). α -Helices are coloured magenta, β -sheets are coloured yellow and loops are shown in cyan..... 22
- Figure 1.17** Superimposition of the six solved crystal structures of Wee1-inhibitor complexes. The pyrrolocarbazole core of the inhibitors (colored orange) shows the same interaction at the ATP-binding site. Clusters of cocrystallized waters are displayed as balls. Hydrogen bonds are shown as dashed green line..... 23
- Figure 1.18** **(A)** Molecular structures of the pyrrolo[3,4-c]carbazole-1,3(2*H*,6*H*)-dione derivatives, and **(B)** Molecular structures of the 2-anilio-6-phenylpyrido[2,3-*d*]pyrimidin-7(8*H*)-ones derivatives representing as Wee1kinase inhibitors..... 25
- Figure 2.1** Two zones (twilight zone and the safe homology modeling zone) of sequence alignments. The homology model would be fine if the length and the percentage sequence identity compared to the template falls into the “safe” zone..... 33
- Figure 2.2** The Needleman-Wunsch algorithm gives the global sequence alignment between two sequences (on the top and on the left of the matrix), (a) scoring the sequence which is identical, (b) and (c) partially complete score matrix, (d) scoring the matrix till complete

(e) traceback route for giving globally optimum alignment. (f)	
Alternate alignment illustrating non-uniqueness.	35
Figure 2.3 Schematic representations for performing a MD simulation.....	40
Figure 3.1. Sequence alignment of DV and WNV NS2B and NS3 domains. Stars indicate residues that are identical whereas dots and colons indicate similar residues. Residues which are drawn italic and underlined are missing residues in the X-ray structures. Residues located in the S1, S2, S3 and S4 pocket are colored cyan, yellow, green and magenta, respectively, and catalytic triad are displayed in bold red.....	54
Figure 3.2. Detailed view of the interaction of the covalently bound peptidic inhibitor (ball and stick) with the catalytic site of WNV NS2B/NS3 protease (ribbon). The covalent bond (distance 1.36 Å) between the C atom of the inhibitor aldehyde warhead and the hydroxyl group (OG) of Ser135 of NS3 is shown.....	56
Figure 3.3. Superimposition of (A) 2FP7 and 2FOM, (B) 2FOM and DV-1, (C) WNV-X and DV-1, (D) WNV-X and DV-2, and (E) DV-1 and DV-2. The individual structures are colored as follows: WNV-X and 2FP7: NS2B blue, NS3 red, inhibitor cyan; 2FOM: NS2B magenta, NS3 white, DV-1: NS2B yellow, NS3 green, inhibitor orange; DV-2: NS2B Orange, NS3 gray, inhibitor magenta.....	59
Figure 3.4. (A) Comparison of the RMSD plots for the MD simulations of WNV-X (black-line) and DV-1 (red line), (B) RMSD plots for the simulations of WNV-X (black line) and DV-2 (green line), and (C) RMSD plots for the simulations of WNV-X (black line) and DV-3 (blue line)	60
Figure 3.5. Comparison of the H-bond occupancy between each residue of the inhibitor and the individual binding pockets of the protease; P1-S1 pocket (A), P2-S2 pocket (B), P3-S3 pocket (C) and P4-S4 pocket (D).....	66
Figure 3.6. Plot of the salt-bridge distance between Arg-P1 and Asp129 from	

- NS3 of WNV-X (black), DV-1 (red) and DV-2 (green)..... 67
- Figure 3.7.** Distance between O δ -Asn152 (NS3) and N ϵ -Arg-P2 of the inhibitor during the simulation of WNV-X (black), DV-1 (red) and DV-2 (green)..... 68
- Figure 3.8.** (A) GRID results: N⁺ probe (magenta contour) at the contour level of -5.25 kcal/mol of WNV-X (A-1) and DV-2 (A-2), (B) OH probe (green contour) at the contour level of -7.50 kcal/mol of WNV-X (B-1) and DV-2 (B-2) and (C) Dry probe (yellow contour) at the contour level of -0.50 kcal/mol of WNV-X (C-1) and DV-2 (C-2). The inhibitor is shown colored cyan..... 71
- Figure 3.9.** (A) RMSD values of the inhibitor Bz-Nle-Lys-Arg-Arg-H bound in WNV-X (black line) and DV-2 (green line). (B) RMSD values of the mutated inhibitor Bz-Nle-Lys-Arg-Ala-H bound to WNV-X (black line) and DV-2 (green line) 71
- Figure 3.10.** Comparing the conformation of the WNV protease bound inhibitor Bz-Nle-Lys-Arg-Arg-H (orange) and the bound inhibitor Bz-Nle-Lys-Arg-Ala-H (yellow) at the end of the 5 ns simulation time..... 73
- Figure 3.11.** Comparing the conformation of the DV protease bound inhibitor Bz-Nle-Lys-Arg-Arg-H (orange) and the bound inhibitor Bz-Nle-Lys-Arg-Ala-H (yellow) at the end of the 5 ns simulation time..... 73
- Figure 3.12.** Footprints of (A) van der Waals interaction, (B) electrostatic interaction, (C) free energy of solvation and (D) total interaction energy of ligand-protein interaction per residue..... 79
- Figure 3.13.** (A) Molecular docking solution of compound 4 (orange balls and stick) with the DV NS2B/NS3 protease. H-bonds are shown as dot lines. (B) GRID result of OH probe (yellow) at the contour level -5.5 kcal/mol, (C) GRID result of N:= probe (N-sp² with lone pair) (cyan) at the contour level -4.0 kcal/mol and (D) The final pharmacophore model derived from docking solution of compound 4 (sphere A, B and C), OH contours of GRID field calculation

- (sphere D) and N:= contours (sphere D and E). Red spheres represent the H-bond acceptor areas, Green sphere means H-bond donor areas and, yellow sphere indicates hydrophobic areas and magenta sphere denotes H-bond acceptor or donor area..... 85
- Figure 3.14.** (A) Docking solution of compound 4; H-bonds are shown as dot lines and ligand is displayed as orange stick. (B) A static pharmacophore model (PH-1); Pharmacophore features are displayed as yellow sphere = Hydrophobic feature, dark red sphere = H-bond acceptor features, dark green sphere = H-bond donor features and gray = excluded volumes in all pictures. For clearly display, some exclude volume spheres were omitted..... 92
- Figure 3.15.** (A) All ten different static pharmacophore models derived from each representative conformation of each cluster projected on the binding pocket of DV NS2B/NS3 protease, and (B) dynamic pharmacophore model; Pharmacophore features are displayed as yellow sphere = Hydrophobic feature, dark red sphere = H-bond acceptor features, dark green sphere = H-bond donor features and gray = excluded volumes in all pictures. For clearly display, some exclude volume spheres were omitted..... 93
- Figure 3.16.** Chemical Structures of hit compounds derived from PH-1..... 96
- Figure 3.17.** Chemical Structures of hit compounds derived from PH-2..... 97
- Figure 4.1.** (A) Alignment 1 derived from the GOLD docking study, (B) Alignment 2 obtained by flexibly aligning the docking poses on the template structure of inhibitor 1, (C) Alignment 3 derived by minimizing the docking poses in the wee1 active site. Inhibitor 1 is shown in green..... 108
- Figure 4.2.** Comparison of the docking solution for **1** (colored cyan) compared with its position in the crystal structure (colored gray, hydrogen-bonds are shown as dashed line, cocrystallized water molecules as red balls and the magnesium ion as green ball)..... 111
- Figure 4.3** (A). GOLD docking solution for the most potent inhibitor **106**

	(orange). Hydrogen bonds are shown as dashed line. (B) . Schematic representation of the interaction of the most potent inhibitor 106 with the residues of the Wee1 binding site. Hydrogen bonds are indicated as arrows.....	112
Figure 4.4	(A) . Correlation between the actual pIC_{50} and the predicted pIC_{50} obtained with the best CoMFA model 10. (B) . Correlation between the actual pIC_{50} and the estimated pIC_{50} of the best CoMFA model 10.....	118
Figure 4.4	(C) Prediction of the external test set containing 30 inhibitors.....	120
Figure 4.5	(A) . CoMFA steric field projected on the Wee1 binding pocket. The docked inhibitor 106 is displayed in ball and stick. The Connolly molecular surface of the enzyme is shown in white. (Color code: favoured, green (contribution level 80%); disfavoured yellow (contribution level 20 %)).....	121
Figure 4.5	(B) . CoMFA electrostatic field overlaid on the docked inhibitor 169 (colored orange). The Connolly molecular surface of the enzyme is shown in white. (Color code: increase in positive charge favoured, blue (contribution level 80%); increase in negative charge favoured red (contribution level 20 %)).....	121
Figure 4.6	Docking solutions of all compounds of pyridopyrimidine derivatives projected on the binding pocket of Wee1 kinase. Hydrogen bonds are shown as dashed line, cocrystallized water molecules are presented as red balls. Ligands are displayed as line whereas important amino acid residues are shown as balls and stick.....	128
Figure 4.7	(A) GOLD docking solution for one of the potent inhibitors, compound 218, (B) Schematic representation of the interaction of compound 218 with the residues at the Wee1 binding site. Hydrogen bonds are indicated as arrows.....	130
Figure 4.8	Correlation between ΔG_{exp} and GoldScore (A) , and ChemScore (B) of all 222 compounds applied in this study.....	130

Figure 4.9. Correlation between ΔG_{exp} and ΔG_{est} of the compounds in the training set of (A) pyrrolocarbazole using model 1, (B) pyridopyrimidine using model 4, and (C) the general model using model 8.....	133
Figure 4.10. Correlation between ΔG_{exp} and ΔG_{pred} of the compounds in the test set of pyrrolocarbazole subset.....	135
Figure 4.11. Correlation between ΔG_{exp} and ΔG_{pred} of the compounds in the test set of pyridopyrimidine subset.....	136
Figure 4.12. Correlation between ΔG_{exp} and ΔG_{pred} of the compounds in the test set of the general model.....	137
Figure 4.13. Correlation between ΔG_{exp} and ΔG_{pred} of the all 222 compounds derived by using LIECE model 20.....	140
Figure 4.14. EF curves of datasets of Chembridge Kinaset decoys with 6 different datasets from active Wee1 kinase inhibitors ranked by LR-MM/PBSA model 6 (A) and LIECE model 20 (B), and the EF curves of datasets of decoys from DUD with 6 different datasets from active Wee1 kinase inhibitors scored by LR-MM/PBSA model 6 (C) and LIECE model 20 (D).....	144
Figure 4.15 Three different pharmacophore models (A) pharmacophore model derived from X-ray structure 2IN6 (PH-1), (B) pharmacophore model generated from the docking solution of compound 106 (PH-2), (C) pharmacophore model (PH-3) built by merging the feature at the region A from PH-1 and PH-2. Pharmacophore feature represent by magenta sphere = H-bond/donor feature, yellow sphere = hydrophobic feature, red sphere or arrow = H-bond acceptor and green sphere or arrow = H-bond donor.....	145
Figure 4.16 EF Curve of the database containing 222 active Wee1 inhibitors and 3,037 hit compounds obtained from pharmacophore search and drug-like property filtering. The database was ranked by GoldScore.....	147

Figure 4.17 Chemical structures of Hit compounds and their predicted binding free energy.....	148
------------------------------------------------------------------------------------------------------	-----

LIST OF CHARTS

	Page
Chart 4.1 Molecular structures of the pyrrolo[3,4-c]carbazole-1,3(2 <i>H</i> ,6 <i>H</i>)-dione derivatives used for the 3D-QSAR study.....	103
Chart 4.2. Molecular structures of the 2-anilio-6-phenylpyrido[2,3- <i>d</i>]pyrimidin-7(8 <i>H</i>)-ones derivatives used in this study.....	125

LIST OF ABBREVIATION

2D	=	Two Dimension
3D	=	Three Dimension
Ala(A)	=	Alanine
Arg(R)	=	Arginine
Asn(N)	=	Asparagine
Bz	=	Benzene
C	=	Capsid Protein
CDK	=	Cyclin Dependent Kinase
CoMFA	=	Comparative Molecular Field Analysis
CoMSIA	=	Comparative Molecular Similarity Indices Analysis
Cys(C)	=	Cysteine
DC	=	Decomposition
DF	=	Dengue Fever
DHF	=	Dengue Hemorrhagic Fever
DSS	=	Dengue Shock Syndrome
DUD	=	Directory of Useful Decoys
DV	=	Dengue Virus
E	=	Envelope Protein
E_{ele}	=	Electrostatic Interaction
ER	=	Endoplasmic Reticulum
E_{vdW}	=	van der Waals Interaction
FEP	=	Free Energy Perturbation
GA	=	Genetic Algorithm
GB	=	Generalized Born
$G_{ele-sol}$	=	Electrostatic Free Energy of Solvation
Gln(Q)	=	Glutamine
Glu(E)	=	Glutamic Acid
Gly(G)	=	Glycine
$G_{nonele-sol}$	=	Non-Electrostatic Free Energy of Solvation
GOLD	=	Genetic Optimisation for Ligand Docking
His(H)	=	Histidine

HTS	=	Hight Throughput Srceening
Ile(I)	=	Isoleucine
K_i	=	Inhibition Constant
Leu(L)	=	Leucine
LIE	=	Linear Interaction Energy
LIECE	=	Linear Interaction Energy with Continuum Electrostatics
LR	=	Linear Response
Lys(K)	=	Lysine
M	=	Membrane
MD	=	Molecular Dynamics
Met(M)	=	Methionine
MM/PBSA	=	Molecular Mechanic/Possion-Boltzmann Surface Area
Nle	=	Norleucine
NS	=	Non-Structural Protein
PDB	=	Protein Data Bank
Phe(F)	=	Phenylalanine
PLS	=	Partial Least Square
Pro(P)	=	Proline
prM	=	Precusor Membrane
QSAR	=	Quantitative Structure-Activity Relationship
RMSD	=	Root Mean Square Deviation
RMSE	=	Root Mean Square of Error
SASA	=	Solvent Accessible Surface Area
SCRs	=	Structurally Conserved Regions
SDEE	=	Standard Error of Estimation
SDEP	=	Standard Error of Prediction
Ser(S)	=	Serine
TI	=	Thermodynamic Integration
Thr(T)	=	Threonine
Trp(W)	=	Tryptophan
Tyr(Y)	=	Tyrosine

Val(V)	=	Valine
VP	=	Vesicle Packets
VRs	=	Variable Regions
VS	=	Virtual Screening
WHO	=	World Health Organization
WNV	=	West Nile Virus
XRMSE	=	leave-one-out cross-validation Root Mean Square of Error

CHAPTER I

INTRODUCTION

1.1. NS2B/NS3 protease of Dengue virus (DV)

1.1.1 The Disease and History

Dengue virus (DV) is a member of genus *Flavivirus*, family Flaviviridae. DV is transmitted by the bite of infected mosquitoes type *Aedes aegypti* and *Aedes albopictus*. [1] Infection of all serotypes of DV (DV-1, DV-2, DV-3 and DV-4) results in a diversity of illnesses ranking from no apparent symptoms, and mild symptoms called dengue fever (DF), to more severe forms, such as dengue hemorrhagic fever (DHF) or dengue shock syndrome (DSS). DF causes a high fever, rash, and muscle and joint pain, whereas DHF and DSS could be fatal because of the hemorrhaging leading to dramatic loss of blood pressure. [2, 3] The first infection of each serotype of DV induces long-life immunity for the infected serotype; however, the second infection of a different type of DV could result in DHF or DSS. [4] Nowadays, there are no specific treatments, drug, or any antiviral therapies, for the DV infection. Development of vaccines for DF and DHF is quite difficult because all of four serotypes can cause the disease, and a second infection of any different serotypes from the protection serotype could result in the more serious disease, such as DSS or DHF. Fortunately, from the considerable effort in developing therapeutic treatment against DV infection, the number of vaccines candidates for all four serotypes are now in clinical trial phase I and phase II testing. [5]

DF is not a new disease. It was first recorded in a Chinese encyclopedia of disease symptoms and remedies published during the Chin Dynasty (265 to 420 A.D.). [2] DF was distributed worldwide in the tropics during the 18th and 19th centuries. The first official record of the outbreak of DHF was in Manila, Philippines in 1953-1954, subsequently in Bangkok, Thailand 1958, and then in Malaysia, Singapore and Vietnam in the 1960s. During this period the epidemic of DHF was located only in Southeast Asia because of the isolation of the Pacific island and the success of the program for controlling urban yellow fever, which is also transmitted by mosquito type *Aedes aegypti*, in Americas. However, *Aedes aegypti* re-invaded in the 1970s in most countries in the Americas as shown in Figure 1.1 because of the collapse of the mosquito control program.

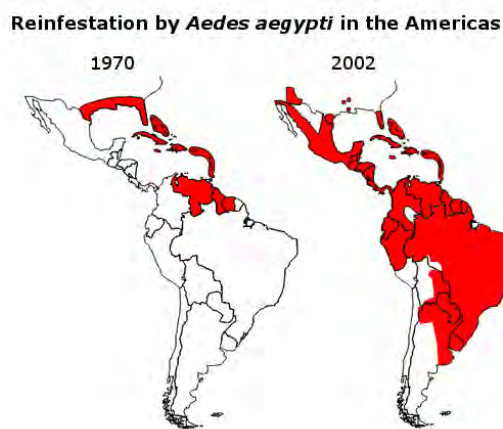


Figure 1.1 *Aedes aegypti* distribution in the Americas in 1970, at the end of the mosquito control program, and in 2002[6]

Furthermore, modern transportation facilitating the movement of people and commodities across continents leads to the spread of the vectors, *Aedes aegypti* and *Aedes albopictus*, and also the viruses. In addition, there are other factors such as global warming, rapid population growth and rural-urban migration that lead to the increasing of the epidemic of DV. Moreover, an insufficient basic urban infrastructure such as an inadequate water supply resulting in storing water in containers close to houses and an enhancing of an amount of solid wastes generate epidemiological conditions for the spreading of DV.[7, 8] Consequently, since the 1980s, both the mosquito vectors and the viruses continue spreading globally, resulting in the growth of the number of DF and DHF cases.[9]

1.1.2 Current Status in the World and in Thailand

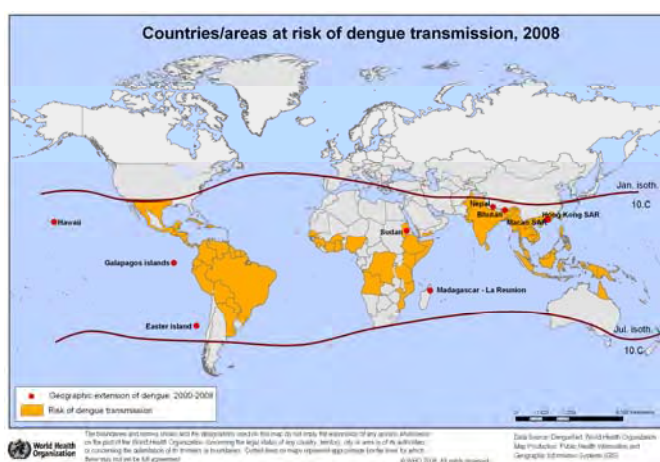


Figure 1.2 High-risk area for the dengue transmission indicated in yellow color[10]

Nowadays, the global epidemic of the dengue virus infections is comparable to malaria. More than 2.5 billion people in more than 100 countries, mostly in tropical and subtropical regions, are living in the risk area of the transmission of the DV[9] as indicated in the map displayed in Figure 1.2. The number of the DF cases reported to the World Health Organization (WHO) is annually around 1 million with 500,000 cases of DHF and 22,000 deaths, mostly among children. However, the number of infected DV including the cases which are not reported to WHO could be estimated at around 50 million in each year. The number of DF cases is increasing significantly and continuously, since the first epidemic in 1950 as the graph showed in Figure 1.3[8].

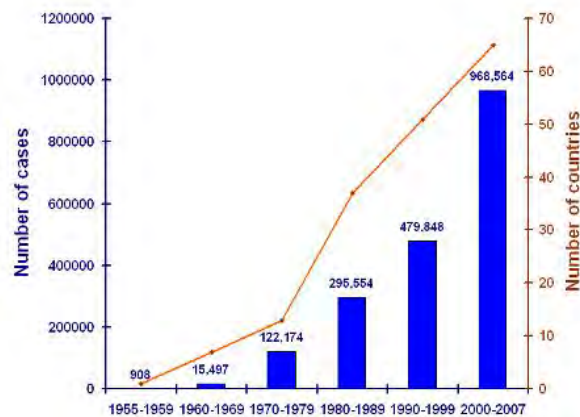


Figure 1.3 Average annual number of DF/DHF cases (bar graph) reported to WHO and average annual number of countries (line) reporting dengue.[8]

The situation of the epidemic of DV infection in Thailand is quite serious. The Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health of Thailand reported that there are approximately 50,000 - 60,000 DV infected cases and 50 – 60 fatality cases in Thailand every year. However, in the year 2008, there were a total of 87,494 cases of DV infection (138.80 infected cases per 100,000 population), and 101 patients died. The number of cases in the year 2008 doubled compared with the last few years. In the year 2007, there were 65,581 infected cases (104.21 cases per 100,000 population), and 95 patients died, whereas 42,456 infected cases (59 case per 100,000 population) and 59 fatal cases were reported in 2006. The distribution of the infected patients according to age shows that most of the patients are children in the 10-14 age group, around 400 cases per 100,000 population,

followed by the 5-9 age group (~315 cases/100,000 population), 15-24 age group (~170 cases/100,000 population), 0-4 age group (~110 cases/100,000 population), 25-34 age group (~55 cases/100,000 population), and older than 35 age group (~20 case/100,000 population).[11]

1.1.3 Structure of Dengue virus

Three dimension (3D) structure of DV was solved in 2002.[12] Both mature and immature particles of the virus are found in infected cells. The mature particle has a diameter of about 500 Å, whereas the value of the immature one is about 600 Å. Both particles contain the well-defined outer glycoprotein shell consisting of 180 copies each of an envelope (E) and membrane protein (prM/M) and internal host derived lipid bilayer.[13] The positive-strand RNA genome of ~11 kb in length of the virus is packaged by the viral capsid protein (C) within the lipid bilayer.[14] The viral RNA is capped by a type I methyl-guanosine at the 5'-end but contains no polyadenylate tail at the 3'-end. The RNA of DV as shown in Figure 1.4 encodes three structural proteins (C, prM and E) which form the components of the virus and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) which are important for the replication cycle of the virus. [13, 15]

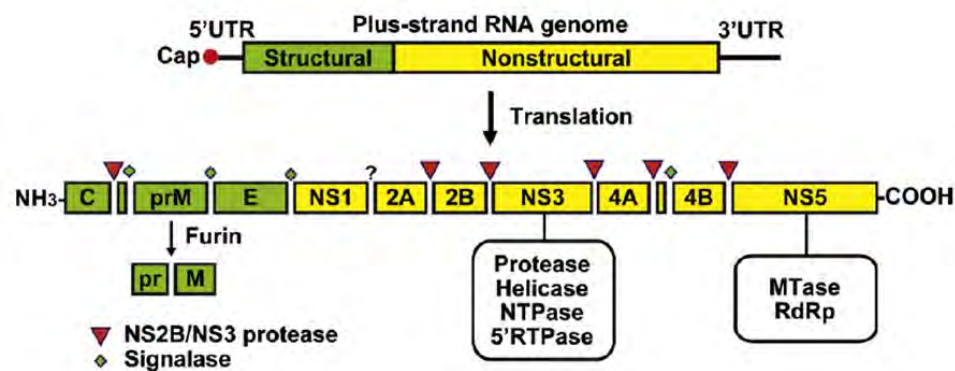


Figure 1.4 Schematic representation of DV genome organization and polyprotein, and sites where host protease (signalase) and DV protease (NS2B/NS3 protease) cleave the polyprotein[14]

1.1.4 Replication Cycle of Dengue Virus

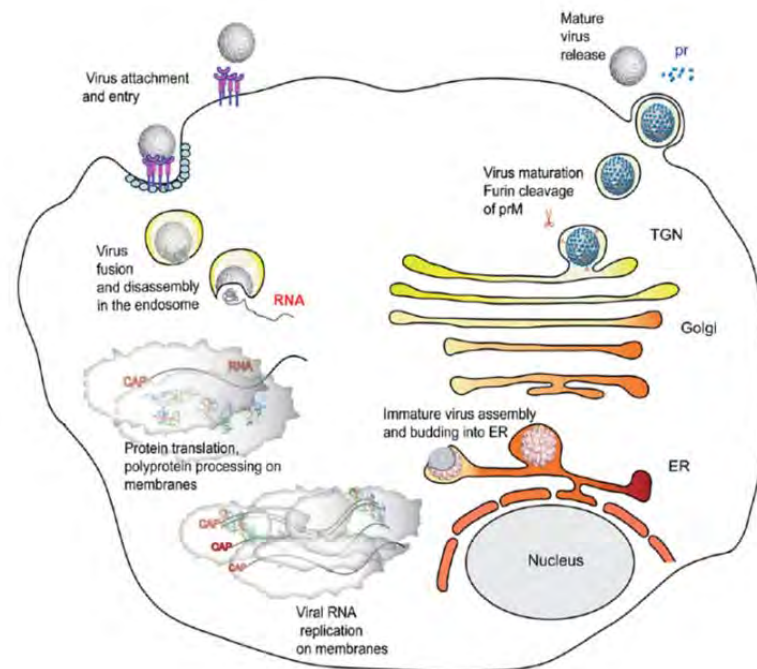


Figure 1.5 The replication cycle of Dengue Virus.[14]

The replication cycle of the DV is displayed in Figure 1.5. First, the E protein mediates the viruses to bind to the cell-surface attachment molecules and receptor of the host cell. Then, the viruses enter into the cell by receptor-mediated endocytosis. Upon the acidic condition of the endosome, the viral envelope glycoproteins (E) mediate the fusion of the host and virus lipid membranes, allowing disassembly of the viruses and release of the viral genomic RNA into the cytoplasm of the host cell.[14] The genomic viral RNA, which functions as mRNA, is then translated directly into the single polyprotein precursor. After that, both the virus-encoded serine protease, NS2B/NS3 protease, and the host-encoded proteases, signalase and furin, cleave co- and post-translationally this single polyprotein precursor into three structural and seven non-structural proteins which are in the order of C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5. The replication of the viral genomic RNA, taken place in the rough endoplasmic reticulum (ER) and in Golgi-derived membranes called vesicle packets (VP)[14], occurs after the viral polymerase NS5 is synthesized and released from the polyprotein precursor. During this process, a dsRNA transient intermediate is formed and separated into its individual strands for accessing to the NS5

polymerase for the next round of the viral replication. [15] Then, the newly synthesized viral RNA exits from the intermembrane space of the double-membrane VPs into the cytoplasm.[16] Next, the assembly of virus develops in the lumen of the rough ER from which an immature virus particle buds into the Golgi. [17] The cleavage of prM to M by furin, which occurs at the trans-Golgi network, together with the conformational rearrangements of E result in the maturation of the virus particle. [18-20] Finally, the mature virus particles exit from the host cell by exocytosis.

1.1.5 Enzyme Target (NS2B/NS3 Protease)

From the replication cycle of DV, both structural proteins (Capsid, Membrane and Envelope protein) and non-structural proteins (NS2B/NS3 protease, NS3 NTPase/Helicase, NS5 Methyltransferase and NS5 RNA-dependent RNA polymerase (RdRp) can be targeted for inhibiting this cycle. Thus, drugs development against DV infection can be developed by targeting these protein targets. However, in this work the focus was set on the NS3 protease as the enzyme target for drug development for DV infection because, at the beginning, only X-ray structures of NS3 proteases of DV were available in the protein data bank. In addition, there are many known inhibitors against NS3 protease of DV, while such information for the other protein targets was insufficient for using as a starting point. Therefore, all available information about DV NS2B/NS3 protease could be a good starting point for searching or developing drugs against NS3 protease of DV.

A full-length of NS3 containing 618 amino acid residues is a multifunctional protein. The N-terminus, one-third of NS3 (approximately 180 residues), functions as a serine protein with a classical serine protease catalytic triad (His51, Asp75 and Ser135), while the C-terminus, two-third of NS3, represents an RNA helicase and RTP/NTPase.[21] The polyprotein precursor, containing three structural proteins (C, M and E) and seven non-structural proteins (NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5), requires the host protease, such as signalase and furin [22], and the viral protease (NS3 protease) for co- and post-translational processing this polyprotein into individual functional protein.[23-27] Therefore, NS3 protease plays a vital role in the replication cycle of DV. However, the enzymatic activity of NS3 protease is enhanced by interacting with the hydrophilic central domain (around 40 amino acid spanning

from residue 49-95 as shown in Figure 1.6) of NS2B, which acts as an essential cofactor.[28] The NS2B/NS3 protease is responsible to cleave the nonstructural protein at the region NS2A/NS2B, NS2B/NS3, NS3/NS4A and NS4B/NS5 and also the internal site within the capsid protein C, NS3 and NS4A [23, 26, 29-32] as shown in Figure 1.6. Mutations in the NS3 cleavage sites of the polyprotein precursor abolish viral infectivity, suggesting that the inhibition of NS3 protease by small molecules may represent an effective antiviral therapy.[33-35]

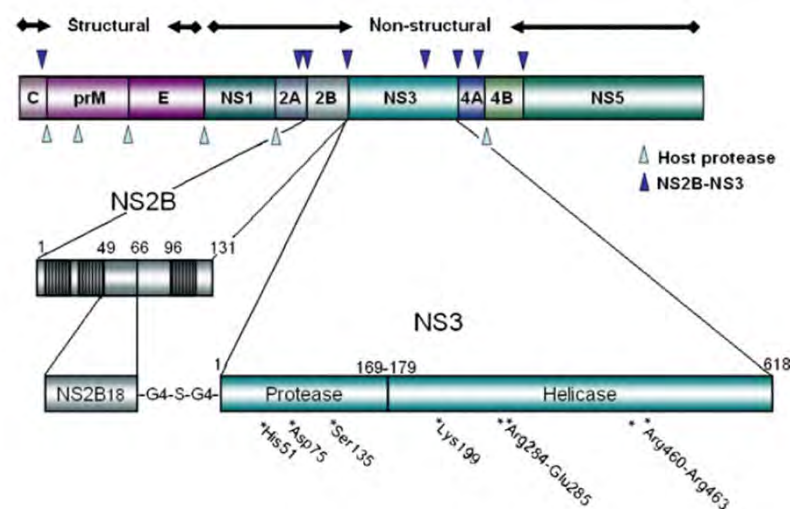


Figure 1.6 Diagram represents the organization of DV polyprotein and NS2B/NS3 protease. An essential cofactor domain of NS2B (residues 49-95) for activating activity of NS3 protease is displayed in the white box. The NS3 protease domain is from residue 1 to 179, whereas NS3 Helicase domain locates at residue 180 to 618 of NS3. Amino acids diagram of the X-ray structure (PDB code 2VBC)[36], a full length NS3 complexed with 18 amino acid residue of NS2B (residue 49-66), linked with G4-S-G4, is also displayed.[36]

NS2B, which is an integral membrane protein[25], comprises of 3 hydrophobic domains flanking a conserved hydrophilic region. A membrane association of the polyprotein precursor rendering the protease sensitive sites in the optimal context for *cis*- and *trans*-cleavages needs these hydrophobic regions of NS2B.[14] However, these hydrophobic regions are not necessary for the *cis*-cleavage of the NS2B-NS3 site *in vivo* as indicated by the abrogation of the membrane

requirement for the NS2B-NS3 cleavage by deletion of these three hydrophobic domains.[25] Whereas, the linking of the hydrophilic central domain of the NS2B viral cofactor with the NS3 protease (amino acid residue 1-169) of DV via a flexible linker (Gly₄-Ser-Gly₄) results in the expression of soluble and active protease.[37]

Recently, two co-crystallized structures of DV NS2B/NS3 protease were resolved. The first X-ray structure (PDB code 2FOM) [38] is the complex between the NS2B (residue 43-96) with the NS3 protease domain. This structure contains missing residues at the residue 77-84 of the NS2B chain. Experimental data [38], NMR spectroscopy, revealed that the C-terminus of NS2B directly interacts with the substrate-binding site of NS3 protease. Thus, this structure is catalytically inactive because the C-terminus of the NS2B of this structure does not interact with the substrate-binding site. The other structure (PDB code 2VBC) [36] is the complex between 18 amino acid residues (residue 49-66) with the full-length of NS3 of DV. This structure shows that the β -strand N-terminus of NS2B (residue 49-66) inserts into the N-terminal β -barrel of the NS3 protease concealing the hydrophobic domain of protease from solvent resulting in the stabilization of this domain. This structure supports the previous work [38] which was indicated that the truncated NS2B part is sufficient to stabilize the NS3 protease. However, the NS3 protease domain of this structure is also inactive because this structure lacks the C-terminal part of the NS2B which is important for interacting with the substrate-binding site as indicated by the NMR spectroscopy [38]. Analysis of these two X-ray structures together with the experimental data, it indicates that the N-terminal part of NS2B performs a chaperone-like role for stabilizing NS3 protease, whereas the C-terminal part of NS2B adopts a totally different conformation between the free form and in complex with the inhibitors or substrates. The NS3 protease domain is wrapped around by the cofactor NS2B as a 'belt-like' structure and the C-terminal part of NS2B is integral as a part of the active site of protease.[13]

The NS2B/NS3 protease is an attractive enzyme target for drug development for DV infection. HIV-1 protease inhibitors which are currently used in the clinical trial as a component of highly active antiretroviral therapy (HAART) are a good example of successful drug development targeting the enzyme protease class. However, the DV protease inhibitors have to be tested to ensure they are not toxic.

That is because the cellular serine protease such as furin also recognizes the dibasic amino acid such as Arg or Lys at the P1 and P2 position as the DV protease.[14] Nevertheless, inhibitors of NS2B/NS3 protease of WNV, a virus which is in the same class as DV and shares high percent sequence identity to DV NS2B/NS3 protease, have been recently identified [39]. These compounds exhibit low IC₅₀ and low EC₅₀ value against this enzyme.[39] Therefore, this result implies that the development of DV protease inhibitors with high biological activity and low toxicity is feasible.

1.1.6 Literature Reviews

1.1.6.1 Protein Structures (Homology Models and X-ray Structure)

There are several X-ray structures of NS3 protease deposited in the protein data bank. Two X-ray structures, DV NS3 protease (PDB code 1BEF) and DV NS3 protease complex with Mung-Bean Bowman-Birk inhibitor (MbBBI) (PDB code 1DF9) [40, 41], were first published in 1999 by Murthy *et al.*. The essential cofactor NS2B is missing in these crystal structures; therefore, the analysis of the interaction between NS2B and NS3 as well as the interaction between the NS2B/NS3 protease and potent inhibitor and the complete structure of NS2B/NS3 protease cannot be deduced and is unclear so far.

A further crystal structure of the DV NS3 protease complex with its cofactor NS2B in the apo-form (PDB code 2FOM) was recently reported.[38] At the same time, the structure of the related WNV NS2B/NS3 protease complex with the peptidic inhibitor, benzoyl-norleucine(P4)-lysine(P3)-arginine(P2)-arginine(P1)-aldehyde (Bz-Nle-Arg-Arg-Arg-H), (PDB code 2FP7) was successfully cocrystallized by the same working group.[38] The structure (2FP7) of this tetrapeptidic inhibitor projected on the binding pocket and its interaction with surrounded residues are displayed in Figure 1.7. Interestingly, these X-ray structures reveals that the conformation of the cofactor NS2B in the inhibitor-complexed WNV protease crystal structure (PDB code 2FP7) differs from the conformation of the inhibitor-free structure of the DV protease (RMSD=11.59 Å), whereas the NS3 domain shows quite similar conformation in both protease complexes (RMSD 2.36 Å as summarized in Table 1.1). It was recently reported that the C-terminal domain of DV NS2B shows a direct interaction with the substrate-binding site of the NS3 protease, which is not observed in the DV crystal

structure in the apo-form (PDB code 2FOM).[42] Thus, the available crystal structure of apo DV NS2B/NS3 protease (PDB code 2FOM) does not represent a suitable target structure for analyzing the binding mode of potential active site inhibitors.

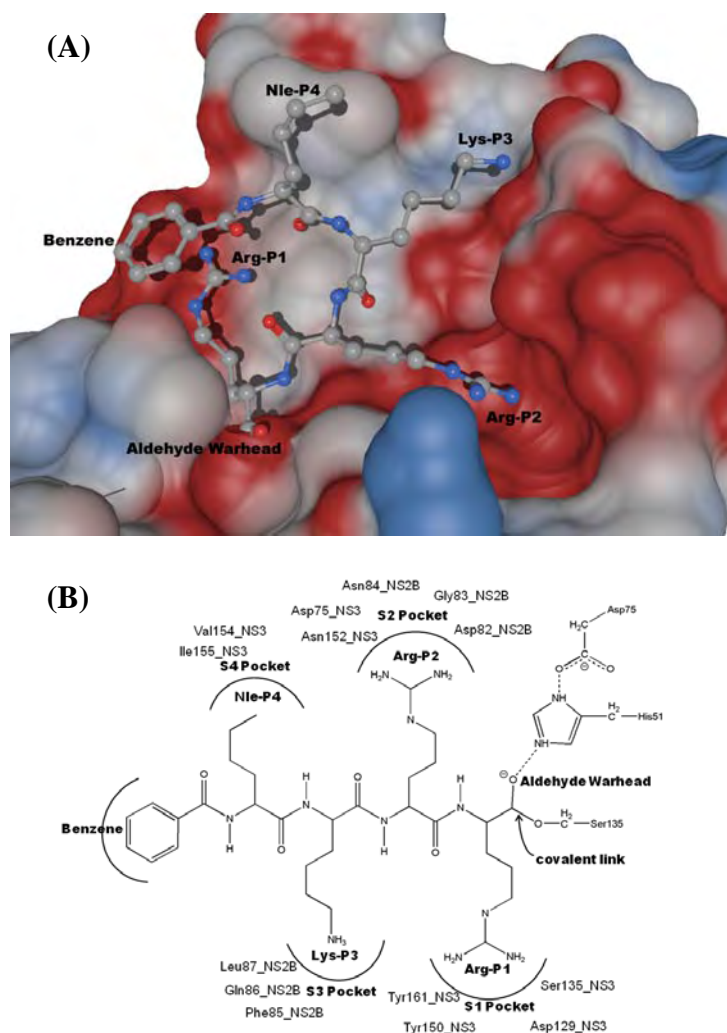


Figure 1.7 (A) Interaction of the tetrapeptidic inhibitor with the WNV NS2B/NS3 protease (taken from the crystal structure 2FP7). (B) Schematic representation of the interaction of the inhibitor with the substrate pockets of WNV NS2B/NS3.

Recently, two novel X-ray structures of WNV NS2B/NS3 protease complex with the inhibitor aprotinin (PDB code 2IJO) and in the apo-form (PDB code 2GGV) were resolved.[42] As shown in Table 1.1 that the structure of the NS2B of 2GGV is different from 2IJO or 2FP7 (RMSD value between the NS2B of 2GGV and 2IJO is 11.71 Å and between 2GGV with 2FP7 is 11.39 Å). On the other hand, the NS2B chain of 2IJO is similar to 2FP7 (RMSD value between these two structure is 0.69 Å).

The reported data supports the earlier finding that the binding of an active site inhibitor results in a rearrangement of the C-terminal domain of NS2B compared to the inhibitor-free form. From these observations [42], the conclusion was drawn that the NS2B/NS3 protease of flavivirus adopts a “productive conformation” in inhibitor-complexed form and a “non-productive conformation” in the apo-form. The new structural data also showed that the WNV protease adopts the same conformation in complex with different kinds of inhibitors (PDB code 2IJO and PDB code 2FP7; RMSD 0.87 Å shown in Table 1.1). Moreover, the cofactor NS2B also shows a similar conformation in both WNV crystal structures. Table 1.1 also shows that the apo-form of DV NS2B/NS3 protease (2FOM) is close to the apo-form of WNV (2GVV), RMSD between these two structures for the NS2B part = 2.60 Å, NS3 protease domain = 2.20 Å and NS2B/NS3 protease = 2.35 Å. Therefore, based on these data and the available experimental data, it is suggested that the active inhibitor-complexed form of DV NS2B/NS3 protease should have the conformation which differ from the apo-form, especially the C-terminus of the NS2B as it is found in the WNV protease.

Besides the experimental structures of DV NS3 protease, there is also the homology model of DV NS2B/NS3 protease. The only one reported homology model for DV NS2B/NS3 protease [43] was generated using the hepatitis C virus (HCV) NS3 protease with cofactor NS4A (PDB code 1JXP) as a template. However, HCV NS3/NS4A protease shows only low sequence identity (14.8%) to DV NS2B/NS3 protease which results in a large deviation of the NS3 protease structure. This could be shown by the later solved NS3 crystal structure of DV (PDB code 2FOM). The C α -atom RMSD between the NS3 protease domain of HCV protease (PDB code 1JXP) and DV (PDB code 2FOM) is 6.57 Å. In addition, the cofactor NS2B of that model was built only for a short 12 amino acid segment (residues 70–81). It was proposed that the 12 amino acids of the modeled NS2B domain rather than the 40 amino acids could be sufficient for activating NS3 protease.[43] However, the recently determined DV NS2B structure (PDB code 2FOM)[38] shows that the residues 51–57 (and not residues 70–81 as proposed by Brinkworth *et al.*) of NS2B form a β -strand which associates with the NS3 protease. In addition, NS3 protease complexed with a truncated NS2B (residues 49–66) resulted in a catalytically inactive enzyme, indicating that the short NS2B is not sufficient for activating NS3 protease.[38] The

NS3 protease requires at least 40 amino acids of NS2B, expanding from residues 49 to 95 [28], for enhancing its activity. Thus, this homology model is not suitable to represent a complexed structure of DV NS2B/NS3 protease.

Table 1.1. RMSD values (Å) of NS2B, NS3, and NS2B/NS3 protease, indicated as the first, the second, and the third values, respectively, between each pair of different X-ray structures.

	Structure	X-ray structure and DV models						
		1BEF	1DF9	2FOM	2FP7	2GGV	2IJO	2VBC
1BEF	DV NS3-protease	- - -	- 0.23 -	- 9.52 -	- 9.15 -	- 9.81 -	- 9.52 -	- 9.20 -
1DF9	DV NS3-protease complexed with Mung-Bean Bowman Birk inhibitor	- 0.23 -	- - -	- 9.51 -	- 9.15 -	- 9.41 -	- 9.51 -	- 9.24 -
2FOM	DV NS2B/NS3 protease	- 9.52 -	- 9.51 -	- - -	11.59 2.36 10.90	2.60 2.20 2.35	10.97 2.41 10.31	- 1.37 -
2FP7	WNV NS2B/NS3 protease in complex with tetrapeptidic inhibitor	- 9.15 -	- 9.15 -	11.59 2.36 10.90	- - -	11.39 1.17 12.07	0.69 0.87 0.72	- 1.79 -
2GGV	WNV NS2B/NS3 protease	- 9.81 -	- 9.41 -	2.60 2.20 2.35	11.39 1.17 12.07	- - -	11.71 0.85 12.61	- 1.30 -
2IJO	WNV NS2B/NS3 protease complexed with Bovine Pancreatic Trypsin Inhibitor	- 9.52 -	- 9.51 -	10.91 2.41 10.31	0.69 0.87 0.72	11.71 0.85 12.61	- - -	- 1.51 -
2VBC	Full length NS3 (NS3 protease-Helicase) of DEN	- 9.20 -	- 9.24 -	- 1.37 -	- 1.79 -	- 1.30 -	- 1.51 -	- - -

*1BEF and 1DF9 have no NS2B part and 2VBC complexed with only 18 residues of NS2B. Therefore, these structures could not be used to superimpose with the NS2B and NS2B/NS3 protease structures with the other structures.

1.1.6.2 Inhibitors of DV NS2B/NS3 protease

Most of the standard serine protease inhibitors are less active against DV NS2B/NS3 protease. Thus, these inhibitors cannot be used to inhibit DV NS2B/NS3 protease. However, aprotinin, the classical serine protease inhibitor, shows activity against this enzyme at the nanomolar concentration ($IC_{50} = 65 \text{ nM}$) that could be due to this inhibitor is a large peptide. Therefore, it could block the substrate to access the binding site by enveloping the enzyme.[37] However, from the drug development point of view, this compound is not a good oral drug candidate because this compound doesn't agree with the drug-like properties. From the desirable effort for investigating small-molecule inhibitors against DV NS2B/NS3 protease [37, 39, 44-50], many compounds including peptidic inhibitors, small peptides, small molecules, and natural products are identified as inhibitors against DV NS2B/NS3 protease and most of these compounds show moderate activity against this enzyme.

Starting with peptidic inhibitors, most of them were designed based on the substrate specification at the P1 and P2 subsite (Arg or Lys) for DV NS2B/NS3 protease. The first substrate-based peptidic inhibitor identified as DV NS2B/NS3 protease inhibitor is Ac-FAAGRR-CHO, $K_i = 16 \mu\text{M}$.[37] Since then many peptides have been developed as DV NS2B/NS3 protease inhibitors. [44, 49, 50] Even though the tetrapeptidic inhibitor with aldehyde warhead (Bz-Nle-Arg-Arg-Arg-H) does not show the most effective activity against DV NS2B/NS3 protease ($K_i = 5.8 \mu\text{M}$), this compound was used as the template to investigate the substitution at each subsite of peptidic inhibitors. That is because the compounds with aldehyde warhead are readily amenable to high-throughput synthesis[50]. Results revealed that the P2 subsite is more important than P1 subsite for interacting with the enzyme. In addition, the replacement at the P1 subsite by neutral groups such as Phe, Phg, (p-Me)Phe or Trp resulted in moderate activity implying the possibility for replacing the cation subsite at P1 by the neutral group. Moreover, the short peptide (dipeptide (Bz-Arg-Arg-H, $K_i = 12.0 \mu\text{M}$) or tri-peptide (Bz-Lys-Arg-Arg-H, $K_i = 1.5 \mu\text{M}$)) are active as well as a tetrapeptide inhibitor (Bz-Nle-Arg-Arg-Arg-H).[49] These results suggest the possibility of development of small-molecule inhibitors against DV NS2B/NS3 protease, and this information is important to assist the designing potent inhibitors for

DV. In addition, one example of patented dipeptide used as viral serine protease inhibitor is shown in Figure 1.8. [47]

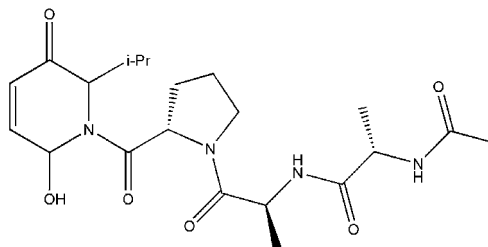


Figure 1.8 An example of patented DV NS3 protease inhibitor

Many small molecules representing as DV NS2B/NS3 protease inhibitors have been developed. The first reported small-molecule inhibitor for this enzyme was a series of compounds containing guanidine groups mimicking Arg, a specific substrate at P1 or P2 subsite for DV NS2B/NS3 protease. Many compounds were tested for their activities against DV NS2B/NS3 protease; however, only three compounds shown in Figure 1.9 are active at the concentration of micromolar whereas the other compounds are inactive or less active. [45]

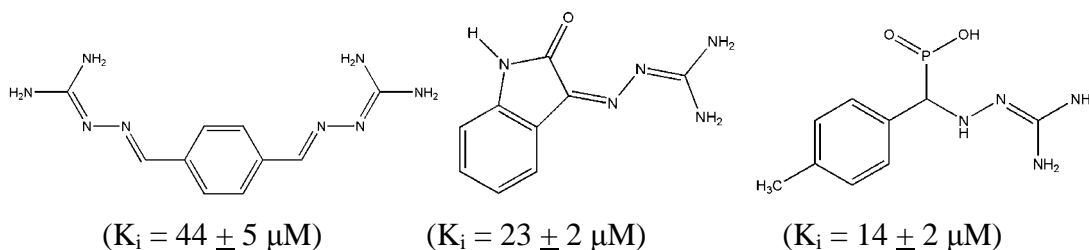


Figure 1.9 Example of some small-molecule inhibitors of DV NS2B/NS3 protease.

Moreover, some natural compounds (flavanones and cyclones) were extracted from fingerroot, *Boesenbergia rotunda* (L.) Mansf. Kulurpfl. (BR) which is a common spice belonging to the ginger family (Zingiberaceae). The cyclohexenyl chalcone derivatives (4-hydroxypanduratin A and panduratin A) show a good inhibitory activity against DV NS2B/NS3 protease (K_i values are $21 \pm 6 \mu\text{M}$ and $25 \pm 8 \mu\text{M}$, respectively). In addition, the kinetic study of these compounds revealed that they bind to DV NS2B/NS3 protease in a competitive mechanism. On the opposite site, the pinostrobin and cardamomin are the non-competitive inhibitors and their K_i

value are quite high (approximate 350 μM). The chemical structures of these compounds are shown in Figure 1.10 [46].

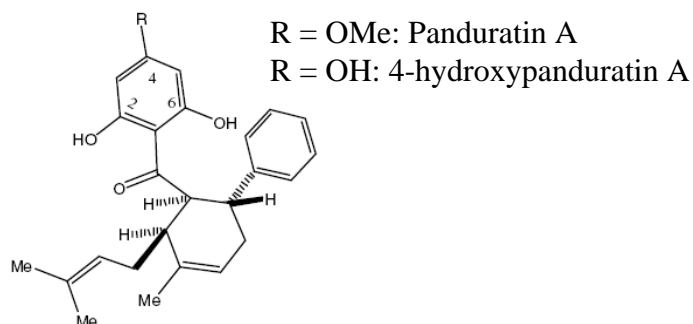
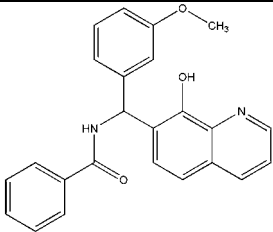
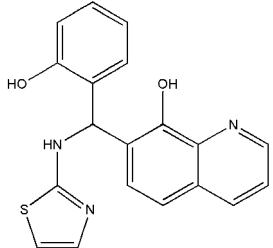
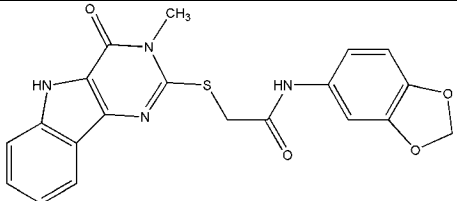


Figure 1.10 Natural compounds as DV NS2B/NS3 protease inhibitor

Furthermore, by applying a high throughput screening (HTS) or a virtual screening (VS) method, many compounds were recently identified as potent DV or WNV NS2B/NS3 protease inhibitors [39]. For example, HTS was performed to screen for novel inhibitors of WNV NS2B/NS3 protease using a database containing approximately 32,000 compounds. In addition, some criteria were applied to reduce the number of hit compounds. Finally, some compounds were identified as WNV NS2B/NS3 protease inhibitor. The compound 1-3 were also tested for their activities against DV NS2B/NS3 protease. Interestingly, compound 3 shows a good inhibitory activity against this enzyme. Table 1.2 shows the structures and their K_i value for WNV and DV NS2B/NS3 protease derived from HTS.[39]

Table 1.2 Chemical Structures and their K_i values of WNV and DV NS2B/NS3 protease inhibitors

Comp. ID	Structure	K_i Value (μM)	
		WNV NS2B/NS3 protease	DV NS2B/NS3 protease
1		3.2 ± 0.3	28.6 ± 5.1
2		3.4 ± 0.6	30.2 ± 8.6
3		37.3 ± 6.4	17.0 ± 4.3

Recently, novel DV NS2B/NS3 protease inhibitors were identified by combining VS method and biological test.[48] A subset of an in-house database containing 2.5 million compounds was first filtered out by using cellular uptake and cell membrane impermeability to ions properties as criteria. Thus, only neutral non-zwitterionic compounds were passed these criteria. Then, only compounds which could be purchased from highly reputable chemical vendors were selected. Consequently, the rest of the compounds were computationally screened using a EUDOC docking program. Then, 20 compounds which gave the lowest energy score were selected as potent inhibitors and thus were purchased to test for their biological activities. Results revealed that only two compounds (ARDP0006 and ARDP009 whose chemical structure are displayed in Figure 1.11) showed antiviral activity in cell culture. In addition, the toxicity was tested and compound ARDP0006 showed no toxicity at the concentration tested.

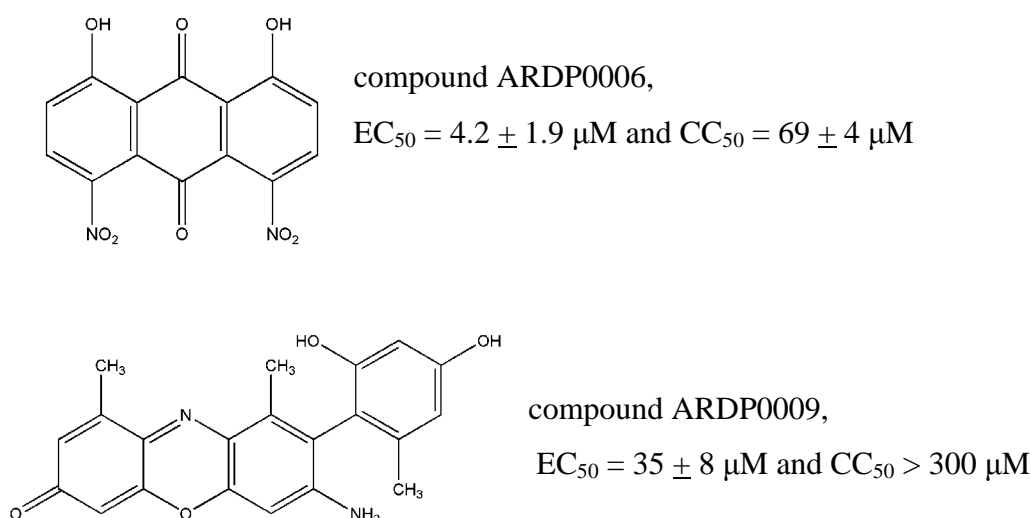


Figure 1.11 Inhibitors of DV NS2B/NS3 protease and their EC_{50} and CC_{50} values

1.1.6.3 Molecular Modeling and Molecular Dynamics Simulation

There is only one published work[51] applying the molecular dynamics (MD) simulation method to elucidate the dynamic motion of DV NS2B/NS3 protease and the interaction between NS2B and NS3 protease. In their work, the apo form (NS3 protease; PDB code 1DF9) and the DV NS3 protease complex with its essential cofactor (NS2B) (PDB code 2FOM) were used to perform MD simulations. Moreover, in their work, molecular docking between substrates and DV NS2B/NS3 protease using these X-ray structures was also performed. In addition, several works using molecular docking to investigate the interaction between inhibitors with DV NS2B/NS3 protease were reported. [44, 52] However, most of these works used the X-ray structures, either 1DEF/1BEF or 2FOM, as a representative conformation of DV protease. As discussed before, the structure of DV NS3 protease, which is not complexed with its essential cofactor (NS2B), deviated substantially from the DV NS2B/NS3 protease. Moreover, the C-terminus of NS2B in the apo form of DV NS3 protease (2FOM) does not interact with the substrate-binding site of the DV NS3 protease, whereas the experimental data[38] showed clearly that this part of NS2B binds directly with the substrate-binding site of the DV NS3 protease. Hence, by applying these X-ray structures to perform MD simulation or molecular docking, it

could lead to misinterpretation of the interaction between NS2B and NS3 protease and the impact of NS2B for enhancing the activity of the protease and also the interaction between inhibitors and NS2B/NS3 protease.

1.2 Wee1 Kinase

1.2.1 Cancer and Cell Cycle

Cancer is a multifaceted disease, but shares a common feature in which a cellular proliferation controlled by the cell division cycle displays uncontrolled growth. This unrestrained growth of cancer cells interferes with the body at the site of their location. Moreover, these cancer cells can also spread to the other organs in the body. The effects can result in severe pain or even death.[53] Therefore, understanding the cell cycle and cellular proliferation could lead us to understand cancer and could find the protein target for drug development for cancer therapy.

The cell cycle process of a typical eukaryotic cell undergoes five sequential phases, which are G₁, S, G₂, M and G₀ as shown in Figure 1.12. The most important phases are the S (synthesis) phase where the DNA replicates itself and the M (mitosis) phase in which chromosomes segregate to generate two genetically identical daughter cells. There are also two gap phases in between these two major cell cycle events, which are the G₁ separates mitosis from the S phase, and the G₂ in between the S phase and the mitosis. In addition, cells can stop cycling after division and then enter into the G₀ phase which is known as a quiescent state.

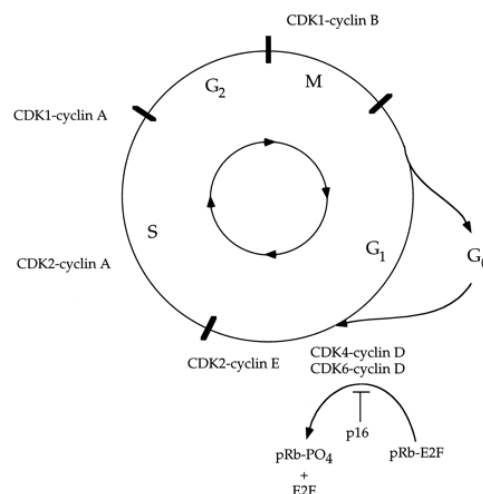


Figure 1.12 A schematic representation shows cell cycle and CDK-cyclin complexes which are important in each checkpoint.[54]

There are two types of mechanisms which are important for controlling the cell cycle. The first type is a cascade of protein phosphorylations involved with the highly regulated cyclin dependent kinases (CDKs) family of serine/threonine kinase. As the name of CDKs, the activity of CDK totally depends on its association with a regulatory subunit which is known as cyclin. From one stage to the next stage of the cell cycle, it is controlled by several CDK-cyclin complexes through the phosphorylation or dephosphorylation mechanism. Each of these CDK-cyclin complexes functions at the specific stages as displayed in Figure 1.13. The cell cycle process and a transition from one phase to the next are additionally regulated by the second type of mechanism which is known as checkpoint control.[53, 54] Checkpoints are defined as a sensor that monitors the cell cycle progression at each step to ascertain whether all conditions are already fulfilled before it goes further into the next step. There are three major cell cycle checkpoints as displayed in Figure 1.13. The first checkpoint in the cell cycle is at the G1/S phase transition which is responsible for DNA damage. Then, when a cell passes this phase, it will be checked again at the G2/M checkpoint monitoring the fidelity of DNA replication and also the sensor for DNA damage as with the G1/S checkpoint. Finally, the spindle checkpoint occurs at the M phase in order to ensure that a functional mitotic spindle is formed correctly.[53]

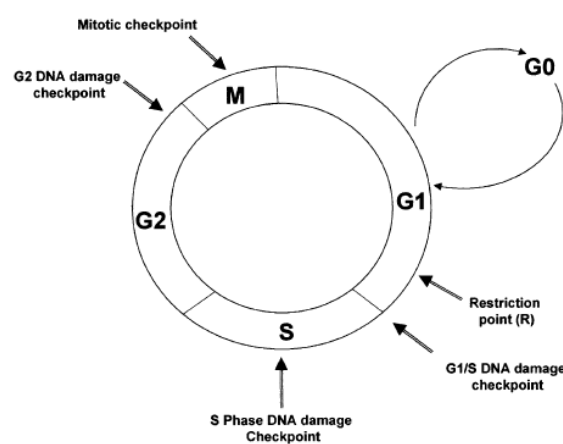


Figure 1.13 Checkpoint at each stage in the cell cycle[53]

These checkpoints not only function by assisting DNA damage to repair itself, but they can also promote cell death in the case of unrepaired cells. By targeting at the

checkpoints after the exposure of cancer cells, the DNA damage cannot be fixed and it results in the death of cancer cell. This strategy is now becoming more interesting in the drug development for cancer therapeutics.[55] Many cancer cells lack a functional p53 signalling pathway, which means that the significant damage-sensitive checkpoint at the G1/S is not controlled. As a result, a cancer cell can pass the G1 phase and process to the next step until it meets the next barrier for the cell cycle, which is the G2 checkpoint. G2 checkpoint abrogation inhibits cancer cells to repair DNA damage and force them to go into the mitosis phase which results in cell death.[55] Therefore, the G2 checkpoint is an interesting therapeutic target for anticancer drug development.

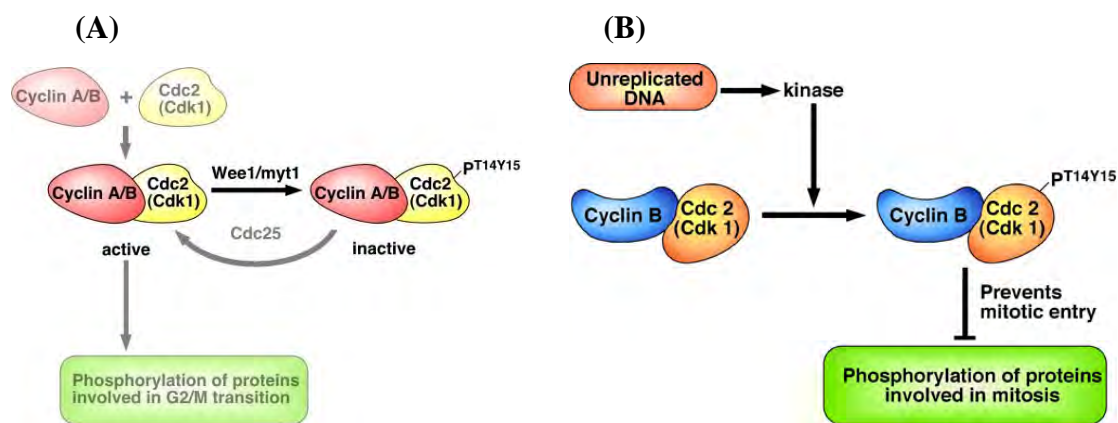


Figure 1.14 A representation of cell cycle at the G2/M checkpoint; (A) CDK1(cdc2)-cyclin B complex is phosphorylated by Wee1 and Myt1 resulting in an inactivated state of this complex, giving time for repairing DNA damage. This process, a phosphorylation of CDK1(cdc2)-cyclin B complex, is important for preventing the DNA damage cell to go into the mitosis phase as displayed in (B) [56]

The process of phosphorylation at the G2/M checkpoint is displayed in Figure 1.14 (A) and (B). Before cells go into the mitosis phase, DNA damage and unreplicated DNA have to be checked at the G2/M checkpoint. In addition, at this phase the genome is also checked to see whether it is replicated once and only once per cell cycle. If cells pass these checkpoints, mitosis of cell occurs as usual; otherwise, the mitosis phase is prevented by specific G2/M checkpoint events. Cyclin-dependent kinase CDK1 (Cdc2) in complex with type B cyclin (cyclin B) play a

significant role for controlling such events. The activity of the CDK1/cyclin B complex is regulated by activating and inhibitory phosphorylation. The activity of the CDK1/cyclin B complex is activated by phosphorylation at Thr161 of CDK1, while phosphorylation at Thr14 and Tyr15 of CDK1 results in inactivation of the CDK1/cyclin B complex. The inactive state of CDK1/cyclin B at the G2 phase is one of the major mechanisms in the cell cycle because it creates a pause, giving time to repair DNA damage before cell undergoes mitosis in the next step.

Wee1, a nuclear kinase, is functionally a serine/threonine kinase. Wee1 kinase is a key regulator for controlling the cell cycle progression. The enzymes in the Wee1 kinase family (Wee1 and Myt1) function for inhibitory inactivation of CDK1 at the G2/M checkpoint. The Wee1 phosphorylates at Tyr15, whilst the dual-specificity Myt1, a membrane bound, function for phosphorylation at both sites (Thr14 and Tyr15). At the late G2 phase, CDK1 is activated by dephosphorylation, which is a function of Cdc25 phosphatase. Then, the cell undergoes mitosis. However, if Wee1 or Myt1 cannot function to phosphorylate at Thr14 and Tyr15 of CDK1, the DNA damage of cancer cells cannot be repaired. Then, these cancer cells are forced to go into mitosis and finally end up with cells death. Therefore, cancer cells cannot replicate themselves. Hence, both of the two enzyme are important for the cell cycle and present as promising targets for drug development for cancer therapy.[57] Developing Wee1 inhibitors as the G2 checkpoint abrogator is a new strategy for designing anticancer drugs enhancing conventional cancer therapy. It was found that PD0166285 (a pyrido[2,3-d]pyrimidine compound) acts as G2 checkpoint abrogator. This compound (Figure 1.15) inhibits the Wee1 in the nanomolar concentration.[58] This finding is a good starting point to develop and search for more potent novel Wee1 kinase inhibitors which could be further developed as drugs for cancer therapy.

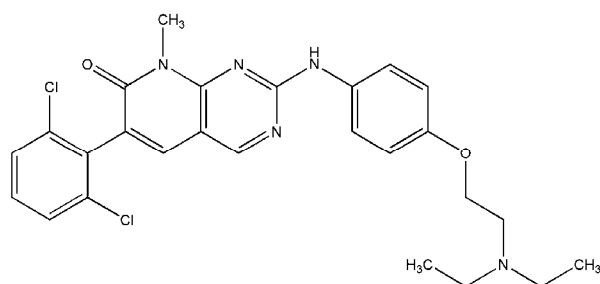


Figure 1.15 Chemical structure of PD0166285

1.2.2 Literature Reviews on Wee1 Kinase

1.2.2.1 Wee1A Kinase Structure

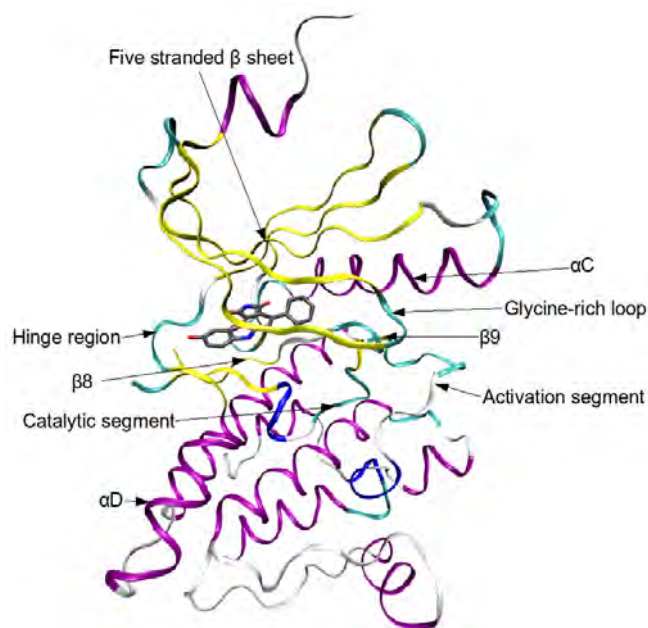


Figure 1.16 Overall structure of the Wee1 kinase (X-Ray Structure of the Wee1/PD0407824 complex, pdb code 1X8B). α -Helices are coloured magenta, β -sheets are coloured yellow and loops are shown in cyan.

Two types of Wee1 (Wee1A and Wee1B), which are different in temporal and spatial expression, are found in humans. [57] The somatic Wee1 kinase is Wee1A and it is functionally a tyrosine kinase; however, in sequence and structure it most closely resembles serine/threonine kinases such as the checkpoint kinase Chk1.[59, 60] The first crystal structure of Wee1 was solved in 2005 and showed that although the ATP binding pocket closely resembles that of other protein kinases, the activation segment contains Wee1-specific features that maintain it in an active conformation.

One example of the X-ray structure of Wee1 kinase complexed with its inhibitor (PDB code 1X8B) is display in Figure 1.16. Wee1A contains 646 amino acid residues comprising three major domains, which are N-terminal domain, a central kinase domain and C-terminal domain. There are two standard lobes kinase fold (N- and C-terminal lobe) at the kinase domain of Wee1A (residue 291-575). [57] The N lobe locating above the active site composing a five-strand anti-parallel β sheet and a typical glycine-rich loop (residues 306-311). Four helix bundles and the catalytic

segment are the components of the C-terminal lobe. A hinge region (residue 377-381) connects a strand $\beta 5$ in the N lobe to helix αD in the C-terminal lobe. The active site, where Wee1 inhibitors bind to the enzyme, locates between these two lobes. A short section of extended polypeptide referred to a catalytic segment and a large loop which is the activation segment position at the active-site region. A catalytic segment (residue 422-433) contains the essential catalytic residue which is Asp426. The activation segment (residue 462-486) is the region between $\beta 8$ and $\beta 9$ at the C-terminal lobe. This activation segment is important for providing a substrate binding platform and controlling the conformation of kinase.[57]

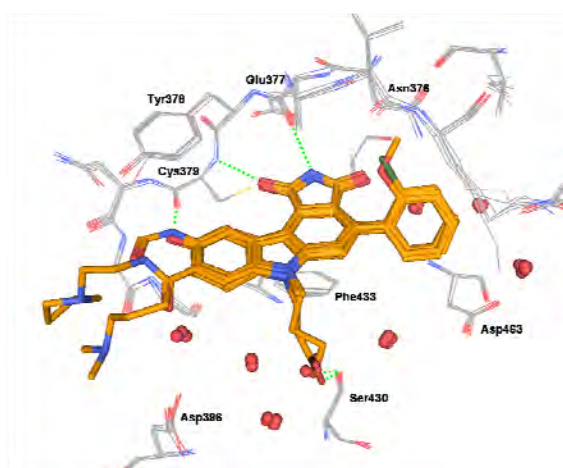


Figure 1.17 Superimposition of the six solved crystal structures of Wee1-inhibitor complexes. The pyrrolocarbazole core of the inhibitors (colored orange) shows the same interaction at the ATP-binding site. Clusters of cocrystallized waters are displayed as balls. Hydrogen bonds are shown as dashed green line.

Wee1A kinase can be cocrystallized only when it binds with its high-affinity inhibitors, implying that inhibitors may assist in stabilizing and/or changing the conformation of the Wee1A into the active state.[57] The complexes provide important information about the interaction with the residues of the binding site. The X-ray structures of Wee1 (Figures 1.17) in complex with six pyrrolo[3,4-*c*]carbazole-1,3(2*H*,6*H*)-dione derivatives were analyzed, and they show the same conformation in all six crystal structures. All cocrystallized inhibitors form several hydrogen bonds between its maleimido ring and the backbone atoms of Glu377 and Cys379 and the

sidechain amide of Asn376. These interactions are well known from other kinase inhibitors, like staurosporine. Two cocrystallized water molecules mediate a hydrogen bond to His350. The phenyl ring of the inhibitors occupies a hydrophobic pocket between the side chains of Val313, Lys328 and Ile374. Stabilization of the planar aromatic system is provided by Ile305, Val313, Ala326 and Phe433.[57]

1.2.2.2 Wee1 Kinase Inhibitors and Molecular Modeling

In the last few years, a variety of Wee1 kinase inhibitors, including phenylpyrrolocarbazoles and phenylpyridopyrimidine whose core structures are shown in Figure 1.18 (A) and (B), respectively, has been described.[61-64] Moreover, the X-ray structures of Wee1 kinase complexed with pyrroloindole as a new class of Wee1 kinase inhibitor (PDB code 3CQE and 3CR0) are recently available in the protein data bank. The compounds in the class of 2-anilino-6-phenylpyrido[2,3-d]pyrimidin-7(8*H*)-ones can be used as the inhibitors for several tyrosine kinase enzymes. These include the receptor kinase, such as EGFr (erbB1), PDGFr, FGFr, and non-receptor kinase (c-Src) [61]. Many of the phenylpyridopyrimidine derivative compounds show high activity against Wee1 kinase ($IC_{50} < 100$ nM). However, the selectivity of the series of these compounds for Wee1 over c-Src is quite low [62]. The inhibition of the c-Src results in many other cellular effect [61]. It was found that the phenylpyrrolocarbazole compounds are equally active as in the phenylpyridopyrimidine derivatives and the selectivity of these compounds for Wee1 over c-Src is quite high. For example, 4-phenylpyrrolocarbazole has $IC_{50} = 97$ nM against the Wee1 kinase but shows inhibitory activity less than 50 μ M against c-Src [62]. Even though the phenylpyrrolocarbazole derivatives are quite high activity against Wee1 kinase and high selective for Wee1 kinase over c-Src, these compounds have some drawback because they lacks sufficient solubility for in vivo development[63]. Therefore, novel inhibitors, which are more potent, selective for Wee1 kinase and have sufficient solubility for in vivo testing, have to be developed.

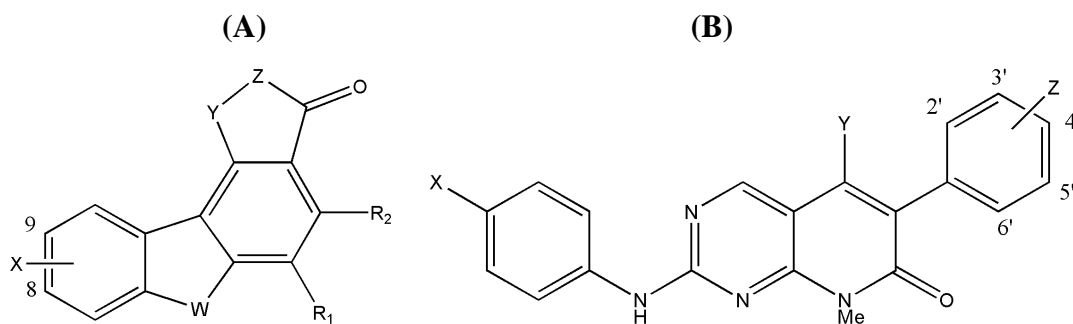


Figure 1.18 (A) Molecular structures of the pyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione derivatives, and (B) Molecular structures of the 2-anilino-6-phenylpyrido[2,3-d]pyrimidin-7(8H)-ones derivatives representing as Wee1 kinase inhibitors

There is only work that reported the 3D-QSAR (CoMFA and CoMSIA) model for the phenylpyrrolocarbazole representing as Wee1 kinase inhibitors [65]. The models were generated from a set of 97 inhibitors of 4-phenylpyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione derivatives. Although, a series of the 4-phenylpyrrolocarbazole compounds shows high inhibitory activity against Wee1 kinase without concomitant C-Src activity, these compounds have some disadvantages about the solubility as discussed before [63]. Recently, a new series of the 9-hydroxy-4-phenylpyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione with a variety of neutral, basic and acidic solubilising groups at the *N*-6 position (w position in Figure 1.18 (A)) and with a variety of solubilising side chains at the C-8 position were synthesized and tested activity against Wee1 kinase. Most of these compounds are potent inhibitor for Wee1 kinase. However, these compounds were not included to construct the previous reported 3D-QSAR model [65]. Thus, a new 3D-QSAR model including these series of Wee1 kinase has to be generated. The derived model can provide the information about the structural requirement for increasing the activity of Wee1 kinase inhibitors and can also guide to synthesis new compounds with no problems about the solubility.

1.3 Scope of this Research Work

In this thesis, several computational approaches were applied to design and find novel lead compounds as inhibitors against DV NS2B/NS3 protease and Wee1 kinase. These compounds could be further developed as drugs for DV infection and

cancer, respectively. The aims of the research work on each enzyme target are as following.

1. DV NS2B/NS3 protease: The main objective of this project is to obtain lead compounds as potent inhibitors for DV NS2B/NS3 protease by applying a virtual screening approach. However, in order to accomplish this, many computational methods had to be employed. Since at the moment no productive conformations of DV NS2B/NS3 protease complex with inhibitors are available in the protein data bank, this complex had to be constructed using a homology model technique. Many models were built using different templates, and these homology models were subsequently validated by performing a molecular dynamics (MD) simulation. In addition, the interaction between NS2B with NS3 protease and also between inhibitor and NS2B/NS3 protease were also investigated. The suitable representative structure of DV NS2B/NS3 protease was determined. The derived structure was next applied for the structure-based design purpose and virtual screening (VS), which was carried out in the hierarchical strategies. The number compounds in the databases were screened by applying several criteria such as, drug-like properties filtering, pharmacophore searching, ranking compounds by using docking score and visual inspection and binding free energy calculation. Finally, some compounds were selected and suggested for biological testing.

Besides obtaining lead compounds for testing their activities against DV NS2B/NS3 protease, the interaction between ligand and enzyme would be very helpful for understanding the interaction in the molecular level. These results are the basic information which can be further applied for drug development against DV infection.

2. Wee1 Kinase: A 3D-QSAR model applying the CoMFA methodology was constructed for a set of pyrrolo[3,4-c]carbazole-1,3(2*H*,6*H*)-dione derivatives representing Wee1 kinase inhibitors. The derived model can be used to describe the relationship between the biological activities of these compounds and their structures. In addition, the obtained model provide information about the structural requirement for the pyrrolocarbazole compounds used as Wee1 kinase inhibitors which can be very useful for designing novel potent inhibitors.

Binding free energy using linear interaction energy with continuum electrostatics (LIECE) method was employed to calculate the binding free energy of Wee1 kinase inhibitors, including pyrrolocarbazole and pyridopyrimidine derivative, of Wee1 Kinase. This method has more advantages than the conventional approach such as linear interaction energy (LIE), MM/PB(GB)SA because it requires less computational demand, fast and can deal with a large amount of compounds in the database with a reasonable real time calculation. Moreover, the calculated binding free energy derived by this method correlated well with the biological data, which can be very helpful to identified lead compounds retrieved from virtual screening. Furthermore, VS was carried out for indentifying lead compounds to test their activities against Wee1 Kinase.

CHAPTER II

THEORY AND COMPUTATIONAL METHODS

Drug discovery and drug development for any disease is a long and expensive process. Therefore, molecular modeling, the method used to mimic the behavior of molecules and molecular systems, is nowadays widely used in the chemical and pharmaceutical industries in order to reduce time and cost. Currently, molecular modeling is regularly accomplished by using computers. The rapid development of computational tools, such as hardware and software, gives impetus to the drug discovery and drug development process.

In this chapter, several molecular modeling techniques, which were applied in this work, are described. Firstly, a 3D-QSAR technique using the CoMFA model approach (used for lead optimization purpose) is briefly explained. Secondly, a short overview of the homology model method, which is applied to construct a protein structure for which its three dimensional (3D) structure is not experimentally available, is given. The derived homology model is helpful in the structure-based drug design strategy. Thirdly, a molecular dynamics simulation providing the basic information about the dynamic motion of protein and the protein-ligand interaction, which assist in the drug development process, is explained. Finally, a virtual screening approach, using several stepwise methods (pharmacophore modeling, pharmacophore searching, molecular docking, docking score ranking and binding free energy calculation), is described.

2.1 Quantitative Structure-Activity Relationship (QSAR)

The concept of the quantitative structure-activity relationship (QSAR) method is to correlate the binding affinities of inhibitors with their physicochemical properties (lipophilicity, polarizability, electronic and steric properties) or their structural properties. The QSAR technique is based on the assumption that compounds which have similar physicochemical/structural properties should also have similar biological binding affinities to the same protein. Thus, the derived QSAR model can also be applied to predict the binding affinities of unknown compounds that have similar structure to the compounds which were used to generate the model. The derived QSAR model is also very helpful as guidance for novel inhibitors synthesis. Hammett[66] was the first person who applied the simple linear regression analysis to

correlate electronic properties of organic acids and bases with their equilibrium constants and reactivity. Based on the Hammett equation, many classical QSAR methodologies were developed. The works of Hansch and Fujita [67], and Free and Wilson [68] are considered to be the beginning of the modern QSAR approach. Hansch and Fujita [67] applied the multiple linear regressions analysis to generate the simple equation to correlate physicochemical properties of compounds with their biological activities. Free and Wilson [68], who worked independently from Hansch and Fujita, also generated the equation to describe the relationship between the biological activity and the additive group. However, neither the 3D structure of inhibitors nor their stereochemistries are considered for generating the classical QSAR model. Actually, the biological affinities of compounds are affected by its interaction with the target protein, and this interaction relies on the 3D structural properties of compounds. Thus, the method called 3D-QSAR, in which the 3D structure of compounds is taken into account, was developed. Several 3D-QSAR approaches have been developed, but the Comparative Molecular Field Analysis (CoMFA) technique is the most popular and often used.

2.1.1 Comparative Molecular Field Analysis (CoMFA)

The 3D-QSAR method named Comparative Molecular Field Analysis (CoMFA) was first developed in 1988 by Cramer et al. [69]. The CoMFA methodology is based on the assumption that a non-covalent interaction between ligand and protein (receptor or enzyme), and the shape of compounds correlate with their biological activities. The noncovalent ligand-protein interaction is normally related with steric and electrostatic interactions which are computed by applying the standard potential energies (van der Waals and coulombic potential function). Therefore, a difference in these interactions in a set of molecules with the protein correlates to a difference in their biological activities. The procedures for constructing the CoMFA model can be categorized into four steps; preparation, interaction energy calculation, statistical analysis and results interpretation.

- Ligand Preparation

A set of molecules used to generate the CoMFA model is first selected. These selected compounds must interact with the same protein and bind in the same

identical manner (the same binding site and the similar binding mode). Then, these compounds are divided into two groups, a training set and a test set. The training set is a set of molecules which is used to generate the model, whereas the test set is applied to validate the reliability and predictive ability of the derived model. Then, the structures of these molecules are constructed. The active conformation of these compounds represents how they bind to the protein. This can be generated by using the related experimental complex structure as a template or by performing molecular docking. Subsequently, these molecules are superimposed by using different molecular alignment techniques. Since the results of CoMFA depend strongly on the alignment of the molecules, this step is very crucial for generating CoMFA model.

- Interaction Energy Calculation

After all the compounds in the training set are superimposed to each other, a box and grid spacing are created to cover the aligned molecules. The default value of the grid spacing (2 Å) and a C-sp3 probe atom with +1 charge are normally used. Prior to calculating the electrostatic and the steric interaction, charges for each molecule must be assigned. Several methods (AM1, PM3, PEOE, Gasteiger etc.) can be used for generating charges for the molecules. Subsequently, the interaction is calculated by placing the probe atom at each grid point, and computing the steric and electrostatic interaction by applying the Lennard-Jones potential function (equation (1)) and coulombic potential function (equation (2)), respectively.

$$E_{vdw} = \sum_{i,j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] \quad (1)$$

A_{ij} = repulsive term coefficient, B_{ij} = attractive term coefficient, and r_{ij} = distance between atom i and probe atom j (Å)

$$E_{ele} = \sum_{i,j} \left[\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] \quad (2)$$

q_i and q_j = atomic charge of atom i and probe atom j , respectively, ϵ = dielectric constant, and r_{ij} = distance between atom i and probe atom j (Å)

- Statistical Values Analysis

To form the basis for a statistical significant model, the method of partial least squares (PLS) regression is used to correlate variations in their biological activities with variations in their interaction fields. The optimum number of PLS components

corresponding to the smallest standard error of prediction (SDEP) is determined by the leave-one-out cross-validation procedure. Using the optimal number of components, the final PLS analysis is carried out without cross-validation to generate a predictive model with a conventional correlation coefficient. Leave-one-out cross-validation is carried out by removing one compound from the training set. The remaining compounds are used to generate the PLS model which is used to predict the activity of the compound which was not included in the model. This process is repeated until each compound was removed once. The small value of the SDEP and the high value of leave-one-out cross-validation correlation coefficient (q^2) indicate the robustness of the predictability of the derived model. In addition, the quality of the model can be determined by other statistical parameters, such as the correlation coefficient (r^2), the standard error of estimation (SDEE) and the F-value. The r^2 value represents how well the model reproduces the input data. A high r^2 value, close to 1, indicates a better fit of the regression. F-value is the ratio of the explained to unexplained variance. A high value of the F-value suggests that the derived model is statistically significant.

Both the r^2 and the q^2 have to be considered together to determine the quality of the derived model since r^2 indicates the reproducibility and q^2 reflects the predictability of the model. However, the leave-one-out cross-validation method might lead to high q^2 values, which do not necessarily reflect a general predictive ability of a model. Therefore, further cross-validation, using five and two groups of approximately the same size in which the objects are assigned randomly, are performed. In this method 80% or 50% of the compounds are randomly selected and a model is generated, which is then used to predict the remaining compounds (leave-20%-out, leave-50%-out). This cross-validation technique, especially the leave-50%-out procedure, has been found to be stricter and to give a better estimate of the robustness of a model than the normal leave-one-out procedure [70-72]. However, it must be stated that there is no qualitative difference between the different cross-validation procedures. All methods are used to examine the internal predictability, and thus they have the problem that they are not able to estimate the external predictability of a QSAR model. Therefore, an external test set is used in addition to test the predictability of the final derived CoMFA model.

- Result Interpretation

The results derived from CoMFA are a regression equation with thousands of coefficients which are used for calculating and predicting biological activities of compounds. Thus, it is difficult and impossible to use a large number of these coefficients for directly interpreting the derived model. In order to facilitate the interpretation and make the model easy to understand, these coefficients are therefore represented by a set of contour maps. These contour maps show regions around the molecule which are favorable and unfavorable areas for a steric group as well as for an electrostatic group. These contour maps are helpful for suggesting how to modify the compound in order to increase the activity.

2.2 Homology Modeling

The three-dimensional (3D) protein structure is very beneficial in structure-based molecular design and in the drug discovery process. In addition, the 3D protein structure can also be applied to study the dynamic motion, mechanisms and other properties such as a ligand-protein interaction. The derived information provides basic knowledge and better understanding on an atomic level which could be useful for drug development. In the last decade, many protein structures were cocrystallized and deposited in the protein data bank (PDB), and the number of experimental protein structures in the PDB is still increasing exponentially. However, there are still lots of proteins that cannot be experimentally determined by X-ray crystallography or NMR. Therefore, the computational technique, called the homology (comparative) model, is applied for building an unknown 3D protein structure based on a related available structure called a template. The overview steps for constructing a homology model of a target protein as follows. First, the template structure(s) which has high amino acid sequence identity to a target protein is selected. Secondly, an amino acid sequence of a template and of a protein target is compared by sequence alignment. Thirdly, the target protein is constructed by taking coordinates of backbone atoms of the template(s). The coordinates of the side chain atoms of the template(s) can be directly used for residue of the target, only if the residue, located at the same position in the target and the template(s), are identical or similar. In the case of residues which are different to the templates, a side chain rotamer library is used to generate their

coordinates. Finally, the obtained homology model is refined by energy minimization in order to relieve the steric strain which is introduced during the model-generating process. Moreover, the derived homology model can also be subjected to molecular dynamics simulation to adjust the side chain and protein conformation. The details of each step are described below.

2.2.1 Template Selection

As the quality of a homology model relies on the template(s), selecting the right template(s) is a crucial step for constructing the model. The search techniques, such as FASTA [73] and BLAST [74], can be applied to search for the proteins which have some degree of sequence similarity to the amino acid sequence of the interested protein, as well as search for proteins for which the 3D structure is available. After obtaining the template candidates, the decision has to be made whether multiple templates or a single template should be used. The accuracy of the derived model is affected by selecting the correct template. This step is the most important step for constructing the homology model. The template can be selected by determining the percent sequence identity value. This value refers to the percentage of the identical amino acids that the template and the target have at the same position in the sequence. The other options to choose the right template(s) include a percent sequence similarity (the percentages of the amino acids at the same position in the two sequences that are similar) and the predicted secondary structure of target and the actual secondary structure of the template.

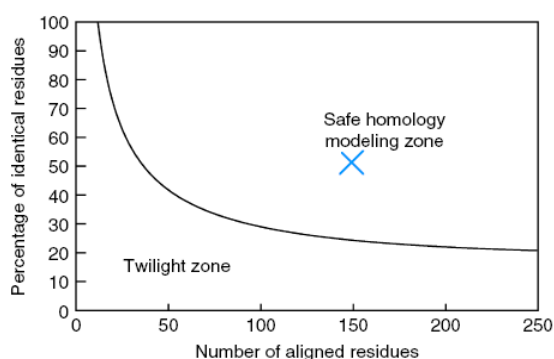


Figure 2.1 Two zones (twilight zone and the safe homology modeling zone) of sequence alignments. The homology model would be fine if the length and the percentage sequence identity compared to the template falls into the “safe” zone.[75]

In general, the target and the template should share at least 30% sequence identity [76]. Nevertheless, the higher the percentage of the sequence identity, the better quality of a homology model is obtained. However, the percent sequence identity selected for constructing a homology model also depends on the length of the amino acid sequence as displayed in Figure 2.1. The shorter the peptide, higher percent sequence identity is required. In addition, to construct the reliable homology model, the sequence identity between two sequences (the target and the template) and the length of the amino acid sequence should fall in the “safe homology modeling zone” as shown in Figure 2.1.

2.2.2 Sequence Alignment

In order to evaluate the similarity between two sequences and to identify which amino acid residues correspond to each other, a method called sequence alignment has to be employed. The Needleman and Wunsch algorithm [77] is widely used to align the amino acid sequences in proteins. An iterative approach is performed in this algorithm starting with all amino acids from both sequences are put in a 2D matrix. Then, the pathway through this array is optimized in such a way to obtain the maximum score, representing the best alignment between these sequences. The strategy for aligning protein sequences using the Needleman-Wunsch algorithm is as follows.

1. A matrix is formed with the two protein sequences; one sequence is placed on the top row and the other one is put on the left hand side column. Then, the score of 1 is given for the identical amino acid and the score of 0 is assigned for mismatched pair as shown in Figure 2.2 (a).
2. The successive summation of cells is completed by starting at the last cell at the lower right corner. Then, the maximum value of two constituent subpaths of the considered cell is added to the cell. For example, as demonstrated in Figure 2.2 (b), the subpaths of cell (R, R) are highlighted in the blackened cells and the maximum value is 1. Thus, this value is added to this cell resulting in that the cell (R, R) has a value of 2. The summation of the scores is continually added to the cells as shown in Figure 2.2 (c) until every cell is added and the scoring of the matrix is complete as displayed in Figure 2.2 (d).

3. A traceback is completed as a final step to generate a sequence alignment, starting from the highest score at top left corner and following the path of the maximum score through the right down as shown in Figure 2.2 (e). In the case where there is more than one solution to follow the maximum score, such as score of 6 (an expression of non-uniqueness), the best alignments of any sublength can be also generated as shown in Figure 2.2 (f).

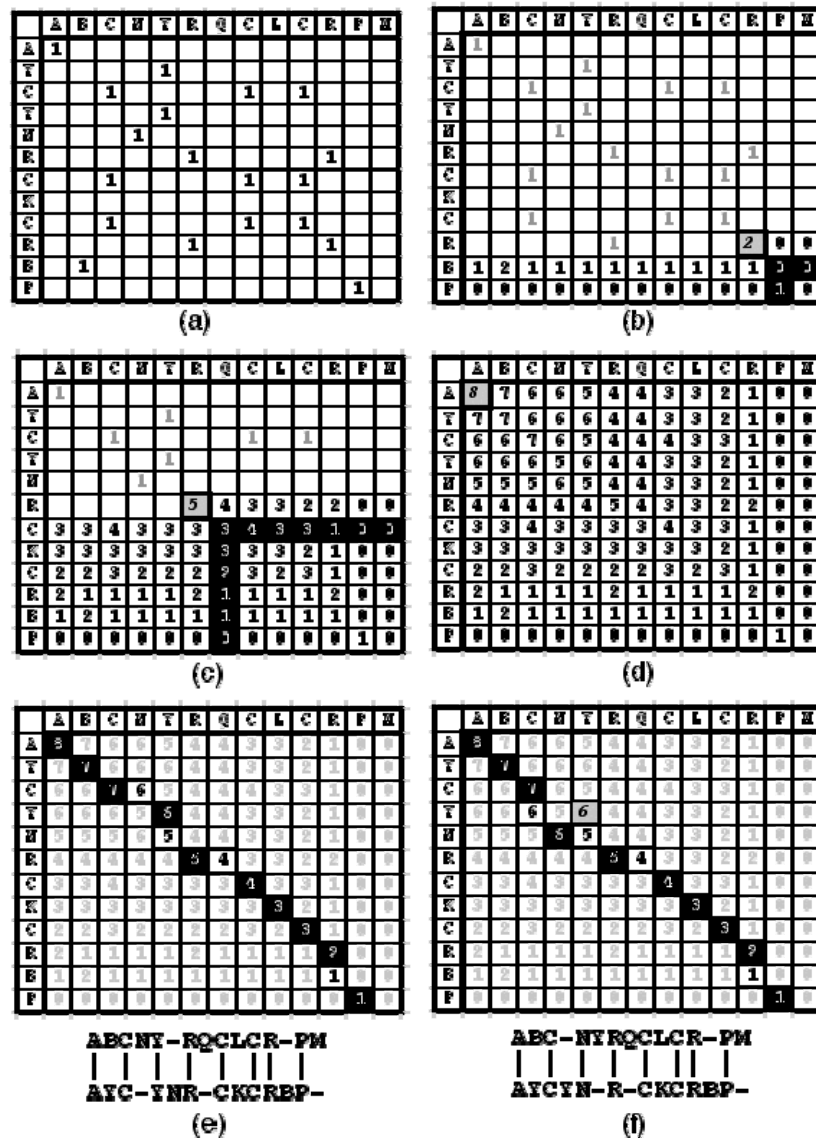


Figure 2.2 The Needleman-Wunsch algorithm gives the global sequence alignment between two sequences (on the top and on the left of the matrix), (a) scoring the sequence which is identical, (b) and (c) partially complete score matrix, (d) scoring the matrix till complete (e) traceback route for giving globally optimum alignment. (f) Alternate alignment illustrating non-uniqueness.[78]

2.2.3 Model Generation

After sequences are completely aligned, the structurally conserved regions (SCRs), which in general correspond to the elements of the secondary structure (alpha-helices and beta-sheets) and to the binding site, are identified. The regions called variable regions (VRs), which are usually the region at the surface or the loop regions, also have to be determined. The conformation of these regions (VRs) in known protein structures may be different from each other. Therefore, in order to construct the VRs of unknown proteins, special techniques have to be used.

At the SCRs region, coordinates of the main chain and side chain atoms are simply copied from the known protein structure to the unknown protein structure, when the amino acid residue type at the same position of the unknown and the template is identical or very similar. On the other hand, when the amino acid residues are of a different type, only the coordinates of backbone atoms of the template are used for the model. A side chain rotamer library in a systematic approach can be applied to explore possible side chain conformations of these residues. A variety of methods can be performed to generate coordinates of residues located at the VRs regions in the unknown structure. If the loop region of the template is well defined, then coordinates of main chain atoms of these residues can be used directly for the corresponding residues in the unknown structure. The coordinates of the side chain of these residues at the VRs can be generated in two different ways; use the same coordinates as the template when the residue type is the same or similar, or generate from the rotamer libraries when the residue type is different. When a loop region of the template cannot be cocrystallized, there will be no coordinates for this region in the template. Therefore, a loop search technique has to be applied to search for a peptidic fragment in the databases that matches with this sequence of the unknown protein structure. Then, coordinates of this peptidic fragment are used for assigning coordinates of residues at this loop region of the unknown protein structure.

2.2.4 Model Evaluation and Refinement

Before applying the derived homology model for further analysis, the quality and the reliability of the obtained model must be checked. There are several programs to accomplish this. For example, PROCHECK[79], 3D-Profler[80, 81], and

PROSA[82] can be applied to evaluate the correctness of the derived homology model. The most popular one among these programs is PROCHECK, a suite of programs that assess the stereochemistry quality of the protein structure. Stereochemical parameters such as phi/psi angles, peptide bond planarity, bond lengths, bond angles, hydrogen-bond geometry, and side-chain conformations reflecting the geometry of the protein can be checked by using the PROCHECK program. The standard values of these parameters, which are described in detail by Morris *et al.*[83], are used to compare with these values derived from the homology model to evaluate the geometry of the residues. The Ramachandran plot [84] is the useful tool for inspecting the distribution of the main chain torsion angles (phi/psi angles), and the results indicate whether the stereochemistry of the backbones are in the acceptable areas. However, if some residues are located at the unacceptable (disallowed) regions, it doesn't mean an error. This could be due to the fact that these residues are located at the active site, and the torsions of these backbones are distorted for binding with ligand or substrate. Nevertheless, these residues have to be checked carefully.

During the homology model building process, many artifacts such as strained peptide bonds, and non-optimum conformation, or even some errors can occur. Therefore, the derived homology model should be subjected to further refinement by performing energy minimization. The aim of energy minimization approach is to find the minimum points on the potential energy surface representing the stable state of the system. Three main different approaches are always used to find the localization of the minimum point representing the local or the global minimum of the potential energy surface. These methods are a non-derivative method (the simplex method), a first and a second derivative method. Steepest descent and conjugated gradient method, which are the first-order minimization algorithms, and the Newton-Rapson method (a second derivative method) are frequently used. The difference between these two approaches is that the first derivative method provides information of the gradient and the new direction for the minimizing step whereas the second derivative method provides information about the curvature of the function. Furthermore, energy minimization together with the molecular dynamic simulation can also be carried out in order to refine the structure.

2.3 Molecular Dynamics (MD) Simulation

One of the computational methods often used for studying the biological system is molecular dynamics (MD) simulation. This method is nowadays routinely applied to elucidate the information such the structural, dynamic, kinetic, and thermodynamic behaviors of biological molecules (protein, membrane or nucleic acid). In addition, the information like the fluctuations, conformation change, and the ligand-protein interaction can also be obtained by performing MD simulation. Results derived from MD simulation, especially the ligand-protein interaction, are the basic information on an atomic level which is very useful for the further drug development.

2.3.1 Theoretical Background

- *Statistical Mechanics*

MD simulations provide information at the microscopic level, such as the coordinates and velocities of each atom. This derived information can be related to the macroscopic properties (the binding free energy of inhibitors with a target protein, the conformational change, energies and mechanism) via statistical mechanics. An experiment is normally done on the macroscopic sample containing sufficient samples (atoms or molecules) with a huge number of conformations. From the statistical mechanics point of view, the average of experimental observations is considered as the ensemble average. Ensemble is defined as a collection of all possible systems that have different microscopic states where the macroscopic or thermodynamic states are identical. Therefore, the ensemble average is carried out by averaging the replicas of the considered systems. To calculate the ensemble average, MD simulations have to pass through all possible states according to the considered thermodynamic state. Consequently, MD simulations compute a time average instead of an ensemble average. However, the time average can be connected to the ensemble average by using the most important axiom in the statistical mechanics-**ergodic hypothesis**. This hypothesis states that the time average is equal to the ensemble average.

The Ergodic Hypothesis

$$\langle A \rangle_{\text{ensemble}} = \langle A \rangle_{\text{time}}$$

Ensemble Average = Time Average

The fundamental concept of this hypothesis is that all possible states of the investigated system can be obtained by allowing that system to evolve in time

indefinitely. Therefore, a MD simulation should be performed long enough in order to get sufficient representative conformations. When the equality of the ergodic hypothesis is satisfied, the related experimental information such as the structural, dynamical, and thermodynamic properties can be computed.

- Newton's Second Law and Classical Mechanics

The MD simulation approach is based on the equation of motion known as Newton's Second Law. The formula is written as equation (3) below:

$$F = m_i a_i , \quad (3)$$

where F_i is the force acting of the particle, and m_i and a_i are the mass and the acceleration of that particle, respectively. Basically, the acceleration of each atom can be computed using equation (3) when the force acting on each atom and the mass are known. Then, the velocities and positions (coordinates) of these atoms can be obtained by integrating this equation of motion. Finally, when trajectories, containing the information of velocity, position and acceleration, are obtained, the average values of the interested properties can be calculated.

The force can also be expressed as the deviation of the potential function (V) as displayed below in equation (4).

$$F = -\nabla_i V \quad (4)$$

By combining equation (3) with (4), the correlation between the deviation of the potential function and the changing of positions as a function of time can be expressed as equation (5) below:

$$F = -\nabla_i V = m_i a_i = m_i \frac{dv_i}{dt} = m_i \frac{d^2 x}{dt^2} \quad (5)$$

Finally, by integrating the relationship between acceleration and velocity ($a_i = \frac{dv_i}{dt}$), the equation between the coordinates, time, velocities and accelerations can be written as $x = at^2 + v_0 t + x_0$. When the position and velocity of atoms are once determined using this equation, these values and the states of the system at any time (future or past) can also be predicted.

Since the biological system is comprised of a large amount of atoms, quantum calculations are not possible at the moment for the entire system. Therefore, the simple empirical potential function, which requires less computational demand and cost, is applied to perform MD simulation. The potential function, $V(R)$, is a sum of the internal terms (bonded interaction) and external terms (non-bonded interaction). The bonded interaction can be obtained by the summation of the bond energy, angle energy and dihedral energy. In contrast, the non-bonded term is comprised of two terms, which are the van der Waals interaction and the electrostatic interaction. Thus, the potential function ($V(R)$) can be written as equation (7):

$$V(R) = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} (1 + \cos[n\phi - \gamma]) + \sum_{i < j}^{\text{atoms}} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right) + \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{4\pi\epsilon R_{ij}}, \quad (7)$$

where K_r , K_θ are force constant of bond stretching and bond bending, respectively, r_{eq} represents the equilibrium of the bond length; θ_{eq} denotes for the equilibrium of the bond angle, V_n is the rotational barrier height, n is the periodicity of rotation, γ stands for the phase angle, A_{ij} and B_{ij} are the coefficients of the van der Waals interaction, q_i , q_j correspond to atomic charge of atom i and atom j , respectively, ϵ is a dielectric constant, and R_{ij} is a distance between atom i and j .

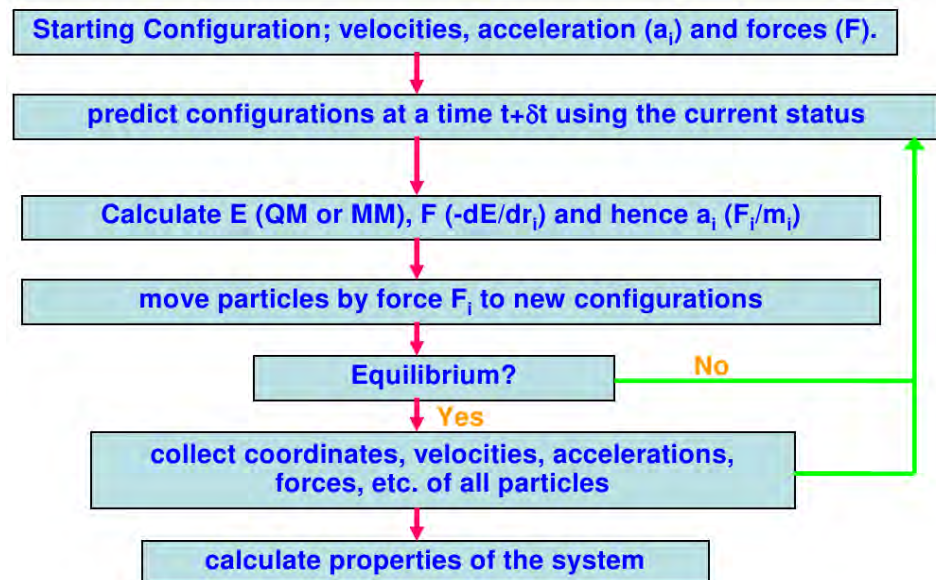


Figure 2.3 Schematic representations for performing a MD simulation

The MD simulation is carried out as in diagram (Figure 2.3) above. However, before running a MD simulation, the energy of the structure must be minimized in order to remove bad contacts and initial strain, which might disturb the MD simulation. According to the equation of motion, in order to obtain the trajectory, first the initial atomic positions (coordinates), velocities and accelerations must be assigned. The coordinates are obtained from the experimental structure (X-ray structure or NMR) or the homology model. On the other hand, the velocity is normally randomly selected from the Maxwell-Boltzman or Gaussian distribution at a considered temperature. Then, the configurations (atomic positions, velocities and accelerations) at the time $t+\delta t$ can be predicted from the current status by using the integration algorithm approximated by a Talyor series expansion as shown in equation (8).

$$\begin{aligned}x(t + \delta t) &= x(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots \\v(t + \delta t) &= v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \dots \\a(t + \delta t) &= a(t) + b(t)\delta t + \dots\end{aligned}\tag{8}$$

There are several numerical algorithms, such as the Verlet algorithm, the Leap-frog algorithm, the Velocity Verlet algorithm and Beeman's algorithm that can be applied for integrating the equation of motion. The details for each of these algorithms are not described here. When the energy of the system and force are calculated, the acceleration can be determined. After that, the particles (atoms) are moved to their new positions according to the force. The equilibrium of the system, measured by the stability of the energy and the root mean square deviation (RMSD) values of the system, is then checked. When the system is not in equilibrium, the configurations of the system are forced to move to a new position until the system becomes equilibrate. Even though the system is already in the equilibrium state, the MD simulation is still performed and the configurations of each step are collected. Finally, the properties of the system can be calculated from the collected trajectories. The longer the simulation, the more trajectories (representative microscopic states) are collected, leading to more accurate results.

2.4 Binding Free Energy Calculation

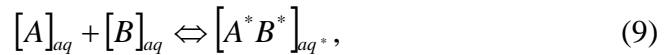
The key to the structure-based drug design is to be able to design ligands and predict how strongly they will bind as noncovalent association with the pharmaceutical targets (enzymes or receptors). Over the years many approaches have been developed and applied for predicting the binding energy of ligands with the target. These approaches include the empirical method, the statistical method, and the rigorous method based on the molecular force field and the interaction energy. Time and computational demand are the main factors that need to be considered for selecting which method should be used. Virtual screening always gives hundreds of candidates. Therefore, the time requirement for calculating and predicting a binding affinity for one compound should be short and the calculation should require less computational demand. However, the methods that use more computational demand and yield more accurate results can also be applied when the experimental complex between the “lead” compound and the target is available. The purpose for performing these approaches for the lead optimization is to investigate a limited number of chemical modifications of these lead compounds.[85]

An empirical method known as the knowledge-based scoring approach is the most rapid method for estimating binding affinities of ligands because this method is based on the simple energy function. However, this method is much less accurate. The details of the scoring function will be discussed later in the molecular docking part. In contrast, the time-consuming approaches for calculating binding affinity of ligands are based on the potential function of molecular mechanics (or force field) using the different sampling conformations derived from MD simulation. By applying these approaches many ligand-protein binding free energy calculation studies showed a good agreement between the calculated and the actual binding affinities of ligands. Nevertheless, there are also several studies that have revealed that these methods cannot be used straightforward for predicting binding affinities as expected. Consequently, these approaches cannot be applied as a black box tool. Thus, it has to be kept in mind that careful attention is needed for setting up the system and interpreting results. [85] Several methods, such as free energy perturbation (FEP), thermodynamic integration (TI), and molecular mechanic/Possion-Boltzmann (Generalized Born) surface area (MM/PB(GB)SA), are used for calculating binding

energy based on the MM calculation and MD simulation. Only the MM/PB(GB)SA method which was applied in this study is reviewed here. In addition, the linear response approach, linear interaction energy with continuum electrostatic (LIECE), is also described.

2.4.1 MM/PB(GB)SA

The main objective of these approaches is to calculate the difference of binding free energy of the noncovalent association of any two different states. These states are normally referred to a ligand-protein complex and a free protein in solution, or the same ligand in the same solution but in the different conformation. The general idea for calculating the difference of binding free energy is shown as the relation below:



where $[A]_{aq}$ and $[B]_{aq}$ are the dynamic structures of molecule A and B free in the solution and $[A^*B^*]_{aq^*}$ refers to the A-B complex structure in the solution. $[A^*]$ and $[B^*]$ represent any structure change and the solvent reorganization upon complex formation is denoted by aq^* .

However, in such a simulation of a biological system, the solvent-solvent interactions contribute more interaction than that from the binding energy to the total energy of the system. In addition, the MD simulation of the three species, $[A^*]$, $[B^*]$, and $[A^*B^*]_{aq^*}$, have to be carried out. Therefore, an excessive amount of time is required till the calculation is converged. Hence, to reduce the time and cost, the binding free energy, computed by the MM/PB(GB)SA approach, is calculated based on the thermodynamic quantities.

Generally, the binding free energy ($\Delta G_{binding}$) of the equation above, $\Delta G_{binding}(A+B \rightarrow AB)$, can be written as the difference between the free energy of the ligand-receptor complex (G_{cplx}) and the unbound receptor (G_{rep}) and the ligand (G_{lig}) as the following:

$$\Delta G_{binding} = G_{cplx(AB)} - (G_{rep(B)} + G_{lig(A)}) \quad (10)$$

The free energy of any molecule X can be separated into two major components, the free energy from solute and from solvent contribution:

$$G(X) = G^{solute}(X) + G^{solvent}(X) \quad (11)$$

The free energy contribution from the solute is obtained by the summation of an average energy of molecular mechanic (E_{MM}) interaction and entropy ($-T\Delta S$) as expressed below:

$$G^{solute}(X) = \langle E_{MM}(X) \rangle - T\Delta S(X), \quad (12)$$

where

$$E_{MM}(X) = E_{int} + E_{ele} + E_{vdw}, \quad (13)$$

and

$$S(X) = S_{trans} + S_{rot} + S_{vib} + S_{config} \quad (14)$$

E_{MM} is normally obtained from the summation of the internal energy (E_{int}), electrostatic energy (E_{ele}), and van der Waals energy (E_{vdw}). T is the temperature of the system (K), and $S(X)$ represents the entropy of the system. This is derived by the summation of the entropic contribution from the translation (S_{trans}), the rotation (S_{rot}), the vibration motion (S_{vib}), and the side chain reorganization (S_{config}). S_{trans} and S_{rot} are obtained from classical statistical mechanics whereas S_{vib} is derived from a normal mode analysis. In the MM/PB(GB)SA approach, the S_{config} is not considered.

The free energy contribution from the solvent consists of two components: the electrostatic ($G_{ele-sol}(X)$) and nonelectrostatic ($G_{nonele-sol}(X)$) contribution. This can be written as:

$$G_{solvent}(X) = G_{ele-sol}(X) + G_{nonele-sol}(X), \quad (15)$$

where $G_{ele-sol}(X)$ is obtained by solving Poisson-Boltzmann (PB) or Generalized Born (GB) equation.

In the PB method, a molecular charge distribution, which is obtained by solving the Poisson-Boltzmann equation, yields the electrostatic potential. The Poisson-Boltzmann equation, as shown below, is simply written in the linearized form because the salt effects are not considered.

$$\nabla \varepsilon(r) \nabla \phi(r) = -4\pi \rho_m(r) + \kappa^2 \varepsilon(r) \phi(r), \quad (16)$$

where $\varepsilon(r)$ is the position-dependent dielectric constant, $\phi(r)$ is the electrostatic potential, $\rho_m(r)$ represents for the molecular charge distribution, and κ is the Debye-Huckel screening parameter to take into account the electrostatic screening effects of (monovalent) salt.

Once the electrostatic potential is determined, the electrostatic contribution to the solvation free energy is computed using the equation, $\sum_i q_i [\phi(r_i) - \phi(r_i)_{vac}]$, where

q_i represent the partial atomic charges at position r_i making up the molecular charge density [$\rho_m(r) = \sum_i \delta(r - r_i)$] and $\phi(r_i)_{vac}$ is the electrostatic potential calculated for the same charge distribution in vacuum.

The electrostatic contribution to the solvation free energy can also be solved by the GB method, which is an analytical approximation to the PB method:

$$\Delta G_{ele-sol} \approx \Delta G_{GB} = -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f_{GB}(r_{ij}, \alpha_i, \alpha_j)} \left(1 - \frac{e^{-kf_{GBij}}}{\epsilon_w} \right), \quad (17)$$

where r_{ij} is the distance between atom i and j , α_x are the effective Born radii of the atoms, and f_{GB} is a smooth function, of which a common form is

$$f_{GB} = \left[r_{ij}^2 + \alpha_i \alpha_j \exp\left(\frac{-r_{ij}^2}{4\alpha_i \alpha_j}\right) \right]^{\frac{1}{2}} \quad (18)$$

$G_{nonele-sol}(X)$ is computed by using a linear function of the solvent accessible surface area (SASA) which is expressed as equation (19) below:

$$G_{nonele-sol}(X) = \gamma \text{SASA}(X) + b \quad (19)$$

where γ is a surface tension parameter, set at $0.00542 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for PB method and $0.0072 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for GB method. $\text{SASA}(X)$ represents the solvent accessible surface area of molecule (X) where the surface area is determined by the Molsurf program, which is based on analytical ideas developed by Mike Connolly [86], and b is a parameterized value which is normally set to 0.00 for both PB and GB approaches.

2.4.2 Linear Interaction Energy with Continuum Electrostatics (LIECE)

Although the methods such as FEP, TI and MM/PB(GB)SA give a good correlation between the predicted and experimental binding affinity, these methods have the disadvantage because of their high computational demand. Thus, these methods can usually not be applied to larger compound collections, which are analyzed in drug discovery and virtual screening. Therefore, faster and less

demanding methods are required to estimate the binding strength of molecules coming out from virtual screening studies.

A variety of linear response (LR) approaches have been developed to overcome the limitation of FEP, TI, and MM/PB(GB)SA techniques. One approach used is the Linear Interaction Energy method (LIE)[87]. LIE estimates the binding free energy of protein-inhibitor complexes by using the following equation:

$$\Delta G = \alpha(\langle E^{\text{vdW}} \rangle_{\text{bound}} - \langle E^{\text{vdW}} \rangle_{\text{free}}) + \beta(\langle E^{\text{ele}} \rangle_{\text{bound}} - \langle E^{\text{ele}} \rangle_{\text{free}}) + \gamma \quad (20)$$

where α and β are coefficients and γ is the constant term which could be additionally added when the absolute binding free energy has to be considered[88]. E^{vdW} and E^{ele} represent van der Waals and electrostatic interaction energies, respectively, between a ligand and its environment (protein bound or free in solution). LIE requires an ensemble sampled over trajectories, representing by $\langle \rangle$ in the equation, which is obtained by means of MD simulation [87] or Monte Carlo[89]. The coefficients (α and β) can be derived by empirical fitting to experimental data. At the beginning, α and β were set to fix values ($\alpha = 0.161$ and $\beta = 0.50$), and they can be transferred among the different protein systems[87]. However, from many studies which investigated the suitable α and β values, it was found that α and β depend on the system studied, the force field applied and the computational methods employed[90]. Even though, LIE is faster than FEP, TI and MM/PB(GB)SA for predicting the binding free energy, the method still has some limitations. Two different MD systems have to be carried out (free ligand in solvent and solvated ligand-protein complex) for one system[91]. Therefore, the number of the compound applied for LIE is limited to less than 100 compounds. The original method is not suitable for several hundreds or thousands of compounds analyzed usually in virtual screening.

Thus, another linear response method termed linear interaction energy with continuum electrostatics (LIECE), which is faster and requires less computational demand than the LIE approach, has been developed [92, 93]. LIECE calculates the binding free energy using different models as shown in equation 21, 22 and 23 below.

a one-parameter model

$$\Delta G = \alpha \Delta E^{\text{vdW}} \quad (21)$$

a two-parameter model

$$\Delta G = \alpha \Delta E^{\text{vdW}} + \beta \Delta G^{\text{ele}} \quad (22)$$

and a three-parameter model

$$\Delta G = \alpha \Delta E^{\text{vdW}} + \beta_1 \Delta E^{\text{ele}} + \beta_2 \Delta G^{\text{sol}} \quad (23)$$

where E^{vdW} and E^{ele} correspond to a van der Waals and an electrostatic interaction energy, respectively. ΔG^{sol} is the free energy of solvation, and ΔG^{ele} is the sum of the electrostatic interaction between the ligand and the protein in the gas phase and the change in the solvation energy upon binding. LIECE is a simple method and more appropriate for virtual screening than LIE because LIECE uses only a single snapshot obtained from energy-minimized complex whereas LIE requires snapshots sampling derived from an explicit water MD simulation (or from Monte Carlo simulation). Moreover, a rigorous treatment of solvation within the continuum electrostatics approximation such as the numerical solution of the Poisson-Boltzmann (PB) equation by the finite-difference technique is applied in LIECE. Thus, this strategy can deal with a large database in reasonable time. For example, a LIECE model of kinase inhibitors were generated from a training set of 165 compounds (CDK2, Lck and p38 inhibitors) and yielded reasonable statistical values for a test set of 128 EGFR inhibitors and 37 EphB4 inhibitors[93]. LIECE requires approximately five minutes per molecule on a single CPU whereas LIE needs at least 6 hours[93]. In addition, novel inhibitors of β -secretase (BACE-1), a membrane-bound aspartic protease, which is a promising target for drug development against Alzheimer's disease [94, 95], EphB4 and CDK2 kinase [93] and West Nile Virus NS2B/NS3 protease[96] were discovered by using this approach.

Another linear response approach is the LR-MM/PBSA[97-99]. A LR-MM/PBSA model is generated by linear-regression based optimization of the coefficients of E^{vdW} , E^{ele} , solvation energy, and solvent accessible surface area (SASA). Then, the derived model is used for predicting the activities of unknown compounds. LIECE and LR-MM/PBSA are related to structure-based scoring function; however, these two methods perform better than docking-based scoring. Recently, docking solutions of Cathepsin B inhibitors were applied to generate a LR-MM/PBSA model. The derived model showed good correlation ($r^2 = 0.919$) and high internal predictivity ($q^2 = 0.887$) [99]. The method is fast and straightforward for calculating binding affinities of unknown compounds in large data sets.

2.5 Virtual Screening

Two approaches that are widely and frequently used today in the drug discovery process are the High-throughput Screening (HTS) and the Virtual Screening (VS) approaches. HTS is done by performing an experimental bioassay to test the activities of numerous compounds against the target protein. The HTS approach is a laborious process that requires lots of time and money. Consequently, the VS technique was developed to assist the drug discovery process for identifying lead compounds, which have notable possibilities to bind to the target protein. Many computational tools are used in the VS strategy for screening compounds in the libraries, which are normally made up of more than a hundred thousand compounds. Therefore, the number of compounds purchased for testing the activity can be reduced by applying the VS approach. This helps in saving time and cost. There are many successful examples [48, 96, 100] using VS method as a part to identify compounds as novel inhibitors for WNV or DV NS2B/NS3 protease. Many different computational chemistry techniques, such as similarity searching, pharmacophore searching, and molecular docking, can be implemented as a VS tools.

A number of compounds in the focused library can be first filtered out by removing unwanted compounds. These compounds include the compounds which contain reactive group, toxic or undesirable functional group. In addition, the drug-like properties according to Lipinski's rule five [101] can also be applied in order to reduce the number of compounds. The rule states that compounds used as orally active drugs in humans should not violate one of the following criteria: molecular weight less than 500 daltons, no more than 5 H-bond donors, less than 10 H-bond acceptors, and log P (an octanol-water partition coefficient) less than 5.

2.5.1 Pharmacophore Model and Pharmacophore Search

A pharmacophore, as defined by IUPAC, is “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response” [102]. The basic idea of pharmacophore in drug discovery is that compounds which bind to the same binding site of the target protein should have similar interaction. Therefore, molecules which have the same essential features for interacting with the

target are expected to inhibit the target protein and have similar biological activity. Typical pharmacophore features include hydrophobic, aromatic, H-bond acceptor, H-bond donor, cation, and anion. In addition, the regions where the ligand cannot be occupied, normally referred to the molecular surface of the target protein, can also be defined as the exclude volume. The exclude volume is helpful for limiting the size of the ligand as it cannot be bigger than the binding pocket of the target protein. A variety of approaches can be employed to generate the pharmacophore model. If the structure of the target protein is not available by either experiment or by the homology model, but a set of active compounds against this target is identified, then a ligand-based pharmacophore model can then be built. This can be accomplished by extracting common features among three-dimensional structures of these compounds [103]. On the other hand, the structure-based pharmacophore model can be generated when a 3D structure of the target protein is available from either experiment (X-ray or NMR) or the homology model. The derived pharmacophore model can then be used to search for molecules sharing the identical features which are situated at the same positions.

2.5.2 GRID – Molecular Field Calculation

GRID [104] is an approach to predict noncovalent interactions between a molecule of known three-dimensional structure and a probe representing chemical features of an inhibitor. In principle, the interaction between the probe and the target (an interested target protein) at each grid point is calculated by using equation (24) as shown below:

$$E_{GRID} = \sum E_{LJ} + \sum E_{EL} + \sum E_{HB} , \quad (24)$$

where E_{LJ} , E_{EL} and E_{HB} represent the Leonnard-Jones energy, the electrostatic term and the hydrogen bond interaction, respectively. The derived results can be visually displayed as three-dimensional contour maps which then are projected on the structure of the target protein. In addition, the derived results suggest the energetically favorable binding regions and can also be applied to generate the structure-based pharmacophore model.

2.5.3 Molecular Docking and Scoring Function

Molecular docking is a computational tool for a structure-based drug design. The molecular docking method is applied to predict a possible binding mode of a ligand with a known 3D-structure target protein. Nowadays this technique is additionally used as an integral part of virtual screening by ranking compounds according to their docking scorings. In the last decade many docking programs using different algorithms and scoring functions have been developed. The main concept common among these different docking programs is that ligands are free to be flexible whereas proteins are normally kept fix or partially fixed during the docking process. A comparison among these programs, algorithms, and scoring functions is beyond the scope of this thesis. Thus, only the algorithm used in the GOLD docking program and the docking scoring function implemented in the GOLD docking program (GoldScore and ChemScore) used in this work are briefly explained.

GOLD (Genetic Optimisation for Ligand Docking) [105, 106] uses a genetic algorithm (GA) to dock ligands into the binding site of the target protein. A set of possible docking solution, known as population, are maintained in the GA. Each possible solutions, termed as a chromosome, contains information about the mapping of a H-bond pair, hydrophobic contact point between ligand and protein, and the conformation around flexible ligand bonds and OH group of protein. The population of the solution is first randomly generated and a fitness score of each chromosome is calculated. Then, the genetic operations such as crossover, mutation and migration are performed in order to get the best possible solution (the highest fitness score). During the docking process GOLD allows ligands to be fully flexible. On the opposite site, the protein is partially flexible which means that the torsion angles of Ser, Thr, Tyr hydroxyl group as well as the NH_3^+ of lysine are rotated to optimize the interaction with the ligand.

Two fitness functions that are normally used in GOLD are GoldScore and ChemScore. GoldScore is comprised of four components (H-bond energy (hb_ext), external (vdW_ext) and internal van der Waals (vdW_int) and ligand torsional strain energy (tor_int)). The equation of the GoldScore fitness function is written as equation (25) shown below:

$$\text{GoldScore} = S_{(hb_ext)} + 1.3750 * S_{(vdW_ext)} + S_{(tor_int)} + 1.0000 * S_{(vdW_int)} \quad (25)$$

Moreover, additional terms such as the internal H-bond, constraint scoring and a covalent term can be included into the GoldScore fitness scoring function when a constraint and a covalent docking are applied in the docking calculation. To correct the protein-ligand hydrophobic contact, the external vdW term is multiplied by 1.375 in the total fitness function. However, the fitness function is applied to predict the binding mode of one particular ligand rather than used to correlate the docking scores and their binding affinities.

ChemScore [107, 108] was generated by empirically fitting from a set of 82 protein-ligand complexes for which their experimental binding affinities were reported. The ChemScore fitness function, as shown in equation (26) below, estimates the free energy of ligand upon the binding with the protein; however, the derived ChemScore value cannot be used as the absolute binding energy or binding affinity. Although the ChemScore fitness function was created by empirically fitting with experimental values, there is no clear indication that this scoring function performs better than GoldScore for predicting binding affinities.

$$\text{ChemScore} = \Delta G_{\text{binding}} + P_{\text{clash}} + C_{\text{internal}} * P_{\text{internal}} + (C_{\text{covalent}} * P_{\text{covalent}} + P_{\text{constraint}}) \quad (26)$$

where P_{clash} is clash penalty and P_{internal} is internal torsion terms. These two terms are included because they militate against close contacts in docking and poor internal conformations. In addition, like in the GoldScore, covalent and constraint scores can extra be added in the fitness function.

A $\Delta G_{\text{binding}}$ in the ChemScore can be computed from this equation:

$$\Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hbond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}}, \quad (27)$$

where each descriptor in the equation represents the physical properties, such as the H-bond, the metal, the lipophilic and the rotatable bonds, that contributes to the free energy. Each parameter is multiplied by a scalar factor (a coefficient) which is derived from the regression,

$$\Delta G_0 = \nu_0, \Delta G_{\text{hbond}} = \nu_1 * P_{\text{hbond}}, \Delta G_{\text{metal}} = \nu_2 * P_{\text{metal}}, \Delta G_{\text{lipo}} = \nu_3 * P_{\text{lipo}}, \Delta G_{\text{rot}} = \nu_4 * P_{\text{rot}}$$

Here, the ν terms are the regression coefficients and P term are the various types of physical contributions to binding.

2.5.4 Similarity Search

Similarity search is one kind of ligand-based virtual screening techniques. The fundamental assumption of this approach is that compounds which have similar chemical structure should also have similar biological activity. Thus, this approach is carried out by searching for compounds in a database that are similar to the query compound. Different approaches (2D fingerprints, 3D pharmacophore) can be used for performing a similarity search. However, the 2D fingerprints, a binary string encoding the substructural fragments, are the most widely used for a similarity-base virtual screening because this method is a computational efficiency and also effective when compared with the other approaches [109]. The Tanimoto coefficient is often used in the 2D fingerprint for indicating the similarity between two molecules (the reference structure and the structure in the database). The simple Tanimoto coefficient equation is expressed as equation, $\frac{c}{a+b-c}$, where a and b represent bits set in their fragment bit-strings and c is bits set being in both of the fingerprints [109].

CHAPTER III

NS2B/NS3 PROTEASE OF DENGUE VIRUS

3.1 Homology modeling and MD simulations of DV NS2B/NS3 Protease

Due to lacking a 3D structure of the inhibitor-bound form of DV NS2B/NS3 protease, a homology model of this complex should be built based on the available crystal structures from WNV and DV protease. The high sequence identity between WNV and DV NS2B/NS3 (56% over both domains) indicates a high similarity between these two structures[110]. Hence, it is reasonable to select the WNV NS2B/NS3 protease for modeling the related DV NS2B/NS3 protease.

In this present work, we have constructed homology models of DV NS2B/NS3 protease complexed with the peptidic inhibitor (Bz-Nle-Lys-Arg-Arg-H) by using different template structures. In addition, molecular dynamics (MD) simulations of the homology models as well as the X-ray structure of WNV NS2B/NS3 protease (2FP7) were carried out in order to investigate the stability and flexibility of the inhibitor-enzyme complex and to examine the interaction between NS2B and NS3 affecting the enzyme structure and inhibitor binding. To our knowledge, this is the first study focusing on the interaction between inhibitors and WNV/DV NS2B/NS3 protease by means of MD simulation. Furthermore, the inhibitor-enzyme interaction was analyzed by applying the GRID approach to identify key residues important for inhibitor binding and protease selectivity.

3.1.1 Materials and Computational Methods

3.1.1.1 Sequence Alignment and Homology Modeling

The amino acid sequences of NS2B and NS3 of DV and WNV were taken from the National Center for Biotechnology Information (NCBI) Database[111]. A modified Needleman and Wunsch [77] method implemented in MOE2006.08[112] and the BLOSUM62 substitution matrix[113] was applied for aligning the sequences.

DV_NS2B (2FOM)	43	GSHMLEADLE	LERAADVWE	EQAEISGSSP	ILSI T I S E D G	S M S I K N E E E E	QTL	95
WNV_NS2B (2FP7)	44	<u>GSHMLETDMW</u>	IERTADITWE	SDAEITGSSE	RVDVRLDDG	<u>NEQLMNDPGA</u>	<u>PWK</u>	96
		*****:*	:**:*:	**	..***:***	::: :::**	::: *	
DV_NS3 (2FOM)	1	<u>AGVLWDVPSP</u>	<u>PPVGKAELED</u>	GAYRIKQKGI	LGYSQIGAGV	YKEGTFHTMW	HVTRGAVLMH	60
WNV_NS3 (2FP7)	1	<u>GGVLWDTPSP</u>	<u>KEYKKGDITTT</u>	GVRIMTRGL	LGSYQAGAGV	MVBGVFHTLW	HTTKGAALMS	60
		.*****:***	*.:	*.***	:::	** * ****	**.:***:*	*.:**.*
DV_NS3 (2FOM)	61	KGKRIEPSWA	DVKKDLISYG	GGWKLEGEWK	EGEEVQVLAL	EPGKNPRAVQ	TKPGLFKTNT	120
WNV_NS3 (2FP7)	61	GEGRLDPYWG	SVKEDRLCYG	GPWKLQHKWN	GHDEVQMIWV	EPGKNVKNVQ	TKPGVFKTPE	120
		:: *	.**:*	:.**	* **:	:::	***** : **	*****:***
DV_NS3 (2FOM)	121	GTIGAVSLDF	SPGTS SGSPIV	DKKGKVVGLY	GNGV V TRSGA	YVSAIANTEK	<u>SIEDNPEIED</u>	180
WNV_NS3 (2FP7)	121	GEIGAVTLDY	PTGTS SGSPIV	DKNGDVIGLY	GNGV I MPNGS	YISAIVQGER	<u>MEEPAPAGFE</u>	180
		* ****:***:	..*****	**:*.*:***	*****: .*:	*:***. : *	* * :	
DV_NS3 (2FOM)	181	<u>DIFRK</u>	185					
WNV_NS3 (2FP7)	181	<u>PEMLRKK</u>	187					
		:	:					

Figure 3.1. Sequence alignment of DV and WNV NS2B and NS3 domains. Stars indicate residues that are identical whereas dots and colons indicate similar residues. Residues which are drawn italic and underlined are missing residues in the X-ray structures. Residues located in the S1, S2, S3 and S4 pocket are colored cyan, yellow, green and magenta, respectively, and catalytic triad are displayed in bold red.

The sequence alignment of DV NS2B/NS3 protease and the template from WNV is shown in Figure 3.1. Experimental data [25, 26, 28, 37] suggest that a central part (40 amino acid residues) of the hydrophilic domain of NS2B is sufficient to act as the essential cofactor for NS3 protease. Since no structural data is available for both terminal parts of the WNV NS2B chain (residue 43 – 48 and 89 – 96) both parts were neglected when in the homology modeling. The homology model of DV NS2B was constructed by using the residues 49-88 of the template structure from WNV NS2B (2FP7). The alignment revealed that this part of WNV NS2B has a sequence identity of 40% (85% sequence similarity) with the corresponding sequence from DV NS2B. For the NS3 protease, electron densities for residues 1-18 of both DV and WNV were not observed in the X-ray structure, due to the flexibility of that region. The entire structure of NS3 protease of DV without the NS2B (1BEF[40] or 1DF9[41]) is substantially different from that complexed with NS2B[38] (RMSD values are shown in Table 1.1). Therefore, the residues 1 to 18 of WNV and DV were also not considered for model building. Only amino acids from 19 to 170 of NS3 of WNV and

DV were used for the alignment and resulted in 57% sequence identity and 76% sequence similarity.

The homology models of DV NS2B/NS3 protease in complex with a peptidic inhibitor were constructed applying the MOE2006.08 program using different template structures for each domain. The first model (DV-1) was generated using the crystal structures of WNV NS2B/NS3 and DV NS2B/NS3 protease as template structures. Coordinates for the NS2B domain and the inhibitor (Bz-Nle-Lys-Arg-Arg-H) were taken from the WNV crystal structure (2FP7), whereas the coordinates of the NS3 domain were directly taken from the DV crystal structure in the apo-form (2FOM). In this way, a complex of DV NS2B (2FP7)/NS3(2FOM) protease with the inhibitor (Bz-Nle-Lys-Arg-Arg-H) (2FP7) was achieved.

The second model (DV-2) was built using the WNV crystal structure 2FP7 as template for both DV protease domains. Missing residues (28 to 32) of the WNV NS3 protease were added by taking the coordinates from the DV NS3 protease (2FOM). In this way, a complex between DV NS2B(2FP7)/NS3(2FP7) protease and the inhibitor (Bz-Nle-Lys-Arg-Arg-H) (2FP7) was derived.

The comparison between different parts (NS2B, NS3protease and NS2B/NS3 protease) of the different X-ray structures and DV models (shown in Table 1.1) was done by using the superpose module implemented in MOE2006.08. The structural alignment was carried out by using a weighted non-linear optimization to determine the solid-body transformation which is required to maximize the superposition of the protein atomic coordinates. The protein superimposition is based on the derived sequence alignment for determining the structurally conserved regions. Mean square distance deviation of corresponding backbone atoms was used to optimize the protein superposition.

3.1.1.2 Stereochemical quality

The geometry and stereochemistry of the models were evaluated by using the program PROCHECK [79]. The stereochemical quality of the generated homology models as well as the WNV crystal structure was evaluated using Ramachandran plots. Results revealed that 92.2% and 7.8% of the residues of WNV-X are located in the most favored regions and the additional allowed regions, respectively. The

Ramachandran plot of DV-1 and DV-2 exhibited that 90.2% and 90.3%, respectively, are located in the most favored region. 9.8% and 9.0% of DV-1 and DV-2 residues, respectively, were found in the additional allowed region. There was no residue of DV-1 located in the disallowed regions whereas just one residue (Lys63_NS3) of DV-2 was observed in the disallowed region. However, this residue is located far from the binding pocket and is not involved in ligand-enzyme interaction. The obtained results indicate that all models possess sufficient stereochemical quality.

3.1.1.3 Protein and Inhibitor Preparation

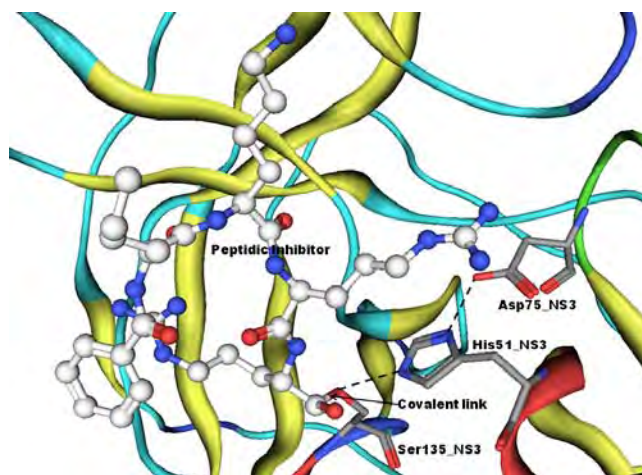


Figure 3.2. Detailed view of the interaction of the covalently bound peptidic inhibitor (ball and stick) with the catalytic site of WNV NS2B/NS3 protease (ribbon). The covalent bond (distance 1.36 Å) between the C atom of the inhibitor aldehyde warhead and the hydroxyl group (OG) of Ser135 of NS3 is shown.

The protease-inhibitor models as well as the X-ray structure of WNV NS2B/NS3 protease in complex with the peptidic inhibitor (2FP7) in which the missing residues (residue 28-32 of NS3) were already added (named as WNV-X for this system) were prepared for the MD simulation. The peptidic inhibitor forms a covalent bond between the C-atom of the aldehyde warhead and the OG atom of Ser135 of the NS3 protease (shown in Figure 3.2)[114]. Therefore, this covalent bond was established in all three protein-inhibitor complexes by removing the atom HG from Ser135 and generating a single bond between the atoms C and OG. Corresponding parameters for the C-O bond were added to the GROMACS topology

file. To mimic the conditions of the tetrahedral intermediate of the bound aldehyde, His51, which is involved in the catalytic triad, was protonated at both δ - and ϵ -nitrogen positions (HIP-form). Parameters of the amino acid norleucine (Nle-P4 of the inhibitor) were generated by considering parameters of the lysine side chain, and parameters for the benzene ring of the inhibitor were created by taking parameters from the benzene ring of phenylalanine.

3.1.1.4 Molecular Dynamics Simulations

MD simulations were performed using GROMACS version 3.3.1[115, 116]. The G43a1 force field (GROMOS96.1 [117]) was applied for all three complexes. Each protein complex was hydrated by 9 Å of the SPC water model[118] applying the triclinic periodic box with 9 Å cutoff. Counter ions, Na⁺ or Cl⁻, were added by replacing water molecules in each system in order to neutralize the system. Subsequently, 10000 steps of steepest descent were applied to minimize the energy of the system using the GROMOS96 force field. The position of backbone atoms was restrained during the first 150 ps with the force constant 1000 kJ/mol in order to allow the adjustment of the solvent molecules. After this, free molecular dynamics simulations were performed using the NPT ensemble applying a time step of 1 fs. The Particle-Mesh-Ewald (PME)[119] method was used to treat the electrostatic interactions. The temperature was kept constant at 300 K by using 0.1 ps coupling time using the Berendsen algorithm [120]. Pressure was maintained using 1×10^5 Pa with a coupling time 0.5 ps and isothermal compressibility of 4.5×10^{-5} Pa⁻¹. The system was equilibrated after 1 ns. For all three complexes a free MD were carried out for 10 ns. The production phase, snapshots from 1-10 ns, was analyzed using subprogram of GROMACS 3.3.1. Results were graphically analyzed using MOE2006.08 and VMD [121].

A comparison of the H-bond occupancy of the protein-inhibitor complexes was done using the GROMACS program. H-bonds were defined by the default values given in the GROMACS program: proton donor–acceptor distance ≤ 3.5 Å and acceptor-donor–H bond angle $\leq 30^\circ$. During the simulation, H-bonds between individual residues could be formed in several ways, such as between inhibitor side chain with enzyme side chain, between side chain and backbone, and between inhibitor backbone and enzyme backbone. Therefore, several H-bonds can occur

between individual residue pairs. The percentage H-bond occupancy (shown in Table 3.1 and Figure 3.5) was calculated taking the snapshots which contained at least one intermolecular H-bond between any atom of the two interacting residues divided by the total number of snapshots.

3.1.1.5 GRID calculations

GRID [104] is an approach to predict noncovalent interactions between a molecule of known three-dimensional structure (i.e., DV NS2B/NS3 protease) and a probe representing chemical features of an inhibitor. The calculations were performed using the GRID program version 22a [122]. The structure of DV and WNV NS2B/NS3 protease was taken from the last snapshot of the MD simulations, the most equilibrated one, of DV-2 and WNV-X, respectively, the inhibitors were removed, and the resulting conformations were used for the GRID analysis. The calculations were carried out on a cube of the size 20 x 20 x 20 Å³ with a spacing of 1 Å covering the inhibitor binding pocket, to search for binding sites complementary to the functional groups of the inhibitors. The following probes were used for the calculations: hydroxyl (OH), positively charged amine (N+) and the hydrophobic probe (DRY). The OH and N+ probes imply the electrostatic favorable region whereas DRY probe indicates the hydrophobic favorable region. Therefore, the selectivity between the Arg compared to Ala at the P1 subsite of WNV and DV could be explained by applying these probes. The calculated GRID contour maps were then projected on the NS2B/NS3 structures of DV and WNV using the MOE program.

3.1.2 Results and Discussion

3.1.2.1 Model Building and Evaluation

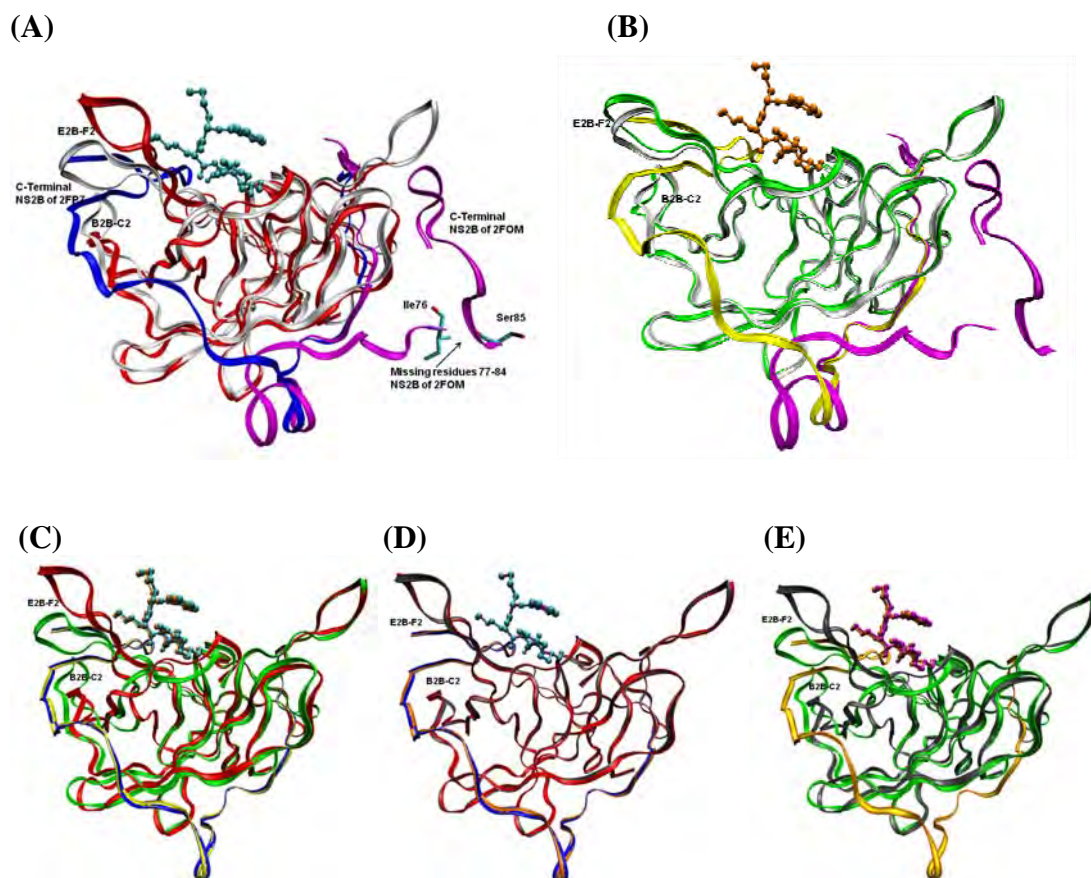


Figure 3.3. Superimposition of (A) 2FP7 and 2FOM, (B) 2FOM and DV-1, (C) WNV-X and DV-1, (D) WNV-X and DV-2, and (E) DV-1 and DV-2. The individual structures are colored as follows: WNV-X and 2FP7: NS2B blue, NS3 red, inhibitor cyan; 2FOM: NS2B magenta, NS3 white, DV-1: NS2B yellow, NS3 green, inhibitor orange; DV-2: NS2B Orange, NS3 gray, inhibitor magenta.

The two DV protease homology models and the WNV protease crystal structure were structurally analyzed by superimposing them on their backbone atoms of NS2B/NS3 protease (Figure 3.3 (C)-(E)). The analysis revealed that WNV-X and DV-2 show a high structural similarity as indicated by a low RMSD value (0.06 Å). Since we used two different templates to model the NS3 domain of the two individual homology models it is obvious that they show larger structural deviations (RMSD = 2.10 Å). The DV NS2B/NS3 models represent two different conformations of the

protease. The major structural differences between DV-1 and DV-2 can be found around two loop regions (Figure 3.3 (C)-(E)). The first loop, denoted as B2B-C2 loop, covers the region from residue 114 to 124, whereas the second loop, named as E2B-F2 loop, involves residues 152 to 164.

In addition, since the X-ray structure of WNV NS2B/NS3 protease complexed with the trypsin inhibitor (2IJO) is available, the influence of inhibitor binding on the conformation of the NS2B/NS3 protease was analyzed. For this purpose a third model of DV NS2B/NS3 (complexed with the trypsin inhibitor) was generated using the 2IJO structure as template. DV-2 and DV-3 shows a high overall structural similarity. Superimposing models DV-2 and DV-3 yielded an RMSD of 0.61 Å indicating that the enzyme adopts approximately the same conformation in the productive form even when bound to a different inhibitor.

In contrast to the former reported DV NS2B/NS3 homology models[43] as discussed before in chapter 1 (section 1.4.1), our model represents the first complete structure of the protease including the complete cofactor NS2B in the productive form. Since our model represents the active productive form of the protease it could be applied for target-based design and screening of small molecule inhibitors.

3.1.2.2 Molecular Dynamics Simulations

3.1.2.2.1 Model Stability and Overall Structure

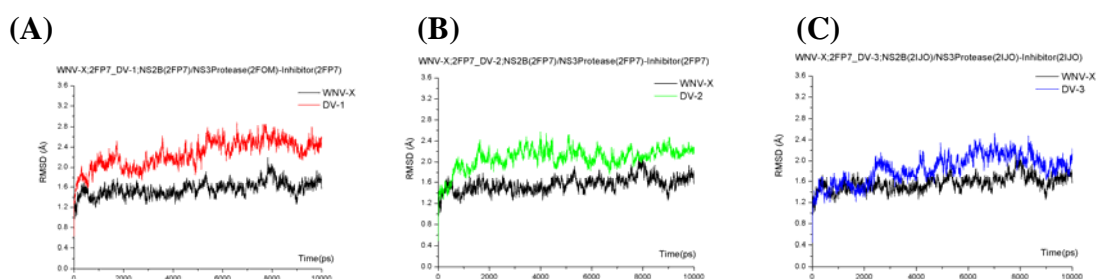


Figure 3.4. (A) Comparison of the RMSD plots for the MD simulations of WNV-X (black-line) and DV-1 (red line), (B) RMSD plots for the simulations of WNV-X (black line) and DV-2 (green line), and (C) RMSD plots for the simulations of WNV-X (black line) and DV-3 (blue line)

The stability of the three protease structures was tested by running MD simulations. In Figure 3.4, the RMSD values between each snapshot with respect to

the starting structure is plotted over a period of 10 ns simulation time. The RMSD values for WNV-X, DV-1, DV-2 and DV-3 remained at ~ 1.50 Å, ~ 2.20 Å, ~ 2.00 Å and ~ 2.00 Å, respectively, indicating the stability of the three systems.

3.1.2.2.2 NS2B-NS3 Interaction

Table 3.1. Comparison of the H-bond occupancy between residues from NS2B and NS3. The number of H-bond between each residue pair are given in the parenthesis, behind the percentage of H-bond occupancy.

H-bond Pair between		% H-bond occupancy			
NS2B	NS3	WNV-X	DV-1	DV-2	DV-3
Residues relating with the mutagenesis data (Leu74, Ile76 and Ile78)					
Val75	Val115	93.3 (1)			
Leu74	Leu115		8.9 (1)	89.2 (1)	82.8 (1)
Val75	Lys117	95.2 (1)			
Leu74	Lys117		76.1 (1)	91.1 (1)	91.8 (1)
Val77	Lys117	92.5 (2)			
Ile76	Lys117		93.2 (1)	94.6 (1)	92.1 (1)
Val77	Pro119	0.00 (0)			
Ile76	Asn119		10.20 (1)	31.30 (1)	0.00 (0)
Leu79	Pro119	0.00 (0)			
Ile78	Asn119		39.03 (3)	32.36 (3)	1.00 (1)
Residues locating within 10 Å around the inhibitor					
Asp81	Lys73	5.61 (4)			
Glu80	Lys73		11.44 (4)	5.72 (3)	0.80 (2)
Asp82	Asn152	0.00 (0)			
Asp81	Asn152		0.00 (0)	2.78 (3)	0.00 (0)
Gly83	Asn152	61.40 (1)			
Gly82	Asn152		0.02 (1)	66.44(1)	58.60 (1)
Asn84	Asn152	1.60 (2)			
Ser83	Asn152		0.00 (0)	1.30 (3)	0.10 (1)
Phe85	Asn152	17.99 (3)			
Met84	Asn152		0.00 (0)	0.89 (3)	0.60 (2)
Phe85	Gly153	5.20 (1)			
Met84	Gly153		0.00 (0)	12.70 (1)	0.00 (0)
Phe85	Ile155	0.00 (0)			
Met84	Val155		22.80 (2)	1.80 (1)	0.10 (1)
Met88	Pro119	0.00 (0)			
Lys87	Asn119		3.66 (3)	7.83 (3)	0.50 (2)
Met88	Pro157	0.00 (0)			
Lys87	Arg157		16.88 (6)	0.00 (0)	0.40 (2)

The C-terminus of NS2B affects the formation of the active site. This has been recently demonstrated by a site directed mutagenesis study[34] reporting that an inefficient autoproteolysis is caused by the mutations of the DV NS2B residues L74A, I76A, I78A and I76A/I78A. Therefore, the interaction between NS2B, NS3 and the inhibitor was focused on these residues (Leu74, Ile76 and Ile78) and other residues within a region of 10 Å around the bound peptidic inhibitor including the residues from NS2B. NS2-NS3B interaction occurs between the loop domains of NS3, the B2B-C2 and E2B-F2 loops, as well as the C-terminus of NS2B.

We first analyzed the H-bond occupancy within the MD simulations of the protease-inhibitor complexes (Table 3.1). The analysis revealed strong H-bonds between Leu74 (NS2B) and the two NS3 residues Lys117 and Leu115. The percentage of H-bond occupancy between Leu74 and Leu115 was 8.9 % for DV-1 and 89.2% for DV-2, and between Leu74 and Lys117 76.1% for DV-1 and 91.1% for DV-2, respectively. The strong hydrogen bonds observed for model DV-2 are in good agreement with the experimental data, whereas such a stable interaction was not observed for the simulation of the DV-1 model. Moreover, the strong H-bonds of Leu74-Leu115/Lys117 and Ile78-Asn119 observed in the DV-2 model confirms the functional role of a $\Phi X_3 \Phi$ motif (two hydrophobic residues separated by three unspecific amino acid residues, corresponding to $^{74}\text{Leu-Ser-Ile-Thr-Ile}^{78}$ of the DV NS2B) in association with the protease [123]. Also here the DV-2 model is in better agreement with the reported experimental data. The low H-bond occupancy of Asp81/Glu80 (NS2) and Lys73 (NS3) (5.61% for WNV-X, 11.44% for DV-1 and 5.72% for DV-2), the absence of a H-bond between Asp82/Asp81 (NS2B) and Asn152 (NS3) also agree well with the experimental data [124] suggesting that Asp81 of NS2B is not involved in a direct interaction with NS3.

Further H-bonds were observed between Ile76 (NS2B) with Lys117 and Asn119 (NS3) for both DV models. The observed occupancy was 10.20% for DV-1 and 31.30% for DV-2 for the Ile76_NS2B with Asn119_NS3 pair, and it was 93.2% and 94.6% for DV-1 and DV-2, respectively, for the Ile76 with Lys117_NS3 pair. The other H-bond interaction was observed between Ile78 (NS2B) and Asn119 (NS3). The observed occupancy was 39.03% for DV-1 and 32.36% for DV-2. According to recent experimental work [124], which suggested a strong and direct interaction

between amino acid residue 50-80 of NS2B with the NS3 domain of DV, the homology models are in agreement with the experimental data. In the original work[124], no details about the interaction between NS2B and NS3 were given. Therefore, the DV homology model shed light on the molecular level of this interaction. The model shows a strong interaction between the β -strand of NS2B and the β -barrel of the NS3 protease (between residues 51-57 (NS2B) and 22-28 (NS3) and between residues 73-75 (NS2B) and 115-117 (NS3)). Furthermore, the MD simulations revealed a strong interaction between Ile76 (NS2B) and Asn119 (NS3) as well as between Ile78 (NS2B) and (Asn119). These results indicate the importance of the identified residues to stabilize and activate the NS3 protease. Furthermore, the observed H-bond interaction between Ile76 and Ile78 from NS2B with Asn119 from NS3 is supported by the site-directed mutagenesis study [34]. In the generated Ala-mutants no H-bond interaction can be established between the residues from NS2B and NS3.

Another strong hydrogen bond which is supported by the experimental data[38] was detected for Gly83/Gly82 and Asn152 in the DV-2 model as well as in the WNV X-ray structure. In the DV-1 model no such interaction was observed during the MD simulation. The DV-1 model lacks the interaction of the C-terminus of NS2B and the substrate-binding site around Asn152 of NS3[40]. The observed H-bond between the C-terminus of NS2B and NS3 is stabilizing the S2 pocket of the NS3 protease. Moreover, the MD simulation of the model DV-3, which was generated using only the crystal structure 2IJO as template, showed exactly the same H-bond interaction pattern as observed for DV-2, except the H-bond interaction between Ile76 and Ile78 (NS2B) with Asn119 (NS3). In general, DV-2 and DV-3, which were generated on the basis of the productive inhibitor-bound form of WNV NS2B/NS3 protease, are in good agreement with the mutagenesis data of DV NS2B/NS3 protease. Furthermore, our simulations indicate that the binding of structurally different inhibitor does not affect the interaction between the C-terminus of NS2B and the NS3 protease.

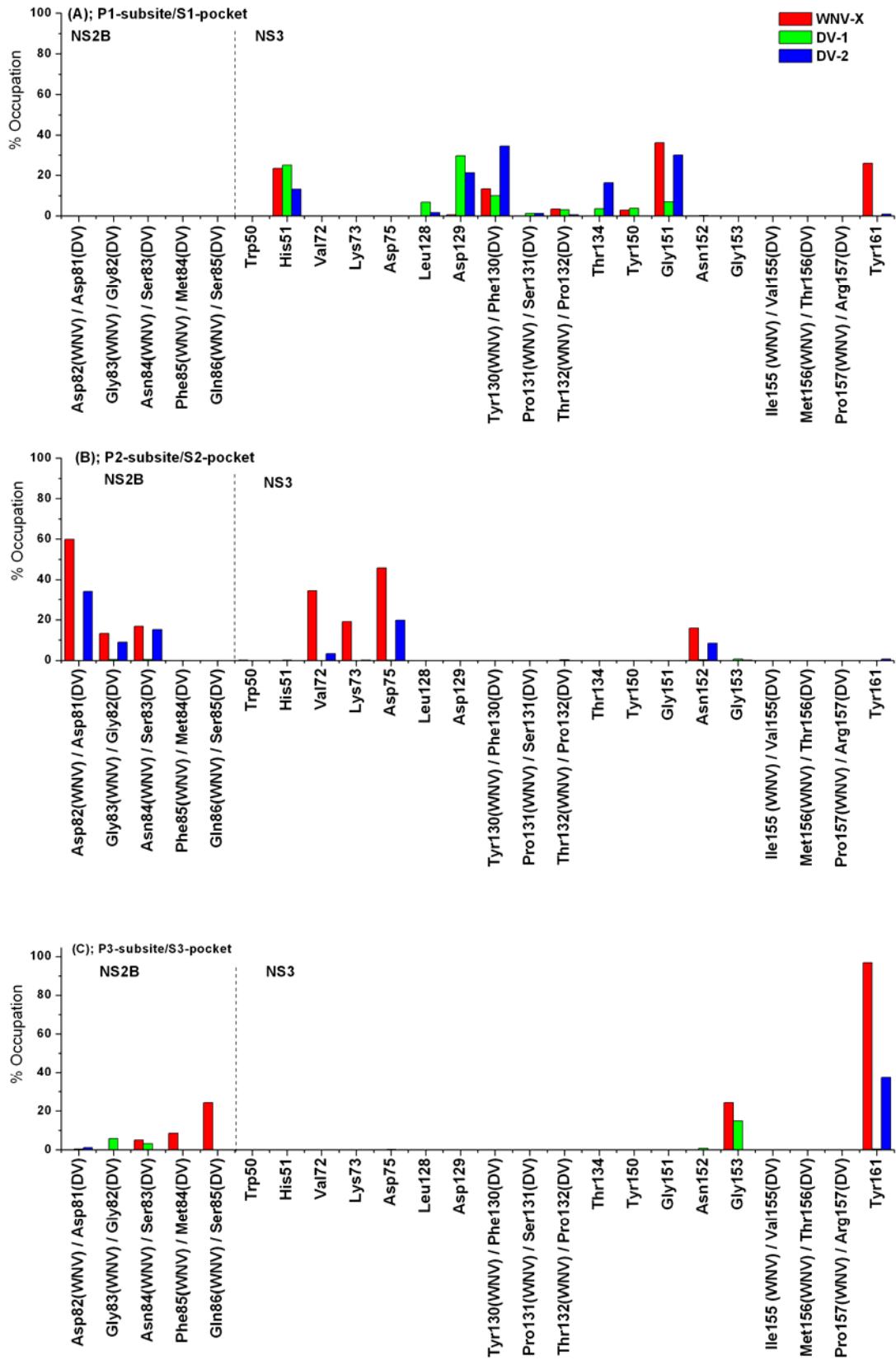
Recently, MD simulation of the X-ray structure of DV NS2B/NS3 protease in the non-productive form (2FOM)[51] was carried out to investigate the interaction of NS2B for stabilizing NS3 protease domain. However, it was stated in the original

publication of WNV and DV protease crystal structures[38] that the cofactor NS2B in the DV crystal structure, especially the C-terminal domain, does not interact with the substrate-binding site. Therefore, the interaction between NS2B and NS3 and its impact for enhancing the activity of the protease could not be deduced from this structure or models built on that structure [38]. In contrast, our DV protease models built on the basis of 2FP7 or 2IJO, where the C-terminal domain of NS2B is bound to the substrate-binding site, are more appropriate to study the influence of the cofactor NS2B, especially the C-terminal domain, for stabilizing NS3 and enhancing the activity.

It was proposed that the cofactor NS2B functions as a molecular chaperone which promotes the folding of the NS3 protease into the active conformation[37]. Our MD simulations of the DV-2 model support this hypothesis. The simulations showed that the strong interaction between NS2B and NS3 protease; especially, between the C-terminus of NS2B and Asn119 and Asn152 of NS3 results in a stable conformation of the loop region (both B2B-C2 and E2B-F2 Loop) of NS3. In addition, the results revealed that DV-2 is in better agreement with the experimental data[34, 38, 123, 124] than DV-1. Thus, the DV-2 represents the NS2B/NS3 protease-inhibitor complex in the active form. Therefore, it can be suggested that the productive inhibitor-bound conformation of the NS2B/NS3 protease-inhibitor complex of WNV and of DV should be similar, as observed for the inhibitor-free form (2GGV and 2FOM, RMSD = 2.20 Å).

3.1.2.2.3 Protease-Inhibitor Interaction

Although the overall sequence identity of WNV and of DV is high, there are also some differences which can be observed in the sequence alignment shown in Figure 3.1. Therefore, the interaction between the peptidic inhibitor and the binding pocket was analyzed for both proteases.



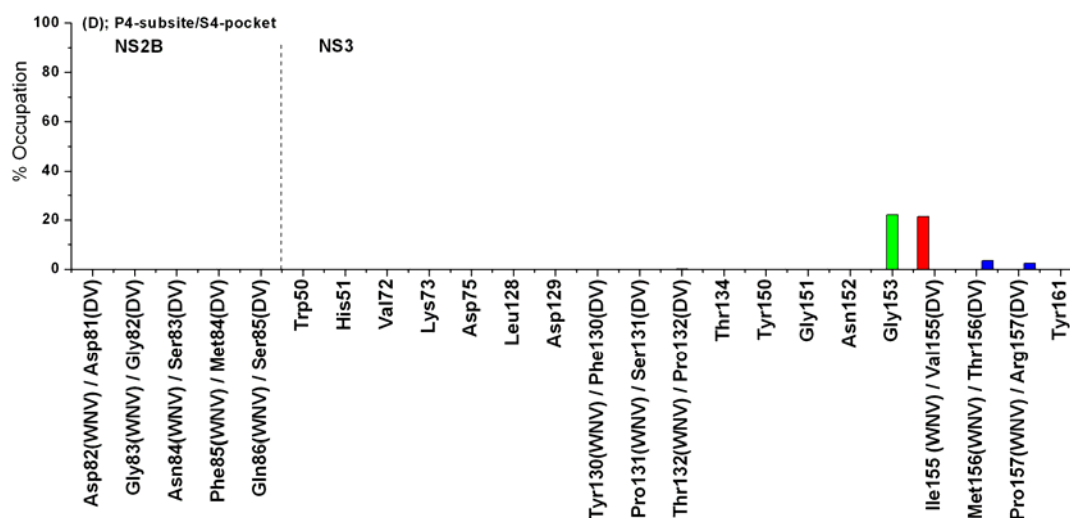


Figure 3.5. Comparison of the H-bond occupancy between each residue of the inhibitor and the individual binding pockets of the protease; P1-S1 pocket (A), P2-S2 pocket (B), P3-S3 pocket (C) and P4-S4 pocket (D).

The H-bond interaction between each amino acid residue of the peptidic inhibitor and the enzyme was investigated for all three complexes. Results are shown in Figure 3.5 (A) – (D). It was reported in literature [125-127] that the P1-S1 pocket interaction of both WNV and DV involves only residues from the NS3 protease domain. The WNV protease crystal structures (2FP7 [38] and 2IJO [42]) reveal that the S1 pocket is formed by Asp129-Ser135, Tyr150, and Tyr161 of the NS3 protease. In agreement with this observation, the MD simulation demonstrated that strong H-bonds between Arg-P1 and the S1 pocket of WNV-X and DV-2 exist (Figure 3.5 (A)). The occupancy of the H-bond between Asp129 and Arg-P1 of WNV-X is low. However, the salt-bridge interaction between both residues was observed throughout the whole simulation time (~ 4 Å, shown in Figure 3.6). Based on the MD simulations it can be suggested that Asp129 of the NS3 protease play a significant role in stabilizing the basic inhibitor P1 residue Arg or Lys[128].

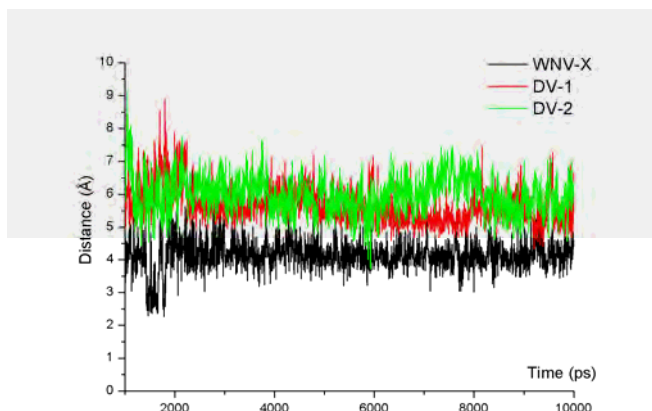


Figure 3.6. Plot of the salt-bridge distance between Arg-P1 and Asp129 from NS3 of WNV-X (black), DV-1 (red) and DV-2 (green)

The MD simulation of WNV-X showed a strong H-bond between Arg-P2 and Asp82, Gly83, Asn84 of NS2B and Asp75, Asn152 of the NS3 protease. These results agree well with the experimental data [38, 42] showing that the S2 pocket is formed by these residues. It was reported that the C-terminus of NS2B, Asp75 and Asn152 (NS3) are part of the S2 pocket of the DV NS3 protease. During the simulation of DV-2 the interaction between Arg-P2 and these residues was observed (Figure 3.5 (B)). The strong H-bond between Gly82 (NS2B) and Asn152 (NS3) resulted in increased stability of the E2B-F2 loop, and thus the interaction between Arg-P2 and Asn152 can occur. Moreover, these results suggest that the C-terminus of NS2B does not only play an important role for stabilizing and preorganizing the NS3 protease but also is essential for the interaction with the P2 residue of the inhibitor.

The results of the MD simulation of WNV-X help to understand the recent mutagenesis data [33] of WNV NS2B/NS3 protease. In this study it was shown that complete inactivation of WNV NS2B/NS3 protease is achieved by the mutations G83A and F85A in NS2B. Our MD simulation of WNV-X reveals H-bonds between Gly83 and Phe85 of NS2B with Asn152 (NS3) (Table 3.1). The H-bonds are quite stable as indicated by the high percentage of H-bond occupancy during the simulation, e.g. 61.40 % for Gly83. In the mutation G83A the H-bond is disfavoured by the introduction of the side-chain of the alanine residue. By losing this hydrogen-bond network between NS3 and NS2B the activity of the enzyme for cleaving a

substrate is totally lost. The stability of the hydrogen bonding network can be observed for WNV-X and DV-2 but not for the DV-1 model (Figure 3.7).

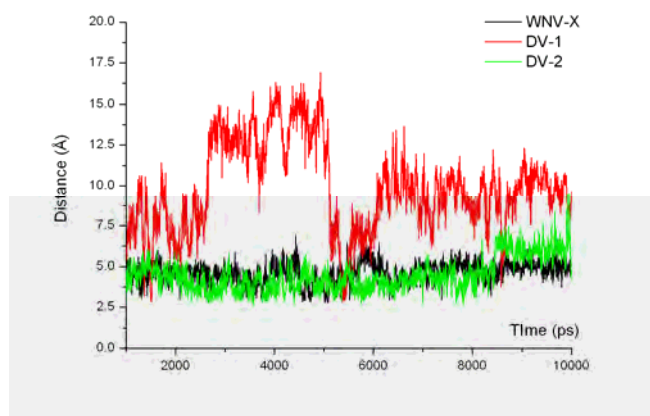


Figure 3.7. Distance between O δ -Asn152 (NS3) and N ϵ -Arg-P2 of the inhibitor during the simulation of WNV-X (black), DV-1 (red) and DV-2 (green).

Figure 3.5 (C) shows that the percentage of occupation for the H-bonds between Lys-P3 of the inhibitor and the protease residues. For WNV-X the H-bonds between the backbone of Lys-P3 and the side chain of Gly153 and Tyr161 of NS3 are observed throughout the simulation time. Interestingly, only the H-bond between Lys-P3 and the Tyr161 was observed for DV-2. Molecular docking study [129] of inhibitors at WNV NS2B/NS3 showed that the Lys-P3 side chain did not bind in the S3 pocket of the enzyme which is largely solvent exposed. From the sequence alignment it can be detected that the residues forming the S3 pocket which are part of the C-terminus of NS2B are not conserved among the flaviviruses. Thus, larger differences in the type of interaction between inhibitors and the S3 pocket can be expected. In addition, the structure-activity relationship of peptidic inhibitors showed that modifying the P3 position of inhibitor from NH to N-CH₃ caused a major loss of the inhibitor's activity for both DV[50] and WNV[130]. These data indicate the significance of the H-bond between the P3 backbone and the protease. The interaction of the N-CH₃ group results in a flip of the C=O backbone group into the opposite direction and results in losing the H-bond interaction between C=O and Tyr161 or Gly153. Furthermore, the replacement of the P3 subsite by a shorter side chain, which resulted in higher potency of the inhibitor, indicating that the P3 side chain is less

important for the inhibitor-enzyme interaction. In agreement with these experimental data [49, 129, 130], our MD simulation results showed that Gly153 and Tyr161 of NS3 of WNV and Tyr161 of DV bind to the backbone of the P3 residue of the inhibitor and are key residues in the S3 pocket.

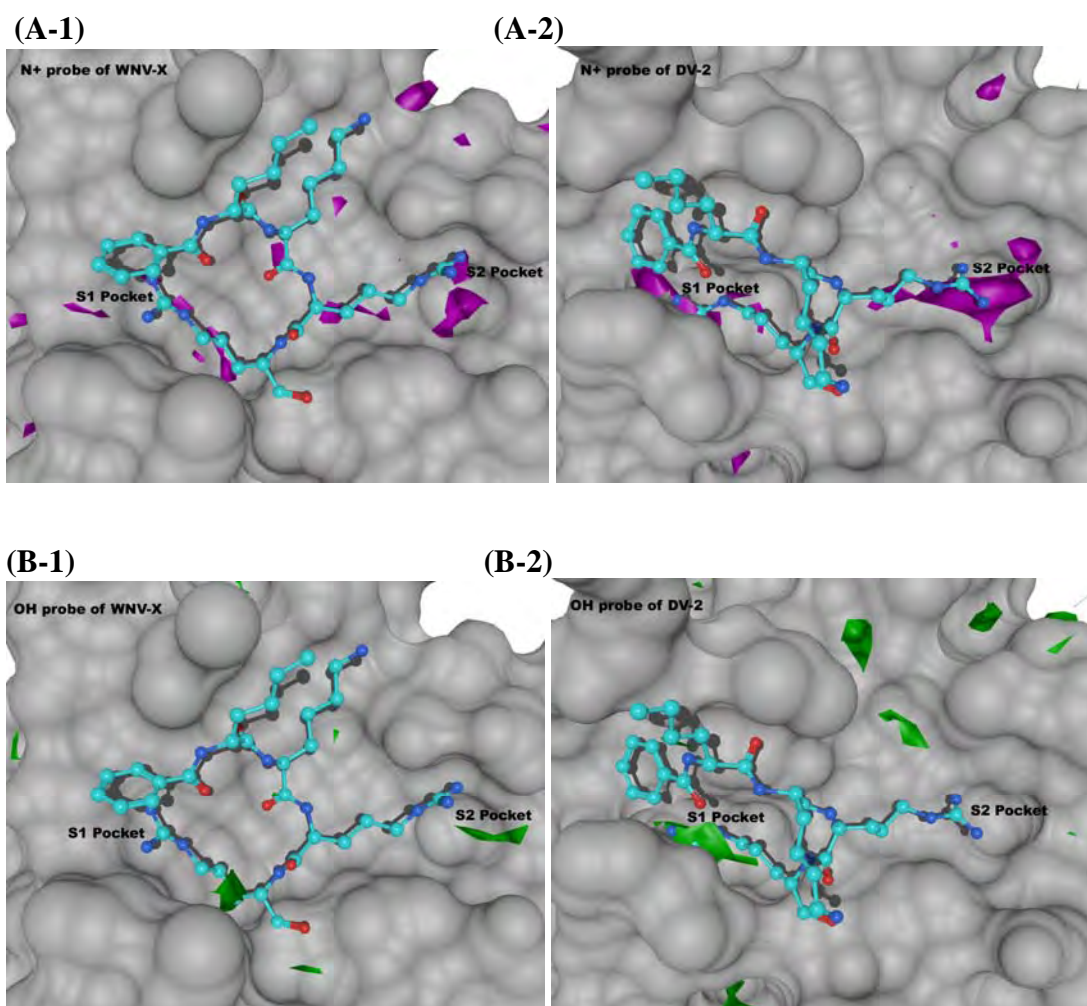
The S4 pocket of the protease is formed by the following residues: Val154, Ile155, Met156 and Pro157 (NS3) in case of WNV, whereas Val154, Val155, Thr156 and Arg157 are the corresponding residues of the S4 pocket in case of the DV protease. In general, the S4 pocket is dominated by hydrophobic residues. This is in agreement with the preference of norleucine or leucine at the P4 position of the peptidic inhibitor [128]. The MD simulation revealed that the hydrogen bond between Nle-P4 of the inhibitor and the S4 pocket is very low (Figure 3.5 (D)) implying that H-bond interactions in this pocket rarely occur.

The visual analysis of the inhibitor-DV2 and inhibitor-WNV-X interactions showed that the inhibitor is interacting in a similar way with the corresponding residues. A series of tetrapeptidic inhibitors (Bz-P4-P3-P2-P1-H)[49, 130] and non-peptidic inhibitor [45] were synthesized and tested against WNV and DV NS2B/NS3 protease. Results showed that the most potent inhibitors of WNV NS2B/NS3 protease show also high activity against DV NS2B/NS3 protease. In addition, less active WNV NS2B/NS3 protease inhibitors are also less active against DV NS2B/NS3 protease. The analysis of the inhibitor-enzyme interaction derived from the MD simulations showed more clearly that the similar activity profile of the inhibitor can be explained by their similar binding mode.

3.1.2.3 Selectivity profile of the inhibitors Bz-Nle-Lys-Arg-Ala-H compared to Bz-Nle-Lys-Arg-Arg-H

In order to rationalize the structure-activity relationships of peptidic WNV and DV protease inhibitors, the structure of WNV and DV NS2B/NS3 protease taken from the last snapshot of the MD simulation of WNV-X and DV-2, respectively, was applied for a GRID field analysis. The calculations were carried out for the inhibitor-free protease and using several chemical probes (hydroxyl (OH), positively charged amine (N^+) and the hydrophobic (DRY) probe). The GRID results shown in Figure 3.8 revealed a favourable field for the N^+ (Figure 3.8 (A)) and for the OH (Figure 3.8

(B)) probe at the S1 pocket. In case of the DV protease the field is larger compared to that of WNV indicating the preference of the DV S1 pocket for a positively charged polar group. In contrast, the DRY field (Figure 3.8 (C)) at the S1 pocket of the WNV protease is larger than that of DV suggesting the preference of a hydrophobic group at this region of the WNV S1 pocket. The GRID results can be used to explain the decreased DV protease activity when changing the inhibitor P1 residue from Arg to Ala (K_i increased from 5.8 μM to 193.0 μM), whereas the activity of the modified inhibitor for WNV is unchanged ($\text{IC}_{50} = 4.1 \mu\text{M}$ vs. 4.6 μM).



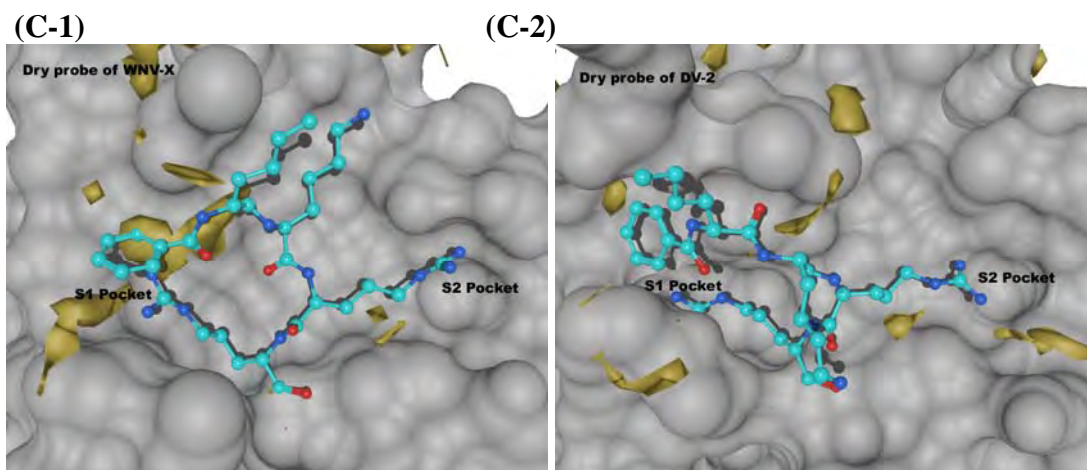


Figure 3.8. (A) GRID results: N^+ probe (magenta contour) at the contour level of -5.25 kcal/mol of WNV-X (A-1) and DV-2 (A-2), (B) OH probe (green contour) at the contour level of -7.50 kcal/mol of WNV-X (B-1) and DV-2 (B-2) and (C) Dry probe (yellow contour) at the contour level of -0.50 kcal/mol of WNV-X (C-1) and DV-2 (C-2). The inhibitor is shown colored cyan.

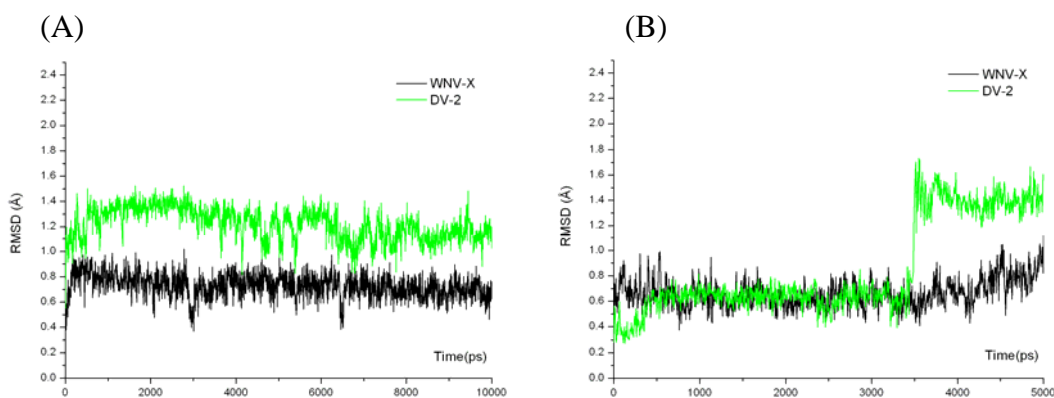


Figure 3.9. (A) RMSD values of the inhibitor Bz-Nle-Lys-Arg-Arg-H bound in WNV-X (black line) and DV-2 (green line). (B) RMSD values of the mutated inhibitor Bz-Nle-Lys-Arg-Ala-H bound to WNV-X (black line) and DV-2 (green line).

Bz-Nle-Lys-Arg-Ala-H shows the highest selectivity among the reported inhibitors for WNV ($K_i = 193.0 \mu\text{M}$ for DV and $IC_{50} = 4.6 \mu\text{M}$ for WNV). The substitution at the P1 position of the inhibitor by hydrophobic amino acids (e.g. mutating Arg to Ala) resulted in decreased activity against DV protease whereas it

does not affect the activity for WNV protease. Therefore, to test the influence of the P1 Arg-Ala mutation MD simulations were carried out for DV and WNV protease in complex with the inhibitor Bz-Nle-Lys-Arg-Ala-H. The complexes of WNV and DV NS2B/NS3 with Bz-Nle-Lys-Arg-Ala-H were constructed by taking the last snapshot from the MD simulation of the WNV-X and DV-2 structures and subsequently modifying the P1 residue from Arg to Ala. Then, these two systems were subjected to MD simulation for 5 ns using the same conditions as the previous MD simulations. The observed interactions of Bz-Nle-Lys-Arg-Arg-H with the WNV enzyme (WNV-X) and the DV enzyme (DV-2) are similar as discussed above. The conformation of the original inhibitor was quite stable in both simulations (WNV-X and DV-2) indicated by the low RMSD values (shown in Figure 3.9 (A)). In contrast, the WNV-selective inhibitor (Bz-Nle-Lys-Arg-Ala-H) gave different results. The conformation of this inhibitor bound to the WNV protease is quite stable whereas it changed the position in the DV system. This is indicated by a considerable increase in the RMSD value of the inhibitor after 3500 ps simulation time (Figure 3.9 (B)). Comparing the structure of the WNV protease complexed with the inhibitors Bz-Nle-Lys-Arg-Ala-H and Bz-Nle-Lys-Arg-Arg-H after 5 ns simulation time (Figure 3.10) showed clearly that the conformation of the two inhibitors in these two complexes is not much different from each other. In contrary, the mutated inhibitor Bz-Nle-Lys-Arg-Ala-H adopts a different conformation at the end of the 5 ns simulation time (Figure 3.11). As a consequence of the mutations Arg to Ala, not only the interaction at the S1 pocket is different, but also the interaction at the S2 and S3 pocket changed during the MD simulation. Especially, the important interactions between Tyr161 (NS3) at the S2 pocket and P2 is lost. The lost interactions are able to explain the decreased activity against DV (K_i 193.0 μ M).

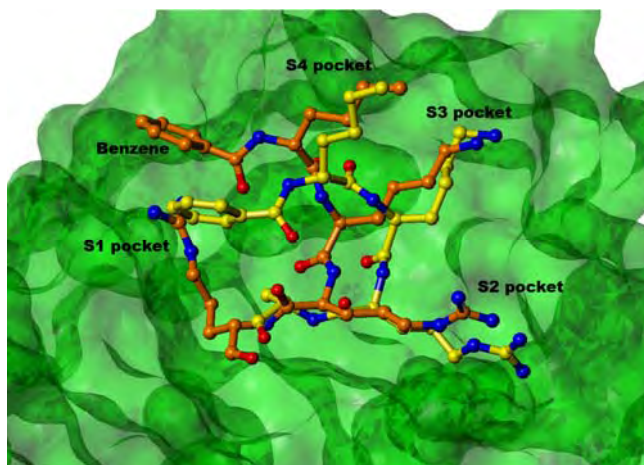


Figure 3.10. Comparing the conformation of the WNV protease bound inhibitor Bz-Nle-Lys-Arg-Arg-H (orange) and the bound inhibitor Bz-Nle-Lys-Arg-Ala-H (yellow) at the end of the 5 ns simulation time.

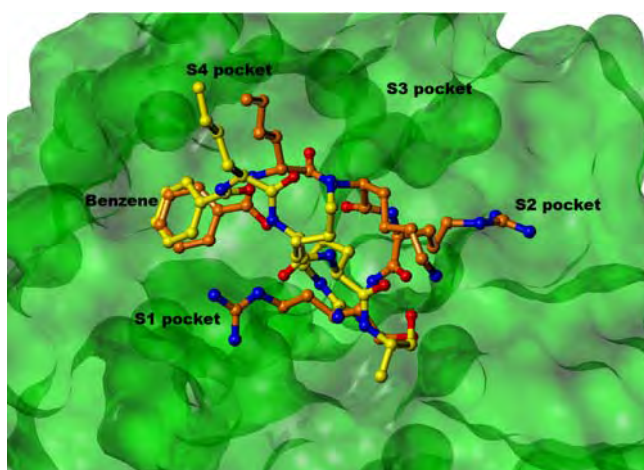


Figure 3.11. Comparing the conformation of the DV protease bound inhibitor Bz-Nle-Lys-Arg-Arg-H (orange) and the bound inhibitor Bz-Nle-Lys-Arg-Ala-H (yellow) at the end of the 5 ns simulation time.

3.2 MD simulation of DV NS2B/NS3 protease complexed with small-molecule inhibitors and Binding Free Energy Calculation

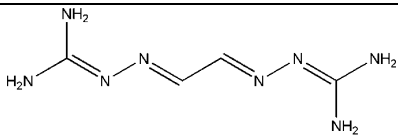
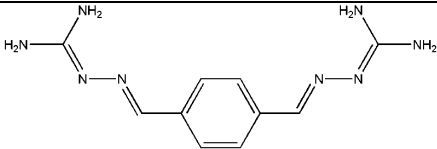
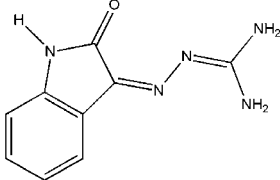
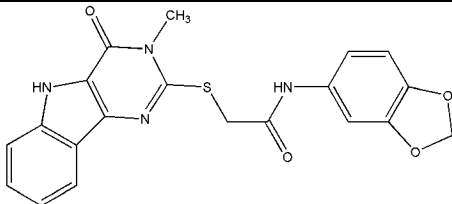
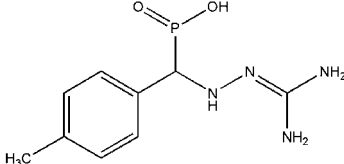
In the previous work, we reported the homology models of DV NS2B/NS3 protease and the MD simulations of this enzyme complex with a tetrapeptidic inhibitor (section 3.1)[131]. In this work, we present the MD simulations of complexes between small-molecule inhibitors with the DV NS2B/NS3 protease. The representative conformation of this enzyme is obtained from the former work [131].

The binding free energies of these compounds with the DV NS2B/NS3 protease are reported. In addition, the information about the interaction between these inhibitors with the enzyme, which was investigated by means of decomposition (DC) energy analysis, is given here.

3.2.1 Material and Computational Methods

3.2.1.1 Molecular docking of known small-molecule inhibitors

Table 3.2. Chemical Structures of the known small-molecule inhibitors and their K_i values against NS2B/NS3 protease of DV and WNV

Comp. ID	Structure	K_i Value (μM)	
		WNV NS2B/NS3 protease	DV NS2B/NS3 protease
1		337 ± 56	423 ± 50
2		35 ± 5	44 ± 5
3		16 ± 2	23 ± 2
4		37.3 ± 6.4	17.0 ± 4.3
5		13 ± 1	14 ± 2

Inhibitors representing the inhibitors of DV NS2B/NS3 protease and showing the binding affinities (K_i) at the micromolar concentration were selected from the literatures[39, 45]. Other compounds which show an inhibitory concentration lower than millimolar, or contain unclear stereochemistry were not included in this study.

The chemical structure of the selected compounds and their K_i values are shown in Table 3.2. These compounds were docked into the binding site of the DV NS2B/NS3 protease using the GOLD4.0 program[105, 106]. The default GOLD parameters were used except that the ‘flood fill radius’ was set to 20 Å around Asn152 of NS3. All types of constraints (H-bond or hydrophobic constraint) were not applied in this study.

3.2.1.2 Molecular Dynamics (MD) Simulation

The docking solutions of known inhibitors and hit compounds retrieved from virtual screening were subjected to MD simulation using the same parameters for all systems. In this study, MD simulations were performed using the AMBER9 program[132]. The AMBER1999SB force field[133] was applied for the protein, whereas the general Amber force field (GAFF)[134] and the restrained electrostatic potential (RESP)[135] charges were assigned for ligands. The RESP charges were calculated at the Hartee-Fock (HF) level of theory with the basis set 6-31G* (HF/6-31G*) using Gaussian03 program[136]. The protein-ligand complexes were then solvated by the explicit TIP3P[137] water molecules in the radius of 9 Å from the molecular surface to the edge of a simulation box (an octahedral box). In order to neutralize the system, counter ions (Na^+ or Cl^-) were added by replacing the water molecules in each system. Prior to performing the MD simulation, two different energy minimization processes have to be carried out in order to relieve bad steric and strain interaction which can cause problems during the MD simulation run. In the first process, the energy of the system was minimized by applying 1,000 steps of steepest descent and followed by 2,000 steps of conjugate gradient. In this process, a protein-ligand complex structure was kept fix by the force constant 500 kcal/mol, whereas the water molecules and counter ions were relaxed for adjusting their positions. Then in the second process, the same protocols as in the first process were also applied, but all atoms in the system were not constrained during the energy minimization process.

After finishing the energy minimization process, the MD simulation was performed by starting at the position-restrained phase which was carried out through the first 100 ps. In this step, the positions of the protein and ligands were restrained with the weak force constant 10 kcal/mol. The temperature was gradually increased

from 0 to 300 K during the first few picoseconds, and then it is kept constant at 300 K by using Langevin dynamics[138] with a collision frequency of 1 ps^{-1} . The constant volume periodic boundary was set. After finishing this step, free molecular dynamics simulations were performed using the NPT ensemble. The free MD simulations were completed by dividing into two phases: the equilibrated phase was carried out from 100 ps – 1ns, the production run was performed from 1ns – 6ns. The same parameters as following were applied in all MD simulations. A time step of 2 fs with SHAKE algorithm[139] to constrain all bond involving hydrogens was used. A cut-off radius was set at 10 \AA for the non-bonded interaction. The Particle-Mesh-Ewald (PME)[140] method was used to treat the electrostatic interactions. Pressure was maintained at 1 bar by an isotropic pressure scaling method using a pressure relaxation time of 2 ps. Temperature was kept fixed at 300 K using the same protocols as in the position-restrained phase. The convergence of the energies, temperature, pressure and global root mean square deviation (RMSD) were checked for indicating the stability of the system.

3.2.1.3 Binding Free Energy Calculation (MM/PBSA)

MD trajectories of each species (complex, protein and ligand) were extracted every 50 ps during 4-6 ns of the complex simulation. These trajectories, totaling 40 snapshots for each species, were used for the binding free energy calculation using MM/PBSA approach. The details of the MM/PBSA method are explained in section 2.4.1. The molecular mechanic energy (the van der Waals (ΔE_{vdW}), the electrostatic (ΔE_{ele}) and the internal energy (E_{int}) interaction in the gas phase) was calculated by using SANDER module in AMBER9. The $\Delta G_{\text{ele-sol}}$ (electrostatic free energy of solvation) was computed by using a numerical solvation of the Poisson-Boltzmann (PB) equation as implemented in the *pbsa* program[141] in AMBER9. All default parameters for the PB solver (which are the grid spacing at 0.5 \AA , dielectric constant of 1.0 and 80.0 for solute and solvent, respectively, and ionic strength at 0 M concentration) were used. The non-electrostatic free energy of solvation ($\Delta G_{\text{nonele-sol}}$) was calculated using the linear function of the solvent accessible surface area (SASA)[142] ($\Delta G_{\text{nonele-sol}} = \gamma \text{SASA} + b$). The SASA was calculated using the solvent probe radius at 1.4 \AA , whereas the default values of γ ($0.00542 \text{ kcal/mol} \cdot \text{\AA}^2$) and b

(0.92 kcal/mol) referring to the experimental solvation parameter were applied. The temperature of the system is set at 300 K.

The entropy of the system (S) is derived by the summation of the entropic contribution from translation (S_{trans}), rotation (S_{rot}), and vibration motion (S_{vib}). S_{trans} and S_{rot} are obtained from classical statistical mechanics, whereas S_{vib} is derived from a normal mode analysis implemented in the NMODE module in Amber 9. The energy of each species (complex, protein, and ligand) was minimized by the SANDER module using a conjugate gradient method. The energy is converged when the root-mean-square-deviation of the elements of the gradient vector was less than 10^{-4} kcal/mol $^{-1}$ Å $^{-1}$. Then, the normal mode analysis was performed.

Decomposition (DC) energy per residue was calculated by using the MM/GBSA approach which means that the Generalized Bond (GB) was used instead of the PB method for calculating the free energy of solvation.

3.2.2 Results and Discussion

3.2.2.1 Binding Free Energy of known inhibitors

Table 3.3. Components of MM/PBSA and binding free energy of known inhibitors

	Compound 1	Compound 2	Compound 3	Compound 4	Compound 5
K_i (μM)	423 \pm 50	44 \pm 5	23 \pm 2	17 \pm 4.3	14 \pm 2
ΔE_{ele}	-125.54	-157.86	-91.43	-14.84	-55.51
ΔE_{vdW}	-17.22	-22.33	-24.96	-32.32	-25.78
ΔE_{MM}	-142.76	-180.19	-116.39	-47.16	-81.29
$\Delta G_{\text{nonele-sol}}$	-1.18	-1.23	-1.19	-1.29	-1.2
$\Delta G_{\text{ele-sol}}$	136.28	171.42	103.57	31.37	68.03
ΔG_{sol}	135.1	170.19	102.38	30.08	66.84
ΔH_{tot}	-7.66	-10	-14.01	-17.09	-14.45
TAS	-18.98	-18.74	-18.13	-17.43	-18.09

The binding free energy of these inhibitors and the components of the binding free energy are summarized in Table 3.3. Results revealed that the order of the enthalpy value (ΔH_{tot}) is compound 1 (-7.66 kcal/mol) > compound 2 (-10.00

kcal/mol) > compound 3 (-14.01 kcal/mol) > compound 5 (-14.45 kcal/mol) > compound 4 (-17.09 kcal/mol). On the other hand, the experimental binding affinities of these inhibitors are in the order as compound 1 ($423 \pm 50 \mu\text{M}$) > compound 2 ($44 \pm 5 \mu\text{M}$) > compound 3 ($23 \pm 2 \mu\text{M}$) > compound 4 ($17 \pm 4.3 \mu\text{M}$) > compound 5 ($14 \pm 2 \mu\text{M}$). By considering the fact that the binding affinity of each compound has also some deviation, therefore the binding affinity of compound 3 can be considered to be active as well as compound 4 and compound 5. Thus, it can be concluded that the relative binding free energies of these compounds showed a satisfied correlation with their experimental values. It is clearly shown in Table 2 that even though the chemical structure of compound 4 is not similar to the other compounds, the entropy value of this compound (-17.43 kcal/mol) is quite close to the values of the other compounds (~ -18 to -19 kcal/mol). These results indicated that nearly the same amount of entropy contributes to the absolute binding free energy, implying that the entropy values of these compounds are not significantly different to change the trend of the relative binding free energy. Therefore, the relative binding free energy can be approximated by considering only the enthalpy term, $\Delta G_{\text{binding}} \approx \Delta H_{\text{tot}}$. Since the entropy calculated from the normal mode analysis requires high computational demand, the relative binding free energy of the hit compounds retrieved from virtual screening can be obtained by considering only the enthalpy term. This will help to reduce time and cost for the calculation. Moreover, these results indicate that the selection of protein conformation and the binding mode of these inhibitors are reliable.

3.2.2.2 Decomposition (DC) Energy per residue Analysis

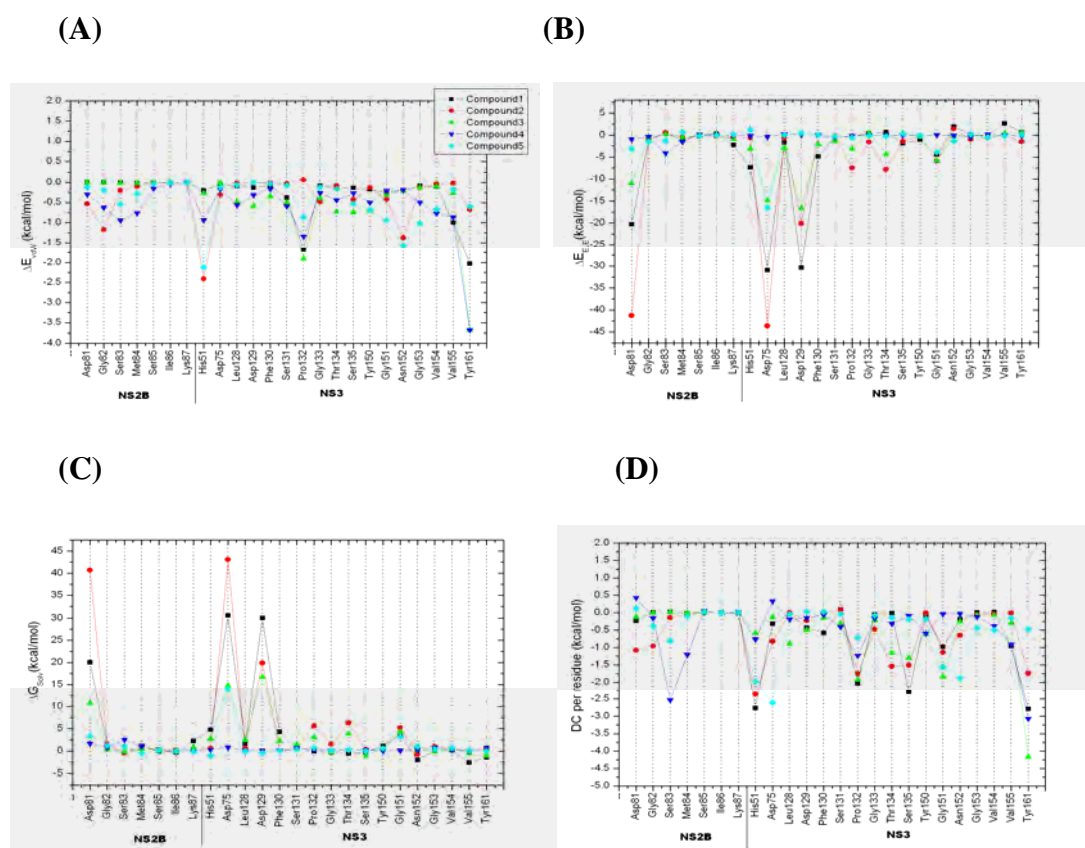


Figure 3.12. Footprints of (A) van der Waals interaction, (B) electrostatic interaction, (C) free energy of solvation and (D) total interaction energy of ligand-protein interaction per residue.

The decomposition (DC) energy per residue as shown in Figure 3.12 (D) clearly demonstrates that the active compounds such as compound 4 and compound 5 can interact with residues from NS2B, such as Ser83 and Met84, and also form NS3 (His51, Asp75, Pro132, Gly151, Asn152 and Tyr161). In the case of compound 4, a neutral compound, the van der Waals interaction provides the most interaction between this compound and the enzyme as displayed in Figure 3.12 (A) and 3.12 (B), 3.12(C) and 3.12 (D). This compound also has an electrostatic interaction with Ser83 of NS2B as shown in Figure 3.12 (B). It is interesting to point out that Pro132 of NS3 located at the S1 pocket, which showed low percent H-bond occupation with the tetrapeptidic inhibitor during the simulation as discussed in the previous work (section 3.1)[131], can interact with all of these small-molecule inhibitors as demonstrated in

Figure 3.12 (D). The main contribution of the interaction energy between Pro132_NS3 with these inhibitors comes from the van der Waals interaction as displayed in Figure 3.12 (A). Thus, these derived results provide additional information from our previous work that Pro132_NS3 is one of the key residues at the S1 pocket for the van der Waals interaction with inhibitor.

Even though the compound 1 has an interaction with important residues such as His51, Ser 135 (two of the catalytic triad), Pro132, Gly151 and Tyr161 of NS3, this compound lacks the interaction with residues at the C-terminus of the NS2B as shown in Figure 3.12 (D). From our previous work (section 3.1), it showed that residues at the C-terminus of the NS2B, such as Asp81, Gly82, and Ser83, are also important for binding with the inhibitor. Therefore, the absence of the interaction between the compound 1 and these residues results in the low activity of this compound. The interaction between compound 2 with Asp81 and Gly82 of NS2B were found as shown in Figure 3.12 (D). Gly82 of NS2B plays an important role for stabilizing the binding pocket of NS3 protease by interacting with Asn152 of NS3. Therefore, the inhibitor, such as compound 2, which can interact with Gly82 of NS2B, can interrupt the activity of the NS3 protease resulting in being as a high active compound. Thus, compound 2 is approximately 10 times more active than compound 1. Surprisingly, DC energy reveals that compound 3, which is quite active ($K_i = 23 \pm 2 \mu\text{M}$), has no interactions with any residue at the C-terminus of the NS2B as demonstrated in Figure 3.12 (D). However, by considering each element (van der Waals, electrostatic interaction and free energy of solvation as shown in Figure 3.12 (A), 3.12 (B) and 3.12 (C), respectively) that contributes to the total energy, it is found that compound 3 has a strong electrostatic interaction with Asp81 of NS2B, Asp75 and Asp129 of NS3 as shown in Figure 3.12 (B), and also a strong van der Waals interaction with residue at the S1 pocket such as Asp129, Pro132, Thr134 and Ser135 of NS3. In addition, from Figure 3.12 (D), it is obviously that compound 3 has the highest interaction with Gly151 and Tyr161 of NS3. From the previous work (section 3.1), a stable H-bond pair (nearly 100% of percent H-bond occupancy) between Tyr161_NS3 with the peptidic inhibitor was observed. Thus, Tyr161_NS3 is important for interacting with inhibitor. The interaction of compound 3 with Gly151_NS3 and also a strong electrostatic interaction with Asp81_NS2B can disturb the Asn82_NS2B –

Asn152_NS3 pair interaction (a main interaction for stabilizing the binding pocket). Therefore, compound 3 is quite active because of the interaction with Asp81_NS2B, Gly151 and Tyr161 of NS3 which can affect the stability of the binding pocket and the activity of the enzyme.

From Figure 3.12 (A) – (D), it can be concluded that the van der Waals interaction as well as the electrostatic interaction are important for interacting with small-molecule inhibitors. These derived results support our former work indicating that residues located at the C-terminus of NS2B as well as residue from NS3 are important for interacting with inhibitors (both peptidic inhibitors and also small-molecule inhibitors). In addition, residues which play important role for interacting with inhibitor are also addressed.

3.3 Virtual Screening for inhibitors of Dengue Virus NS2B/NS3 protease

A stepwise virtual screening (VS) was carried out by starting from using three different pharmacophore models. First, the docking solution of the most selective compound (compound 4) together with the GRID field calculation was used to generate a structure-based pharmacophore model. A second pharmacophore model (a static model) was built by directly using the docking pose of this compound. The last pharmacophore model, a dynamic model, was created by using a representative conformation of this compound in complex with the DV NS2B/NS3 protease. The static pharmacophore model was created by a conventional method, using a single crystal structure or the complex obtained from the molecular docking. The approach has some disadvantages because the flexibility of ligand and protein is not taken into account. Thus, the dynamic pharmacophore model has been developed to resolve this problem and this approach was successful in applying in virtual screening[143-145]. The dynamic pharmacophore model was generated by using a series of representative conformations of the ligand-protein complex derived from different snapshots of MD simulation. Consequently, all of these three pharmacophore models (the structure-based, the static and the dynamic model) were applied for searching for compounds in databases that match with these models. The number of compounds was subsequently filtered by applying drug-like properties and molecular docking. Hits compounds were then visually inspected and selected for calculating binding free energy. The

derived binding free energy of the hit compounds were compared with the values of the known inhibitors. Finally, lead compounds as potent inhibitors for NS2B/NS3 protease of DV are suggested for testing biological activities.

3.3.1 Material and Computational Methods

3.3.1.1 Databases

Muti-conformational databases (Chembridge, NCI, Maybridge and drug-like ZINC database) were screened for potent hit compounds for the DV NS2B/NS3 protease. Numbers of compounds and total conformations, indicated in the parenthesis of each database, are as following, Chembridge 405,494 compounds (41,829,578 conformations), NCI 884,439 compounds (55,874,026 conformations), Maybridge 102,419 compounds (2,463,388 conformations) and drug-like subset of ZINC database 1,697,950 compounds (168,586,567 conformations).

3.3.1.2 Pharmacophore Model and Pharmacophore Search

Virtual screening was carried out in a hierarchical strategy starting from the pharmacophore search in these focused databases. Even though compound 4 is not the most active compound against the DV NS2B/NS3 protease among these potent inhibitors, it is the most selective compound for the DV NS2B/NS3 protease because it is two-fold effective against this enzyme of DV than that of WNV. This is in contrast to compound 5 which is the most active compound against the DV NS2B/NS3 protease but less selective, nearly the same activity for the DV and the WNV NS2B/NS3 protease. Therefore, the docking solution of compound 4 was applied for generating pharmacophore models. The structure-based pharmacophore model was built by using the docking solution of compound 4 together with the GRID field interaction. The hydroxyl (OH) probe and N-sp² with lone pair (N:=) probe, implying the H-bond donor areas and H-bond acceptor regions, respectively, were used for the calculations. On the other hand, the docking solution of compound 4 was directly applied to generate the static pharmacophore model using the LigandScout software[146]. LigandScout is a program for developing a structure-based pharmacophore model based on protein-ligand structures which could be obtained from the X-ray structure or docking solution. Beside the important features of ligands

(i.e., H-bond donor, H-bond acceptor, and hydrophobic), excluded volumes on the basis of the surrounding amino acid residues were also generated. LigandScout generated the excluded volumes on the basis of C α atoms. The pharmacophore was exported as a PH4 file and imported into the MOE2008.10 program[112] for performing pharmacophore searching.

The dynamic pharmacophore model was created by using snapshots from 4-6 ns of the MD simulation of compound 4. These snapshots were first clustered into ten groups by using the K-mean clustering analysis algorithm. The representative conformation of each cluster was then used to generate the static pharmacophore model. Then, ten different static pharmacophore models were obtained and were subsequently projected onto the binding pocket of the enzyme. Finally, the dynamic pharmacophore model was built by creating features to cover all sub-features derived from these ten static pharmacophore models. The derived pharmacophore models were applied for searching the potential hits in the multi-conformational database using pharmacophore search module implemented in MOE2008.10

3.3.1.3 Database Filtration

Due to multi-conformational databases were already generated and stored in our clusters, we first started the virtual screening by applying a pharmacophore search approach. This is a fast technique for screening compounds in the databases. Then, a number of hit compounds in databases were filtered by using the properties of drug-like compounds. Lipinski's rule of five [101], stating that the orally active drug should have these properties; H-bond donor < 5, H-bond acceptor < 10, molecular weight < 500, and logP < 5, was applied for this objective. In addition, the molecules containing the reactive group, based on the reactive group descriptor in the QuaSAR module implemented in MOE, were also removed from the databases.

3.3.1.4 Molecular Docking and Energy Minimization of docking solution

The hits derived from pharmacophore search and database filtration were subsequently subjected to the molecular docking calculation using two consecutive docking runs. The first docking run was carried out by using 10,000 operations of the Genetic Algorithm (GA) run. This protocol speeds up the docking calculation for the

large amount of compounds in a reasonable computational time. However, the binding mode of compounds is less accurate. Therefore, the top 200 compounds ranked by GoldScore from the first step of the docking run were consequently subjected to the second step of the docking run. In this step, 100,000 operations of the GA run and the search efficiency of 100% were used. This protocol leads to a more accurate binding mode of compounds. The 'flood fill radius' of 20 Å around Asn152 of NS3 was used and no constraints were applied.

After the second step of molecular docking was completed, the first rank (the highest GoldScore among ten solutions) of each compound was selected. Then, the docking solutions of these 200 compounds were refined by employing energy minimization using the MOE2008.10 program. AM1-BCC charges were assigned for all ligands whereas the Amber force field 99 charge were applied for the DV NS2B/NS3 protease. During the energy minimization process, the DV NS2B/NS3 protease was constrained by tethering heavy atoms by the force constant 100 kcal/mol whereas the ligand structures were relaxed. The energy-minimized docked structures were subsequently visually inspected for ligand interaction compared to the DC (decomposition) energy analysis (section 3.2.2.2) and ligand-protein interaction derived from our previous MD simulation (section 3.1.2.2.3). Finally, the hit compounds were selected. The complexes of these hit compounds with DV NS2B/NS3 protease were subjected to MD simulation and binding free energy calculation using the MM/PBSA method. The derived relative binding free energies of the hit compounds were compared to the values of known inhibitors.

3.3.1.5 MD simulation and Binding Free Energy Calculation

MD simulations of the hit compounds complex with DV NS2B/NS3 protease were employed using the same protocols as described in the section 3.2.1.2. In addition, binding free energies of these hits were also computed using the MM/PBSA method and the same protocols as in section 3.2.1.3.

3.3.2 Results and Discussion

3.3.2.1 Structure-based Pharmacophore Model

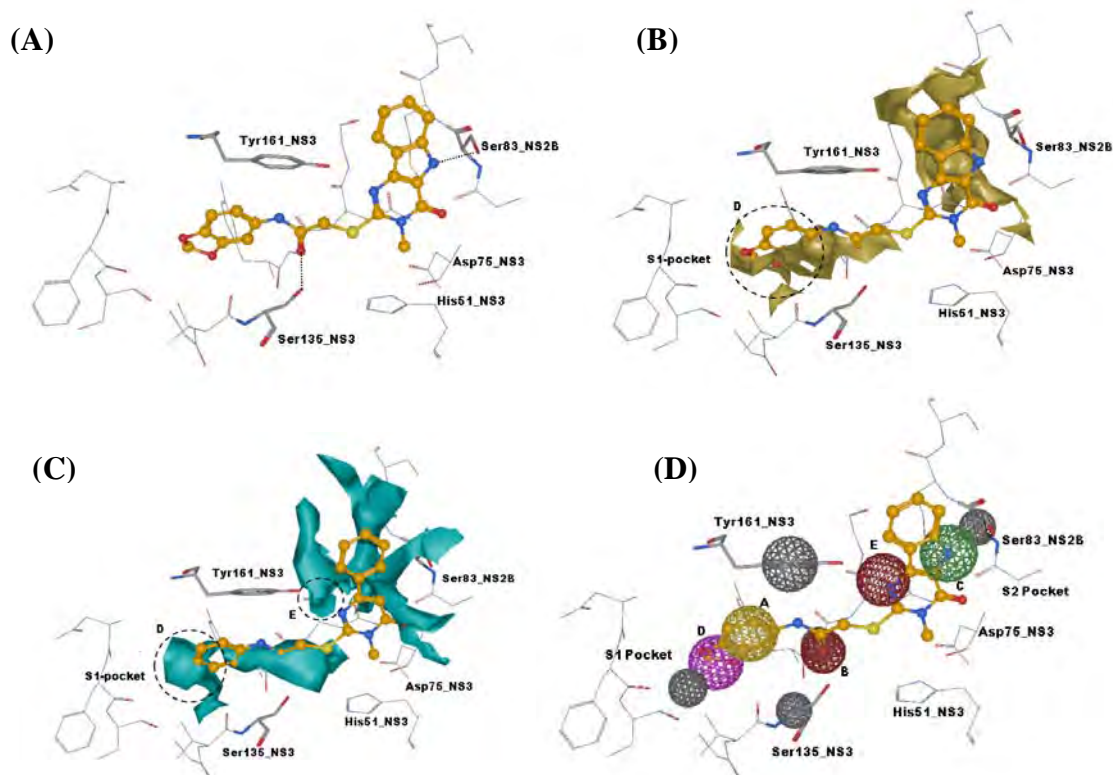


Figure 3.13. (A) Molecular docking solution of compound 4 (orange balls and stick) with the DV NS2B/NS3 protease. H-bonds are shown as dot lines. (B) GRID result of OH probe (yellow) at the contour level -5.5 kcal/mol, (C) GRID result of N:= probe (N-sp² with lone pair) (cyan) at the contour level -4.0 kcal/mol and (D) The final pharmacophore model derived from docking solution of compound 4 (sphere A, B and C), OH contours of GRID field calculation (sphere D) and N:= contours (sphere D and E). Red spheres represent the H-bond acceptor areas, Green sphere means H-bond donor areas and, yellow sphere indicates hydrophobic areas and magenta sphere denotes H-bond acceptor or donor area.

Comparison the activity of these compounds for DV and WNV NS2B/NS3 protease, it was found that compound 4 is the most selective compound for DV NS2B/NS3 protease among these five compounds. K_i value of compound 4 against this enzyme of WNV is $37.2 \pm 6.4 \mu\text{M}$ which is approximate two times higher than the K_i value of DV ($17.0 \pm 4.3 \mu\text{M}$). While, The K_i value of the other compounds

against NS2B/NS3 protease of DV and WNV are nearly the same. Therefore, the docking solution and the MD simulation of compound 4 were selected to generate pharmacophore models. Docking solution of compound 4 (Figure 3.13 (A)) into the binding pocket of DV NS2B/NS3 protease shows H-bond between this inhibitor with Ser83_NS2B and Ser135_NS3 (one of the catalytic triad). In addition, π - π interaction between the benzene ring of inhibitor with side chain of the Tyr161_NS3 was also observed. Based on the docking result, three features of the structure-based pharmacophore model was created as shown in Figure 3.13 (D). First, The yellow sphere (region A), which means the hydrophobic feature, was created to demonstrate the π - π interaction between the inhibitor and Tyr161_NS3. Second, the dark red sphere representing the H-bond acceptor feature was built at the area (region B) where the compound forms H-bond with Ser135_NS3. Lastly, the dark green sphere (region C) illustrating for the H-bond donor feature was generated to represent the H-bond between the compound and Ser83_NS2B.

From the MD simulations of the previous works (section 3.1 and section 3.2), results indicate that the interaction between the S1 pocket of enzyme with the P1 subsite of inhibitor is one of the important protein-ligand interactions. Moreover, the H-bond between the inhibitor with Tyr161_NS3 is quite stable during the simulation, indicating by the high percent H-bond occupancy (nearly 100%) between the peptidic inhibitor with this residue. This result suggested that the H-bond interaction between inhibitor with Tyr161_NS3 is essential. These interactions were not found in the docking pose of compound 4. Therefore, the GRID field calculation using hydroxyl (OH) probe and N-sp2 with lone pair probe (N:=) probe was additionally computed. The OH and N:= probe indicate the H-bond donor and H-bond acceptor region, respectively. Thus, the pharmacophore features representing the interaction at the S1 pocket and also with Tyr161_NS3 residue can be generated by applying these probes. The contours of GRID these field calculations, which were projected onto the binding pocket of DV NS2B/NS3 protease and also compared with the docking solution of compound 4, are displayed in Figure 3.13 (B), H probe, and Figure 3.13 (C), N:= probes. Results revealed that the OH and N:= probe were found at S1 pocket of the enzyme. These results indicated that inhibitors should contain H-bond donor or acceptor features at this region for interacting with residues at the S1 pocket of the

enzyme. Thus, the magenta sphere (region D as shown in Figure 3.13 (D)) was generated to represent H-bond donor or acceptor at this pocket. The N:= probe was also observed between the pyrimidine ring and Tyr161_NS3. This result implied that H-bond interaction between inhibitors with side chain of Tyr161_NS3 is one of the essential interactions between inhibitors and this residue. Thus, inhibitors should have a feature that they can form H-bond with the side chain of Tyr161_NS3. Consequently, the dark red sphere at region E (displayed in Figure 3.12 (D)) was generated to demonstrate the H-bond acceptor feature of inhibitors. In addition, exclude volumes displayed as gray contours were also generated. Exclude volume which were generated on the basis C α atom relates with the conformation of the protein and surrounding amino acid residues. To clearly display, only some of excluded volumes are shown. Then, this structure-based pharmacophore model, containing five pharmacophore features, was applied to screen compounds from the multi-conformation database.

3.3.2.1.1 Virtual Screening and Binding Free Energy Calculation

Table 3.4. Number of Hits derived from each steps of virtual screening using the structure-based pharmacophore model

Number of Hits derived from each step	Database			
	Chembridge	Maybridge	NCI	ZINC (drug-like subset)
Pharmacophore Search	9,899	1,247	128,194	23,168
Drug like properites	5,826	588	11,170	14,972
Docking Step 1	5,826	588	11,170	14,972
Docking step 2	200	200	200	200
MM/PBSA calculation	5	5	5	5

Results, which were derived from virtual screening, are summarized in Table 3.4. These results demonstrated that 9,899 (2.44%) compounds, 1,247 (1.22%) compounds, 128,194 (14.49%) compounds and 23,168 (1.36%) compounds, which were retrieved from Chembridge, Maybridge NCI, and drug-like ZINC database, respectively, are in agreement with the structure-based pharmacophore model. The hit compounds should also follow Lipinski's rule of five and they should contain no reactive groups. Thus, compounds that violate these properties were filtered out. By

applying these filtering, 5,826, 588, 11,170 and 14,972 compounds were obtained from Chembridge, Maybridge, NCI, and drug-like ZINC database, respectively. These hits were further docked into the representative conformation of DV NS2B/NS3 protease. Molecular docking of these hit compounds was carried out in two consecutive steps as explained in the method part (section 3.3.1.4). After finishing the second docking run, the top-ranked docking solution (the highest docking score) of each compound was selected. After that, the energy of these selected compounds complex with the enzyme was minimized in the partial fix of binding pocket of the enzyme. The results of energy-minimized docked structure were then visually inspected. The hit compounds, which show interaction with important residues, were selected for performing MD simulation and calculating binding free energy using MM/PBSA approach. The essential residues for interacting with inhibitor include residues at the C-terminus of NS2B, residues located at the S1, S2 pocket and also Gly151, Asn152, Gly153 and Tyr161 of NS3. Finally, 5 compounds from each database were selected as potential hit compounds. Thus, the total number of selected hit compounds was 20 compounds. These compounds in complex with DV NS2B/NS3 protease were subjected to MD simulation and binding free energy calculation. Table 3.5 shows the chemical structures of these hit compounds and also their binding free energies. As discussed in section 3.2.2.1, the relative binding free energy can be approximated by considering only the enthalpy term. Thus, these values of these hits were computed only the enthalpy term, and the entropy values were neglected. The binding free energies ($\Delta G_{binding}$) of these hits were compared with these values of known inhibitors. As shown in Table 3.5, compound **6641412**, **7658769**, **JFD02062**, **ZINC02069320** and **ZINC05693897** gave the binding free energy lower than the lowest binding free energy (-17.09 kcal/mol) of the known inhibitor (compound 4). Thus, these compounds could be potent as the known inhibitors and they are suggested for testing their binding affinities for DV NS2B/NS3 protease.

Table 3.5. $\Delta G_{\text{binding}}$ of the hit compounds derived from virtual screening using the structure-based pharmacophore model

Cpd.-ID (Database)	Chemical Structure	ΔE_{ele}	ΔE_{vdW}	$\Delta G_{\text{nonele-sol}}$	$\Delta G_{\text{ele-sol}}$	$\Delta G_{\text{binding}} \approx \Delta H_{\text{tot}}$
6050086 (Chembridge)		-31.53	-32.19	-1.24	48.44	-16.53
6641412 (Chembridge)		-44.26	-38.53	-1.33	51.60	-32.53
6691885 (Chembridge)		-10.28	-42.12	-1.39	40.92	-12.88
6984064 (Chembridge)		-29.89	-25.69	-1.23	45.02	-11.79
7658769 (Chembridge)		-43.50	-47.96	-1.34	52.56	-40.24
BTB05499 (Maybridge)		-177.50	-35.12	-1.28	204.82	-9.07

Table 3.5. (Continue)

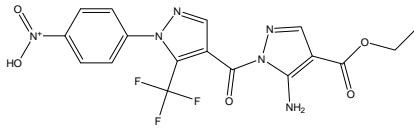
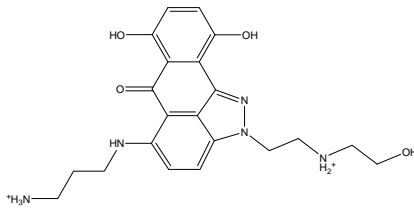
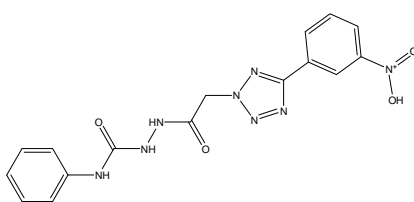
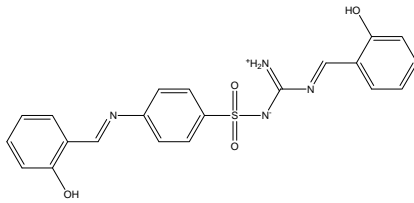
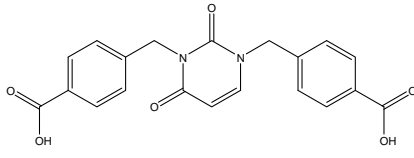
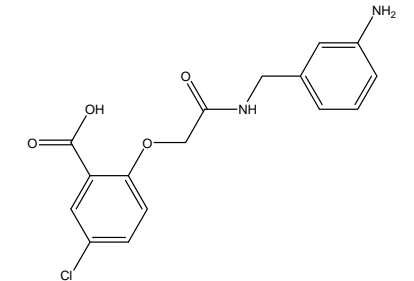
GK03608 (Maybridge)		-107.06	-44.09	-1.33	138.19	-14.30
JFD02062 (Maybridge)		-203.84	-28.51	-1.27	211.74	-21.88
RJF01462 (Maybridge)		-58.45	-42.67	-1.31	101.49	-0.94
S04863 (Maybridge)		-25.28	-23.18	-1.18	38.43	-11.20
161070 (NCI)		-42.12	-11.51	-1.12	44.27	-10.49
212126 (NCI)		6.08	-17.86	-1.17	3.17	-9.78

Table 3.5. (Continue)

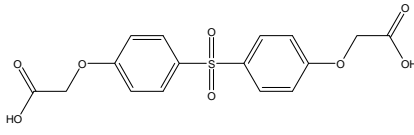
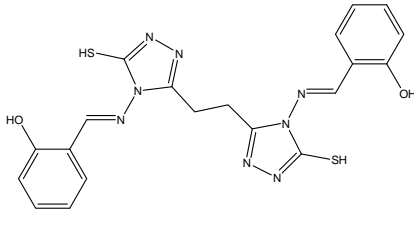
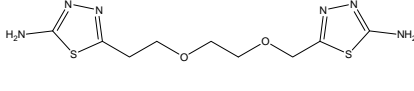
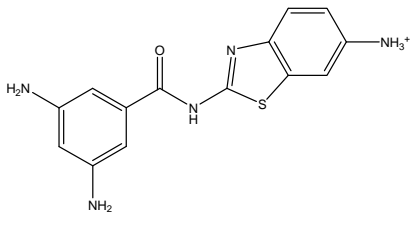
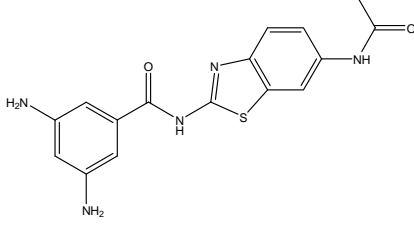
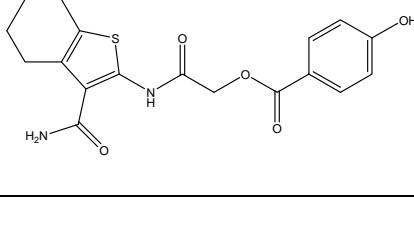
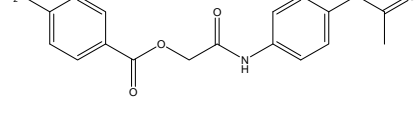
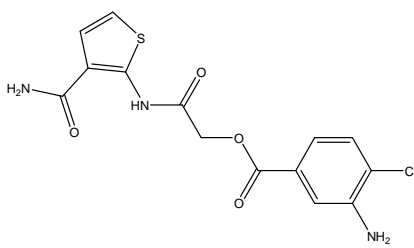
39645 (NCI)		4.41	-17.29	-1.15	7.16	-6.87
713329 (NCI)		-17.15	-38.53	-1.28	44.73	-12.23
720751 (NCI)		-15.64	-30.22	-1.25	32.04	-15.08
ZINC00328297 (ZINC)		-60.44	-24.87	-1.22	78.69	-7.84
ZINC02069320 (ZINC)		-134.95	-32.72	-1.28	135.53	-33.41
ZINC05693897 (ZINC)		-16.76	-35.16	-1.31	35.00	-18.23
ZINC05709590 (ZINC)		-91.73	-15.31	-1.15	101.40	-6.79

Table 3.5. (Continue)

ZINC06421909 (ZINC)		-15.22	-23.27	-1.19	28.45	-11.33
------------------------	-----------------------------------------------------------------------------------	--------	--------	-------	-------	--------

Hits which gave enthalpy values (ΔH_{tot}) lower than the lowest enthalpy value of the known inhibitor (-17.09 kcal/mol) are highlighted in red.

3.3.2.2 Comparison between Static and Dynamic Pharmacophore model

3.3.2.2.1 Pharmacophore Models (static and dynamic model) Generation

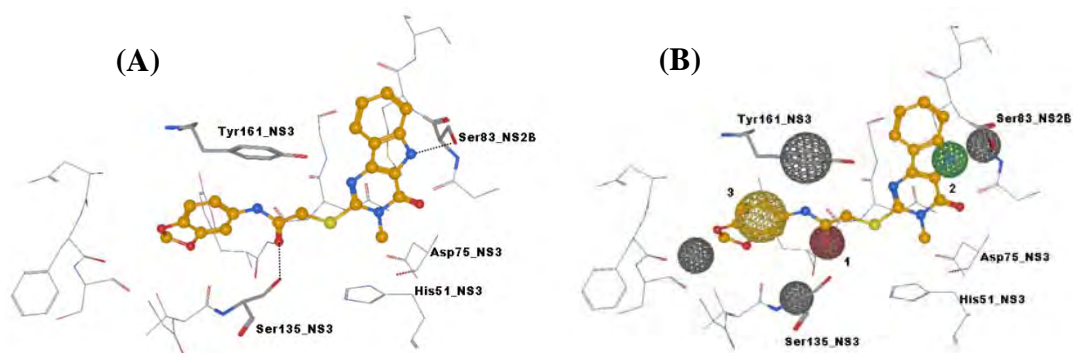


Figure 3.14. (A) Docking solution of compound 4; H-bonds are shown as dot lines and ligand is displayed as orange stick. (B) A static pharmacophore model (PH-1); Pharmacophore features are displayed as yellow sphere = Hydrophobic feature, dark red sphere = H-bond acceptor features, dark green sphere = H-bond donor features and gray = excluded volumes in all pictures. For clearly display, some exclude volume spheres were omitted.

The static pharmacophore model, which is displayed in Figure 3.14 (A), was generated based solely on the docking solution of compound 4. Three pharmacophore features as displayed in Figure 3.14 (B) were generated to represent the static pharmacophore model. The details for generating these features and this model were already explained in section 3.3.2.1. This derived static pharmacophore model (displayed in Figure 3.14 (B)) is named as PH-1.

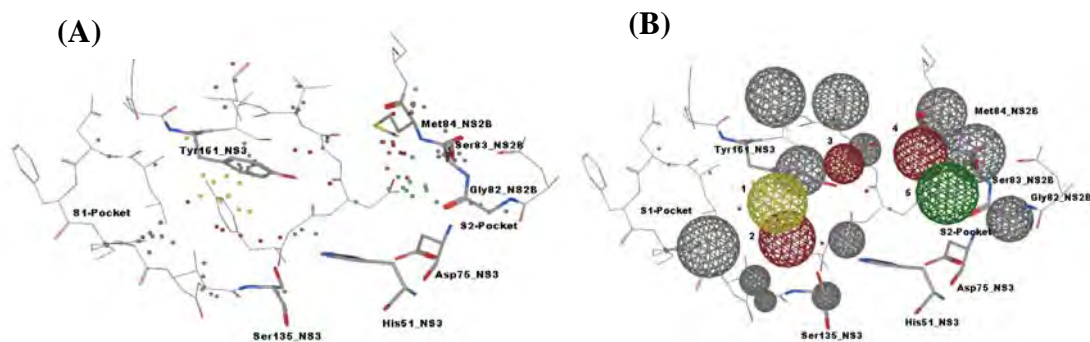


Figure 3.15. (A) All ten different static pharmacophore models derived from each representative conformation of each cluster projected on the binding pocket of DV NS2B/NS3 protease, and (B) dynamic pharmacophore model; Pharmacophore features are displayed as yellow sphere = Hydrophobic feature, dark red sphere = H-bond acceptor features, dark green sphere = H-bond donor features and gray = excluded volumes in all pictures. For clearly display, some exclude volume spheres were omitted.

A dynamic pharmacophore model was generated by taking trajectories during 4 ns to 6 ns of the MD simulation of compound 4. These trajectories were first clustered and a representative complexed structure of each cluster was then applied to create a static pharmacophore model using the same method and the same program (LigandScout) as for generating the PH-1. Ten different static pharmacophore models were obtained. These ten static pharmacophore models were then projected on the binding pocket of the enzyme. All points as shown in Figure 3.15 (A) represent these ten static pharmacophore models. Five pharmacophore features of a dynamic model, which is shown in Figure 3.15 (B), were then built as follows. A yellow sphere was built as the first feature by creating a sphere to cover all small yellow points near Tyr161 of NS3. A yellow sphere and points represent a hydrophobic feature implying the π - π interaction between the inhibitor and the side chain of Tyr161_NS3. Second, a dark red sphere was created to cover all small red points near Ser135. This feature represents a H-bond acceptor of the inhibitor with this residue. Third, a small dark sphere was made and this sphere corresponds to a H-bond acceptor between the inhibitor with the side chain of Tyr161 of NS3. Fourth, the dark red sphere

representing a H-bond acceptor feature between inhibitor with Ser83 and Met84 of NS2B was built. Finally, the last dynamic pharmacophore feature is illustrated by a dark green sphere close to Gly82 and Ser83 of NS3. This feature which is a H-bond donor feature represents a H-bond interaction between the inhibitor with these residues. The given name PH-2 is used to refer to the final dynamic pharmacophore model as displayed in Figure 3.15 (B).

3.3.2.2.2 Virtual Screening

Table 3.6. Number of hits derived from each steps of Virtual Screening using PH-1 and PH-2 model

Number of Hits in each step	Pharmacophore model 1 (PH-1)				Pharmacophore model 2 (PH-2)			
	Database				Database			
	CB	MB	NCI	ZINC	CB	MB	NCI	ZINC
PH Search	114,667	17,453	306,289	445,063	45,005	5,209	222,920	163,632
Similarity	6,038	4,469	125,170	4,155	7,534	2,307	119,205	6,986
drug like	3,005	1,503	2,355	2,385	3,771	1,103	1,039	3,791
Docking step 1	3,005	1,503	2,355	2,385	3,771	1,103	1,039	3,791
Docking step 2	200	200	200	200	200	200	200	200
MM/PBSA	2	3	3	2	2	3	3	2

CB = Chembridge and MB = Maybridge

Virtual screening (VS) was carried out as a stepwise approach. The number of compound in the multiconformational databases (Chembridge, Maybridge, NCI and ZINC) is first filtered by searching compounds in databases that match with the pharmacophore models (PH-1 and PH-2). Table 3.6 shows a number of compounds retrieved from each step of virtual screening method. From Table 3.6, it is obvious that the PH-2 performs better than PH-1 for reducing the amount of compounds in the databases. For instance, the numbers of compounds that pass with the PH-1 model are 114,667 (28.27%) compounds from Chembridge database, 17,453 (17.04%) compounds from Maybridge database, 306,289 (34.63%) compounds from NCI database, and 445,063 (26.21%) compounds from drug-like ZINC database. On the other hand, 45,005 (11.00%) compounds, 5,209 (5.09%) compounds, 222,920 (25.20%) compounds and 163,632 (9.64%) compounds, which were retrieved from Chembridge, Maybridge NCI, and drug-like ZINC database, respectively, are the number of compounds that match with the PH-2 model. The high number of

compounds that pass with the PH-1 filter may also contain false-positive compounds. These results indicate that the dynamic pharmacophore model performs much better than the static pharmacophore model for filtering the number of compounds in databases which can also help to reduce the amount of false-positive compounds. The compounds that match with the PH-1 and PH-2 were subsequently screened by searching for the compounds that are similar to compound 4. Then, the drug-like properties according to the Lipinski's rule of five[101] was applied in order to reduce the number of the hit compounds in the databases. After this, these hit compounds were docked into the representative conformation of the DV NS2B/NS3 protease.

Molecular docking of these hit compounds, which are total 9,284 and 9,704 compounds in all databases that pass with PH-1 or PH-2, respectively, was performed in two consecutive steps. First, these hits were docked into the representative conformation of the DV NS2B/NS3 protease using the protocol for virtual screening as discussed in the method part (section 3.3.1.4). Then, the docking results of these hit compounds in this step were ranked according to their GoldScore. The top 200 compounds in each database were selected for the second docking step by applying the protocol that leads to obtaining more accurate binding modes as explained before in section 3.3.1.4. In both steps of molecular docking, ligands were allowed to be flexible, whereas the protein was kept fixed. Direct investigating into the ligand-protein interaction derived from molecular docking could lead to misinterpreted results. Thus, the docking solutions derived from the second step of the docking run were subsequently energy minimized in the partial fix of binding pocket of the enzyme. The results of the energy-minimized docking solution of these hit compounds were then visually inspected for the ligand interaction. Results from the ligand-protein interaction in section 3.1.2.2.3 and from the DC energy analysis (section 3.2.2.2) indicate that residues at the C-terminus of NS2B, His51_NS3, Asp75_NS3, residues at the S1 pocket (Leu128-Thr134 of NS3), Ser135, Gly151, Asn152, Gly153 and Tyr161 of NS3 play an important role for interacting with inhibitors. Therefore, hit compounds which show interaction with these residues were selected. Finally, hits (20 compounds-10 compounds from each of the pharmacophore searching (PH-1 and PH-2)) were derived. The chemical structures of hit compounds are shown in Figure 3.16 and Figure 3.17.

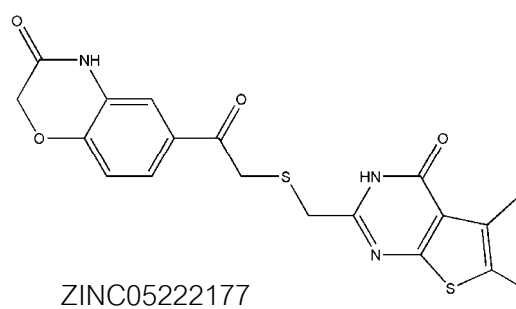
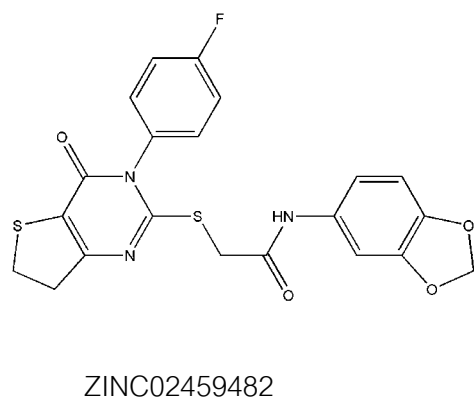
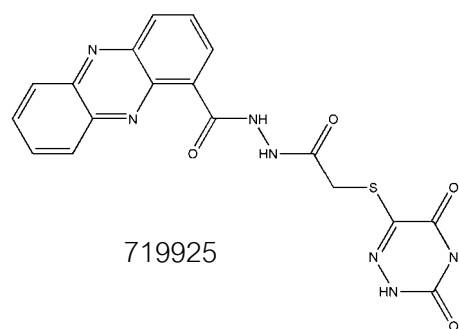
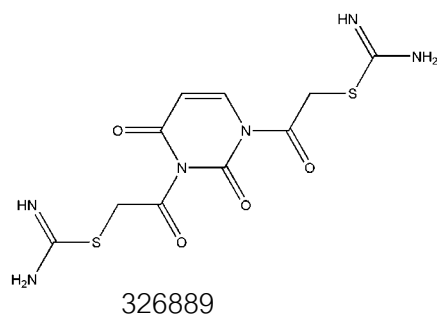
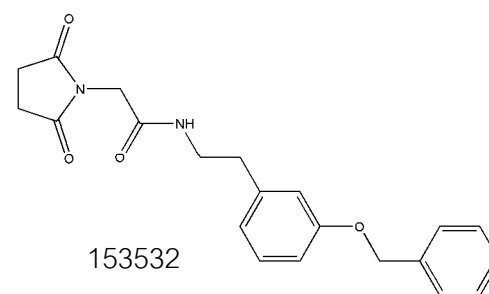
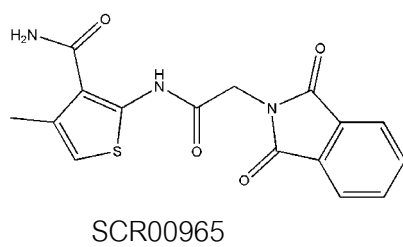
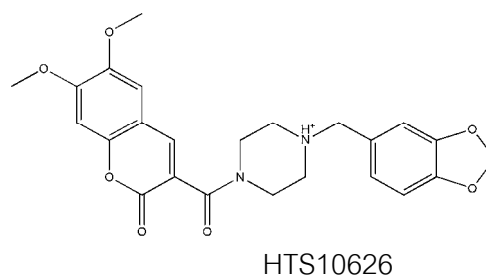
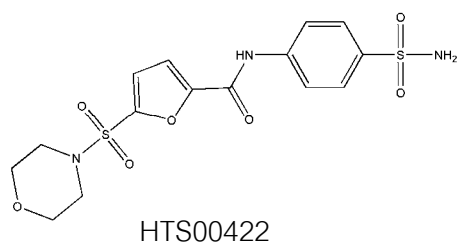
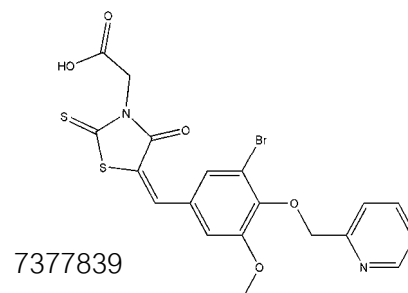
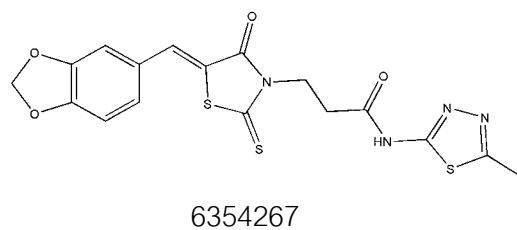
Figure 3.16. Chemical Structures of hit compounds derived from PH-1

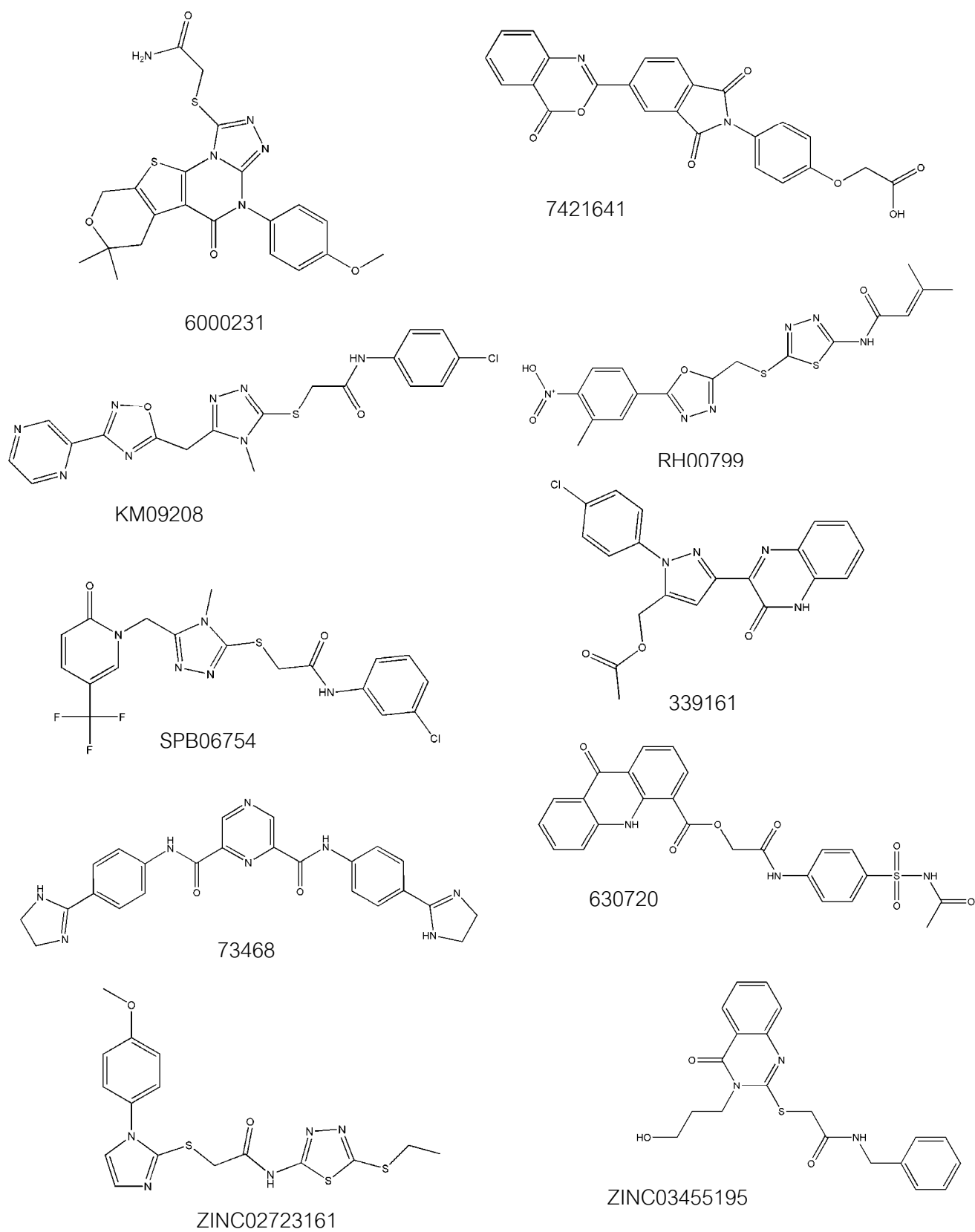
Figure 3.17. Chemical Structures of hit compounds derived from PH-2

Table 3.7. $\Delta G_{\text{binding}}$ of hit compounds derived from PH-1

Compound-ID	Database	ΔE_{ele}	ΔE_{vdW}	$\Delta G_{\text{nonele-sol}}$	$\Delta G_{\text{ele-sol}}$	$\Delta G_{\text{binding}} \approx \Delta H_{\text{tot}}$
6354267	Chembridge	-10.50	-44.55	-1.34	37.70	-18.70
7377839	Chembridge	-8.67	-39.51	-1.30	28.98	-20.50
HTS00422	Maybridge	-28.25	-33.21	-1.28	41.16	-21.58
HTS10626	Maybridge	-53.62	-43.59	-1.34	77.86	-20.69
SCR00965	Maybridge	-12.15	-25.68	-1.20	29.74	-9.30
153532	NCI	-19.31	-23.75	-1.23	42.85	-1.45
326889	NCI	-150.37	-24.82	-1.21	156.30	-20.11
719925	NCI	-32.17	-34.50	-1.28	48.40	-19.55
ZINC02459482	ZINC	-24.46	-52.02	-1.37	54.07	-23.78
ZINC05255177	ZINC	-10.08	-37.18	-1.28	27.27	-21.26

Red and bold indicate the compound giving binding free energy lower than the lowest enthalpy values of known inhibitors (-17.09 kcal/mol).

Table 3.8. $\Delta G_{\text{binding}}$ of hit compounds derived from PH-2

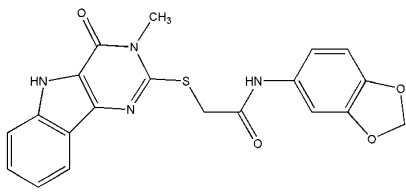
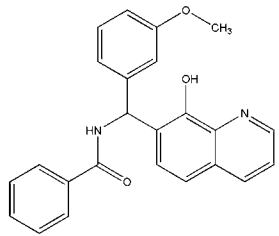
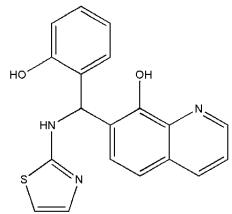
Compound-ID	Database	ΔE_{ele}	ΔE_{vdW}	$\Delta G_{\text{nonele-sol}}$	$\Delta G_{\text{ele-sol}}$	$\Delta G_{\text{binding}} \approx \Delta H_{\text{tot}}$
6000231	Chembridge	-18.28	-33.85	-1.26	41.39	-12.00
7421641	Chembridge	-18.72	-49.25	-1.33	42.44	-26.86
KM09208	Maybridge	-14.13	-35.75	-1.27	28.72	-22.43
RH00799	Maybridge	-139.97	-43.63	-1.35	163.63	-21.32
SPB06754	Maybridge	-19.64	-44.15	-1.37	40.07	-25.08
339161	NCI	-17.13	-42.85	-1.33	37.99	-23.31
630720	NCI	-8.52	-57.20	-1.41	33.70	-33.43
73468	NCI	-17.80	-53.36	-1.41	47.72	-24.85
ZINC02723161	ZINC	-22.26	-36.55	-1.33	42.06	-18.08
ZINC03455195	ZINC	-15.79	-31.10	-1.26	33.51	-14.64

Red and bold indicate the compound giving the enthalpy value lower than the lowest enthalpy values of known inhibitors (-17.09 kcal/mol).

Next, the complexes between hit compounds and the DV NS2B/NS3 protease were subjected to MD simulations and binding free energy calculations using the same protocols as applied for the known small-molecule inhibitors (section 3.2). As discussed before, the enthalpy term of the known inhibitors gave a good correlation with their experimental binding affinities. By including the entropy term that contributes to the binding free energy, the relative binding free energies of these compounds were not changed. In addition, the entropy calculation requires high computational time and cost. Therefore, the binding free energies of these hit compounds were calculated using only the enthalpy term. Then, the binding free energies of these hits were compared with these values of known inhibitors. Binding free energies of hits are summarized in Table 3.7 and 3.8. The compounds which gave

binding free energies lower than those of the known inhibitors are highlighted in bold red, indicating that these compounds may be potent inhibitors for the DV NS2B/NS3 protease. Moreover, results revealed that hits which were derived from PH-1 gave binding free energies in the range of -1.45 to -23.78 kcal/mol. On the other hand, the binding free energies of hit compounds, which were retrieved from PH-2, were in the range of -12.00 to -33.43 kcal/mol. These results pointed out that hit compounds retrieved from the dynamic pharmacophore gave binding free energy lower than those obtained from the static pharmacophore model, implying that hit compounds derived from the dynamic model can be more potent than those derived from the static model. The dynamic pharmacophore model is more advantageous than a static pharmacophore model because the flexibility of the protein and the ligand were also taken into account for generating a dynamic pharmacophore model, whereas this thing was not considered in a static pharmacophore model.

Table 3.9. Tanimoto Coefficient of hits compared to known small-molecule inhibitors. K_i values of these compounds against the DV NS2B/NS3 are also reported.

Compound name and their chemical structure			
			
	Compound 4	Compound 6	Compound 7
Compound-ID	$K_i = 17.0 \pm 4.3 \mu\text{M}$	$K_i = 28.6 \pm 5.1 \mu\text{M}$	$K_i = 30.2 \pm 8.6 \mu\text{M}$
Hits derived from PH-1			
6354267	0.75	0.36	0.42
7377839	0.66	0.46	0.44
HTS00422	0.50	0.29	0.32
HTS10626	0.60	0.45	0.35
SCR00965	0.65	0.45	0.48
153532	0.54	0.45	0.33
326889	0.55	0.34	0.35
719925	0.56	0.43	0.45
ZINC02459482	0.75	0.35	0.44
ZINC05255177	0.73	0.50	0.53
Hits derived from PH-2			
6000231	0.69	0.41	0.43
7421641	0.67	0.50	0.43
KM09208	0.63	0.34	0.39
RH00799	0.54	0.32	0.40
SPB06754	0.60	0.31	0.36
339161	0.61	0.46	0.39
630720	0.54	0.42	0.35
73468	0.50	0.37	0.39
ZINC02723161	0.65	0.37	0.45
ZINC03455195	0.68	0.42	0.43

By considering the chemical structure of these hit compounds, it is found that compound **6354267**, **HTS0042**, **719925**, **ZINC02459482**, **ZINC05255177**, **KM09208**, **RH00799**, **SPB06754**, **630720**, and **ZINC02723161** have a core structure similar to compound 4. Moreover, the Tanimoto coefficient (a measure of the similarity between two compounds) of these hits compared to the known active inhibitors[39] (compound 4, 6 and 7) was also calculated and are summarized in

Table 3.9. The chemical structure of these known inhibitors and their K_i value are shown in Table 3.9. The compound 6 and compound 7 were not applied for the MD simulation because the stereochemistry of these two compounds is unclear. The Tanimoto coefficient supports the earlier visual inspection that compound **6354267**, **ZINC02459482**, **ZINC05255177**, **KM09208**, and **ZINC02723161** are structurally closer to compound 4. In addition, the Tanimoto coefficient point out that compound **ZINC05255177** is also quite similar to compound 6 and 7. In the case of compound **7421641**, even though it is not clearly seen that the structure of this compound resembles compound 4, 6 and 7, the Tanimoto coefficient reveals that this compound has some chemical structure or functional groups that are similar to compound 4 and 6. In addition, the binding free energy of compound **7421641** is quite low. Therefore, this compound might be an active compound for DV NS2B/NS3 protease. In conclusion, compound **6354267**, **ZINC02459482**, **ZINC05255177**, **7421641**, **KM09208**, and **ZINC02723161** might be active as well as compound 4. For the rest of the compounds (**HTS0042**, **719225**, **RH00799**, **SPB06754**, and **630720**), even though they are not really similar to the compound 4, they still contains the similar functional group compared the compound 4, compound 6 or compound 7. Therefore, it is quite interesting to test the biological activities of these compounds. In summary, by considering the relative binding free energies of hits compared with the values of known inhibitors and their chemical structures, compound **6354267**, **HTS0042**, **719225**, **ZINC02459482**, **ZINC05255177**, **7421641**, **KM09208**, **RH00799**, **SPB06754**, **630720**, and **ZINC02723161** are proposed to test the biological activities against NS2B/NS3 protease of DV.

CHAPTER IV

WEE1 KINASE

4.1 3D-QSAR (CoMFA) model of checkpoint Wee1 kinase inhibitors

Several studies have been published in which the combination of receptor-based methods and 3D-QSAR was successfully applied for the design and prediction of bioactive compounds[70-72]. The three-dimensional structure of a target protein, along with a docking protocol is used to guide alignment selection for comparative molecular field analysis. It is quite appealing to combine the accuracy of a receptor-based alignment with the computational efficiency of a ligand-based method. Receptor structures, either experimentally resolved or obtained by homology modeling, can provide important information that is critical for an alignment in CoMFA[69], while QSAR can provide better prediction of binding energies[70].

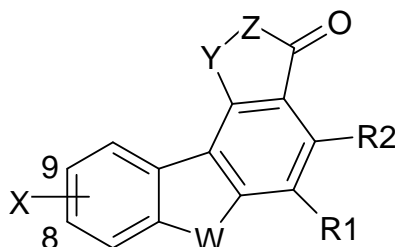
In the present work we applied this receptor-based 3D-QSAR technique to a set of 174 Wee1 kinase inhibitors which has been recently developed [62-64]. The crystal structure of the catalytic domain of Wee1 together with an automatic docking program was used to determine the molecular alignment of the ligands.

4.1.1 Material and Computational Methods

4.1.1.1 Inhibitor Data Set

A set of pyrrolo[3,4-c]carbazole-1,3(2*H*,6*H*)-dione derivatives representing inhibitors of the Kinase Wee1[62-64] was used to generate a 3D-QSAR model applying the CoMFA methodology. All inhibitors were developed and tested in the same laboratory using the same assay condition, a prerequisite for generating reliable QSAR models. Inhibitors with no clear IC₅₀ value (indicated with > 50 μM) and inhibitors with undefined stereochemistry, were omitted. The 174 inhibitors were divided in training and test set (Chart 4.1). In this way 144 ligands were randomly selected to generate the CoMFA model. The IC₅₀ value of each inhibitor was converted into pIC₅₀ (-logIC₅₀) in order to use the data as dependent variable in the CoMFA model. The structures and the activity values (IC₅₀ and pIC₅₀) of the studied inhibitors are displayed in Chart 4.1.

Chart 4.1 Molecular structures of the pyrrolo[3,4-c]carbazole-1,3(2*H*,6*H*)-dione derivatives used for the 3D-QSAR study.



Training Set (144 compounds)

Cpd	W	X	Y	Z	R1	R2	Activity	
							IC ₅₀ (μ M)	pIC ₅₀
3	NH	9-OH	CO	NH	H	I	2.3	5.63
4	NH	8-OH	CO	NH	H	Ph	0.31	6.51
5	O	9-OH	CO	NH	H	Ph	0.43	6.37
6	S	9-OH	CO	NH	H	Ph	0.078	7.11
8	NH	9-OH	CO	NH	Me	Ph	0.13	6.89
9	NH	9-OH	CO	NH	Et	Ph	1.6	5.79
10	NH	9-OH	CO	NH	Ph	Me	9.7	5.01
11	NH	9-OH	CO	NH	Ph	Ph	2.3	5.63
12	NH	9-OH	CO	NH	Ph	H	4.0	5.39
13	NH	9-OH	CH ₂	NH	H	Ph	37	4.43
14	NH	9-OH	CO	N-NH ₂	H	Ph	3.9	5.40
15	NH	9-OH	CO	NH	H	2-ClPh	0.011	7.96
16	NH	9-OMe	CO	NH	H	2-ClPh	0.64	6.19
17	NMe	9-OH	CO	NH	H	2-ClPh	0.057	7.24
19	NH	9-OH	CO	NH	H	2-FPh	0.33	6.48
21	NH	9-OH	CO	NH	H	2-IPh	0.013	7.89
22	NH	9-OH	CO	NH	H	2-MePh	0.15	6.82
23	NH	9-OH	CO	NH	H	2-EtPh	0.51	6.29
24	NH	9-OH	CO	NH	H	2-CF ₃ Ph	0.58	6.24
25	NH	9-OH	CO	NH	H	2-CH ₂ OHPH	0.45	6.35
26	NH	9-OH	CO	NH	H	2-CNPh	0.19	6.72
27	NH	9-OH	CO	NH	H	2-COMePh	0.83	6.08
28	NH	9-OH	CO	NH	H	2-Ph-Ph	0.57	6.24
31	NH	9-OH	CO	NH	H	2-OEtPh	0.26	6.59
32	NH	9-OH	CO	NH	H	2-SMePh	0.033	7.48
33	NH	9-OH	CO	NH	H	2-SOMePh	0.22	6.66
34	NH	9-OH	CO	NH	H	2-NO ₂ Ph	0.047	7.33
35	NH	9-OH	CO	NH	H	2-NH ₂ Ph	0.21	6.68
36	NH	9-OH	CO	NH	H	3-FPh	0.22	6.66
37	NH	9-OH	CO	NH	H	3-ClPh	0.055	7.26
38	NH	9-OH	CO	NH	H	3-MePh	0.23	6.64
39	NH	9-OH	CO	NH	H	3-CH ₂ OHPH	0.87	6.06
40	NH	9-OH	CO	NH	H	3-CH ₂ NH ₂ Ph	4.4	5.35
41	NH	9-OH	CO	NH	H	3-CNPh	0.18	6.74
42	NH	9-OH	CO	NH	H	3-COMePh	4.3	5.36
45	NH	9-OH	CO	NH	H	3-OMePh	0.62	6.21
46	NH	9-OH	CO	NH	H	3-NO ₂ Ph	0.30	6.52
47	NH	9-OH	CO	NH	H	3-NH ₂ Ph	0.070	7.15
48	NH	9-OH	CO	NH	H	4-FPh	16	4.80
49	NH	9-OH	CO	NH	H	4-ClPh	0.73	6.14
52	NH	9-OH	CO	NH	H	4-CNPh	1.8	5.74

53	NH	9-OH	CO	NH	H	4-COMePh	3.6	5.44
54	NH	9-OH	CO	NH	H	4-OHPh	0.067	7.17
55	NH	9-OH	CO	NH	H	4-OMePh	12	4.90
57	NH	9-OH	CO	NH	H	4-SO ₂ MePh	1.1	5.95
58	NH	9-OH	CO	NH	H	4-NH ₂ Ph	0.15	6.82
59	NH	9-OH	CO	NH	H	2-Cl, 3-ClPh	0.028	7.55
60	NH	9-OH	CO	NH	H	2-Cl, 3-OHPh	0.012	7.55
61	NH	9-OH	CO	NH	H	2-Cl, 3-NH ₂ Ph	0.021	7.92
62	NH	9-OH	CO	NH	H	2-Cl, 4-OHPh	0.023	7.64
63	NH	9-OH	CO	NH	H	4-Cl, 3-NH ₂ Ph	0.024	7.62
64	NH	9-OH	CO	NH	H	2-Cl, 5-ClPh	0.49	6.31
66	NH	9-OH	CO	NH	H	2-Cl, 5-NH ₂ Ph	0.020	7.70
67	NH	9-OH	CO	NH	H	2-Cl, 6-ClPh	0.028	7.55
68	NH	9-OH	CO	NH	H	2-Cl, 6-OHPh	0.045	7.35
69	NH	9-OH	CO	NH	H	2-Cl, 6-OMePh	0.015	7.82
70	NH	9-OH	CO	NH	H	2-Br, 4-NH ₂ Ph	0.020	7.70
71	NH	9-OH	CO	NH	H	2-Br, 6-BrPh	0.035	7.46
72	NH	9-OH	CO	NH	H	2-Me, 3-MePh	0.27	6.57
73	NH	9-OH	CO	NH	H	2-Me, 5-MePh	0.96	6.02
74	NH	9-OH	CO	NH	H	2-Me, 6-MePh	0.075	7.12
75	NH	9-OH	CO	NH	H	2-OMe, 4-NH ₂ Ph	0.019	7.72
76	NH	9-OH	CO	NH	H	2-OMe, 5-NH ₂ Ph	0.11	6.96
77	NH	9-OH	CO	NH	H	2-OMe, 6-OMePh	0.027	7.57
78	NH	9-OH	CO	NH	H	2-OMe, 6-FPh	0.029	7.54
79	NH	9-OH	CO	NH	H	2,6-diCl, 3-OHPh	0.018	7.74
81	NH	9-OH	CO	NH	H	2-thienyl	0.14	6.85
82	NH	9-OH	CO	NH	H	3-thienyl	0.042	7.38
83	NH	9-OH	CO	NH	H	2-pyrrolyl	0.18	6.74
84	NH	9-OH	CO	NH	H	3-pyrrolyl	0.038	7.42
87	NEt	9-OH	CO	NH	H	2-ClPh	0.050	7.30
88	N- <i>n</i> -Pr	9-OH	CO	NH	H	2-ClPh	0.063	7.20
89	N- <i>i</i> -Pr	9-OH	CO	NH	H	2-ClPh	0.053	7.28
90	N- <i>n</i> -Bu	9-OH	CO	NH	H	2-ClPh	0.059	7.23
91	N(CH ₂) ₂ <i>i</i> -Pr	9-OH	CO	NH	H	2-ClPh	0.15	6.82
92	N- <i>n</i> -pentyl	9-OH	CO	NH	H	2-ClPh	0.17	6.77
93	NH	9-OH	CO	NH	H	H	0.097	7.01
94	NH	9-OH	CO	NH	H	2,6-diClPh	0.028	7.55
95	NMe	9-OH	CO	NH	H	H	0.14	6.85
96	NMe	9-OH	CO	NH	H	2-ClPh	0.057	7.24
97	N <i>n</i> -Bu	9-OH	CO	NH	H	2-ClPh	0.059	7.23
98	N(CH ₂) ₂ OH	9-OH	CO	NH	H	H	0.025	7.60
99	N(CH ₂) ₂ OH	9-OH	CO	NH	H	2-ClPh	0.045	7.35

101	N(CH ₂) ₃ OH	9-OH	CO	NH	H	H	0.2	6.70
102	N(CH ₂) ₃ OH	9-OH	CO	NH	H	2-ClPh	0.009	8.05
103	N(CH ₂) ₃ OH	9-OH	CO	NH	H	2,6-diClPh	0.007	8.15
104	N(CH ₂) ₃ OH	9-OH	CO	NH	H	2-OMePh	0.03	7.52
105	N(CH ₂) ₂ CONH ₂	9-OH	CO	NH	H	H	0.021	7.68
106	N(CH ₂) ₂ CONH ₂	9-OH	CO	NH	H	2-ClPh	0.006	8.22
107	N(CH ₂) ₂ CONH ₂	9-OH	CO	NH	H	2,6-diClPh	0.33	6.48
108	N(CH ₂) ₂ CN	9-OH	CO	NH	H	2-ClPh	0.015	7.82
109	N(CH ₂) ₂ COOMe	9-OH	CO	NH	H	2-ClPh	0.03	7.52
110	N(CH ₂) ₃ CN	9-OH	CO	NH	H	2-ClPh	0.033	7.48
112	N(CH ₂) ₂ NMe ₂	9-OH	CO	NH	H	H	0.2	6.70
113	N(CH ₂) ₂ NMe ₂	9-OH	CO	NH	H	2-ClPh	0.096	7.02
114	N(CH ₂) ₂ NMe ₂	9-OH	CO	NH	H	2,6-diClPh	0.17	6.77
116	N(CH ₂) ₂ Nmorpholide	9-OH	CO	NH	H	2-ClPh	0.064	7.19
117	N(CH ₂) ₂ Nmorpholide	9-OH	CO	NH	H	2,6-diClPh	0.11	6.96
118	N(CH ₂) ₂ Nimidazolide	9-OH	CO	NH	H	H	0.23	6.64
119	N(CH ₂) ₂ Nimidazolide	9-OH	CO	NH	H	2-ClPh	0.092	7.04
120	N(CH ₂) ₂ Nimidazolide	9-OH	CO	NH	H	2,6-diClPh	0.12	6.92
121	N(CH ₂) ₃ NHMe	9-OH	CO	NH	H	H	0.28	6.55
122	N(CH ₂) ₃ NHMe	9-OH	CO	NH	H	2-ClPh	0.069	7.16
123	N(CH ₂) ₃ NHMe	9-OH	CO	NH	H	2,6-diClPh	0.11	6.96
125	N(CH ₂) ₃ NMe ₂	9-OH	CO	NH	H	2-ClPh	0.1	7.00
126	N(CH ₂) ₃ NMe ₂	9-OH	CO	NH	H	2,6-diClPh	0.14	6.85
127	N(CH ₂) ₃ Nmorpholide	9-OH	CO	NH	H	H	0.29	6.54
128	N(CH ₂) ₃ Nmorpholide	9-OH	CO	NH	H	2-ClPh	0.071	7.15
129	N(CH ₂) ₃ Nmorpholide	9-OH	CO	NH	H	2,6-diClPh	0.064	7.19
132	N(CH ₂) ₃ Nimidazolide	9-OH	CO	NH	H	2,6-diClPh	0.059	7.23
133	N(CH ₂) ₃ N(4-Mepiperazine)	9-OH	CO	NH	H	H	0.3	6.52
134	N(CH ₂) ₃ N(4-Mepiperazine)	9-OH	CO	NH	H	2-ClPh	0.082	7.09
135	N(CH ₂) ₃ N(4-Mepiperazine)	9-OH	CO	NH	H	2,6-diClPh	0.062	7.21
136	N(CH ₂) ₃ NHPh	9-OH	CO	NH	H	H	0.093	7.03
139	N(CH ₂) ₂ CONH(CH ₂) ₂ NMe ₂	9-OH	CO	NH	H	H	0.17	6.77
140	N(CH ₂) ₂ CONH(CH ₂) ₂ NMe ₂	9-OH	CO	NH	H	2-ClPh	0.035	7.46
141	N(CH ₂) ₂ CONH(CH ₂) ₂ NMe ₂	9-OH	CO	NH	H	2,6-diClPh	0.014	7.85
142	N(CH ₂) ₂ COOH	9-OH	CO	NH	H	H	0.023	7.64
143	N(CH ₂) ₂ COOH	9-OH	CO	NH	H	2-ClPh	0.009	8.05
144	N(CH ₂) ₂ COOH	9-OH	CO	NH	H	2,6-diClPh	0.39	6.41
145	N(CH ₂) ₂ CONHSO ₂ Me	9-OH	CO	NH	H	2-ClPh	0.012	7.92
146	N(CH ₂) ₂ CONHSO ₂ Ph	9-OH	CO	NH	H	2-ClPh	0.007	8.15
147	N(CH ₂) ₂ C-tetrazole	9-OH	CO	NH	H	2-ClPh	0.021	7.68
148	N(CH ₂) ₂ S-triazole	9-OH	CO	NH	H	2-ClPh	0.024	7.62
149	N(CH ₂) ₂ SO-triazole	9-OH	CO	NH	H	2-ClPh	0.009	8.05
152	N(CH ₂) ₃ C-tetrazole	9-OH	CO	NH	H	2-ClPh	0.016	7.80
153	NH	8-(CH ₂) ₄ -NMe ₂ , 9-OH	CO	NH	H	2-ClPh	0.049	7.31
154	NH	8-(CH ₂) ₄ -Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.05	7.30
155	NH	8-(CH ₂) ₄ -Nmorph, 9-OH	CO	NH	H	2-ClPh	0.037	7.43
156	NMe	8-(CH ₂) ₄ -NMe ₂ , 9-OH	CO	NH	H	2-ClPh	0.034	7.47
157	NMe	8-(CH ₂) ₄ -Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.036	7.44
159	N(CH ₂) ₂ OH	8-(CH ₂) ₄ -Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.024	7.62
160	N(CH ₂) ₂ OH	8-(CH ₂) ₄ -Nmorph, 9-OH	CO	NH	H	2-ClPh	0.019	7.72
161	NH	8-O(CH ₂) ₃ -NMe ₂ , 9-OH	CO	NH	H	2-ClPh	0.026	7.59
163	NH	8-O(CH ₂) ₃ -Nmorph, 9-OH	CO	NH	H	2-ClPh	0.026	7.59

164	NMe	8-O(CH ₂) ₃ -NMe ₂ , 9-OH	CO	NH	H	2-ClPh	0.058	7.24
165	NMe	8-O(CH ₂) ₃ -Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.075	7.12
167	N(CH ₂) ₂ OH	8-O(CH ₂) ₃ -NMe, 9-OH ₂	CO	NH	H	2-ClPh	0.018	7.74
168	N(CH ₂) ₂ OH	8-O(CH ₂) ₃ -Npyrrol, 9-OH	CO	NH	H	2-ClPh	0.024	7.62
169	N(CH ₂) ₂ OH	8-O(CH ₂) ₃ -Nmorph, 9-OH	CO	NH	H	2-ClPh	0.015	7.82
170	NMe	8-S(CH ₂) ₃ -Npyrro, 9-OH	CO	NH	H	2-ClPh	0.02	7.70
171	NMe	8-SO(CH ₂) ₃ -Npyrro, 9-OH	CO	NH	H	2-ClPh	0.033	7.48
173	NMe	8-SO ₂ NH(CH ₂) ₂ -Npyrro, 9-OH	CO	NH	H	2-ClPh	0.046	7.34
174	NMe	8-CONH(CH ₂) ₂ -Npyrro, 9-OH	CO	NH	H	2-ClPh	0.015	7.82

Test Set (30 compounds)

Cpd	W	X	Y	Z	R1	R2	Activity	
							IC ₅₀ (μM)	pIC ₅₀
1	NH	9-OH	CO	NH	H	Ph	0.097	7.01
2	NH	9-OH	CO	NH	H	H	4.0	5.39
7	NMe	9-OH	CO	NH	H	Ph	0.26	6.59
18	O	9-OH	CO	NH	H	2-ClPh	0.033	7.48
20	NH	9-OH	CO	NH	H	2-BrPh	0.023	7.64
29	NH	9-OH	CO	NH	H	2-OHPh	0.060	7.22
30	NH	9-OH	CO	NH	H	2-OMePh	0.024	7.62
43	NH	9-OH	CO	NH	H	4-Biphenyl	40	4.30
44	NH	9-OH	CO	NH	H	3-OHPh	0.089	7.05
50	NH	9-OH	CO	NH	H	4-MePh	3.3	5.48
51	NH	9-OH	CO	NH	H	4-CH ₂ OHPH	1.2	5.92
56	NH	9-OH	CO	NH	H	4-SMe	29	4.50
65	NH	9-OH	CO	NH	H	2-Cl, 5-OHPh	0.042	7.38
80	NH	9-OH	CO	NH	H	2,6-diCl, 4-OHPH	0.049	7.31
85	NH	9-OH	CO	NH	H	4-pyridyl	0.82	6.09
86	NH	9-OH	CO	NH	H	3-pyridyl	0.58	6.24
100	N(CH ₂) ₂ OH	9-OH	CO	NH	H	2,6-diClPh	0.008	8.10
111	N(CH ₂) ₃ OMe	9-OH	CO	NH	H	2-ClPh	0.027	7.57
115	N(CH ₂) ₂ Nmorpholide	9-OH	CO	NH	H	H	0.14	6.85
124	N(CH ₂) ₃ NMe ₂	9-OH	CO	NH	H	H	0.36	6.44
130	N(CH ₂) ₃ Nimidazolide	9-OH	CO	NH	H	H	0.11	6.96
131	N(CH ₂) ₃ Nimidazolide	9-OH	CO	NH	H	2-ClPh	0.054	7.27
137	N(CH ₂) ₃ NHPh	9-OH	CO	NH	H	2-ClPh	0.074	7.13
138	N(CH ₂) ₃ NHPh	9-OH	CO	NH	H	2,6-diClPh	0.067	7.17
150	N(CH ₂) ₂ SO ₂ triazole	9-OH	CO	NH	H	2-ClPh	0.019	7.72
151	N(CH ₂) ₂ COOH	9-OH	CO	NH	H	2-ClPh	0.013	7.89
158	NMe	8-(CH ₂) ₄ -N-morph, 9-OH	CO	NH	H	2-ClPh	0.03	7.52
162	NH	8-O(CH ₂) ₃ -N-pyrrol, 9-OH	CO	NH	H	2-ClPh	0.036	7.44
166	NMe	8-O(CH ₂) ₃ -N-morph, 9-OH	CO	NH	H	2-ClPh	0.057	7.24
172	NMe	8-SO ₂ (CH ₂) ₃ -N-pyrrol, 9-OH	CO	NH	H	2-ClPh	0.16	6.80

4.1.1.2 Ligand Preparation

The 3D structures of the inhibitors were built based on the X-ray structure of 9-hydroxy-4-phenyl-6H-pyrrolo[3,4-C]carbazole-1,3-dione (ligand number 1), which was cocrystallized with human Wee-1 kinase (PDB code 1X8B) [63]. The molecular structures of the inhibitors were generated using Sybyl7.2 [147]. The structures were energy minimized using the MMFF94s force field[148] and the BFGS method[149-152], until the default derivative convergence criterion of 0.05 kcal/mol was reached. All compounds were generated in the protonation state at pH 7.1 using the MOE protonate 3D method[112].

4.1.1.3 Molecular Docking

To predict the appropriate binding orientation of Wee1 kinase inhibitors the molecular docking program GOLD (version 3.2[105, 106]), was employed to generate an ensemble of docking conformations. To test whether GOLD is able to reproduce the experimental data we selected the six available Wee1-inhibitor complexes from the PDB databank (entries 1X8B, 2ZZW, 2IN6, 2IO6, 3BI6, and 3BIZ). The protein structures were prepared for docking using the Sybyl7.2 software[147]. The original ligand and all ions were removed from the Wee1 protein complexes. The default GOLD parameters were used except that the ‘flood fill radius’ was set to 20 Å around Cys379.

In the Wee1 crystal structures a variety of cocrystallized water molecules are found in close proximity to the bound inhibitors (Figure 1.18). Since the docking of the six inhibitors into the corresponding protein structures resulted in low accuracy (high RMSD values for the top-ranked docking pose) when no water molecules or constraints were used, we tested a variety of further docking settings. First, we used the water toggeling mode within the GOLD program. Eight clusters of cocrystallized water molecules (shown as red balls in Figure 1.18) which were observed in close proximity to the inhibitors in the Wee1 X-ray structures were considered as potential interaction sites for the docked inhibitors.

As second setting, we considered two hydrogen bonds to the Wee1 hinge backbone region to be relevant for inhibitor binding. Therefore two protein hydrogen bond constraints were defined in GOLD to the backbone oxygen of Glu377 and to the

backbone NH of Cys379. The ligands were scored based on the fitness function 'GoldScore'. GOLD was run to save up to 10 top-ranked docking solutions for the ligands. The results were visually analyzed using MOE2006.8[112].

Both docking settings resulted in one favourable cluster of docking poses in case of the reference inhibitor **1** (RMSD values between 0.347 and 0.520 for the 10 top-ranked docking poses), whereas for the more flexible inhibitors several individual clusters were detected. In the case of several clusters we selected the ligand conformation which showed the highest GoldScore.

4.1.1.4 Ligand Alignment Method

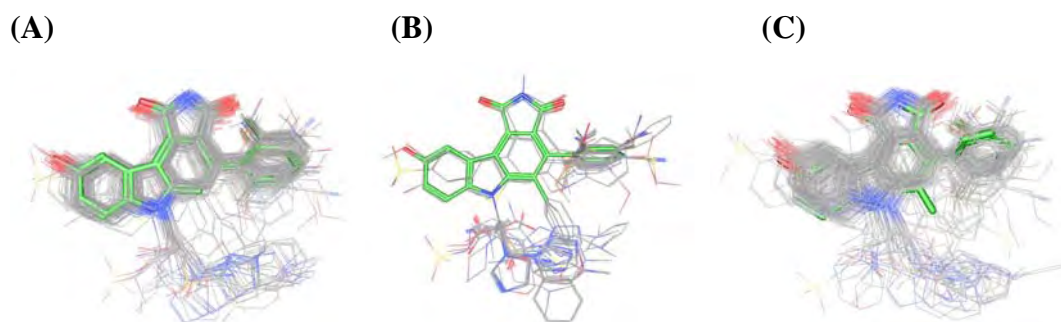


Figure 4.1. (A) Alignment 1 derived from the GOLD docking study, (B) Alignment 2 obtained by flexibly aligning the docking poses on the template structure of inhibitor 1, (C) Alignment 3 derived by minimizing the docking poses in the wee1 active site. Inhibitor 1 is shown in green.

The result of a CoMFA model strongly depends on the quality of the ligand alignment. However, when studying a large number of ligands there is no certain preferable alignment method. Therefore, we tested three different alignment procedures in order to identify the most efficient alignment approach for this data set. The first alignment (Alignment 1 as shown in Figure 4.1 (A)) was derived using the top-ranked GOLD docking pose for each inhibitor.

For the ligand-based alignment 2, displayed in Figure 4.1 (B), a flexible alignment was carried out using the structure of compound **1** derived from the crystal structure 1X8B.pdb and using the flexible alignment tool within MOE2006.8. In this

procedure, a ligand-based alignment was derived using the similarity function as criteria to align each molecule onto the conformation of the potent inhibitor **1** as template structure.

Alignment 3, illustrated in Figure 4.1 (C), was derived by refining the GOLD docking solutions using the Amber force field 99 [153]. The minimization of the Wee1-inhibitor complexes was performed using the MOE2006.08 program. PEOE charges [154] were assigned for the inhibitors whereas the Amber force field 99 charges were applied for the kinase. During the energy minimization process, the kinase was constrained by tethering heavy atoms (force constant 100 kcal/mol) whereas the inhibitor structures were relaxed.

4.1.1.5 CoMFA Model

CoMFA calculations were done by applying the default settings. The standard CoMFA fields performing the Lennard-Jones potential for the steric and Coulomb potential for the electrostatic were used. A cut-off value for the fields was set at 30 kcal/mol. Both steric and electrostatic interactions at each grid point were calculated by applying the C-sp³ probe atom with +1 charge. Three different grid spacing values, 1Å, 1.5Å and 2Å, were employed in order to investigate the influence of the grid spacing. The Sybyl7.2 program [147] was used for generating the CoMFA models.

To extend the scope of CoMFA the region focusing approach was applied. Region focusing is a procedure to refine CoMFA models by contributing the weight to the lattice points. Several region focusing procedures such as StDev*Coefficients (StDev.), Discriminant power (Disc.), Sigma field (Sigma) and PLS coefficient (PLS) as implemented in Sybyl 7.2 were tested.

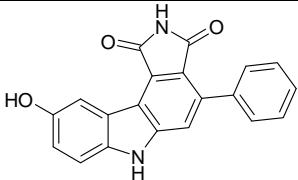
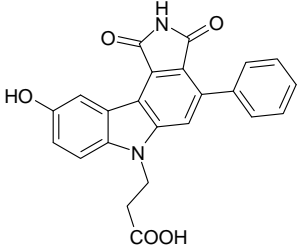
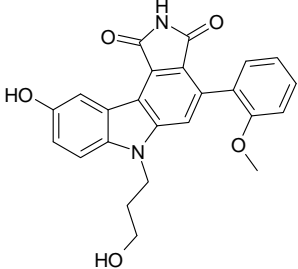
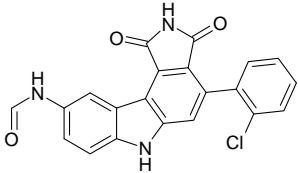
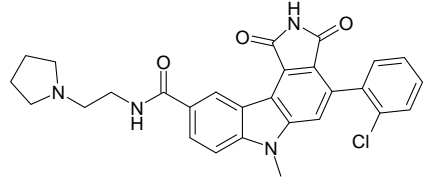
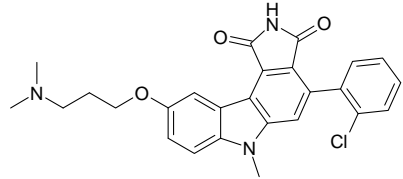
4.1.1.6 Partial Least Square (PLS) Analysis

The method of PLS as explained in section 2.1 was employed in this studies. The leave-one-out, leave-20%-out, and leave-50%-out cross-validation were applied to test a predictive ability of a model. An external test set containing 30 randomly selected inhibitors was used in addition to test the predictivity of the final CoMFA model.

4.1.2 Results and Discussion

4.1.2.1 Ligand Docking

Table 4.1: RMSD values between crystal structure and the top-ranked docking pose using different docking settings.

Cocrystallized inhibitor	PDB code	RMSD water (Å)	RMSD constr. (Å)
	1X8B	0.437	0.434
	2IN6	1.210	0.793
	2IO6	1.257	0.585
	2Z2W	0.373	0.431
	3BI6	0.637	0.504
	3BIZ	0.445	0.491

RMSD water: RMSD observed for the docking using the toggling water mode in GOLD

RMSD constr.: RMSD observed for the docking using two hydrogen bond constraints.

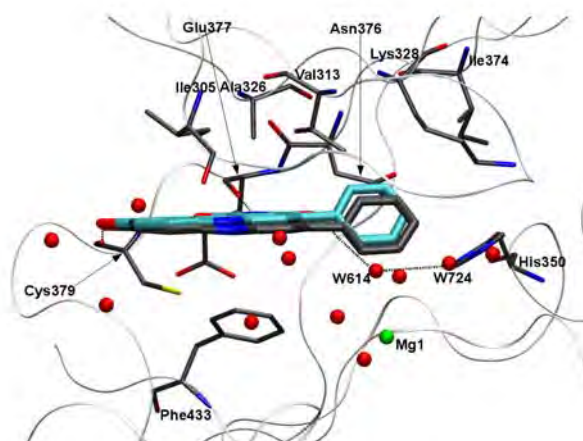


Figure 4.2. Comparison of the docking solution for **1** (colored cyan) compared with its position in the crystal structure (colored gray, hydrogen-bonds are shown as dashed line, cocrystallized water molecules as red balls and the magnesium ion as green ball).

We first analyzed, whether GOLD is able to reproduce the position of the cocrystallized inhibitors observed in the six Wee1 X-ray structures. Docking of the inhibitors into the Wee1 ATP binding site, using two hydrogen bonding constraints and without considering the cocrystallized water molecules, yielded slightly lower root-mean square deviation (RMSD) values compared to the docking when ordered water molecules were considered to be part of the protein (toggeling water mode in GOLD) (Table 4.1). Superimposition between the docking solution, using these hydrogen bonding constraints and toggeling water mode, and the cocrystallized structure (PDB code 1X8B) is also shown in Figure 4.2. Results show clearly that by applying these protocols for docking, we can reproduce the X-ray structure. Therefore we used these two hydrogen bond constraints for all 174 pyrrolocarbazole derivatives. Due to the structural similarity of the analyzed data set, it is likely that all active compounds possessing the pyrrolocarbazole moiety show a similar interaction at the Wee1 ATP-binding pocket. The superimposition of the 174 compounds (Figure 4.1 (A)) reveals a well conserved overall conformation, particularly for the pyrrolocarbazole ring, with slight variations in the orientation of the attached side chains. Beside the four hydrogen bonds of the pyrrolocarbazole to Asn376 (sidechain amide), Glu377 (backbone CO), Cys379 (backbone NH and CO), the most potent

inhibitor **106** forms an additional hydrogen bond between the N6-substituent and the backbone CO of Ser430 (Figure 4.3 (A) and 4.3 (B)).

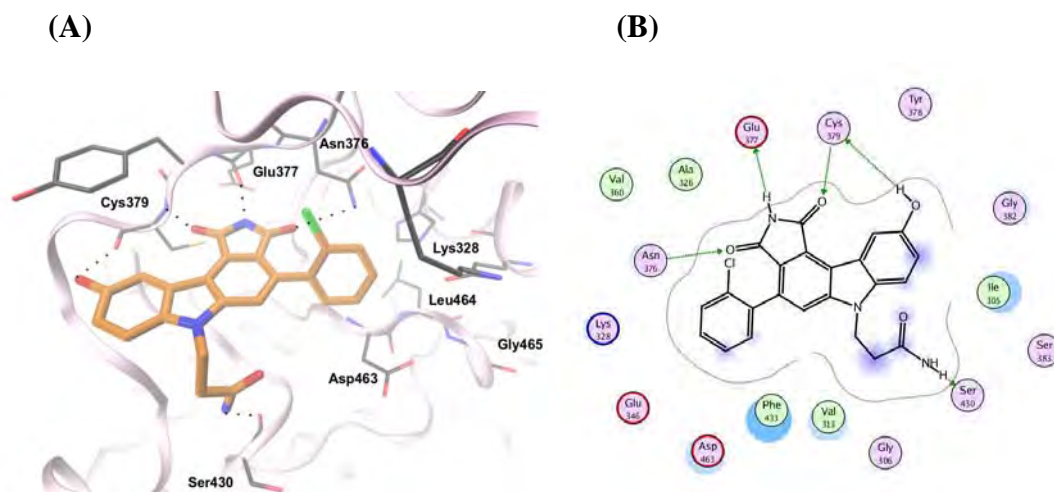


Figure 4.3 (A). GOLD docking solution for the most potent inhibitor **106** (orange). Hydrogen bonds are shown as dashed line. (B). Schematic representation of the interaction of the most potent inhibitor **106** with the residues of the Wee1 binding site. Hydrogen bonds are indicated as arrows.

Subsequently we analyzed whether the docking scores (top-score and average score) can be correlated with the biological data of the inhibitors. Only a low correlation was observed for the studied data set of 174 inhibitors ($r^2 = 0.31$). To counteract the tendency of larger molecules to produce better docking scores, the calculated scores were divided by the square root of the number of heavy atoms[155, 156]. However, this resulted in no significant improvement of the correlation coefficient ($r^2 = 0.34$).

4.1.2.2 CoMFA Models

In this work, statistical models linking the biological activities to the different descriptors were built by means of partial least-square (PLS) regression, the degree of correlation of experimental vs. predicted values was expressed in terms of the square of the correlation coefficient (r^2), indicating the fraction of explained variance, and the internal predictability was measured in terms of cross validated r^2 , hereafter referred to as q^2_{LOO} , after cross-validation using the leave-one-out method. A first CoMFA model was generated by using 144 randomly selected compounds as training

set (Chart 4.1). Analyzing the CoMFA model ($q^2_{LOO} = 0.474$, SDEP = 0.558, $r^2 = 0.706$ and SDEE = 0.417) we found that ligands **13**, **40**, **48**, **107** and **144** have high residual values (actual pIC₅₀ vs. predicted pIC₅₀). Their residual values are 2.0 times higher as the SDEP of the model indicating that these five ligands are being truly outliers. We visually analyzed the docking solutions of the outliers and tried to find an explanation for that. In case of compound **13**, which is the only compound without a second carbonyl group at the pyrrolidine ring, the 3D-QSAR model is not able to cover this difference. In addition, a variety of different possible binding orientations were observed for compound **13** in the docking study. This variability was only observed for compound **13**, whereas all other inhibitors show the conserved binding mode due to the further hydrogen bonding possibilities. Compound **40** is the only inhibitor which possesses an aliphatic amino group (-CH₂NH₂) attached to the phenyl ring, and it is not clear whether the amino group is protonated or not when interacting at the buried hydrophobic part of the binding pocket. To test the influence of the protonation state we generated a variety of different CoMFA models (all aliphatic amines protonated, all amines neutral, only the amino group of **40** neutral) but compound **40** was an outlier in all models. In case of compound **48** it is not clear why the Fluoro substituent in position 4 results in a complete loss of activity, also the QSAR model is not able to explain this discrepancy. Similar observations were made for compound **107** and **144**. It was interesting to see that the five compounds were also outliers in the CoMFA model derived from the ligand-based alignment (data not shown).

Table 4.2. Statistics of generated CoMFA models using different alignment methods.

Model	Grid Spacing (Å)	Leave-one-out cross-validation			Non cross-validation			Field contribution	
		q^2	Comp.	SDEP	r^2	SDEE	F-value	Steric	Elec.
Alignment 1 (docking-based alignment)									
1	1.0	0.549	6	0.475	0.773	0.337	75.590	0.546	0.454
2	1.5	0.685	6	0.397	0.836	0.286	113.017	0.570	0.430
3	2.0	0.624	6	0.433	0.801	0.316	89.120	0.546	0.454
Alignment 2 (ligand-based flexible alignment)									
4	1.0	0.506	4	0.488	0.625	0.430	56.319	0.565	0.435
5	1.5	0.502	4	0.495	0.614	0.436	53.769	0.569	0.431
6	2.0	0.502	5	0.497	0.656	0.413	51.118	0.561	0.439
Alignment 3 (alignment from protein-inhibitor minimization)									
7	1.0	0.470	5	0.513	0.682	0.397	57.509	0.486	0.514
8	1.5	0.462	5	0.517	0.672	0.403	54.991	0.485	0.515
9	2.0	0.462	5	0.517	0.676	0.401	55.890	0.482	0.518

Therefore, we removed the five outliers and generated a CoMFA model for the remaining 139 inhibitors. A variety of CoMFA models were generated based on three different alignments and testing different CoMFA settings. Table 4.2 summarizes the statistics derived for the different models. The best CoMFA model was obtained from the docking-based alignment 1 (model 2, $q^2_{LOO} = 0.685$, SDEP = 0.397, $r^2 = 0.836$ and SDEE = 0.286), whereas the other two alignment procedures resulted in lower q^2 values. In addition, we applied a region focusing strategy to analyze whether reducing the number of variables is able to increase the predictive ability of the CoMFA model 2. The resulting CoMFA model of the 139 ligands with PLS region focusing method (model 10 exhibited in Table 4.3) gave the model with the highest $q^2_{LOO} = 0.764$.

Table 4.3. Statistics of the resulting CoMFA models.

Model	RF	Leave-one-out cross-validated			Non cross-validated			Field contribution	
		q^2	Comp.	SDEP	r^2	SDEE	F-value	Steric	Elec.
Receptor-based CoMFA model									
2	-	0.685	6	0.397	0.836	0.286	113.017	0.570	0.430
CoMFA model, Region Focusing Method									
10	PLS	0.764	6	0.342	0.870	0.254	148.973	0.478	0.522

The actual pIC_{50} , the predicted pIC_{50} and the estimated pIC_{50} of each inhibitor, which was calculated from the CoMFA model 10, are compared and listed in Table 4.4. Actual pIC_{50} versus predicted pIC_{50} and estimated pIC_{50} values are plotted on the graph in the Figure 4.4 (A) and 4.4 (B), respectively. The cross-validation demonstrates that the predicted pIC_{50} values are in good agreement with the actual pIC_{50} . The leave-one-out cross-validation method might lead to high q^2 values, which do not necessarily reflect a predictivity of a model. Therefore, further cross-validation, using five and two groups of approximately the same size in which the objects were assigned randomly, was performed. These models have been shown in the literature to provide a better quantitative estimate of the robustness of a QSAR model. Thus, q^2 values for leave-20%-out (0.747) and leave-50%-out (0.737) were determined. The high q^2 values of the both cross-validation procedures, which are in the same range as the leave-one-out value, indicates that even with a limited number of molecules robust QSAR models could be derived. In addition, CoMFA model 10 was also applied to predict the activity of the external test set containing 30 inhibitors not used for model generation (Table 4.5). The prediction shows a good correlation between the actual and the predicted pIC_{50} values with $r^2_{pred} = 0.790$ (Figure 4.4 (C)). In addition, CoMFA models 4 and model 7 were also used to predict the activities of the test set but resulted in much lower predictive r^2 values ($r^2_{pred} = 0.551$ and 0.501 , respectively).

Table 4.4 List of actual pIC_{50} , predicted pIC_{50} (derived from leave-one-out cross-validation), estimated pIC_{50} values and residual for the 139 training set compounds as derived from the CoMFA model 10.

Cpd	Actual pIC_{50}	Pred. pIC_{50}	Res.	Est. pIC_{50}	Res.
3	5.63	5.81	-0.18	5.71	-0.08
4	6.51	6.87	-0.36	6.80	-0.29
5	6.37	6.77	-0.40	6.62	-0.25
6	7.11	6.72	0.39	6.83	0.28
8	6.89	6.32	0.57	6.42	0.47
9	5.79	6.36	-0.57	6.17	-0.38
10	5.01	5.39	-0.38	5.12	-0.11
11	5.63	5.64	-0.01	5.60	0.03

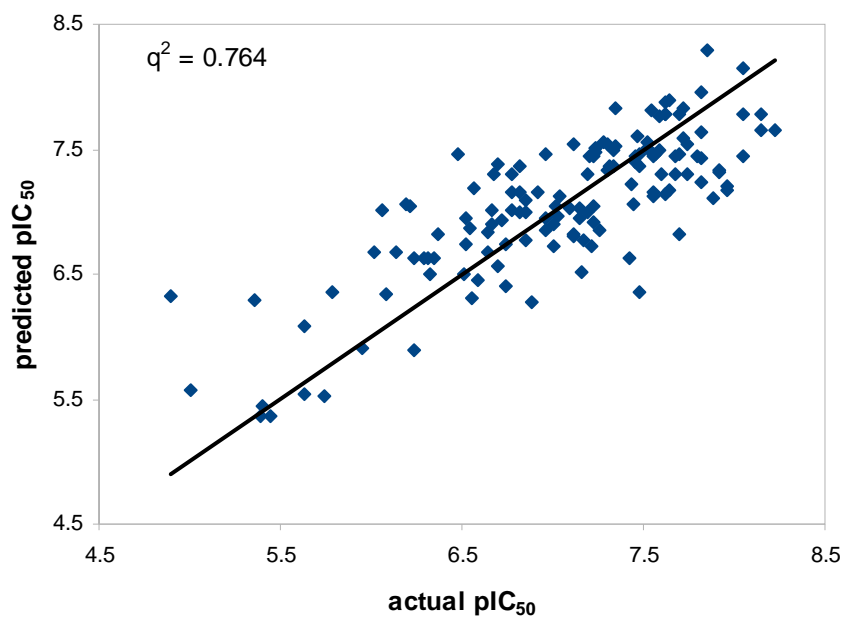
Cpd	Actual pIC_{50}	Pred. pIC_{50}	Res.	Est. pIC_{50}	Res.
90	7.23	6.94	0.29	7.02	0.21
91	6.82	7.12	-0.30	7.01	-0.19
92	6.77	7.33	-0.56	7.16	-0.39
93	7.01	7.01	0.00	7.01	0.00
94	7.55	7.35	0.20	7.38	0.17
95	6.85	7.15	-0.30	7.10	-0.25
96	7.24	7.57	-0.33	7.55	-0.31
97	7.23	7.55	-0.32	7.50	-0.27

Cpd	Actual pIC ₅₀	Pred. pIC ₅₀	Res.	Est. pIC ₅₀	Res.	Cpd	Actual pIC ₅₀	Pred. pIC ₅₀	Res.	Est. pIC ₅₀	Res.
12	5.39	5.17	0.22	5.17	0.22	98	7.60	7.33	0.27	7.39	0.21
14	5.40	5.08	0.32	5.08	0.32	99	7.35	7.72	-0.37	7.63	-0.28
15	7.96	7.29	0.67	7.36	0.60	101	6.70	7.37	-0.67	7.22	-0.52
16	6.19	6.79	-0.60	6.52	-0.33	102	8.05	7.52	0.53	7.63	0.42
17	7.24	7.60	-0.36	7.53	-0.29	103	8.15	7.51	0.64	7.68	0.47
19	6.48	6.58	-0.10	6.60	-0.12	104	7.52	7.47	0.05	7.47	0.05
21	7.89	7.35	0.54	7.47	0.42	105	7.68	7.30	0.38	7.47	0.22
22	6.82	7.39	-0.57	7.29	-0.47	106	8.22	7.57	0.65	7.73	0.49
23	6.29	6.45	-0.16	6.40	-0.11	108	7.82	7.63	0.19	7.64	0.18
24	6.24	6.58	-0.34	6.41	-0.17	109	7.52	7.55	-0.03	7.53	-0.01
25	6.35	6.85	-0.50	6.64	-0.29	110	7.48	7.40	0.08	7.42	0.07
26	6.72	7.34	-0.62	7.28	-0.56	111	6.70	6.53	0.18	6.61	0.09
27	6.08	6.41	-0.33	6.23	-0.15	113	7.02	7.01	0.01	6.97	0.05
28	6.24	6.02	0.22	6.08	0.16	114	6.77	6.97	-0.20	6.86	-0.09
31	6.59	6.35	0.24	6.44	0.15	116	7.19	7.32	-0.13	7.12	0.07
32	7.48	7.36	0.12	7.41	0.07	117	6.96	6.85	0.12	6.82	0.14
33	6.66	6.75	-0.09	6.59	0.07	118	6.64	6.88	-0.24	6.80	-0.16
34	7.33	7.77	-0.44	7.60	-0.27	119	7.04	7.24	-0.20	7.16	-0.12
35	6.68	7.35	-0.67	7.26	-0.58	120	6.92	7.11	-0.19	7.01	-0.09
36	6.66	6.86	-0.20	6.87	-0.21	121	6.55	6.56	-0.01	6.55	0.00
37	7.26	6.81	0.45	6.89	0.37	122	7.16	6.58	0.58	6.86	0.30
38	6.64	6.52	0.12	6.57	0.07	123	6.96	7.19	-0.23	6.78	0.18
39	6.06	6.50	-0.44	6.35	-0.29	125	7.00	6.72	0.28	6.78	0.22
41	6.74	6.37	0.37	6.44	0.30	126	6.85	6.98	-0.13	6.83	0.02
42	5.36	6.25	-0.89	5.66	-0.30	127	6.54	6.79	-0.25	6.72	-0.18
45	6.21	6.38	-0.17	6.31	-0.10	128	7.15	7.10	0.05	7.18	-0.03
46	6.52	7.05	-0.53	6.97	-0.45	129	7.19	7.09	0.11	7.19	0.00
47	7.15	7.14	0.01	7.17	-0.02	132	7.23	6.99	0.24	7.07	0.16
49	6.14	6.75	-0.61	6.65	-0.51	133	6.52	6.76	-0.24	6.65	-0.13
52	5.74	5.78	-0.04	5.77	-0.03	134	7.09	6.90	0.19	6.99	0.10
53	5.44	5.30	0.14	5.27	0.17	135	7.21	7.00	0.21	7.12	0.09
54	7.17	6.94	0.23	6.98	0.19	136	7.03	7.12	-0.09	7.22	-0.19
57	5.95	6.27	-0.32	6.10	-0.15	140	7.46	7.41	0.05	7.68	-0.22
58	6.82	6.97	-0.15	6.99	-0.17	141	7.85	8.01	-0.16	8.05	-0.20
59	7.55	7.16	0.39	7.25	0.30	142	7.64	7.89	-0.25	7.75	-0.11
60	7.55	7.16	0.39	7.37	0.18	143	8.05	8.16	-0.11	8.08	-0.03
61	7.92	7.18	0.74	7.28	0.64	145	7.92	7.38	0.54	7.63	0.30

Cpd	Actual pIC ₅₀	Pred. pIC ₅₀	Res.	Est. pIC ₅₀	Res.	Cpd	Actual pIC ₅₀	Pred. pIC ₅₀	Res.	Est. pIC ₅₀	Res.
62	7.64	7.14	0.50	7.25	0.39	146	8.15	8.04	0.11	8.31	-0.16
63	7.62	7.21	0.41	7.30	0.32	147	7.68	7.45	0.23	7.52	0.16
64	6.31	6.86	-0.55	6.77	-0.46	148	7.62	7.75	-0.13	7.74	-0.12
66	7.70	7.00	0.70	7.14	0.56	149	8.05	7.71	0.34	7.94	0.11
67	7.55	7.46	0.09	7.49	0.06	152	7.80	7.47	0.33	7.57	0.24
68	7.35	7.52	-0.17	7.50	-0.15	153	7.31	7.39	-0.08	7.31	0.00
69	7.82	7.42	0.40	7.55	0.27	154	7.30	7.39	-0.09	7.31	-0.01
70	7.70	7.44	0.26	7.51	0.19	155	7.43	7.28	0.15	7.27	0.16
71	7.46	7.25	0.21	7.29	0.17	156	7.47	7.66	-0.19	7.44	0.03
72	6.57	7.24	-0.67	7.06	-0.49	157	7.44	7.21	0.23	7.34	0.11
73	6.02	6.23	-0.21	6.02	0.00	159	7.62	7.74	-0.12	7.71	-0.09
74	7.12	6.84	0.29	6.85	0.27	160	7.72	7.57	0.15	7.58	0.14
75	7.72	7.93	-0.21	7.92	-0.20	161	7.59	7.55	0.04	7.61	-0.02
76	6.96	6.97	-0.01	6.96	0.00	163	7.59	7.66	-0.07	7.67	-0.08
77	7.57	7.38	0.19	7.51	0.06	164	7.24	7.43	-0.19	7.37	-0.13
78	7.54	7.73	-0.19	7.72	-0.18	165	7.12	7.55	-0.43	7.29	-0.17
79	7.74	7.44	0.30	7.61	0.13	167	7.74	7.50	0.24	7.72	0.02
81	6.85	6.90	-0.05	6.91	-0.06	168	7.62	7.68	-0.06	7.75	-0.13
82	6.32	6.92	-0.60	7.05	-0.73	169	7.82	7.94	-0.12	7.97	-0.15
83	6.74	6.62	0.12	6.68	0.06	170	7.70	7.69	0.01	7.69	0.01
84	7.42	6.67	0.75	6.88	0.54	171	7.48	7.28	0.20	7.72	-0.24
87	7.30	7.56	-0.26	7.54	-0.24	173	7.34	7.53	-0.19	7.31	0.03
88	7.20	7.50	-0.30	7.44	-0.24	174	7.82	7.62	0.20	7.96	-0.14
89	7.28	7.57	-0.29	7.51	-0.23						

*Cpd = Compound Number, Pred. pIC₅₀ = Predicted pIC₅₀, Est. pIC₅₀ = Estimated pIC₅₀ and Res. = Residual Value

(A)



(B)

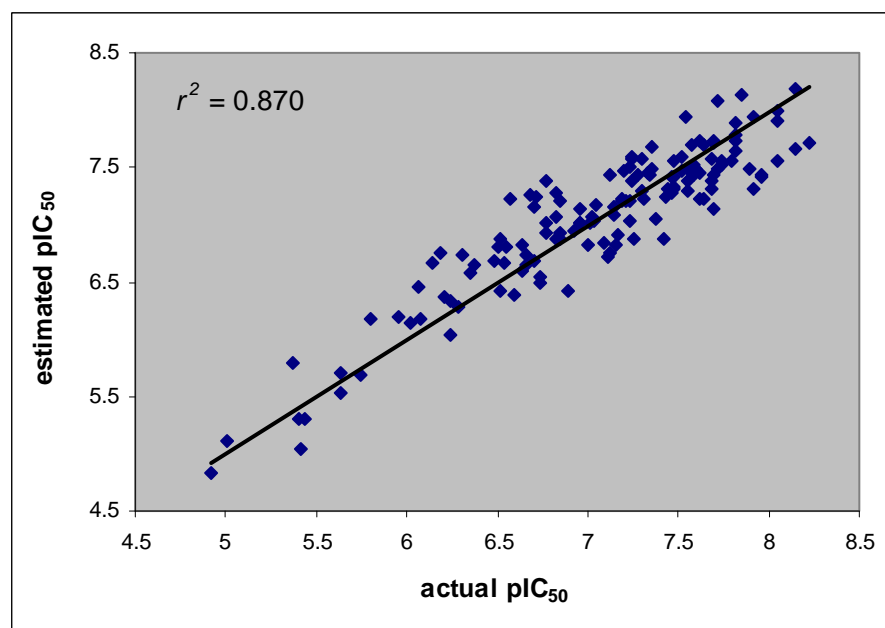


Figure 4.4 (A). Correlation between the actual pIC_{50} and the predicted pIC_{50} obtained with the best CoMFA model 10. (B). Correlation between the actual pIC_{50} and the estimated pIC_{50} of the best CoMFA model 10.

Table 4.5. Actual pIC₅₀, predicted pIC₅₀ and residual values of the 30 test set compounds.

Cpd	Actual pIC ₅₀	Predicted pIC ₅₀	Residual
1	7.01	6.98	0.03
2	5.39	5.92	-0.53
7	6.59	7.17	-0.58
18	7.48	7.18	0.30
20	7.64	7.47	0.17
29	7.22	7.44	-0.22
30	7.62	7.73	-0.11
43	4.30	5.37	-1.07
44	7.05	7.03	0.02
50	5.48	5.70	-0.22
51	5.92	6.64	-0.72
56	4.50	5.29	-0.79
65	7.38	7.39	-0.01
80	7.31	7.37	-0.06
85	6.09	6.70	-0.61
86	6.24	6.76	-0.53
100	8.10	7.51	0.58
111	7.57	7.57	0.00
115	6.85	6.75	0.10
124	6.44	6.81	-0.37
130	6.96	6.79	0.17
131	7.27	7.15	0.12
137	7.13	7.38	-0.25
138	7.17	7.23	-0.05
150	7.72	7.68	0.04
151	7.89	7.63	0.25
160	7.52	7.47	0.05
162	7.46	7.03	0.42
166	7.24	7.55	-0.31
172	6.80	7.56	-0.77

(C)

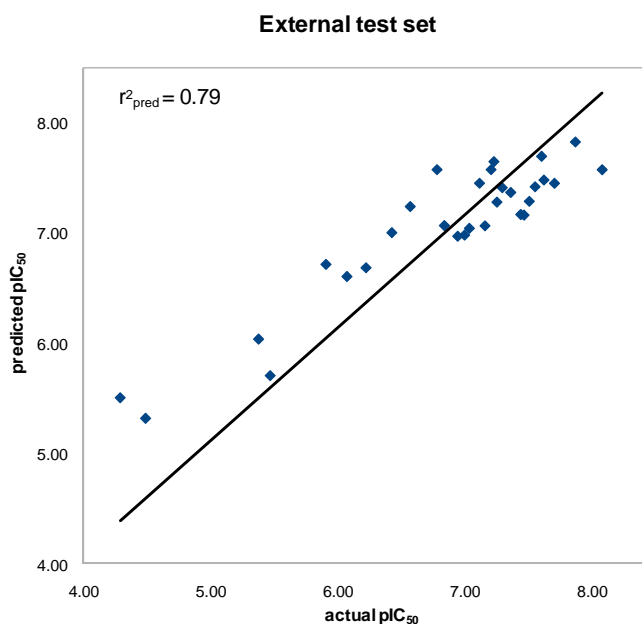


Figure 4.4 (C) Prediction of the external test set containing 30 inhibitors.

4.1.2.3 CoMFA Graphical Contour Plot

The subsequent PLS analysis, using the energy fields as descriptors and the biological activity as the dependent variable, can highlight the relative importance for affinity of certain types of interaction and certain regions around the compounds. Since the structure of the Wee1 kinase is known, the results obtained by the 3D-QSAR analysis were compared with the geometry and properties of the binding pocket. It is necessary to note, that in general, such comparison should be attempted carefully. In receptor-based 3D-QSAR models, the variance observed in the field values reflects both the structural diversity of the ligands and their diverse location within the binding site. The PLS coefficient contour maps can by no means be regarded as a set of low resolution picture of the binding site, since the contour maps reflect only those regions in space, where the ligand-probe interaction energy is correlated with a variance of the biological activity. However, it provides an opportunity to interpret features indicated in the contour maps with respect to the protein environment and to check whether the variance in the field values corresponds to regions known to be important for inhibitor binding. We superimposed the coefficient contour maps and the Wee1 ATP binding pocket. Figures 4.5 (A) and 4.5

(B) show the plot of the PLS coefficients for the steric and electrostatic field, respectively.

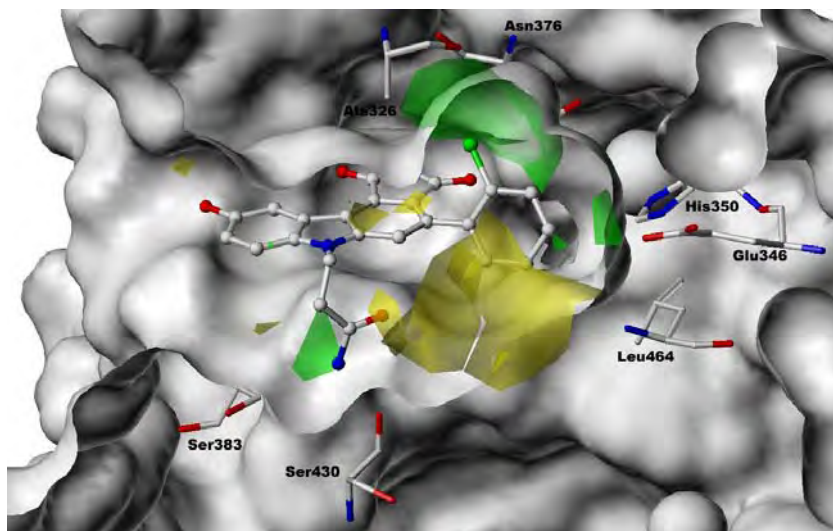


Figure 4.5 (A). CoMFA steric field projected on the Wee1 binding pocket. The docked inhibitor **106** is displayed in ball and stick. The Connolly molecular surface of the enzyme is shown in white. (Color code: favoured, green (contribution level 80%); disfavoured yellow (contribution level 20 %).

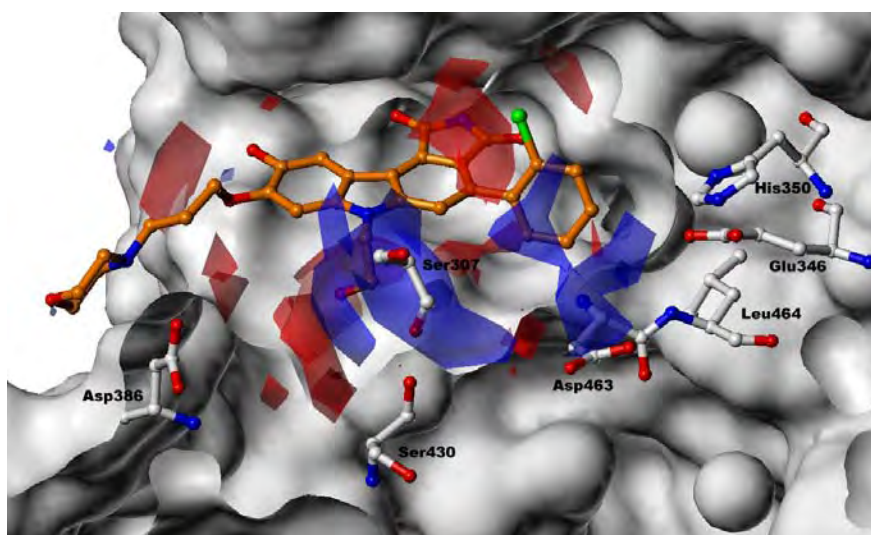


Figure 4.5 (B). CoMFA electrostatic field overlaid on the docked inhibitor **169** (colored orange). The Connolly molecular surface of the enzyme is shown in white. (Color code: increase in positive charge favoured, blue (contribution level 80%); increase in negative charge favoured red (contribution level 20 %).

The steric interaction is represented by green and yellow contours, in which green-coloured regions indicate areas where increased steric bulk is associated with enhanced activity, and yellow regions suggest areas where increased steric bulk is unfavourable. Electrostatic interaction is indicated by red and blue contours, among which blue-coloured regions show areas where more positively charged groups are favoured, and red regions highlight areas where groups with more negative partial charges are favoured.

A large favourable steric field (green) is observed around Ala326 and Asn376 next to the ortho position of the phenyl ring. Potent inhibitors without further activity-increasing substituents at the pyrrolocarbazole core, such as **21**, **29**, **30**, **32**, **34**, and **69** possess substituents at the ortho position of the phenyl ring. A second smaller favourable PLS region is found nearby the para position of the phenyl ring. However, the size of the hydrophobic pocket is restricted by His350, Glu346 and Leu464. Compounds with too large substituents at position 3, 4 and 5 of the phenyl ring (**42**, **43**, **55**, **56** and **57**) are able to interact with the hydrophobic pocket but show no optimal position to make the important hydrogen bonds to the hinge backbone region. Based on these results we suggest that adding a smaller electronegative substituent in position 6 of the phenyl ring would be favourable when combined with the 2-Cl substituent. The prediction of the fluoro analog of **106** (2-Cl, 6-FPhe) resulted in a higher pIC₅₀ value compared to the original inhibitor. The same positive effect was observed when adding the 6-F group to the phenyl ring of compound **103**.

An additional favourable region is located nearby Ser430 and Ser383. This result agrees well with the experimental data which show that adding a substituent to the nitrogen atom of the pyrrolo ring results in highly active inhibitors. The unfavourable regions are located above and behind the phenyl substituent, indicating that the conformational preference of the phenyl ring is important for high activity.

A favourable electrostatic field (Figure 4.5 (B), colored blue) for positively charged groups is found close to the positions 4 and 5 of the phenyl ring, indicating that more positively charged groups can favourably interact with the surrounding residues Glu346 and Asp463. Glu346 and Asp463 mainly contribute to the negative electrostatic potential in that region of the binding pocket. This observation is in agreement with the experimental data; compounds **44**, **47**, **54**, **58**, **65**, **66**, **70**, **75** and

80 bearing a substituent at position 4 or 5 of the phenyl with partial positive charge show higher pIC_{50} values compared to compounds **3, 6, 8, 10-15, 42, 43, 59, 60, 64** and **75** which possess neutral or more negatively charged substituents. Based on the contour plots, we suggest that inhibitors with 2-Cl, 4-OH, 6-FPhe or 2-F, 4-OH, 6FPhe ring attached to the pyrrolocarbazole core represent interesting compounds for future synthesis.

A second favourable electrostatic field for positively charged groups can be observed nearby Ser307 and Ser430. The docking of the potent inhibitor **106** showed that the amide group in the sidechain of **106** is donating a hydrogen bond to Ser430 (Figure 4.3 (A)). In addition, the highly active inhibitors **147, 148, 149, 150** and **152**, show also a hydrogen bond with Ser307. These results imply that the side chain of the phenylpyrrolocarbazole core play as key role in the interaction with Wee1. Inhibitors **122, 130-132, 136, 137** and **138** which contain partially positive charged groups in this region show high pIC_{50} values, whereas on the other site compounds **117, 118, 124-126, 127** and **133** bearing a neutral substituent at this position show lower pIC_{50} values. From the GOLD docking and the established CoMFA model, we suggest that partial positively charged groups at this position are favourable to establish strong interactions with Ser307 and Ser430.

A favourable electrostatic field for electronegative substituents (Figure 4.5 (B), colored red) is found nearby Tyr378 and Cys379. This is in agreement with the observation that the 9-OH substituent is important for the interaction with these two amino acid residues. Another favourable electrostatic field is observed above and below the planar pyrrolocarbazole ring system. These two fields result from the slightly different position of the aromatic system as derived from the docking study.

4.2 Structure-based prediction of activities of Wee1 Kinase inhibitors using the linear response MM/PBSA approach

In the present work, we used the LR-MM/PBSA approach to establish a model for predicting the binding free energies of Wee1 kinase inhibitors. Instead of using an ensemble of protein-ligand complexes derived from MD simulation[98], a single protein-inhibitor complex derived from refined docking solutions was used for generating the LR equation. In addition, the LIECE model[93] generated originally

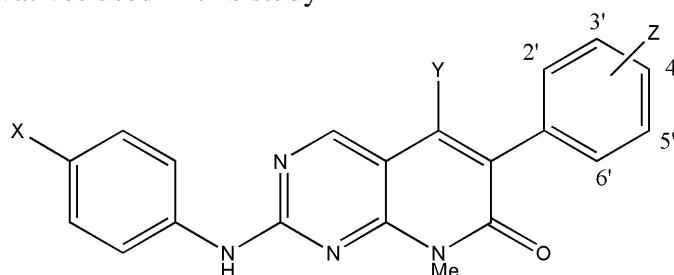
for CDK2, Lck and p38 inhibitors were applied to evaluate whether it could also be applied to predict the binding free energies of other series of kinase inhibitors (i.e. Wee1 kinase inhibitors used in this study) which were not used to set up the original model.

4.2.1 Material and Computational methods

4.2.1.1 Inhibitor data set

A set of pyrrolo[3,4-*c*]carbazole-1,3(2*H*,6*H*)-dione derivatives [62-64] (174 compounds) and of 2-anilio-6-phenylpyrido[2,3-*d*]pyrimidin-7(8*H*)-ones derivatives [61] (48 compounds) which represent Wee1 kinase inhibitors were considered in this study. The IC₅₀ values of inhibitors were converted into the experimental binding free energy ($\Delta G_{\text{exp.}}$) by using equation $\Delta G_{\text{exp.}} = -RT\ln(\text{IC}_{50})$. The structures of the inhibitors together with biological activities (IC₅₀ and pIC₅₀) and the ΔG_{exp} are shown in Chart 4.1 and Chart 4.2. The compounds were randomly separated into training set and test set. For the pyrrolocarbazole dataset, the same training set (144 compounds) and test set (30 compounds) as in our previous work (section 4.1)[157] were used. For the pyridopyrimidine derivatives 38 compounds were selected as training set and 10 compounds as test set.

Chart 4.2. Molecular structures of the 2-anilio-6-phenylpyrido[2,3-*d*]pyrimidin-7(8*H*)-ones derivatives used in this study



Test Set (38 compounds)

Cpd.	X	Y	Z	Activity		ΔG_{exp}
				IC ₅₀ (μ M)	pIC ₅₀	
175	O(CH ₂) ₂ NEt ₂	H	2',6'-diCl	0.165	6.78	-9.31
176	H	H	2',6'-diCl	2.6	5.59	-7.66
177	H	H	2',6'-diF	9.7	5.01	-6.88
179	H	H	2',6'-diMe	0.99	6.00	-8.24
180	H	H	2',6'-diCF ₃	41	4.39	-6.02
181	H	H	2',6'-diOH	27	4.57	-6.27
182	H	H	2'-Cl, 6'-F	2.4	5.62	-7.71
183	H	H	2'-Cl, 6'-Me	1.9	5.72	-7.85
186	H	H	2'-Me, 6'-Br	4.5	5.35	-7.34
189	H	H	2',6'-diCl, 3'-CH ₂ OH	3.5	5.46	-7.49
190	H	H	2',6'-diCl, 3'-CH ₂ NH ₂	31	4.51	-6.19
191	H	H	2',6'-diCl, 3'-COOH	3.2	5.49	-7.54
192	H	H	2',6'-diCl, 3'-CONH ₂	8.6	5.07	-6.95
193	H	H	2',6'-diCl, 3'-OH	0.074	7.13	-9.79
194	H	H	2',6'-diCl, 3'-NH ₂	2.6	5.59	-7.66
195	H	H	2',6'-diCl, 4'-Cl	8.6	5.07	-6.95
196	H	H	2',6'-diCl, 4'-OH	0.22	6.66	-9.14
198	H	H	2',6'-diCl, 4'-NH ₂	3.7	5.43	-7.45
199	H	H	2',6'-diCl, 4'-NHAc	36	4.44	-6.10
201	H	H	2',6'-diMe, 3',5'-diOMe	33	4.48	-6.15
202	H	H	2',6'-diMe, 3',5'-diOH	0.14	6.85	-9.41
203	CH ₂ CONH ₂	H	2',6'-diCl	0.12	6.92	-9.50
205	(CH ₂) ₄ CONH ₂	H	2',6'-diCl	0.26	6.59	-9.04
206	OCH ₂ CONH ₂	H	2',6'-diCl	0.25	6.60	-9.06
207	O(CH ₂) ₂ NEt ₂	H	2',6'-diMe	0.99	6.00	-8.24
208	O(CH ₂) ₃ COOH	H	2',6'-diCl	0.086	7.07	-9.70
209	(CH ₂) ₃ COO(CH ₂) ₂ Nmorph	H	2',6'-diCl	0.095	7.02	-9.64
210	(CH ₂) ₃ COO(CH ₂) ₂ NMe ₂	H	2',6'-diCl	0.124	6.91	-9.48
211	(CH ₂) ₃ COO(CH ₂) ₂ Npip	H	2',6'-diCl	0.142	6.85	-9.40
212	(CH ₂) ₃ COOH	H	2',6'-diCl	0.032	7.49	-10.29
213	CH ₂ CH(NH ₂)COOH	H	2',6'-diCl	0.09	7.05	-9.67
214	(CH ₂) ₃ tetrazole	H	2',6'-diCl	0.069	7.16	-9.83
215	O(CH ₂) ₂ NEt ₂	H	2',6'-diCl, 3'-OH	0.15	6.82	-9.36
218	O(CH ₂) ₃ COOH	H	2',6'-diCl, 4'-OH	0.04	7.40	-10.15
219	H	Me	2'-Cl	0.41	6.39	-8.77
220	O(CH ₂) ₂ NEt ₂	Me	2'-Cl	0.55	6.26	-8.59
221	H	Me	2',6'-diMe	1.2	5.92	-8.13
222	O(CH ₂) ₂ NEt ₂	Me	2',6'-diMe	0.54	6.27	-8.60

Test Set (10 compounds)

Cpd.	X	Y	Z	Activity		$\Delta G_{exp.}$
				IC ₅₀ (μ M)	pIC ₅₀	
178	H	H	2',6'-diBr	0.41	6.39	-8.77
184	H	H	2'-Cl, 6'-OMe	3.4	5.47	-7.50
185	H	H	2'-Cl, 6'-OH	1.5	5.82	-7.99
187	H	H	2'-OMe, 6'-OH	11	4.96	-6.80
188	H	H	2',6'-diCl, 3'-Me	50	4.30	-5.90
197	H	H	2',6'-diMe, 4'-OH	0.58	6.24	-8.56
200	H	H	2',6'-diCl, 3',5'-diOH	0.14	6.85	-9.41
204	(CH ₂) ₂ CONH ₂	H	2',6'-diCl	0.19	6.72	-9.22
216	O(CH ₂) ₃ COOH	H	2',6'-diCl, 3'-OH	0.04	7.40	-10.15
217	O(CH ₂) ₂ NEt ₂	H	2',6'-diCl, 4'-OH	0.08	7.10	-9.74

4.2.1.2 Ligand preparation and docking

Molecular docking of the pyrrolo[3,4-*c*]carbazole-1,3(2*H*,6*H*)-dione derivatives and of 2-anilino-6-phenylpyrido[2,3-*d*]pyrimidin-7(8*H*)-ones derivatives was carried out as described in our previous work (section 4.1.1.3)[157] where we could show that GOLD[105, 106] is able to correctly predict the binding model of Wee1 kinase inhibitors. All ligands were generated using Sybyl 7.2[147] and energy minimized using the MMFF94s force field[148] and the BFGS optimization method[149-152] with a convergence criteria of 0.05 kcal/mol. Protonation states of all ligands were assigned for pH7.1 using the protonate-3D method in MOE2008.10[112]. The compounds were docked into the X-ray structure of Wee1 kinase (PDB code 1X8B) using program GOLD (version 3.2)[105, 106]. The binding site was defined by a setting a radius of 20 Å around Cys379. As in our previous docking of Wee1 kinase inhibitors[157] the influence of including eight co-crystallized water molecules within the binding pocket was considered using the 'water toggle-mode' within GOLD

4.2.1.3 Energy minimization of docking poses

All docking poses were energy minimized using a combination of steepest descent, and conjugate gradient algorithm with a room mean square of gradient at 0.001 kcal/mol with the MOE2008.10[112] program. AM1-BCC charges[158] were assigned for inhibitors whereas the Amber99[153] force field was applied for the protein. The non-bonded cutoff was set at 16 Å. All heavy atoms of Wee1 kinase were tethered with a force constant 100 kcal/mol, whereas the inhibitors were relaxed

during the energy minimization process. The energy-minimized protein-inhibitor complex was subsequently used for calculating the van der Waals and electrostatic interaction energy as well as the solvation energy.

4.2.1.4 Interaction energy calculation

The minimized complexes were prepared for calculating the interaction energy using the LEaP module in AMBER10[159] by applying AM1-BCC[158] charges. The parameters from the general Amber force field (GAFF)[134] were used for the ligands while the Amber1999SB force field was used for Wee1 kinase. Then, van der Waals (E_{vdW}) and electrostatic (E_{ele}) interaction between ligand and protein in gas phase were calculated with an infinite cutoff using the SANDER module in AMBER10. The electrostatic free energy of solvation ($G_{ele-sol}$) was calculated with numerical solvation of the Poisson-Boltzmann (PB) equation as implemented in the *pbsa* program[141] in AMBER10. Default parameters for the PB solver such as a grid spacing at 0.5 Å, dielectric constants of 1.0 for solute and 80.0 for implicit PB solvent, solvent probe radius at 1.4 Å were used. The non-electrostatic free energy of solvation ($G_{nonele-sol}$) was calculated using the linear function of the solvent accessible surface area (SASA)[142]; $\Delta G_{nonele-sol} = \gamma SASA + b$ where γ and b were set at the default values ($\gamma = 0.00542$ kcal/mol*Å² and $b = 0.92$ kcal/mol).

4.2.1.5 Binding free energy calculation using LR-MM/PBSA and LIECE models

Two different approaches were applied for calculating the binding free energies of Wee1 kinase inhibitors. First, binding free energies were calculated by applying the LR-MM/PBSA method using the two following equations:

a four-parameter model:

$$\Delta G = \alpha \Delta E_{vdW} + \beta \Delta E_{ele} + \beta_1 \Delta G_{ele-sol} + \beta_2 \Delta G_{nonele-sol} + \gamma \quad (1)$$

a three-parameter model:

$$\Delta G = \alpha \Delta E_{vdW} + \beta \Delta E_{ele} + \beta_1 \Delta G_{sol} + \gamma \quad (2)$$

ΔG_{sol} is the summation of $\Delta G_{ele-sol}$ and $\Delta G_{nonele-sol}$. The coefficients (α , β , β_1 , and β_2) of LR-MM/PBSA model were generated by empirical fitting with the experimental binding free energy (ΔG_{exp}). Second, LIECE models taken from the literature[93]

were applied for predicting the binding free energies of the same Wee1 kinase inhibitors. These LIECE models were originally generated for predicting CDK2, Lck and p38 inhibitors using the CHARMM22 force field and the PBEQ method within CHARMM. The entropy contribution was not taken into account for both methods because of its high computational demand. Since only the relative binding free energies should be analyzed, this approach is acceptable. In addition, the previous work[93] has shown that an improvement could not be obtained when including the the entropy contribution.

4.2.2 Results and discussion

4.2.2.1 Binding mode of pyridopyrimidine derivatives

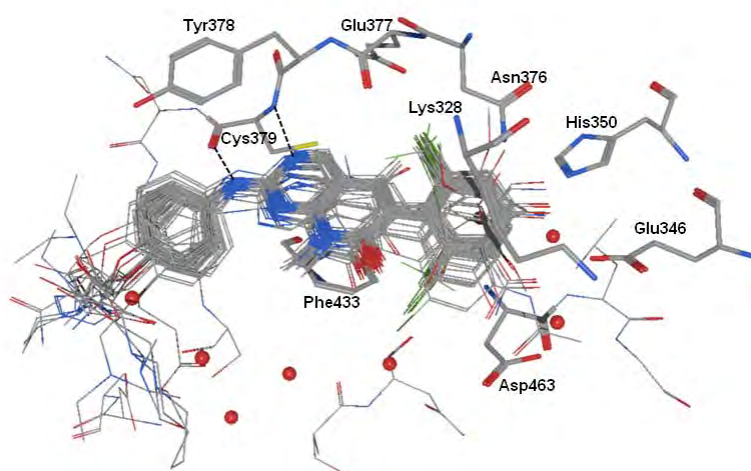


Figure 4.6. Docking solutions of all compounds of pyridopyrimidine derivatives projected on the binding pocket of Wee1 kinase. Hydrogen bonds are shown as dashed line, cocrystallized water molecules are presented as red balls. Ligands are displayed as line whereas important amino acid residues are shown as balls and stick.

In a previous docking study we could already show that GOLD is able to correctly reproduce the available X-ray structures of Wee1 in complex with inhibitors [157]. Therefore in the next step we analyzed the docking solutions of all pyridopyrimidine derivatives studied in the present work. Most compounds gave only one binding mode (Figure 4.5), showing a bidentate H-bond between N-3 and N-2 of the aminopyrimidine ring with backbone atoms (NH and CO) of Cys379, locating at

the hinge region. In this binding mode, the attached phenyl ring is interacting with the hydrophobic pocket surrounded by His350, Lys328, Glu346 and Asp463 as displayed in Figure 4.6. This orientation is also observed in the crystal structure of compound 1 in 1X8B.pdb. Therefore, the docking solutions which showed this binding mode were selected.

Superimposition of the selected docking solution of all compounds is shown in Figure 4.6. All compounds show a conserved binding mode, including a bidentate H-bond interaction with Cys379 and the interaction with residues of the hydrophobic pocket. In addition, the interaction of an active compound (e.g. compound 218) was investigated. Besides of the H-bond interaction with Cys379, the OH at position 4 of phenyl ring form H-bond with Glu346 as showed in Figure 4.7 (A) and 4.7 (B). Moreover, π - π interaction between ligand and Phe433 was also found. The cocrystallized water molecules located at the binding pocket are also involved in a H-bond network between ligand and enclosed residue such as Ile305 (Figure 4.7 (A)). We then examined whether the calculated docking scores, (GoldScore and ChemScore) show a correlation with the experimental binding data. However, neither GoldScore nor ChemScore show significant correlation with ΔG_{exp} , as shown in Figure 4.8 (A) and 4.8 (B) ($r^2 = 0.14$ for GoldScore, and $r^2 = 0.02$ for ChemScore). Therefore, we applied the LR-MM/PBSA approach to predict the binding strength of inhibitors.

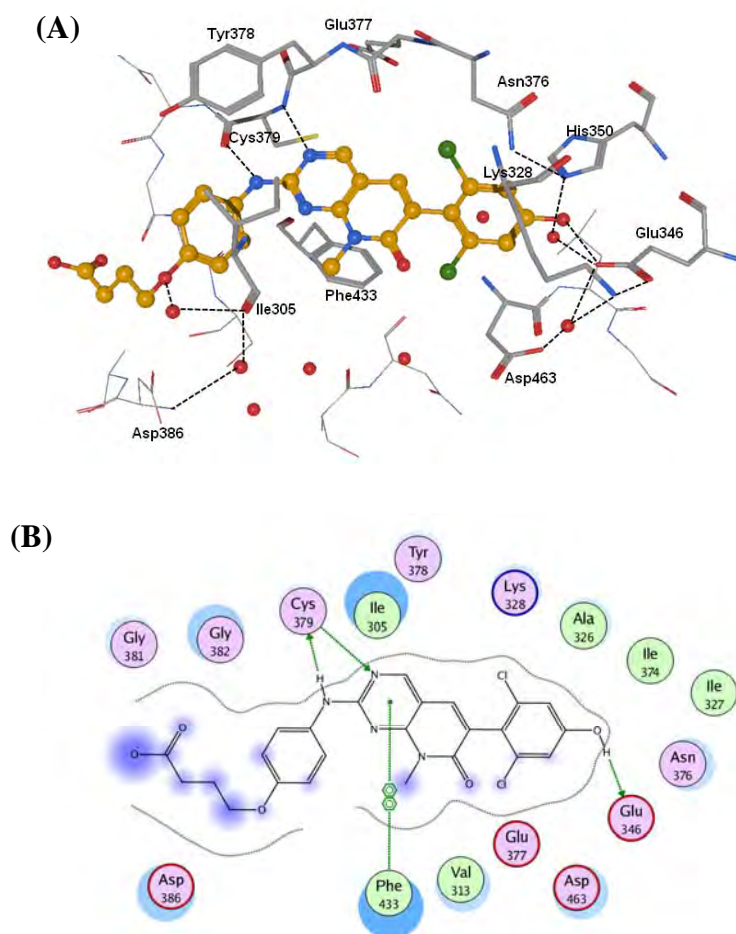


Figure 4.7. (A) GOLD docking solution for one of the potent inhibitors, compound 218, (B) Schematic representation of the interaction of compound 218 with the residues at the Wee1 binding site. Hydrogen bonds are indicated as arrows.

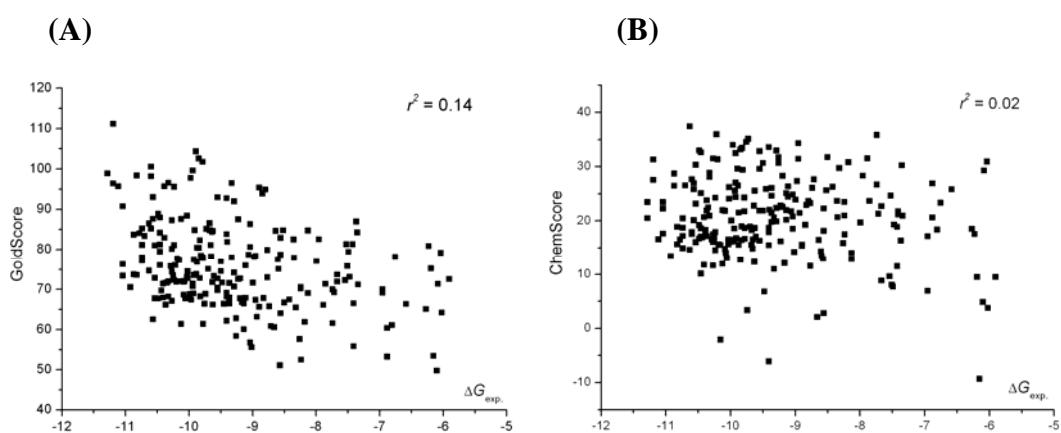


Figure 4.8. Correlation between ΔG_{exp} and GoldScore (A), and ChemScore (B) of all 222 compounds applied in this study

4.2.2.2 LR-MM/PBSA based prediction

The LR-MM/PBSA method was used to generate three different models: *i.*) model for the pyrrolocarbazole subset, *ii.*) model for the pyridopyrimidine subset and *iii.*) a general model for both subsets. A first LR-MM/PBSA model was generated using the four parameters, ΔE_{vdW} , ΔE_{ele} , $\Delta G_{\text{ele-sol}}$ and $\Delta G_{\text{nonele-sol}}$. Nevertheless, we found that the coefficient of the term $\Delta G_{\text{nonele-sol}}$ contributed quite small to the models. So we used only one descriptor for the solvation energy term which is ΔG^{sol} , a summation of $\Delta G_{\text{ele-sol}}$ and $\Delta G_{\text{nonele-sol}}$.

Table 4.6. Statistical Values of training set of LR-MM/PBSA model

Model	Num. of comps.	outliers	Statistical values			
			RMSE	r^2	XRMSE	q^2
Pyrrolocarbazole model						
1	144	-	0.61	0.64	0.64	0.62
2	140	13, 48, 71, 107	0.55	0.67	0.57	0.65
Pyridopyrimidine model						
3	38		0.86	0.52	0.97	0.41
4	35	177, 181, 199	0.65	0.67	0.76	0.58
General Model						
5	182	-	0.74	0.61	0.77	0.58
6	175	13, 48, 71, 107, 177, 181, 199	0.64	0.67	0.66	0.64

Num. of comps. = Number of compounds, RMSE = root mean square of error, XRMSE = root mean square of error derived from leave-one-out cross validation

The LR-MM/PBSA model for the pyrrolocarbazole subset was generated using 144 compounds as a training set which are the same as in the section 4.1. The statistical value obtained from this data set (model 1) are $r^2 = 0.64$, root mean square of error (RMSE) = 0.61 kcal/mol, $q^2_{\text{LOO}} = 0.62$ and leave-one-out cross validated root mean square of error (XRMSE) = 0.64 kcal/mol as summarized in Table 4.6. We found that compound **13, 48, 71** and **107** have high residual values ($\Delta G_{\text{exp.}} - \Delta G_{\text{est.}}$), and their residual values are 2.0 time higher than RMSE value of the model 1. Compound **13, 48,** and **107** were also the outlier as in our previous work (section 4.1)[157] where a 3D-QSAR model was established on the same data set. Therefore, a reduced model was generated by removing these four outliers. However, the statistical values derived from this data set (model 2) showed only a slight

improvement ($r^2 = 0.67$, RMSE = 0.55 kcal/mol, $q^2_{LOO} = 0.65$ and XRMSE = 0.57 kcal/mol). Therefore, the suitable model for the pyrrolocarbazole subset is model 1. The derived equation (model 1) of the pyrrolocarbazole subset is shown below.

$$\Delta G = 0.10877\Delta E_{vdW} + 0.06786\Delta E_{ele} + 0.06239\Delta G_{sol} - 10.91832 \quad (3)$$

The LR-MM/PBSA model for the pyridopyrimidine subset was built in the same way as described for the pyrrolocarbazole subset. A three-parameter model was initially generated using 38 compounds as a training set (chart 4.2). The correlation between observed and calculated data (model 3) yielded the following values: $r^2 = 0.52$, RMSE = 0.86 kcal/mol, $q^2_{LOO} = 0.41$ and XRMSE = 0.91 kcal/mol. Thus, the model for the pyridopyrimidine derivatives shows lower accuracy in calculating the binding data. We also removed for this data set the outliers (compounds **177**, **181** and **199**) having residual values higher than 2.0 times of the RMSE value. We analyzed why these compounds are outliers by visual inspection of their docking solutions. In case of compounds **177** and **181**, it is not clear why a fluoro substitution at position 2 (compound **177**) or a hydroxy group (compound **181**) results in lower activities. In contrast, several compounds such as **175**, **176**, **178**, **179**, **182**, **183**, **184**, and **185** contain an electron withdrawing group (-Cl, -F, -Br, or -OH) at the same position but show high activity. Therefore, the docking solutions and the calculated binding free energies of these compounds are not able to explain this behavior. Finally, a reduced model from the remaining 35 compounds was generated. The obtained model (model 4) gave significant improvement of the statistical values ($r^2 = 0.67$, RMSE = 0.65 kcal/mol, $q^2_{LOO} = 0.58$ and XRMSE = 0.76 kcal/mol). The final equation for this reduced data set is shown below:

$$\Delta G = 0.09395\Delta E_{vdW} - 0.00187 \Delta E_{ele} + 0.00726\Delta G_{sol} - 8.60966 \quad (4)$$

In the last step, an unified model was built using all 182 compounds (144 pyrrolocarbazoles and 38 pyridopyrimidines, respectively). Also here, the seven outliers were omitted. The statistical values of the final model 6 are: $r^2 = 0.67$, RMSE = 0.64 kcal/mol, $q^2_{LOO} = 0.64$ and XRMSE = 0.66 kcal/mol. The r^2 value indicates a good correlation between the $\Delta G_{exp.}$ and $\Delta G_{est.}$ In addition, the obtained q^2_{LOO} value suggests the robustness of the model for predicting novel compounds. The derived unified model is shown in equation 5 below:

$$\Delta G = 0.10505\Delta E_{vdW} + 0.05267\Delta E_{ele} + 0.05176 \Delta G_{sol} - 10.74762 \quad (5)$$

The correlation between the $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{est.}}$ derived from LR-MM/PBSA model 1, 4 and 6 are shown in Figure 4.9 (A), 4.9 (B) and 4.9 (C), respectively. A good correlation was observed in all three models, and most compounds show low residual values in the range of ± 1.0 kcal/mol (indicated by the dashed lines in Figure 4.9 (A), (B) and (C)).

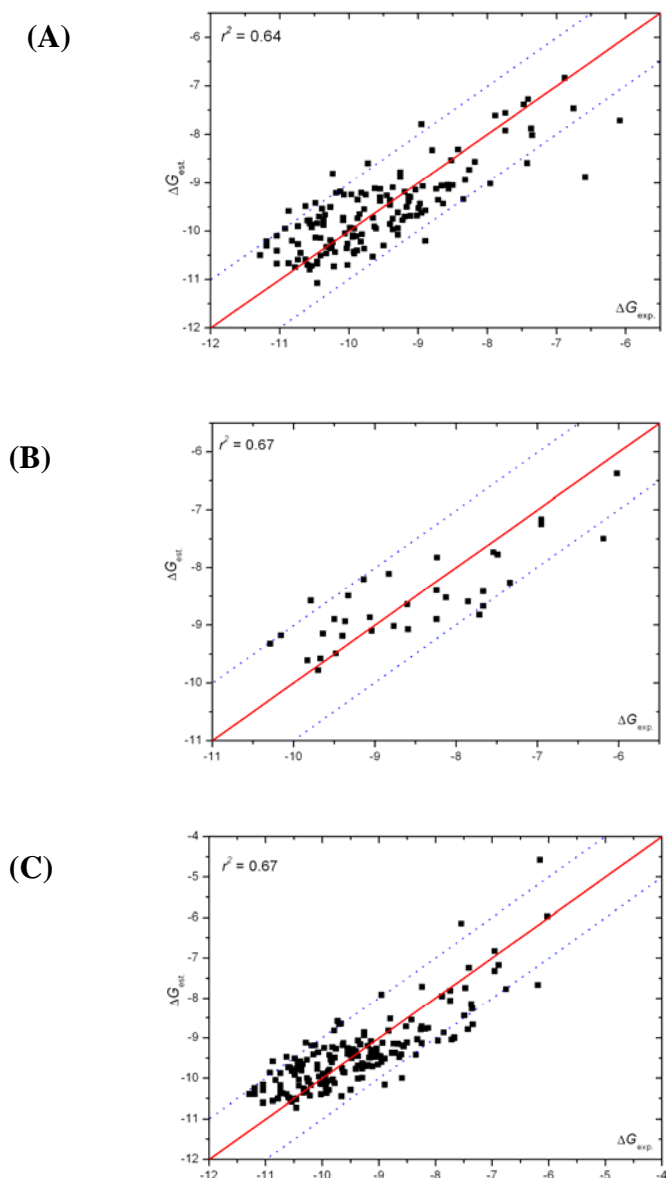


Figure 4.9. Correlation between $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{est.}}$ of the compounds in the training set of (A) pyrrolocarbazole using model 1, (B) pyridopyrimidine using model 4, and (C) the general model using model 6

In the next step, the LR-MM/PBSA models (model 1, 4 and 6) were used to predict the binding free energies of external test compounds. Therefore, 30 pyrrolocarbazoles and 10 pyridopyrimidines were predicted. The $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{pred.}}$ and residual values of each compound corresponding to the test set of pyrrolocarbazole subset, the pyridopyrimidine subset, and the general model are listed in Table 4.7, 4.8 and 4.9, respectively. $\Delta G_{\text{exp.}}$ shows a good correlation with $\Delta G_{\text{pred.}}$, yielding $r^2 = 0.54$, 0.42 and 0.57 for pyrrolocarbazole subset, pyridopyrimidine subset, and the unified data set as plotted in Figure 4.10, 4.11 and 4.12, respectively. The r^2 value of the test set of the pyridopyrimidine subset was found to be quite low. However, by considering the residual value ($\Delta G_{\text{exp.}} - \Delta G_{\text{pred.}}$) of all compounds in the test set, we found that only compound **187** shows a high residual value (1.90, displayed as red point in Figure 4.11), whereas the remaining compounds yielded $\Delta G_{\text{pred.}}$ close to $\Delta G_{\text{exp.}}$. Thus, this result shows that there is only one compound in the test set of pyridopyrimidine subset that was not correctly predicted. The derived result indicates the reliability of the model for predicting and ranking the relative binding free energies of related compounds. In addition, the results of the pyrrolocarbazole subset, and the unified data set demonstrate that $\Delta G_{\text{pred.}}$ of the test set compounds are in a good agreement with the actual values. In summary, the results derived for the test sets point out that the derived models are reliable and robust.

Table 4.7. Comparison between $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{pred.}}$, derived from model 1, of compounds in the test set of pyrrolocarbazole data set

Cpd.	$\Delta G_{\text{exp.}}$	$\Delta G_{\text{pred.}}$	Residual
1	-9.62	-9.28	-0.34
2	-7.41	-9.47	2.06
7	-9.04	-9.53	0.49
18	-10.27	-9.89	-0.38
20	-10.48	-10.35	-0.13
29	-9.91	-9.16	-0.75
30	-10.46	-9.73	-0.72
43	-6.04	-8.27	2.24
44	-9.68	-9.00	-0.68
50	-7.52	-8.17	0.65
51	-8.13	-8.88	0.76
56	-6.23	-7.57	1.34
65	-10.12	-9.51	-0.62
80	-10.03	-8.48	-1.55
85	-8.35	-8.80	0.45
86	-8.56	-8.94	0.38
100	-11.11	-10.16	-0.95
111	-10.39	-10.59	0.20
115	-9.41	-9.13	-0.28
124	-8.84	-9.79	0.95
130	-9.55	-9.82	0.27
131	-9.97	-10.26	0.28
137	-9.79	-10.33	0.55
138	-9.85	-9.93	0.09
150	-10.60	-10.26	-0.34
151	-10.82	-9.97	-0.85
158	-10.32	-10.75	0.43
162	-10.23	-10.69	0.46
166	-9.94	-11.37	1.43
172	-9.33	-9.45	0.12

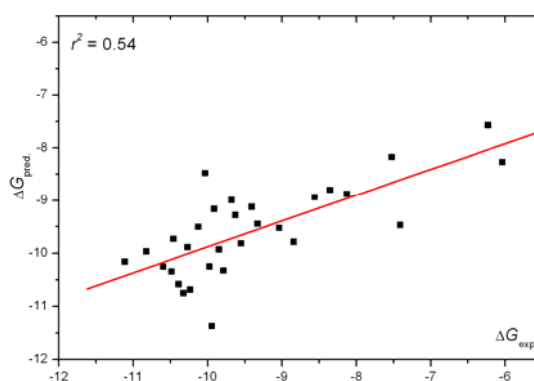
**Figure 4.10.** Correlation between $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{pred.}}$ of the compounds in the test set of pyrrolocarbazole subset.

Table 4.8. Comparison between $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{pred.}}$, calculated from model 4, of compounds in the test set of pyridopyrimidine data set

Cpd.	$\Delta G_{\text{exp.}}$	$\Delta G_{\text{pred.}}$	Residual
178	-8.77	-7.67	-1.09
184	-7.50	-8.29	0.79
185	-7.99	-8.73	0.73
187	-6.80	-8.70	1.90
188	-5.90	-6.96	1.05
197	-8.56	-7.97	-0.59
200	-9.41	-8.49	-0.92
204	-9.22	-9.03	-0.19
216	-10.15	-9.60	-0.55
217	-9.74	-8.68	-1.06

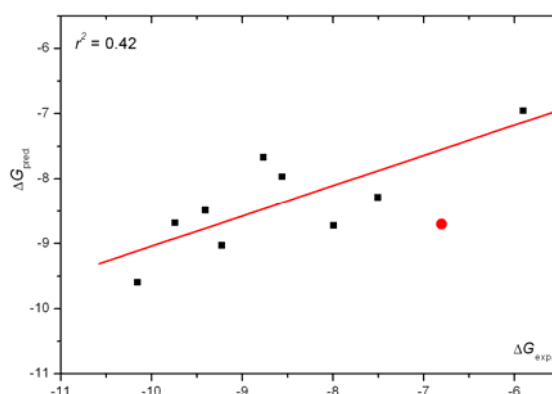


Figure 4.11. Correlation between $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{pred.}}$ of the compounds in the test set of pyridopyrimidine subset.

Table 4.9. Comparison between $\Delta G_{\text{exp.}}$ and predicted $\Delta G_{\text{pred.}}$, calculated from model 6, of compounds in the test set of the general model

Cpd.	$\Delta G_{\text{exp.}}$	$\Delta G_{\text{pred.}}$	Residual
1	-9.62	-9.27	-0.36
2	-7.41	-9.30	1.89
7	-9.04	-9.48	0.45
18	-10.27	-9.81	-0.45
20	-10.48	-10.19	-0.29
29	-9.91	-9.19	-0.72
30	-10.46	-9.69	-0.76
43	-6.04	-8.42	2.38
44	-9.68	-9.05	-0.63

50	-7.52	-8.42	0.90
51	-8.13	-8.85	0.72
56	-6.23	-7.81	1.58
65	-10.12	-9.48	-0.64
80	-10.03	-8.64	-1.39
85	-8.35	-8.89	0.54
86	-8.56	-9.03	0.47
100	-11.11	-10.09	-1.02
111	-10.39	-10.49	0.11
115	-9.41	-9.19	-0.21
124	-8.84	-9.68	0.84
130	-9.55	-9.89	0.34
131	-9.97	-10.21	0.24
137	-9.79	-10.38	0.60
138	-9.85	-10.13	0.29
150	-10.60	-10.31	-0.28
151	-10.82	-10.08	-0.74
158	-10.32	-10.44	0.12
162	-10.23	-10.48	0.25
166	-9.94	-10.93	0.99
172	-9.33	-9.34	0.02
178	-8.77	-7.95	-0.81
184	-7.50	-8.33	0.83
185	-7.99	-8.75	0.76
187	-6.80	-8.60	1.79
188	-5.90	-6.86	0.96
197	-8.56	-8.81	0.25
200	-9.41	-8.47	-0.93
204	-9.22	-9.42	0.19
216	-10.15	-8.40	-1.76
217	-9.74	-9.85	0.11

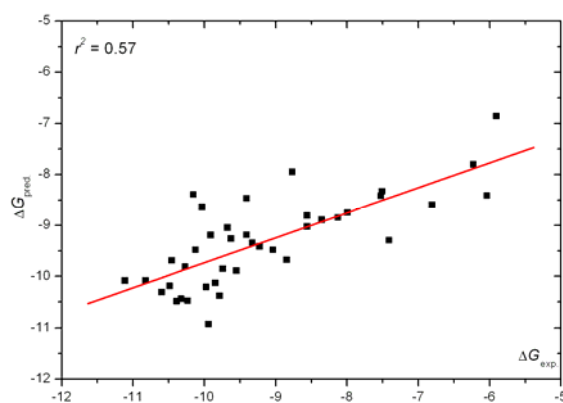


Figure 4.12. Correlation between ΔG_{exp} and ΔG_{pred} of the compounds in the test set of the general model

4.2.2.3 Linear interaction energy with continuum electrostatics (LIECE)

models

Table 4.10. Statistical values of results derived from applying LIECE models

Model	Data Set of Inhibitor of which Protein used for generating Model	α	β or β_1	β_2	r^2	RMSE
One-Parameter Model						
7	CDK2	0.2338	-	-	0.39	0.96
8	Lck	0.277	-	-	0.39	0.96
9	p38	0.2377	-	-	0.39	0.96
10	CDK2+ Lck	0.251	-	-	0.39	0.96
11	CDK2+ p38	0.2383	-	-	0.39	0.96
12	Lck+ p38	0.2513	-	-	0.39	0.96
13	CDK2+ Lck+ p38	0.2463	-	-	0.39	0.96
Two-Parameter Model						
14	CDK2	0.2866	0.0525	-	0.51	0.85
15	Lck	0.2735	0.0046	-	0.40	0.95
16	p38	0.2699	0.0264	-	0.46	0.90
17	CDK2+ Lck	0.3072	0.0657	-	0.54	0.84
18	CDK2+ p38	0.2632	0.0235	-	0.45	0.91
19	Lck+ p38	0.3033	0.0508	-	0.51	0.86
20	CDK2+ Lck+ p38	0.2898	0.0442	-	0.50	0.86
Three-Parameter Model						
21	CDK2	0.2395	0.075	0.0294	0.41	0.94
22	Lck	0.2446	0.1528	0.0076	0.27	1.05
23	p38	0.1827	0.1584	-0.0013	0.24	1.07
24	CDK2+ Lck	0.3118	0.044	0.062	0.50	0.87
25	CDK2+ p38	0.219	0.0439	0.0006	0.36	0.98
26	Lck+ p38	0.2939	0.1186	0.0584	0.43	0.92
27	CDK2+ Lck+ p38	0.2961	0.0325	0.0454	0.48	0.88

First we analyzed whether the already published parameters (Table 4.10) by Kolb P. *et al.*[93], can also be used to predict the ΔG_{exp} values of the studied Wee1 kinase inhibitors. Those parameters were generated from datasets of CDK2, Lck and p38 inhibitors using the CHARMM22 force field and the PBEQ method included in the CHARMM program. The Wee1 inhibitors can thus be considered as test set to evaluate the efficiency of the original LIECE models/parameters.

All one-parameter models (model 7-13, shown in Table 4.10) using different α values taken from the work of Kolb *et al.* gave more or less the same correlation coefficient ($r^2 = 0.39$) and RMSE values (0.96 kcal/mol). The high similarity of the results of model 7-13 is due to the small deviation of the α coefficient among the different models (0.2338 -2.770). In general, all one-parameter models (model 7-13) yielded quite low correlation coefficients and high RMSE values. Interestingly, the correlation increases obviously when the electrostatic term is taken into account. For example, the two-parameter model (model 14-20), in which the electrostatic free energy (ΔG_{ele}) is included, yielded r^2 values in the range of 0.40-0.54 and RMSE values between 0.84 and 0.95 kcal/mol. The three-parameter models (model 21-27), in which ΔG_{ele} is decomposed into two terms (electrostatic interaction and solvation energy), showed lower r^2 values (0.24-0.50) and higher RMSE values (0.87 to 1.07 kcal/mol).

By applying the LIECE models for predicting the binding free energies of the 222 Wee1 kinase inhibitors, we observed similar accuracy as in the published work[93]. The two-parameter model (model 20) based on three different kinases gave slightly higher correlation coefficients ($r^2 = 0.50$) than the three-parameter model (model 27, $r^2 = 0.48$) generated from the same data set of kinases. However, the highest correlation ($r^2 = 0.54$) between $\Delta G_{\text{pred.}}$ and $\Delta G_{\text{exp.}}$ was observed by using model 17, a two-parameter model using two proteins (CDK2 and Lck data set) to generate the model. This might be due to the chemical similarity of the inhibitors of CDK2 and Lck, which were applied for generating the LIECE model, and Wee1 kinase inhibitors, used in this study. However, two-parameter model based on three kinases (model 20) was generated using inhibitors covering more chemical space of the kinase inhibitors than model 17. In addition, r^2 derived from model 20 and model 17 are not much significant different. Therefore, the LIECE model 20 is more appropriate to use as a general model to predict binding free energies of other kinase inhibitors. Noted that even though the correlation between the $\Delta G_{\text{exp.}}$ and $\Delta G_{\text{pred.}}$ of Wee1 kinase inhibitors using model 20 shows a good correlation ($r^2 = 0.50$, RMSE = 0.86 kcal/mol), most $\Delta G_{\text{pred.}}$ values derived from model 20 are positive and have high residual values as demonstrated in Figure 4.13. This might be due to the difference between the force field and method that was used in this study and for which applied

to generate the LIECE models. These results imply that these LIECE models can be applied to rank a relative binding free energy (not the absolute binding free energy) or can be called as LIECE score for ranking other kinases inhibitors which can be useful for ranking compounds retrieved from virtual screening for kinase inhibitors.

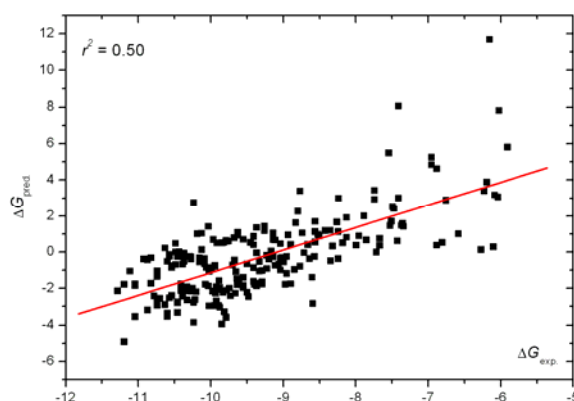


Figure 4.13. Correlation between ΔG_{exp} and ΔG_{pred} of the all 222 compounds derived by using LIECE model 20.

4.2.2.4 Enrichment Study

In the previous section, the correlation between the estimated binding free energies derived from LR-MM/PBSA and LIECE model, with their experimental data was investigated. Beside the good correlation between the estimated and the experimental values, the usefulness and the effectiveness of these models for virtual screening were also examined. Therefore, enrichment factors, showing the fraction of the active compounds recovered in respect to the fraction of the screened database, were calculated. This approach tests the ability of these derived models to discriminate true hits from decoys of the databases. Two different databases were used for the validation study: the first database contains 980 randomly selected compounds from the Chembridge KINAsset collection (Chembridge Corporation, San Diego, USA). This kinase library is a computationally selected collection of 11,000 compounds utilizing a ligand-based pharmacophore selection method and thus represents a kinase-focussed library. The second dataset contains 980 compounds randomly selected from the directory of useful decoys (DUD)[160]. Here we took all decoy sets available for kinase inhibitors which includes EGFR, FGFR1, P38, PDGFR, SRC, TK, VEGFR2 [160]. Then, these two databases were filled with six

different datasets of active Wee1 kinase inhibitors. One dataset contains the 20 most active inhibitors and the other 5 datasets were generated by randomly selecting 20 actives from the large dataset of 222 Wee1 inhibitors. In total, 12 different datasets, each of them contain 1,000 compounds, were applied for the enrichment study. The DUD provides a well-defined, unbiased set of active and decoy compounds, and the chemical structures of the KINAsset library have been designed using known kinase inhibitor scaffolds. To discriminate actives from inactive compounds with similar chemical structure is a challenge and non-trivial task in virtual screening. Therefore, the chosen compounds represent suitable datasets to evaluate the performance of LR-MM/PBSA and LIECE models. The enrichment factor (EF)[161] for each dataset was calculated according to the formula shown below;

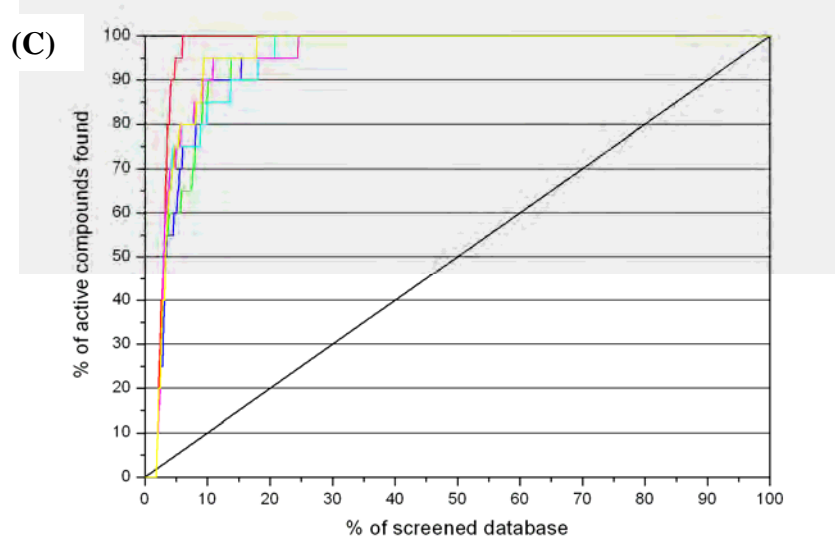
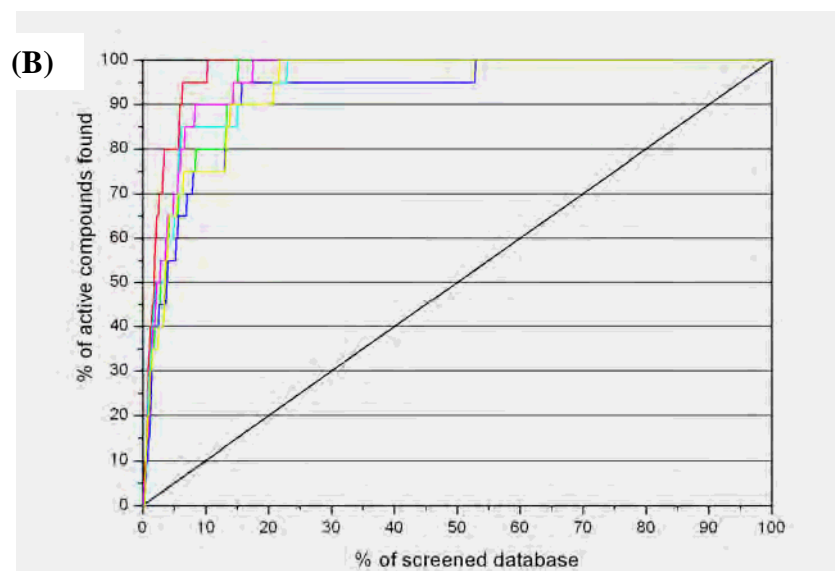
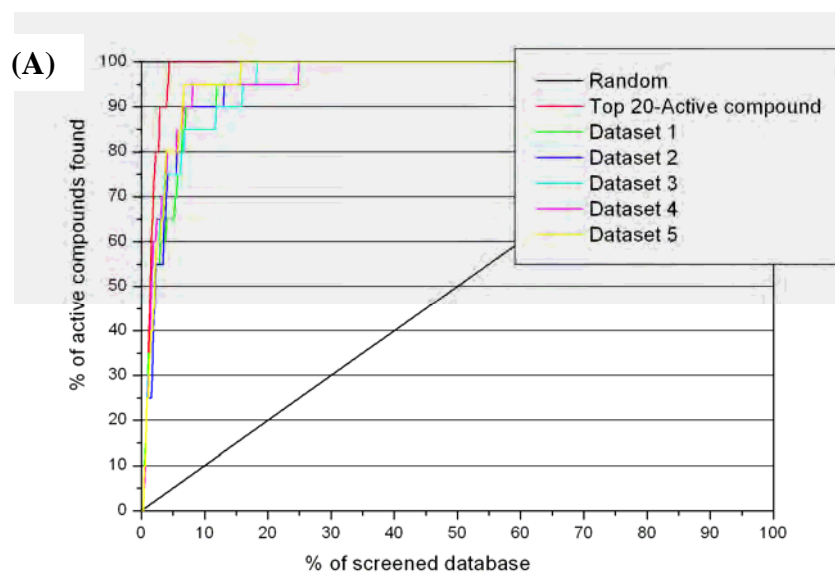
$$EF = \{N_{\text{total}}/N_{\text{sampled}}\} * \{\text{Hits}_{\text{sampled}}/\text{Hits}_{\text{total}}\}$$

where N_{total} and N_{samples} represent the total number of compounds in the database (1,000 compounds) and the number of the screened compounds, respectively. $\text{Hits}_{\text{sampled}}$ and $\text{Hits}_{\text{total}}$ are the number of found actives and the total number of actives (20 compounds). The selection of compounds from the set of known active inhibitors of Wee1 kinase can lead to artificially high EF values. Thus, to avoid this bias, different databases containing different active compounds were used to calculate the EF value, and the average value of EF was also considered. The enrichment factors were considered at 5% and 10% of the screened databases. Note that the maximum values of EF are 20 and 10 at the 5% and 10% of the screened databases, respectively. When these maximum values are obtained, it means that all active compounds are recovered at the considered fraction of the screened database.

Table 4.11. Enrichment Factor (EF) of each dataset, containing decoys from Chembridge KINAsset with 6 different sets of active Wee1 kinase inhibitors and also decoys from DUD combining with the same datasets from active Wee1 kinase inhibitors, at 5% and 10% of screened database using LR-MM/PBSA model 6 and LIECE model 20

	LR-MM/PBSA (Model 6)		LIECE (Model 20)	
	5%	10%	5%	10%
Chembridge Kinaset enriched with Wee1 inhibitors				
Top 20 actives	20.0	10.0	16.0	9.5
Dataset 1	13.0	9.0	13.0	8.0
Dataset 2	15.0	9.0	11.0	7.5
Dataset 3	15.0	8.5	13.0	8.5
Dataset 4	16.0	9.5	14.0	9.0
Dataset 5	16.0	9.5	13.0	7.5
Average	15.8	9.3	13.3	8.3
DUD enriched with Wee1 inhibitors				
Top 20 actives	19.0	10.0	16.0	9.5
Dataset 1	12.0	8.5	13.0	8.0
Dataset 2	12.0	9.0	11.0	7.5
Dataset 3	15.0	8.5	12.0	8.5
Dataset 4	15.0	9.0	13.0	9.0
Dataset 5	15.0	9.5	13.0	7.5
Average	14.7	9.1	13.0	8.3

The enrichment factors are summarized in Table 4.11. Results revealed that most of the top 20 actives were recovered within 5% of the database by using the LR-MM/PBSA model 6. On the other hand, the LIECE model 20 was able to identify 16 active compounds within the first 5%. At 10% of the screened database, it was found that all active compounds were recovered by using the LR-MM/PBSA model 6, and 19 compounds were received by using LIECE model 20. In general, the LR-MM/PBSA model yielded slightly better EFs compared to the LIECE model when considering all 12 datasets. Furthermore, the derived results indicate that most of the actives can be found within the first 10% of database screened. The percentage of the screened database versus the percentage of the actives found are displayed in Figure 4.14 (A) – (D). From Figure 4.14, it can be clearly seen that all depicted EF curves for model 6 and model 20 are well above the curve for random selection. The obtained results imply that the LR-MM/PBSA model and the LIECE model show a good performance for ranking and separating active compounds from decoys.



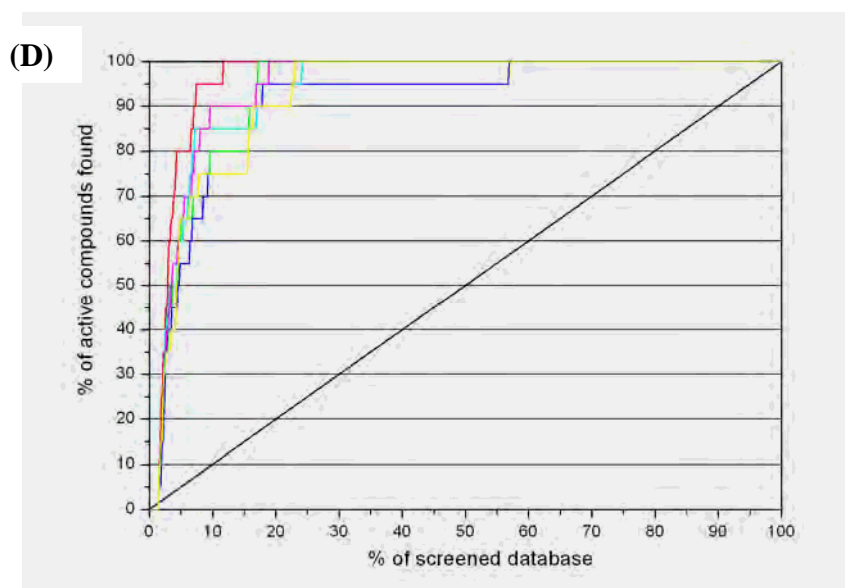


Figure 4.14. EF curves of datasets of Chembridge Kinaset decoys with 6 different datasets from active Wee1 kinase inhibitors ranked by LR-MM/PBSA model 6 (A) and LIECE model 20 (B), and the EF curves of datasets of decoys from DUD with 6 different datasets from active Wee1 kinase inhibitors scored by LR-MM/PBSA model 6 (C) and LIECE model 20 (D)

4.3 Virtual Screening for novel Wee1 Kinase inhibitors

4.3.1 Materials and Computational Methods

Virtual screening was carried out using the same approach as in section 3.3. The pharmacophore model was also generated using LigandScout program [146]. The derived model was then used to search for compounds in the Chembridge database that match with this pharmacophore model. The hit compounds were further screened by using Lipinski's rule of five [101], ranking by docking score and calculating binding free energy LR-MM/PBSA model. The Lipinski's rule of five was already explained in section 3.3.1.3. Molecular docking was performed using all the same protocols as described in section 4.1.1.3 and 4.2.1.1. Finally, binding free energy of these hit compounds were computed using the same approach as explained in section 4.2. The LR-MM/PBSA model 6 (the general model) was used to predict and rank the binding free energy of the hit compounds.

4.3.2 Results and Discussion

4.3.2.1 Pharmacophore Models Generation

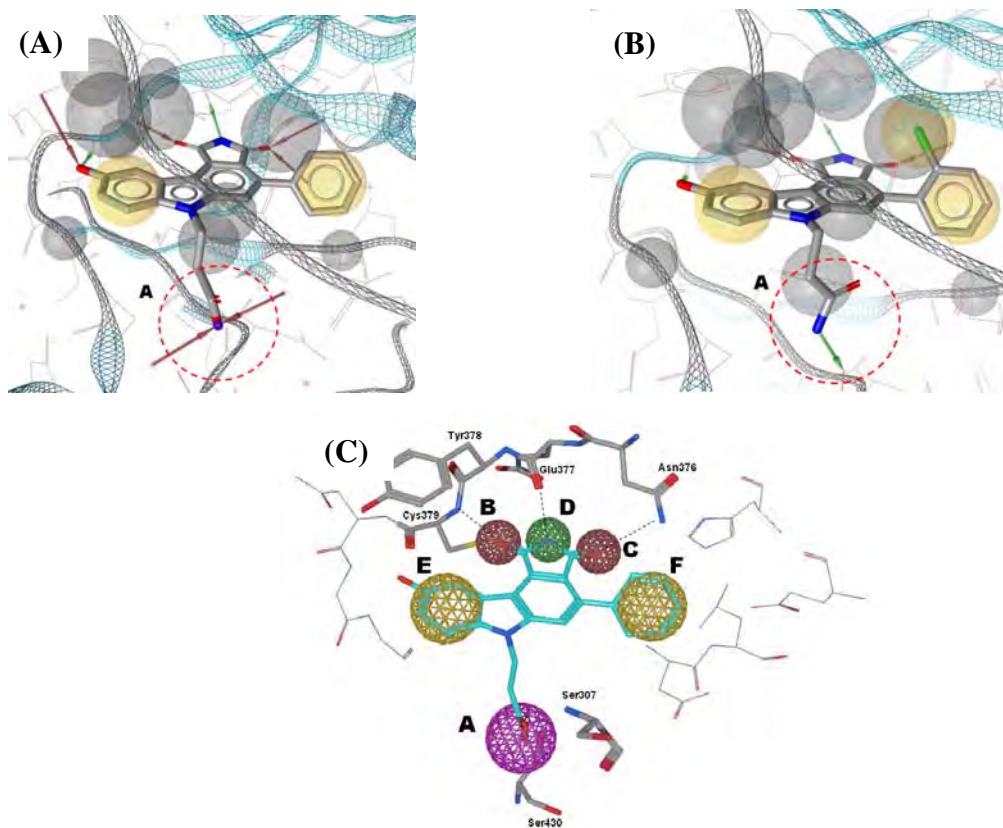


Figure 4.15 Three different pharmacophore models (A) pharmacophore model derived from X-ray structure 2IN6 (PH-1), (B) pharmacophore model generated from the docking solution of compound 106 (PH-2), (C) pharmacophore model (PH-3) built by merging the feature at the region A from PH-1 and PH-2. Pharmacophore feature represent by mangeta sphere = H-bond/donor feature, yellow sphere = hydrophobic feature, red sphere or arrow = H-bond acceptor and green sphere or arrow = H-bond donor

Among the six X-ray structures of the pyrrolocarbazole-Wee1 kinase complexes (PDB code 1X8B, 2ZZW, 2IN6, 2IO6, 3BI6 and 3BIZ), 2IN6 is the most active inhibitor among these six structures. The PDB structure 2IN6 (compound 103, $pIC_{50} = 7.64$) was then used to generated pharmacophore model using LigandScout program. By using this cocrystallized structure, pharmacophore feature at the region A as shown in Figure 4.15 (A) is represented as the H-bond acceptor. However, by applying the docking solution of the most active inhibitor (compound 106, $pIC_{50} =$

8.22) to create pharmacophore model, result reveals that the H-bond donor feature was created at the region A (displayed in Figure 4.15 (B)) to represent the important pharmacophore feature for Wee1 kinase inhibitors. Therefore, a pharmacophore feature at this region was merged, and the H-bond acceptor/donor feature (magenta sphere) was built at the region A as shown in Figure 4.15 (C). Note that the pharmacophore model generated from the PDB 2IN6 is named PH-1 (Figure 4.15 (A)). The model built from the docking solution of compound 106 is called PH-2 (Figure 4.15 (B)), and the pharmacophore model created by merging the feature at region A from PH-1 and PH-2 is named PH-3 (Figure 4.15 (C)). These three pharmacophore models were applied to evaluate the performance of these models for recovering the active compounds. The multi-conformation database containing all of 174 active compounds of pyrrolocarbazole dataset was applied for this objective. Results as summarized in Table 4.12 showed that 60 active compounds were recovered by using PH-3, whereas 43 and 29 compounds were obtained by using PH-1 and PH-2, respectively. Moreover, results also revealed that all of active compounds (7 compounds) which contain $pIC_{50} > 8.00$ were retrieved by applying PH-3. On the other hand, by using P-1, 5 compounds that have $pIC_{50} > 8.0$ were recovered and 3 compounds having $pIC_{50} > 8.00$ were obtained by using PH-2. From these results, it can be concluded that PH-3 performs better than PH-1 and PH-2.

Table 4.12 Number of hit compounds derived from each pharmacophore model

Pharmacophore Model	Number of Hit founds	% Hits	pIC_{50} of Hits	Hit of compound $pIC_{50} > 8.00$ (7compounds)
PH1	43	24.57	6.40-8.22	5
PH2	29	16.67	6.40-8.22	3
PH3	60	34.48	6.40-8.22	7

Therefore, this pharmacophore model (PH-3) was used later in the virtual screening. The magenta sphere (implying H-bond donor or acceptor area) at the region A of PH-3 as shown in Figure 4.15 (C) indicates the interaction between inhibitor with Ser430 or Ser307. Two red spheres, meaning the H-bond acceptor feature, demonstrate the H-bond interaction with residues located at the hinge region.

For example, sphere B denotes the H-bond between inhibitor and Cys379 and sphere C illustrate the H-bond between inhibitor and Asn376. The green sphere D (a H-bond acceptor feature) presents the inhibitor-Glu377 interaction. Yellow sphere E and F represent the hydrophobic interaction between inhibitor and Wee1 kinase.

4.3.2.2 Virtual Screening

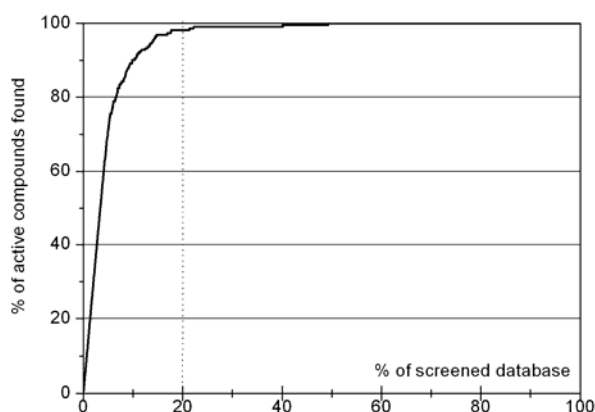
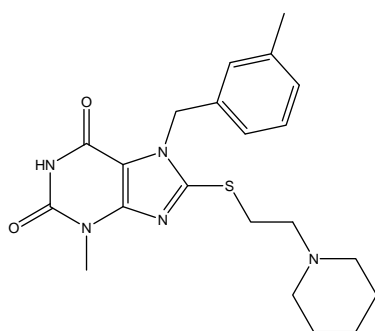


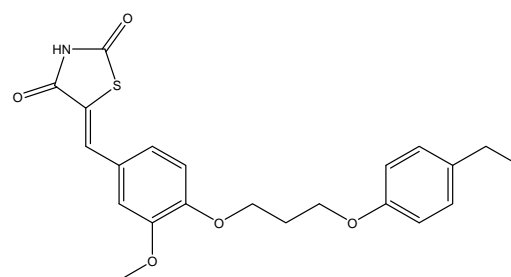
Figure 4.16 EF Curve of the database containing 222 active Wee1 inhibitors and 3,037 hit compounds obtained from pharmacophore search and drug-like property filtering. The database was ranked by GoldScore.

Multiconformation Chembrige database, containing 405,494 compounds and 41,829,578 conformations, was used in the VS study. Results show that 3,980 compounds pass with the PH-3. The number of the hit compounds was further screened by using LOF, which is resulting in that 3,037 compounds are in agreement with this rule. These compounds were subsequently docked into the binding pocket of Wee1 kinase using the same protocols as in section 4.1.1.3 and 4.2.1.2. These hit compounds and the known inhibitors, pyrrolocarbazole and pyridopyrimidine dataset (totally 222 compounds), were ranked according to their docking scores (GoldScore). ROC curve as displayed in Figure 4.16 shows that after 20% of the screened database and ranked by GoldScore, most of the active compounds were recovered. This result suggests that among the top 20% of the ranked database, they may also contain some other active compounds. Therefore, top-rank 500 compounds were selected and applied for binding free energy calculation using LR-MM/PBSA method as explained in section 4.2. The LR-MM/PBSA model 8 (the general model generated from both

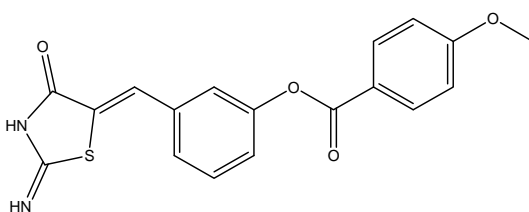
subsets) was used to calculate the predicted binding free energy of these hit compounds. These hit compounds were then ranked according to their binding free energies and were also visual inspected for their interaction with Wee1 kinase. Finally, some hit compounds as shown in Figure 4.17 are proposed as potent inhibitors for Wee1 kinase.



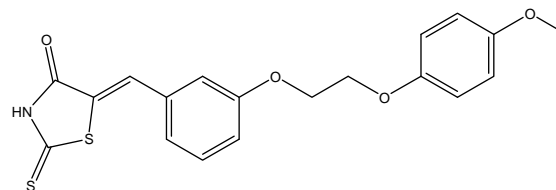
Cpd. ID 9001327 ($\Delta G_{\text{pred.}} = -9.89$ kcal/mol)



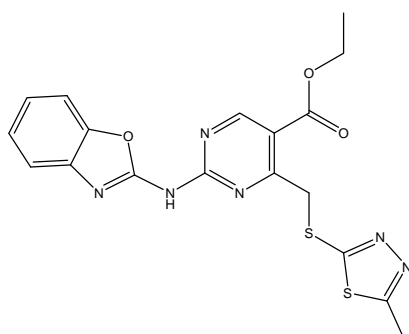
Cpd. ID 6637182 ($\Delta G_{\text{pred.}} = -9.58$ kcal/mol)



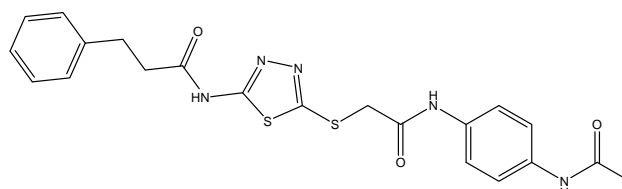
Cpd. ID 6208874 ($\Delta G_{\text{pred.}} = -9.43$ kcal/mol)



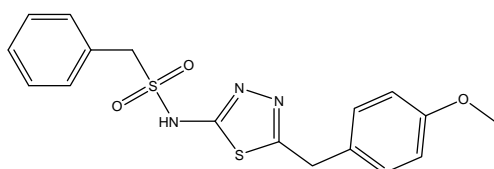
Cpd. ID 6377276 ($\Delta G_{\text{pred.}} = -9.24$ kcal/mol)



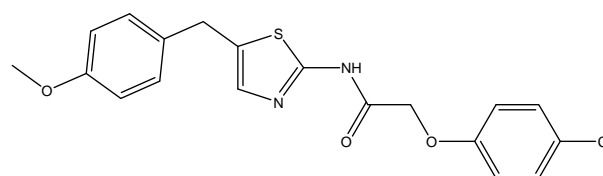
Cpd. ID 7771055 ($\Delta G_{\text{pred.}} = -9.15$ kcal/mol)



Cpd. ID 7754261 ($\Delta G_{\text{pred.}} = -9.00$ kcal/mol)



Cpd. ID 7166121 ($\Delta G_{\text{pred.}} = -8.80$ kcal/mol)



Cpd. ID 5849759 ($\Delta G_{\text{pred.}} = -8.33$ kcal/mol)

Figure 4.17 Chemical structures of Hit compounds and their predicted binding free energy

CHAPTER V

CONCLUSION

5.1 DV NS2B/NS3 protease

We have generated and validated a homology model of the inhibitor-bound form of DV NS2B/NS3 protease. The DV-2 homology model which was built based solely on the X-ray structure of the homolog WNV NS2B/NS3 protease-inhibitor complex gave results which are in nice agreement with the experimental data. On the other site, the DV-1 model, which was based on two different templates, was not able to explain all site-directed mutagenesis data. Thus, we suggest that the DV-2 model represents the DV NS2B/NS3 protease in the productive form. The obtained model and the MD simulations showed high similarity to the results obtained for the WNV protease crystal. The MD simulations revealed that strong interactions between the C-terminal domain of NS2B and NS3 assist the stability of the loop regions of the NS3 protease. This interaction also involved the binding of the Arg-P2 residue of the inhibitor and the residues of the S2 pocket. These results indicate that the C-terminal domain of NS2B is not only important for binding to the NS3 protease but also plays a significant role for the interaction with the P2 residue of the inhibitor.

In addition, our work shows that the interaction of the inhibitor with the S1 pocket involves only residues from the NS3 domain. On the other hand, both residues from the C-terminal domain of NS2B as well as Asp75 and Asn152 from NS3, are important for maintaining the interaction with the P2 residue of the inhibitor at the S2 pocket. At the S3 pocket, the main interactions are observed between the backbone of the P3 residue and Gly153 as well as Tyr161 from NS3. The interaction of the inhibitor at the S4 pocket of the NS3 is maintained only by hydrophobic interaction. The detailed analysis of the interaction of two peptidic inhibitors with the productive conformation of DV NS2B/NS3 protease, as well as the comparison with the related complex from WNV, provided useful information concerning the selectivity of the inhibitor Bz-Nle-Lys-Arg-Ala-H.

The productive form of DV protease can now be used as starting point for structure-based design in order to identify potent and drug-like inhibitors. The process for indentifying novel inhibitors for DV NS2B/NS3 protease was started by docking small-molecule inhibitors of DV NS2B/NS3 protease into the representative

conformation of this enzyme. Then, the binding free energies of these inhibitors were computed using MM/PBSA approach. The ligand-protein interaction was also investigated by applying decomposition (DC) energy analysis method. The calculated binding free energies of these known inhibitors gave a good correlation with their binding affinities. Interestingly, only the enthalpy value can be used to correlate with the experimental data by considering as a relative value. Therefore, the entropy calculation can be neglected. This can be very helpful for calculating the binding free energies of compounds retrieved from virtual screening in order to reduce to calculation time. The DC energy analysis reveals that residues from NS2B as well as from NS3 are important for interacting with inhibitors. In addition, results show that not only the electrostatic interaction that mainly contributes to the total energy, the van der Waals interaction is also important as well for interacting with inhibitors. Moreover, Pro132 located at the S1 pocket is a key residue for interacting as van der Waals interaction with inhibitors. In addition, residues which are important for interacting with inhibitors are addressed.

Virtual screening (VS) was carried out by starting from using three different pharmacophore models (a structure-based model, a static and a dynamic pharmacophore model). These models were applied to search for compounds in the multi-conformational databases that pass with these models. A number of hit compounds were then filtered by using several techniques. Finally, the binding free energies of hit compounds derived from VS (40 compounds) were calculated. Hit compounds which were gave binding free energy lower than the lowest binding free energy of know inhibitor were suggested for testing their biological activities against DV NS2B/NS3 protease. In addition, results revealed that the binding free energies of hits obtained from a dynamic pharmacophore model are lower than those retrieved from the static pharmacophore model. These results indicate that the dynamic pharmacophore model could lead to obtain inhibitors which could be more potent than those derived by using the conventional method such as the static pharmacophore model.

5.2 Wee1 Kinase

The 3D-QSAR and molecular docking study was initially undertaken to explore the structural features needed for high inhibitory activity at the Wee1 kinase. The availability of crystal structures of Wee1 complexed with an inhibitor enabled us to establish and analyze a model of the binding site. The good match of predicted and experimental structures gave confidence that the docking method is able to provide relevant information about the inhibitor interaction. The docking study revealed that the most potent inhibitors (**103**, **106** and **146**) establish hydrogen bonds to the backbone atoms of the hinge region (Asn376, Glu377, Cys379). In addition the sidechain group of the most potent inhibitors interacts with Ser307 or Ser430. Inhibitors bearing a substituted tetrazole or triazole ring in the sidechain are also able to form a hydrogen bond to Ser307, resulting in good inhibitory activity.

Using a combination of receptor-based alignment and 3D-QSAR yielded a significant and predictive model, indicated by the high cross-correlation coefficient and the low SDEP value. Beside the docking-based alignment, two other alignment methods were tested which resulted in models with lower predictive quality. It was shown that the docking scores could not be used to establish a reliable QSAR model, whereas the receptor-based 3D-QSAR model gave a significant correlation and can be used to point out which interaction sites in the binding pocket might be responsible for the variance in biological activities.

We also applied molecular docking for studying the interaction and binding mode between pyridopyrimidine derivatives with Wee1 kinase. Docking solutions show a conserved binding mode, which is the bidentate H-bond between N-3 of pyrimidine ring with NH backbone atom of Cys379 and 2-NH atom of ligand with CO backbone atom of the same residue. In addition, the substituted phenyl ring interacts with residues at the hydrophobic pocket such as His350, Lys328, Glu346 and Asp463. The molecular docking of a high activity compound revealed additionally a π - π interaction between the pyrimidine ring and the Phe430. Moreover, water molecules locating at the binding pocket are important for mediating H-bond network between the side chain of ligand and residues at the binding pocket. No correlation between docking scores, either applying GoldScore or ChemScore, with

experimental binding free energies of compounds was observed. Therefore, the linear response approach such as LR-MM/PBSA and LIECE were performed.

The LR-MM/PBSA model based on the energy-minimized docked complexes was developed for predicting the relative binding free energies of Wee1 kinase inhibitors. All of the derived models (model for pyrrolocarbazole, pyridopyrimidine and the general model for both chemical spaces) gave a significant and predictive model, indicated by the high correlation coefficient, leave-one-out cross-correlation coefficient and low RMSE of the training set. In addition, the obtained model was applied to predict the binding free energies of compounds in the external test set and resulted in satisfied statistical values. This approach has more advantages because it is simple, fast, less computational demand because only the single snapshot got from the energy minimization of the docked complex was used. In addition, this method is similar to docking scoring function, which is a linear equation of the interaction energy. However, this approach performs more efficient because the energy-minimized docked complexes were used for the interaction energy calculation and a linear equation of specific coefficients of each parameter for individual protein was generated. Moreover, results show obviously that the correlation between the $\Delta G_{\text{exp.}}$ with $\Delta G_{\text{est.}}$ derived from the LR-MM/PBSA model is higher than the correlation between the $\Delta G_{\text{exp.}}$ and the docking score.

The LIECE models which were generated from data sets of inhibitors of CDK2, Lck and p38 were applied to correlate the predicted and experimental binding affinities of inhibitors of Wee1 kinase. The two-parameter model of the LIECE model based on three proteins gave slightly higher correlation than the three-parameter model generated from the same data set. However, it has to be noted that $\Delta G_{\text{exp.}}$ of these Wee1 kinase inhibitors calculated from this LIECE model have quite high residual values. This might be due to force fields and methods that we have used in our study are different from the original publication. However, a good correlation ($r^2 = 0.50$) derived by using this LIECE model suggests that it is possible to use this LIECE model for calculating and then ranking the relative binding free energies of other kinase inhibitors which could be helpful for ranking compounds retrieved from virtual screening for kinase inhibitors. In addition, the derived LR-MM/PBSA and LIECE models were also analyzed for their performance to separate actives from

decoys. The obtained results showed that most of known active compounds were identified among the first 10% of the database screened and are ranked correctly by these models. A stepwise virtual screening for Wee1 kinase inhibitors was also performed by applying pharmacophore searching, molecular docking and binding free energy calculation using LR-MM/PBSA model (the general model). Finally, candidate compounds for testing biological activities against Wee1 kinase are proposed.

REFERENCES

- [1] Billoir, F., de Chesse, R., Tolou, H., de Micco P., Gould, E. A., and de Lamballerie, X. Phylogeny of the genus flavivirus using complete coding sequences of arthropod-borne viruses and viruses with no known vector. J. Gen. Virol. 81 (2000): 781-790.
- [2] Gubler, D. J. Dengue and dengue hemorrhagic fever. Clin. Microbiol. Rev. 11 (1998): 480-496.
- [3] Gubler, D. J, and Clark, G. G. Dengue/dengue hemorrhagic fever: the emergence of a global health problem. Emerg. Infect. Dis. 1 (1995): 55-57.
- [4] Alvarez, M., and others. Dengue hemorrhagic Fever caused by sequential dengue 1-3 virus infections over a long time interval: Havana epidemic, 2001-2002. Am. J. Trop. Med. Hyg. 75 (2006): 1113-1117.
- [5] Edelman, R. Dengue vaccines approach the finish line. Clin. Infect. Dis. 45 (2007): S56-S60.
- [6] eMedicine Specialties (Infectious Disease), Dengue: Multimedia. Available from: <http://emedicine.medscape.com/article/963213-media> [2010, February 19].
- [7] Zuo, Z. and others. Mechanism of NS2B-mediated activation of NS3pro in dengue virus: molecular dynamics simulations and bioassays. J. Virol. 83 (2009): 1060-1070.
- [8] Impact of Dengue, Average annual number of DF/DHF cases reported to WHO and average annual number of countries reporting dengue. Available from: <http://www.who.int/csr/disease/dengue/impact/en/index.html> [2010, February 19].
- [9] Gubler, D. J. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends Microbiol. 10 (2002): 100-103.
- [10] Global Health Observatory Map Gallery: Countries/areas at risk of dengue transmission, 2008. Available from: <http://gamapservr.who.int/mapLibrary/app/searchResults.aspx> [2010, February 19].
- [11] Annual Epidemiological Surveillance Report 2007. Available from: http://203.157.15.4/Annual/ANNUAL2550/Annual_index.html [2010, February 19].
- [12] Kuhn, R. J. and others. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell 108 (2002): 717-725.

- [13] Perera, R., and Kuhn, R. J. Structural proteomics of dengue virus. Curr. Opin. Microbiol. 11 (2008): 369-377.
- [14] Sampath, A., and Padmanabhan, R. Molecular targets for flavivirus drug discovery. Antiviral Res. 81 (2009): 6-15.
- [15] Lescar, J., and others. Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. Antiviral Res. 80 (2008): 94-101.
- [16] Uchil, P.D., and Satchidanandam, V. Architecture of the flaviviral replication complex. Protease, nuclease, and detergents reveal encasement within double-layered membrane compartments. J. Biol. Chem. 278 (2003): 24388-24398.
- [17] Mackenzie, J.M., and Westaway, E.G. Assembly and maturation of the flavivirus Kunjin virus appear to occur in the rough endoplasmic reticulum and along the secretory pathway, respectively. J. Virol. 75 (2001): 10787-10799.
- [18] Li, L., and others. The flavivirus precursor membrane-envelope protein complex: structure and maturation. Science 319 (2008): 1830-1834.
- [19] Mukhopadhyay, S., Kuhn, R.J., and Rossmann, M.G. A structural perspective of the flavivirus life cycle. Nat. Rev. Microbiol. 3 (2005): 13-22.
- [20] Yu, I.M., and others. Structure of the immature dengue virus at low pH primes proteolytic maturation. Science. 319 (2008): 1834-1837.
- [21] Li, H., Clum, S., You, S., Ebner, K.E., and Padmanabhan, R. The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. J. Virol. 73 (1999): 3108-3116.
- [22] Stadler, K., Allison, S.L., Schalich, J., and Heinz, F.X. Proteolytic activation of tick-borne encephalitis virus by furin. J. Virol. 71 (1997): 8475-8481.
- [23] Chambers, T.J., and others. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. U.S.A. 87 (1990): 8898-8902.
- [24] Chambers, T.J., Hahn, C.S., Galler, R., and Rice, C.M. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. 44 (1990): 649-688.
- [25] Clum, S., Ebner, K.E., and Padmanabhan, R. Cotranslational membrane insertion of the serine proteinase precursor NS2B-NS3(Pro) of dengue virus type 2

is required for efficient *in vitro* processing and is mediated through the hydrophobic regions of NS2B. J. Biol. Chem. 272 (1997): 30715-30723.

- [26] Falgout, B., Pethel, M., Zhang, Y.M., and Lai, C.J., Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. 65 (1991): 2467-2475.
- [27] Keelapang, P. and others. Alterations of pr-M cleavage and virus export in pr-M junction chimeric dengue viruses. J. Virol. 78 (2004): 2367-2381.
- [28] Falgout, B., Miller, R.H., and Lai, C.J. Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. J. Virol. 67 (1993): 2034-2042.
- [29] Lin, C., Amberg, S.M., Chambers, T.J., and Rice, C.M., Cleavage at a novel site in the NS4A region by the yellow fever virus NS2B-3 proteinase is a prerequisite for processing at the downstream 4A/4B signalase site. J. Virol. 67 (1993): 2327-2335.
- [30] Lobigs, M. Flavivirus premembrane protein cleavage and spike heterodimer secretion require the function of the viral proteinase NS3. Proc. Natl. Acad. Sci U.S.A. 90 (1993): 6218-6222.
- [31] Preugschat, F., and Strauss, J.H., Processing of nonstructural proteins NS4A and NS4B of dengue 2 virus *in vitro* and *in vivo*. Virology 185 (1991): 689-697.
- [32] Yamshchikov, V.F., and Compans, R.W. Processing of the intracellular form of the west Nile virus capsid protein by the viral NS2B-NS3 protease: an *in vitro* study. J. Virol. 68 (1994): 5765-5771.
- [33] Chappell, K.J., Stoermer, M.J., Fairlie, D.P., and Young, P.R., Mutagenesis of the West Nile virus NS2B cofactor domain reveals two regions essential for protease activity. J. Gen. Virol. 89 (2008): 1010-1014.
- [34] Niyomrattanakit, P., Winoyanuwattikun, P., Chanprapaph, S., Angsuthanasombat, C., Panyim, S., and Katzenmeier, G. Identification of residues in the Dengue Virus type 2 NS2B cofactor that are critical for NS3 protease activation. J. Virol. 78 (2004): 13708-13716.
- [35] Radichev, I., and others. Structure-based mutagenesis identifies important novel determinants of the NS2B cofactor of the West Nile virus two-component NS2B-NS3 proteinase. J. Gen. Virol. 89 (2008): 636-641.
- [36] Luo, D., Xu, T., Hunke, C., Grüber, G., Vasudevan, S.G., and Lescar, J. Crystal structure of the NS3 protease-helicase from dengue virus. J. Virol. 82 (2008): 173-183.

- [37] Leung, D., and others. Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. J. Biol. Chem. 276 (2001): 45762-45771.
- [38] Erbel, P., and others. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. Nat. Struct. Mol. Biol. 13 (2006): 372-373.
- [39] Mueller, N.H., Pattabiraman, N., Ansarah-Sobrinho, C., Viswanathan, P., Pierson, T.C., and Padmanabhan, R. Identification and biochemical characterization of small-molecule inhibitors of west nile virus serine protease by a high-throughput screen. Antimicrob. Agents. Chemother. 52 (2008): 3385-3393.
- [40] Murthy, H.M., Clum, S., and Padmanabhan, R. Dengue Virus NS3 serine protease. Crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects. J. Biol. Chem. 274 (1999): 5573-5580.
- [41] Murthy, H.M., Judge, K., DeLucas, L., and Padmanabhan, R., Crystal structure of Dengue virus NS3 protease in complex with a Bowman-Birk inhibitor: implications for flaviviral polyprotein processing and drug design. J. Biol. Chem. 301 (2000): 759-767.
- [42] Aleshin, A.E., Shiryayev, S.A., Strongin, A.Y., and Liddington, R.C. Structural evidence for regulation and specificity of flaviviral protease and evolution of the Flaviviridae fold. Protein. Sci. 16 (2007):795-806.
- [43] Brinkworth, R.I., Fairlie, D.P., Leung, D., and Young, P.R. Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. J. Gen. Virol. 80 (1999): 1167-1177.
- [44] Chanprapaph, S., and others. Competitive inhibition of the dengue virus NS3 serine protease by synthetic peptides representing polyprotein cleavage sites. Biochem. Biophys. Res. Commun. 330 (2005): 1237-1246.
- [45] Ganesh, V.K., Muller, N., Judge, K., Luan, C.H., Padmanabhan, R., and Murthy, K.H. Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. Bioorg. Med. Chem. 13 (2005): 257-264.
- [46] Kiat, T.S., Pippen, R., Yusof, R., Ibrahim, H., Khalid, N., and Rahman, N.A. Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoids of fingerroot, *Boesenbergia rotunda* (L.), towards dengue-2 virus NS3 protease. Bioorg. Med. Chem. Lett. 16 (2006): 3337-3340.

- [47] Stevens, A.J., Gahan, M.E., Mahalingam, S., and Keller, P.A. The Medicinal Chemistry of Dengue Fever. J. Med. Chem. (2009): DOI: 10.1021/jm900652e.
- [48] Tomlinson, S.M., Malmstrom, R.D., Russo, A., Mueller, N., Pang, Y.P., and Watowich, S.J. Structure-based discovery of dengue virus protease inhibitors. Antiviral. Res. 82 (2009): 110-114.
- [49] Yin, Z., and others. Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. Bioorg. Med. Chem. Lett. 16 (2006): 40-43.
- [50] Yin, Z., and others. Peptide inhibitors of Dengue virus NS3 protease. Part 1: Warhead. Bioorg. Med. Chem. Lett. 16 (2006): 36-39.
- [51] Zuo, Z., and others. Mechanism of NS2B-mediated activation of NS3pro in dengue virus: molecular dynamics simulations and bioassays. J. Virol. 83 (2009): 1060-1070.
- [52] Othman, R., and others. Docking of noncompetitive inhibitors into dengue virus type 2 protease: understanding the interactions with allosteric binding sites. J. Chem. Inf. Model. 48 (2008): 1582-1591.
- [53] Garrett, M.D. Cell cycle control and cancer. CURRENT SCIENCE 81 (2001): 515-522.
- [54] Collins, K., Jacks, T., and Pavletich, N.P. The cell cycle and cancer. Proc. Natl. Acad. Sci. U. S. A. 94 (1997): 2776-2778.
- [55] Bucher N., and Britten C.D. G2 checkpoint abrogation and checkpoint kinase-1 targeting in the treatment of cancer. Br. J. Cancer. 98 (2008): 523-528.
- [56] Reactome – a curated knowledgebase of biological pathways. Available from: <http://www.reactome.org>. [2010, February 19].
- [57] Squire, C.J., Dickson, J.M., Ivanovic, I., and Baker, E.N. Structure and inhibition of the human cell cycle checkpoint kinase, Wee1A kinase: an atypical tyrosine kinase with a key role in CDK1 regulation. Structure 13 (2005): 541-550.
- [58] Wang, Y., and others. Radiosensitization of p53 mutant cells by PD0166285, a novel G(2) checkpoint abrogator. Cancer Res. 61 (2001): 8211-8217.
- [59] Vastrik, I., and others. Reactome: a knowledge base of biologic pathways and processes. Genome Biol. 8 (2007): R39.

- [60] Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. The protein kinase complement of the human genome. Science 298 (2002): 1912-1934.
- [61] Palmer, B.D., and others. Structure-activity relationships for 2-anilino-6-phenylpyrido[2,3-d]pyrimidin-7(8H)-ones as inhibitors of the cellular checkpoint kinase Wee1. Bioorg. Med. Chem. Lett. 15 (2005): 1931-1935.
- [62] Palmer, B.D., and others. 4-Phenylpyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione inhibitors of the checkpoint kinase Wee1. Structure-activity relationships for chromophore modification and phenyl ring substitution. J. Med. Chem. 49 (2006): 4896-4911.
- [63] Smaill, J.B., and others. Synthesis and structure-activity relationships of N-6 substituted analogues of 9-hydroxy-4-phenylpyrrolo[3,4-c]carbazole-1,3(2H,6H)-diones as inhibitors of Wee1 and Chk1 checkpoint kinases. Eur. J. Med. Chem. 43 (2008): 1276-1296.
- [64] Smaill, J.B., and others. Synthesis and structure-activity relationships of soluble 8-substituted 4-(2-chlorophenyl)-9-hydroxypyrrrolo[3,4-c]carbazole-1,3(2H,6H)-diones as inhibitors of the Wee1 and Chk1 checkpoint kinases. Bioorg. Med. Chem. Lett. 18 (2008): 929-933.
- [65] Yi, P., Fang, X., and Qiu, M. 3D-QSAR studies of Checkpoint Kinase Wee1 inhibitors based on molecular docking, CoMFA and CoMSIA. Eur. J. Med. Chem. 43 (2008): 925-938.
- [66] Hammett, P.L. Physical Organic Chemistry (2nd edition), New York: McGraw-Hill. 1970.
- [67] Hansch, C., and Fujita, T. ρ - σ - π Analysis; method for the correlation of biological activity and chemical structure. J. Am. Chem. Soc. 86 (1964): 1616-1626.
- [68] Free, S.M. Jr., and Wilson, J.W. A mathematical contribution to structure-activity studies. J. Med. Chem. 7 (1964): 395-399.
- [69] Cramer, R.D. III, Patterson, D. E., and Bunce, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. 110 (1988): 5959-5967.
- [70] Sippl, W. Methods and principles in medicinal chemistry. Pharmacophores and Pharmacophore Concepts, Langer, T., Hofmann R. (Eds.), Publisher, New York (2006): 223-249.

- [71] Lushington, G.H., Guo, J.X., and Wang, J.L. Whither combine? New opportunities for receptor-based QSAR. Curr. Med. Chem. 14 (2007): 1863-1877.
- [72] Frédérick, R., and Denny, W.A. Phosphoinositide-3-kinases (PI3Ks): combined comparative modeling and 3D-QSAR to rationalize the inhibition of p110alpha. J. Chem. Inf. Model. 48 (2008): 629-638.
- [73] Pearson, W.R. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183 (1990): 63-98.
- [74] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 215 (1990): 403-410.
- [75] Rost, B. Twilight zone of protein sequence alignments. Protein Eng. 12 (1999): 85-94.
- [76] Baker, D., and Sali, A., Protein structure prediction and structural genomics. Science 294 (2001): 93-96.
- [77] Needleman, S.B., and Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48 (1970): 445-453
- [78] The Needleman-Wunsch algorithm for sequence alignment. Available from: http://sepwww.stanford.edu/public/docs/sep112/bob2/paper_html/node3.html. [2010, February 19].
- [79] Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M., PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26 (1993): 283-291.
- [80] Lüthy, R., Bowie, J.U., and Eisenberg, D. Assessment of protein models with three-dimensional profiles. Nature 356 (1992): 83-85.
- [81] Bowie, J.U., Lüthy, R., and Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 253 (1991): 164-170.
- [82] Sippl, M.J. Recognition of errors in three-dimensional structures of proteins. Proteins 17 (1993): 355-362.
- [83] Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. Stereochemical quality of protein structure coordinates. Proteins 12 (1992): 345-364.

- [84] Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7 (1963): 95-99.
- [85] Brandsdal, B.O., Osterberg, F., Almlöf, M., Feierberg, I., Luzhkov, V.B., and Åqvist, J. Free energy calculations and ligand binding. Adv. Protein. Chem. 66 (2003): 123-158.
- [86] Connolly, M.L. Analytical molecular surface calculation. J. Appl. Cryst. 16 (1983): 548-558.
- [87] Åqvist, J., Medina, C., and Samuelsson, J.E. A new method for predicting binding affinity in computer-aided drug design. Protein Eng. 7 (1994): 385-391.
- [88] Carlsson, J., Boukharta, L., and Åqvist, J., Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. J. Med. Chem. 51 (2008): 2648-2656.
- [89] Jones-Hertzog, D. K., and Jorgensen, W.L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. J. Med. Chem. 40 (1997): 1539-1549.
- [90] van Lipzig, M.M., and others. Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. J. Med. Chem. 12 (2004): 1018-1030.
- [91] Stjernschantz, E., Marelus, J., Medina, C., Jacobsson, M., Vermeulen, N.P., and Oostenbrink, C. Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method. J. Chem. Inf. Model. 46 (2006): 1972-1983.
- [92] Huang, D., and Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. J. Med. Chem. 47 (2004): 5791-5797.
- [93] Kolb, P., Huang, D., Dey, F., and Caflisch, A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. J. Med. Chem. 51 (2008): 1179-1188.
- [94] Huang, D., and others. Discovery of cell-permeable non-peptide inhibitors of beta-secretase by high-throughput docking and continuum electrostatics calculations. J. Med. Chem. 48 (2005): 5108-5111.

- [95] Huang, D., Lüthi, U., Kolb, P., Cecchini, M., Barberis, A., and Caflisch, A. In silico discovery of beta-secretase inhibitors. J. Am. Chem. Soc. 128 (2006): 5436-5443.
- [96] Ekonomiuk, D., and others., Discovery of a non-peptidic inhibitor of west nile virus NS3 protease by high-throughput docking. PLoS Negl. Trop. Dis. 3 (2009): e356.
- [97] Zhou, Z., and Madura, J.D. Relative free energy of binding and binding mode calculations of HIV-1 RT inhibitors based on dock-MM-PB/GS. Proteins 57 (2004): 493-503.
- [98] Zhou, Z., Bates, M., and Madura, J.D. Structure modeling, ligand binding, and binding affinity calculation (LR-MM-PBSA) of human heparanase for inhibition and drug design. Proteins 65 (2006): 580-592.
- [99] Zhou, Z., Wang, Y., and Bryant, S.H. Computational analysis of the cathepsin B inhibitors activities through LR-MMPBSA binding affinity calculation based on docked complex. J. Comput. Chem. 30 (2009): 2165-2175.
- [100] Ekonomiuk D., and others. Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. J. Med. Chem. 52 (2009): 4860-4868.
- [101] Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug. Deliv. Rev. 46 (2001): 3-16.
- [102] Wermuth, G., Ganellin, C.R., Lindberg, P., and Mitscher, L.A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure and Applied Chemistry 70 (1998): 1129-1143.
- [103] Guene, O.F., Pharmacophore perception, development, and use in drug design. International University Line. 2000.
- [104] Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28 (1985.): 849-857.
- [105] Jones, G., Willett, P., and Glen, R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J. Mol. Biol. 245 (1995): 43-53.
- [106] Jones, G., Willett, P., Glen, R.C., Leach, A.R., and Taylor, R., Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267 (1997): 727-748.

- [107] Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V., and Mee, R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J. Comput. Aided. Mol. Des. 11 (1997): 425-445.
- [108] Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R., and Eldridge, M.D., Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins. **33** (1998): 367-282.
- [109] Willett, P., Similarity-based virtual screening using 2D fingerprints. Drug Discov. Today 11 (2006): 1046-1053.
- [110] Shiryayev, S.A., Aleshin, A.E., Ratnikov, B.I., Smith, J.W., Liddington, R.C., and Strongin, A.Y. Expression and purification of a two-component flaviviral protease resistant to autocleavage at the NS2B-NS3 junction region. Protein Expr. Purif. 52 (2007): 334-339.
- [111] Wheeler, D.L., and others. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 36 (2008): D13-21.
- [112] MOE 2006.08. Chemical Computing Group Inc. 2006
- [113] Henikoff, S., and Henikoff, J.G. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U. S. A. 89 (1992): 10915-10919.
- [114] Wilmouth R.C., and others. X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. Nat. Struct. Biol. 8 (2001): 689-684.
- [115] Lindahl, E., Hess, B., and van der Spoel, D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. J. Mol. Model. 7 (2001): 306-317.
- [116] van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. GROMACS: fast, flexible and free. J. Comput. Chem. 26 (2005): 1701-1718.
- [117] van Gunsteren, W.F., and others. Biomolecular Simulation: The GROMOS96 manual and user guide. Hochschulverlag AG an der ETH Zürich, Switzerland. 1996.
- [118] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., and Hermans, J. Interaction models for water in relation to protein hydration. In: Intermolecular Forces, D. Reidel Publishing Company Dordrecht. 1981.

- [119] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. A smooth particle mesh Ewald method. J. Chem. Phys. 103 (1995): 8577-8593.
- [120] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Dinola, A., and Haak, J.R., Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81 (1984): 3684-3690.
- [121] Humphrey, W., Dalke, A., and Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph. 14 (1996): 33-38, 27-28.
- [122] GRID22a., Moldiscovery Ltd. Pinner, Middlesex, UK. 2004.
- [123] Butkiewicz, N., and others. Virus-specific cofactor requirement and chimeric hepatitis C virus/GB virus B nonstructural protein 3. J. Virol. 74 (2000): 4291-4301.
- [124] Melino, S., and others. The active essential CFNS3d protein complex; A new perspective for the structural and kinetic characterization of the NS2B-NS3pro complex of dengue virus. FEBS. J 273 (2006): 3650-3662.
- [125] Chappell, K.J., and others. Site-directed mutagenesis and kinetic studies of the West Nile virus NS3 protease identify key enzyme-substrate interactions. J. Biol. Chem. 280 (2005): 2896-2903.
- [126] Nall, T.A., and others. Enzymatic characterization and homology model of a catalytically active recombinant West Nile virus NS3 protease. J. Biol. Chem. 279 (2004): 48535-48542.
- [127] Yusof, R., Clum, S., Wetzel, M., Murthy, H.M., and Padmanabhan, R. Purified NS2B/NS3 serine protease of Dengue virus type 2 exhibits cofactor NS2B dependence for cleavage for substrates with dibasic amino acids in vitro. J. Biol. Chem. 275 (2000): 9963-9969.
- [128] Li, J., and others., Functional profiling of recombinant NS3 proteases from all four serotypes of Dengue Virus using tetrapeptide and octapeptide Substrate Libraries. J. Biol. Chem. 280 (2005): 28766-28774.
- [129] Chappell, K.J., Stoermer, M.J., Fairlie, D.P., and Young, P.R. Insights to substrate binding and processing by West Nile Virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies. J. Bio. Chem. 281 (2006): 38448-38458.
- [130] Knox, J.E., and others. Peptide inhibitors of West Nile NS3 protease: SAR study of tetrapeptide aldehyde inhibitors. J. Med. Chem. 49 (2006): 6585-6590.

- [131] Wichapong, K., Pianwanit, S., Sippl, W., and Kokpol, S. Homology modeling and molecular dynamics simulations of Dengue virus NS2B/NS3 protease: insight into molecular interaction. J. Mol. Recognit. (2009): DOI:10.1002/jmr.977.
- [132] Case D.A., and others. AMBER 9, University of California, San Francisco. 2006.
- [133] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 65 (2006): 712-725.
- [134] Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. Development and testing of a general amber force field. J. Comput. Chem. 25 (2004): 1157-1174.
- [135] Bayly, C.I., Cieplak, P., Cornell, W., and Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges the RESP model. J. Phys. Chem. 97 (1993): 10269-10280.
- [136] Frisch M.J., and others. Gaussian 03, Gaussian, Inc., Wallingford CT. 2004.
- [137] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R. W., and Klein, M.L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79 (1983): 926-935.
- [138] Pastor R.W., Brooks B.R., and Szabo A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. Mol. Phys. 65 (1988): 1409-1419.
- [139] Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23 (1977): 327-341.
- [140] Darden, T., York, D., and Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J. Chem. Phys. 98 (1993): 10089-10092.
- [141] Luo, R., David, L., and Gilson, M.K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. J. Comput. Chem. 23 (2002): 1244-1253.
- [142] Sitkoff, D., Sharp, K.A., and Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. 98 (1994): 1978-1988.
- [143] Carlso, H.A., and others. Developing a dynamic pharmacophore model for HIV-1 integrase. J. Med. Chem. 43 (2000): 2100-2114.

- [144] Deng, J., Lee, K.W., Sanchez, T., Cui, M., Neamati, N., and Briggs, J.M. Dynamic receptor-based pharmacophore model development and its application in designing novel HIV-1 integrase inhibitors. J. Med. Chem. 48 (2005): 1496-1505.
- [145] Deng, J., Sanchez, T., Neamati, N., and Briggs, J.M. Dynamic pharmacophore model optimization: identification of novel HIV-1 integrase inhibitors. J. Med. Chem. 49 (2006): 1684-1692.
- [146] Wolber, G., and Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J. Chem. Inf. Model. 45 (2005): 160-169.
- [147] Sybyl 7.2. Tripos Inc., South Hanley, St. Louis, MO, 2007.
- [148] Halgren, T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J. Comput. Chem. 17 (1996): 490-519.
- [149] Broyden, C.G. The Convergence of a class of double-rank minimization algorithms. J. Inst. Math. Appl. 6 (1970): 76-90.
- [150] Fletcher, R. A New approach to variable metric algorithms. Comput. J. 13 (1970): 317-322.
- [151] Goldfarb, D. A family of variable metric method updates derived by variational means. Math. of Comput. 24 (1970): 23-26.
- [152] Shanno, D.F. Conditioning of quasi-newton methods for function minimization Math. of Comput. 24 (1970): 647-656.
- [153] Ponder, J.W., and Case, D.A. Force fields for protein simulations. Adv. Protein Chem. 66 (2003): 27-85.
- [154] Gasteiger, J., and Marsili, M. Iterative partial equalization of orbital electronegativity - A rapid access to atomic charges. Tetrahedron. 36 (1980): 3219-3228.
- [155] Pan, Y., Huang, N., Cho, S., and MacKerell, Jr.A.D. Consideration of molecular weight during compound selection in virtual target-based database screening. J. Chem. Inf. Comput. Sci. 43 (2003): 267-72.
- [156] Carta, G., Knox, A.J., and Lloyd, D.G. Unbiasing scoring functions: a new normalization and rescoring strategy. J. Chem. Inf. Model. 47 (2007): 1564-1571

- [157] Wichapong, K., Lindner, M., Pianwanit, S., Kokpol, S., and Sippl, W. Receptor-based 3D-QSAR studies of checkpoint Wee1 kinase inhibitors. Eur. J. Med. Chem. 44 (2009): 1383-1395.
- [158] Jakalian, A., Jack, D. B., and Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J. Comput. Chem. 23 (2002): 1623-1641.
- [159] Case, D.A., and others. AMBER 10, University of California, San Francisco. 2008.
- [160] Huang, N., Shoichet, B.K., and Irwin, J.J. Benchmarking sets for molecular docking. J. Med. Chem. 49 (2006): 6789-67801.
- [161] Pearlman, D.A., and Charifson, P.S. Improved scoring of ligand-protein interactions using OWFEG free energy grids. J. Med. Chem. 44 (2001): 502-511.

APPENDIX

APPENDIX: WORKSHOPS AND CONFERENCES

WORKSHOPS and ACTIVITIES;

1. Participated in “The 2nd Thai Summer School of Computational Chemistry”, Chiang Mai University, Chiang Mai, Thailand, 2004.
2. Participated in “The 2nd Asian Pacific Conference on Theoretical and Computational Chemistry”, Chulalongkorn University, Bangkok, Thailand, 2005.
3. Participated in “The 10th Annual National Symposium on Computational Science and Engineer (ANSCSE 10)”, Chiang Mai University, Chiang Mai, Thailand, 2006.
4. Participated in “The 3rd Thai Summer School of Computational Chemistry”, Burapa University, Chonburi, Thailand, 2006.
5. Participated in “The Training Course on Molecular Design and Computer-assisted Combinatorial Chemistry”, ICS-UNIDO, **Trieste, Italy**, June 30 – July 5, 2008.
6. Participated in “The eCheminfo Training Workshop on Drug Discovery Design and Planning Methods”, Medical Sciences Teaching Laboratory, Oxford University, **Oxford, England**, July 20-24, 2009.
7. Participated in “The 5th Summer School on Drug Design, University of Vienna, **Vienna, Austria**, September 13-18, 2009.
8. Participated as the instructor for the workshop “The 6th Thai Summer School of Computational Chemistry”, Mahasarakarm University, Mahasarakham, Thailand, 2009. (<http://www.inteligand.com/news/>)

CONFERENCES and MEETINGS

Poster Presentation;

1. **Kanin Wichapong**, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; 3D-QSAR Studies on tetra-peptidic inhibitors of West Nile Virus NS2B/NS3 protease; The 21st Darmstädter Molecular Modelling Workshop, University Erlangen-Nuernberg, **Erlangen, Germany**, May 15-16, 2007
2. **Kanin Wichapong**, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; Molecular docking studies on inhibitors of Dengue NS2B/NS3 protease; The 21st

Darmstädter Molecular Modelling Workshop, University Erlangen-Nuernberg, **Erlangen, Germany**, May 15-16, 2007

3. **Kanin Wichapong**, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; Molecular docking studies on inhibitors of Dengue NS2B/NS3 protease; The 4th Polish-German Symposium, Martin-Luther University Halle-Wittenberg, **Halle(Saale), Germany**, June 6, 2007
4. **Kanin Wichapong**, Somsak Pianwanit, Wolfgang Sippl and Sirirat Kokpol; NS2B/NS3 protease of Dengue Virus-Inhibitor interaction as studied by molecular dynamics simulation and molecular docking; Pure and Applied Chemistry Conference 2008 (PACCON 2008), Sofitel Centara Grand Bangkok, **Bangkok, Thailand**, January 30 –February 1, 2008.
5. **Kanin Wichapong**, Somsak Pianwanit, Wolfgang Sippl and Sirirat Kokpol; Molecular Studies and Drug Design for NS2B/NS3 protease of Dengue Virus and of West Nile Virus; International Conference Drug Design and Discovery for Developing Countries, **Trieste, Italy**, July 3-5, 2008.
6. **Kanin Wichapong**, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; Molecular modeling studies on Dengue and West Nile virus NS2B/NS3 protease-inhibitor interaction; The fourth German Conference on Chemoinformatics, **Goslar, Germany**, November 9-11, 2008.
7. **Kanin Wichapong**, Somsak Pianwanit, Wolfgang Sippl, and Sirirat Kokpol; Search for Novel Inhibitors of Dengue Virus NS2B/NS3 Protease by Combining Virtual Screening and MM-PBSA Binding Free Energy Calculations; The eCheminfo Training Workshop on Drug Discovery Design and Planning Methods”, Medical Sciences Teaching Laboratory, Oxford University, **Oxford, England**, July 20-24, 2009.
8. **Kanin Wichapong**, Somsak Pianwanit, Wolfgang Sippl and Sirirat Kokpol; Molecular dynamics simulation and virtual screening for inhibitors of Dengue Virus NS2B/NS3 protease; The 238th ACS National Meeting, **Washington DC, U.S. A.**, August 16-20, 2009.
9. **Kanin Wichapong**, Somsak Pianwanit, Wolfgang Sippl and Sirirat Kokpol; Search for Novel Inhibitors of Dengue Virus NS2B/NS3 Protease by Combining

Virtual Screening and MM-PBSA Binding Free Energy Calculations; The 5th Summer School on Drug Design, **Vienna, Austria**, September 13-18, 2009.

Oral Presentation;

1. **Kanin Wichapong**, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; Molecular dynamics simulations of homology models of protease of dengue virus complexed with its inhibitor; The 33rd Congress on Science & Technology, Thailand (STT33), Walailak University, **Nakhon Si Thammarat, Thailand**, October 18-20, 2007.
2. **Kanin Wichapong**, Somsak Pianwanit, Wolfgang Sippl and Sirirat Kokpol; Insight into the role of Cofactor NS2B for activating NS3 Protease of Dengue and of West Nile Virus and for interacting with inhibitors; The 12th Annual Symposium on Computational Science and Engineering (ANSCSE12), Ubon Rajathanee University, **Ubonratchathani, Thailand**, March 27-29, 2008.
3. **Kanin Wichapong**, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; Generating Homology Models of NS2B/NS3 protease of Dengue Virus for Molecular Docking; The 9th Ph.D.- RGJ Congress, Pattaya, **Chonburi, Thailand**, April 4-6, 2008

VITAE

Name; Kanin Wichapong

Date of Birth; May 2nd, 1982

Place of Birth; Roi-Et, Thailand

Nationality; Thai

E-mail address; kanin@daad-alumni.de, kanin.wichapong@gmail.com

EDUCATION;

2001 - 2005; Bachelor Degree (2nd class Honor B.Sc.) in Chemistry;

Faculty of Science, Chulalongkorn University, Bangkok, Thailand

2005 – Present; Ph.D. Candidate, Physical Chemistry, Computational Chemistry Unit Cell, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

SCHOLARSHIPS;

June 2005 – Present; Royal Golden Jubilee Ph.D. Program, Thailand Research Fund

June 2008 – September 2009; Deutscher Akademischer Austausch Dienst (DAAD)

PUBLICATIONS;

1. Kanin Wichapong, Marc Linder, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; Receptor-based 3D-QSAR studies of checkpoint Wee1 kinase inhibitors. *Eur. J. Med. Chem.* 44 (2009), 1383-1395.
2. Kanin Wichapong, Somsak Pianwanit, Wolfgang Sippl and Sirirat Kokpol; Homology Modeling and Molecular Dynamics Simulations of Dengue Virus NS2B/NS3 Protease: Insight into Molecular Interaction. *J. Mol. Recognit.* **In press.**
3. Kanin Wichapong, Michael Lawson, Somsak Pianwanit, Sirirat Kokpol and Wolfgang Sippl; Structure-based prediction of activities of Wee1 Kinase inhibitors using linear response approach; in preparation
4. Kanin Wichapong, Somsak Pianwanit, Wolfgang Sippl, and Sirirat Kokpol; Dynamic Pharmacophore model of Dengue Virus NS2B/NS3 protease inhibitors; in preparation