



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

February 2021

Ph.D. Dissertation

**Computational Ligand Modeling for
Kinases and COVID-19 Main Protease
using Various Methods: Docking,
Molecular Dynamics, Free Energy
Calculation, 3D-QSAR, and Virtual
Screening**

Graduate School of Chosun University

Department of Biomedical Sciences

Seketoulie Keretsu

**Computational Ligand Modeling for
Kinases and COVID-19 Main Protease
using Various Methods: Docking,
Molecular Dynamics, Free Energy
Calculation, 3D-QSAR, and Virtual
Screening**

다양한 방법론을 사용한 카이네이즈와
Covid19 주단백질 효소에 대한 계산과학적 리간드
모델링: 도킹, 분자동역학, 자유에너지 계산,
3차원 QSAR 및 가상검색

February 25, 2021

Graduate School of Chosun University

Department of Biomedical Science

**Computational Ligand Modeling for
Kinases and COVID-19 Main Protease
using Various Methods: Docking,
Molecular Dynamics, Free Energy
Calculation, 3D-QSAR, and Virtual
Screening**

Advisor: Prof. Seung Joo Cho

*This dissertation is submitted to the Graduate School of
Chosun University in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Science*

October 2020

Graduate School of Chosun University

Department of Biomedical Sciences

Seketoulie Keretsu

Ph. D. Dissertation of

Seketoulie Keretsu is certified by

Chairman (Chosun Univ.): Prof. Song Yub Shin

Committee Members:

Chosun Univ. : Prof. Eun Ae Kim

Chosun Univ. : Prof. Ho Joong Kim

Chonnam Nat. Univ. : Prof. Jeong Hyun Dam

Chosun Univ. : Prof. Seung Joo cho

December 2020

Graduate School of Chosun University

CONTENTS

CONTENTS	I
ABBREVIATIONS	VIII
LIST OF TABLES	X
LIST OF FIGURES	XIII
ABSTRACT (KOREAN)	XVIII
ABSTRACT (ENGLISH)	XX

PART I

<i>Overview of Protein Kinases</i>	2-4
1. Introduction.....	2
2. Structure.....	2
3. Classification of Kinases.....	3
4. Kinase as Drug Target.....	3
5. Kinase Inhibitors.....	4

Part II

Docking and 3D-QSAR Studies of Hydrazone and Triazole Derivatives for Selective Inhibition of GRK2 over ROCK26-13

1. Introduction.....	6
2. Methodology	7
2.1. Dataset.....	7
2.2. Molecular Docking.....	8
2.3. CoMFA and CoMSIA.....	12
2.4. Model Validation.....	13
3. Results and discussion	
3.1. Molecular Docking.....	14
3.2. 3D-QSAR.....	16
3.2.1.CoMFA and CoMSIA models for GRK2.....	17
3.2.2.CoMFA and CoMSIA models for ROCK2.....	20
3.3 Contour map analysis.....	21
3.3.1. GRK2 CoMFA contour maps.....	22

3.3.2 GRK2 CoMSIA contour maps.....	24
3.3.3. ROCK2 CoMFA contour maps.....	25
3.3.4. ROCK2 CoMSIA contour maps.....	26
4. Conclusion.....	29

PART III

<i>Computational study of paroxetine-like inhibitors reveals new molecular insight to inhibit GRK2 with selectivity over ROCK1</i>	41-73
---	-------

1. Introduction.....	41
2. Methodology.....	44
2.1. Dataset.....	44
2.2. Protein preparation.....	48
2.3. Molecular Docking.....	49
2.4. Molecular Dynamics (MD) simulation.....	50
2.5. Free energy calculation.....	50
2.6. 3D-QSAR	51
2.7 Model validation.....	52

3. Results.....	53
3.1. <i>Molecular Docking</i>	53
3.2. <i>Molecular Dynamics (MD) Simulation</i>	55
3.3. <i>MM/PBSA based free energy calculations</i>	60
3.4. <i>3D-QSAR</i>	61
3.5. <i>Contour Map Analysis</i>	64
4. Discussions.....	66
5. Conclusion.....	73

PART IV

Molecular Modelling Study of c-KIT/PDGFR α Dual Inhibitors for the Treatment of Gastrointestinal Stromal Tumors.....76-111

1. Introduction.....	76
2. Methodology.....	78
2.1. <i>Data Preparation</i>	78
2.2. <i>Molecular Docking</i>	79
2.3. <i>Molecular Dynamics Simulation</i>	79

2.4. Evaluation of Binding Energy.....	80
2.5. 3D-QSAR.....	81
3. Results.....	82
3.1. Molecular Docking.....	87
3.2. Molecular Dynamics Simulation.....	87
3.3. Evaluation of Binding Energy.....	89
3.4. 3D-QSAR.....	93
3.5. Analysis of Contour Map.....	95
3.6. Designed Compounds.....	97
4. Discussion.....	101
5. Conclusion.....	112
 PART V	
 <i>Rational Approach toward COVID-19 Main Protease Inhibitors via Molecular Docking, Molecular Dynamics Simulation and Free Energy Calculation.....</i>	
1. Introduction.....	115
2. Methods.....	118

2.1. <i>Data preparation</i>	118
2.2. <i>Virtual Screening</i>	119
2.2.1. <i>Surflex-Dock</i>	119
2.2.2. <i>Autodock vina</i>	120
2.3. <i>Molecular docking</i>	120
2.4 <i>Molecular dynamics simulation</i>	120
2.5. <i>Calculation of binding free energy</i>	121
3. Results	122
3.1. <i>Virtual screening</i>	124
3.2. <i>Molecular docking</i>	127
3.3. <i>Molecular Dynamics Simulation</i>	128
3.4. <i>Calculation of binding free energy</i>	129
4. Discussion	132
5. Conclusion	142
 PART VI	
<i>Conclusion</i>	144

REFERENCES 147-163

APPENDIX

A. List of Publications.. 163-164

B. Acknowledgement. 165

ABBREVIATIONS

3D-QSAR	Three-dimensional quantitative structure-activity relationship
ADMET	Absorption, distribution, metabolism, excretion and toxicity
BE	Binding Energy
BS SD	Bootstrap standard deviation
CADD	Computer-aided drug discovery
c-KIT	Stem cell factor receptor
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity indices analysis
FDA	Food and Drug Administration
FGFR	Fibroblast growth factor receptor
FLT3	Fms like tyrosine kinase 3
GISTs	Gastrointestinal stromal tumors
GRK2	G-protein coupled receptor kinase 2
ICC	Interstitial cells of Cajal
MD	Molecular dynamics
MM/PBSA	Molecular mechanics energies combined with the Poisson–Boltzmann and surface area continuum solvation

ONC	Optimal number of components
PDB	Protein data bank
PDGFRa	Platelet derived growth factor receptor alpha
RAF1	Rapidly accelerated fibrosarcoma 1
RET	Rearranged during transfection
RMSD	Root mean square deviation
SASA	Solvent accessible surface area
SEE	Standard error of estimation
VEGFR	Vascular endothelial growth factor receptor

LIST OF TABLES

PART II

Table 1. Structure of the hydrazone and triazole derivatives and their pIC ₅₀ values for GRK2 and ROCK2.....	9
Table 2. Statistical results of the CoMFA and CoMSIA models for GRK2.....	17
Table 3. Statistical results of the CoMFA and CoMSIA models for ROCK2.....	18
Table 4. Experimental and predicted pIC ₅₀ values with their residuals of selected CoMFA and CoMSIA model derived from GRK2.....	30
Table 5. Experimental and predicted pIC ₅₀ values with their residuals of selected CoMFA and CoMSIA model for ROCK2.....	35

PART III

Table 1. Comparison of the residues at the adenine subsite, polyphosphate subsite, ribose subsite and hydrophobic subsite for GRK2, ROCK1 and ROCK2.....	42
Table 2. Structure of the paroxetine-like derivatives and their pIC ₅₀ values for GRK2 and ROCK1.....	45
Table 3. The energy contribution of the various energetic terms (van der Waals energy, electrostatic energy, polar solvation energy, and non-polar solvation energy/SASA) to the total binding energy.....	60
Table 4. Statistical results of the CoMFA models for GRK2 and ROCK1.....	63

Table 5. Experimental and predicted pIC_{50} values with their residuals of CoMFA for GRK2.....69

Table 6. Experimental and predicted pIC_{50} values with their residuals of CoMFA for ROCK1.72

PART IV

Table 1. Structure of the pyrazolopyridine derivatives and their pIC_{50} values for c-KIT and PDGFR α83

Table 2. The energy contribution of the various energetic terms (Van der Waals energy, electrostatic energy, polar solvation energy, and non-polar solvation energy/SASA) to the total binding energy during the binding of imatinib and compound 14 with c-KIT and PDGFR α90

Table 3. Residues that showed a high contribution to the total binding energy during the MD simulations of **Compound14-c-KIT**, **Imatinib-c-KIT**, **Compound14-PDGFR α** , and **Imatinib-PDGFR α** . The energy values of the residues are given in kJ/mol.....92

Table 4. Statistical results of the CoMFA and CoMSIA models for c-KIT and PDGFR α . q^2 : cross-validated correlation coefficient; ONC: Optimal number of components; r^2 : non-cross-validated correlation coefficient; SEE: Standard Error of Estimation; F value: F-test value; r^2 ; BS- r^2 : Bootstrapping r^2 mean; BS-SD: Bootstrapping Standard deviation; r^2_{pred} : predictive..... 93

Table 5. The chemical structures and the predicted pIC_{50} values of the newly designed compounds for c-KIT and PDGFR α98

Table 6. The energy contributions of the various energetic terms (Van der Waals energy, electrostatic energy, polar solvation energy, and non-polar solvation energy/SASA) to the total binding energies are shown for the designed compounds.....99

Table 7. The experimental/actual and predicted pIC_{50} values with their residuals for the CoMFA and CoMSIA for c-KIT.....106

Table 8. The experimental/actual and predicted pIC_{50} values with their residuals for the CoMFA and CoMSIA for PDGFR α108

Table 9. The predicted ADMET values and synthetic accessibility values for the 8 designed compounds.....109

PART V

Table 1. The PubChem IDs, total scores (Surflex-Dock), autodock vina and autodock binding energies (kcal/mol), and MM-PBSA based binding energies (kJ/mol) of compounds N3, 13b, and the 32 selected compounds125

Table 2. Energy contributions of the various energetic terms to the total binding energies of the inhibitors with 3CL^{pro}.134

Table 3. Residues with a high contribution to the total binding energy during the MD simulations of the complexes 441243-3CL^{pro}, 451415-3CL^{pro}, 446837-3CL^{pro}, 53361968-3CL^{pro}, 46178275-3CL^{pro}, 9828551-3CL^{pro}, 644196-3CL^{pro}, 134815261-3CL^{pro}, 15942730-3CL^{pro}, and 132531950-3CL^{pro}. The energy values of the residues are in kJ/mol.135

Table 4. Structures of the 10 compounds selected on basis of binding energy (MM-PBSA)137

Table 5. The predicted ADMET values for the 10 selected compounds.....139

Table 6. The binding energies, ZINC compound IDs, and the distributor/vendor names and vendor compound IDs of the 10 compounds selected on basis of high MM-PBSA-based binding energy evaluation.140

LIST OF FIGURES

PART II

Figure 1. (a) The docked conformation of the most active compound (compound 93) inside the active site of GRK2. (b) Docked conformation of the most active compound (compound 60) inside the active site of ROCK2... 15

Figure 2. (a) Substructure from template compound (compound 93) used for the alignment of dataset. (b) Alignment of the dataset compounds used in the GRK2 model development. (c) Substructure from template compound (compound 60) used for the alignment of dataset. (d) Alignment of the dataset compounds used in the GRK2 model development. 19

Figure 3. Standard coefficient contour maps obtained from GRK2 CoMFA analysis with the template compound (compound 93) as reference. 22

Figure 4. Standard contour maps obtained from GRK2 CoMSIA analysis with the template compound (compound 93) as reference. 23

Figure 5. Standard coefficient contour maps obtained from ROCK2 CoMFA analysis with the template compound (compound 60) as reference. 25

Figure 6. Standard contour maps obtained from ROCK2 CoMSIA analysis with the template compound (compound 60) as reference. 27

PART III

Figure 1. The alignment of the amino acid sequences in the kinase domains of GRK2 and ROCK1. Identical and positive matches in the sequences are highlighted by red and green colours respectively.....49

Figure 2. (a) Alignment of the dataset compounds used in the CoMFA model development for GRK2. (b) Alignment of the dataset compounds used in the CoMFA model development for ROCK1.....52

Figure 3. The docked conformation of the most active compound for ROCK1 (compound 11) inside the active site of ROCK1. H-bond interactions were represented as yellow dotted lines.....54

Figure 4. The RMSD diagrams for the 40 ns MD simulation runs. MD production run for each protein-ligand complex was performed once only.....56

Figure 5. H-bond interactions between the compound 11 (salmon), 17 (green) and 47 (magenta) with GRK2 and ROCK1. Snapshots were collected after 40 ns simulations.57

Figure 6. Hydrophobic interactions between compound 11, 17, and 47 with GRK2 and ROCK1. Snapshots were collected after 40 ns simulations.....59

Figure 7. The energy contributions (in kJ/mol) of the key residues to the total binding energy; (a) Interaction of compound 11 (brown), 17 (green) and 47 (magenta) with GRK2. (b) Interaction of compound 11 (brown), 17 (green) and 47 (magenta) with ROCK1.62

Figure 8. Standard coefficient contour maps obtained from GRK2 and ROCK1 CoMFA analyses65

Figure 9. RMSD values of pyridine ring (compound 11), dimethoxybenzene ring (compound 17) and pyrazole ring (compound 47) inside GRK2 and ROCK1 for 40 ns MD simulations. MD production run for each protein-ligand complex was performed once only.67

PART IV

Figure 1. H-bond interactions of imatinib and compound 14 with c-KIT and PDGFR α from the MD simulations.....89

Figure 2. Contour maps generated based on the CoMFA and CoMSIA models for c-KIT and PDGFR α with Compound 14 used as a reference.96

Figure 3. H-bond interactions of the designed compounds with c-KIT and PDGFR α from the MD simulations. H-bond interactions are represented by yellow dotted lines and residues forming H-bonds are shown in purple color. (a) D18-c-KIT (b) D23-c-KIT (c) D25-c-KIT (d) D28-c-KIT (e) D32-c-KIT (f) D39-c-KIT (g) D44-c-KIT (h) D45-c-KIT (i) D18-PDGFR α (j) D23-PDGFR α (k) D25-PDGFR α (l) D28-PDGFR α (m) D32-PDGFR α (n) D39-PDGFR α (o) D44-PDGFR α (p) D45-PDGFR α101

Figure 4. Showing the hydrophobic interactions of the inhibitors with c-KIT and PDGFR α103

Figure 5. (a) The hydrophobic surface of the c-KIT binding site residues; Hydrophobic residues from the α C-helix and the catalytic loops are shown in stick representation (b) Binding interactions between compound 31 and c-KIT. (c) Binding interactions between compound 31 and PDGFR α104

Figure 6. H-bond interactions and residue contact map from the MD simulation of imatinib and compound 14.106

Figure 7. The ligand RMSD plots from the MD simulations. (a) Imatinib-c-KIT (b) Compound 14-c-KIT (c) Compound 31-c-KIT (d) D18-c-KIT (e) D23-c-KIT (f) D25-c-KIT (g) D28-c-KIT (h) D32-c-KIT (i) D39-c-KIT (j) D44-c-KIT (k) D45-c-KIT (l) Imatinib-c-KIT/T670I (m) Imatinib-PDGFR α (n) Compound 14- PDGFR α (o) Compound 31- PDGFR α (p) D18-PDGFR α (q) D23-PDGFR α (r) D25-PDGFR α (s) D28-PDGFR α (t) D32-PDGFR α (u) D39-PDGFR α (v) D44-PDGFR α (w) D45-PDGFR α (x) Imatinib-PDGFR α /T674I (y) Compound 14-c-KIT α /T670I (z) Compound 14- PDGFR α /T674I.....110

PART V

Figure 1. Structure of the inhibitor N3 in complex with 3CL^{PRO} as given in 6LU7. The protein and ligand were shown in gray and magenta color. The mesh (cyan) representation between domain 1 and domain 2 represents the area to be searched by surflex dock (protomol).123

Figure 2. Showing the H-bond interactions of the inhibitors N3 and 13b with 3CL^{PRO} from the MD simulation studies.124

Figure 3. Showing the H-bond interactions of the inhibitors N3 and 13b with 3CL^{PRO} from the molecular docking studies. H-bond interactions were represented by yellow dotted lines and residues forming H-bonds were shown in purple color. (a) Binding interactions between N3 and 3CL^{PRO}. (b) Binding interactions between 13b and 3CL^{PRO}.127

Figure 4. Showing the hydrophobic interactions of the inhibitors N3 and 13b with 3CL^{PRO}.130

Figure 5. Residues that showed high contribution to the total binding energy during the MD simulation of N3-3CL^{PRO} (6LU7) and 13b-3CL^{PRO} (6Y2F) complex.131

Figure 6. Showing the H-bond interactions of the inhibitors with 3CL^{PRO}. The protein and ligand and ligand were shown in gray and green color respectively.132

Figure 7. Least Square Fit Root Mean Square Deviation (RMSD) of the protein (red) and ligand (black) from the 50 ns MD simulations. (a) N3-3CL^{pro} (b) 13b-3CL^{pro} (c) 441243-3CL^{pro} (d) 451415-3CL^{pro} (e) 644196-3CL^{pro} (f) 53361968-3CL^{pro} (g) 46178275-3CL^{pro} (h) 9828551-3CL^{pro} (i) 446837-3CL^{pro} (j) 134815261-3CL^{pro} (k) 15942730-3CL^{pro} and (l) 132531950-3CL^{pro}133

초록
다양한 방법론을 사용한 카이네이즈와 Covid19 주단백질
효소에 대한 계산과학적 리간드
모델링: 도킹, 분자동력학, 자유에너지 계산, 3차원 QSAR 및
가상검색

세계투래 캐릿수

지도교수: 조승주, Ph.D.

의과학과

조선대학교 대학원

단백질카이네이즈는 다양한 세포 신호전달과정에 중요한 역할을 하는 효소이다. 다양한 세포의 과정에 중요한 역할을 하기 때문에, 카이네이즈의 작용을 방해하는 것은 다양한 암과 자가면역질환과 관련되어 있다. 카이네이즈는 중요한 신약개발 타깃이다. 현재까지 미국식품안전청에서 52개의 카이네이즈억제제가 약품의 용도로 승인되었다. 카이네이즈 억제제의 개발의 성공에도 불구하고 부작용 때문에 실제사용은 제한적이다. 따라서, 효능이 우수하고 선택적인 카이네이즈 억제제의 개발은 신약설계에서 중요한 분야이다. 단백질카이네이즈의 억제를 일으키는 구조적이고 물리화학적인 요인을 이해하기 위하여, G-단백질과 연결된 수용체 카이네이즈2(GRK2), 줄기세포인자 수용체(c-KIT), 혈소판유래 성장인자수용체 알파(PDGFR α)등을 계산연구에 대한 타깃으로 선정하였다. 분자 도킹과 분자동력학 시뮬레이션을 이용하여, 단백질과

리간드 결합 상호작용을 연구하였다. 3차원 구조활성상관관계 모델을 만들어서 화합물의 구조와 억제능력간의 관계를 연구하였다. 3차원 QSAR 모델로부터 얻을 수 있는 컨투어 맵을 사용하여 다양한 치환체에 대한 정보를 알아볼 수 있었다. 계산연구의 결과는 더 강력하고 선택적인 억제제를 만드는데 중요한 방향성을 제공한다.

2019년 12월의 COVID19의 발발과 관련하여 과학계에서는 이를 치료해야하는 도적적인 과업에 직면하게 되었다. COVID19를 일으키는 원인물질인 SARS-COV-2의 3CL^{pro} 단백질은, 바이러스의 증식에 대단히 중요한 역할을 하기 때문에 중요한 타깃으로 확인되었다. 따라서, 3CL^{pro} 억제제는 COVID19를 치료하는데 효과적일 것이라고 추정되고 있다. 우리는 이 작용점에 대하여, 단백질 억제제 데이터베이스인 MEROPS를 사용하여 가상검색을 실행하였다. 가상검색의 결과, 32개의 화합물이 강한 결합력을 보였으며, 지속적인 실험으로 검증할 필요가 있다고 생각되었다. 최종적으로 높은 결합에너지를 가지는 15개의 가능성이 큰 3CL^{pro} 억제제를 제안할 수 있었다. 이들 중, Saquinavir (인증된 HIV-1치료제)와 3개의 다른 연구중의 약품들, 즉, aclarubicin, TMC-310911, Faldaprevir 등이 potential 3CL^{pro} 의 억제에 가능성이 큰 것으로 나타났다. Aclarubicin은 화학적 항암요법에 쓰이는 anthracycline 약이고, TMC-130911은 항바이러스 약이다. Faldaprevir은 현재 임상시험중인 C형 간염에 쓸 약이다. 이 화합물들에 대한 계속적이고 실험적인 검증이 필요하다. 본 연구의 결과는 카이네이즈와 COVID 19 주 단백질 분해효소에 대한 억제제의 구조적인 요인을 이해하는데 많은 도움을 주었다.

Abstract

Computational Ligand Modeling for Kinases and COVID-19 Main Protease using Various Methods: Docking, Molecular Dynamics, Free Energy Calculation, 3D-QSAR, and Virtual Screening

Seketoulie Keretsu

Advisor: Prof. Seung Joo Cho

Department of Biomedical sciences

Graduate School of Chosun University

Protein kinases are enzymes that play important role in various cell signaling pathways. Because of their important role in various cellular processes, the dysregulation of kinases has been associated with various cancers and autoimmune diseases. Hence, kinases are regarded as important drug targets. So far 52 kinase inhibitors have been approved by the Food and Drug Administration (FDA) for pharmaceutical use. Despite the success in the development of kinase inhibitors, their use has been limited due to their off-target activity. Hence, the development of potent and selective kinase inhibitors has been an interesting area in drug design. To understand the structural and physicochemical properties that drive the inhibition of protein kinases, the G-protein coupled receptor kinase 2 (GRK2), stem cell factor receptor (c-KIT), and platelet derived growth factor receptor alpha (PDGFR α)

have been selected as model targets for computational study. Molecular docking and molecular dynamics simulation were performed to study the protein-ligand binding interactions. 3D-Quantitative structure-activity relationship (3D-QSAR) models were developed to study the relationship between the structure of the compounds and their inhibitory activities. Contour maps generated based on the 3D-QSAR models provided crucial information regarding various favorable and unfavorable substituents. The result of the computational study provided valuable insights that could be used as guidelines in the future development of potent and selective inhibitors.

Following the outbreak of the COVID-19 pandemic in December 2019, the scientific community was faced with the challenging task to develop therapy against the disease. The 3CL^{pro} protease of the SARS-COV-2, the causative agent of COVID-19, has been identified as a drug target due to its unique role in the viral replication. Hence, 3CL^{pro} inhibitors were considered to be promising agents for COVID-19 treatment. We have performed virtual screening of the protease inhibitor database MEROPS for potential 3CL^{pro} inhibitors. Based on the virtual screening, 32 compounds that showed high binding energy values were further accessed for pharmaceutical use. We found 15 potential 3CL^{pro} inhibitors with high binding affinity. Among them, Saquinavir (an approved drug for HIV-1 treatment) and three other investigational drugs namely aclarubicin, TMC-310911, and Faldaprevir could be suggested as potential 3CL^{pro} inhibitors. Aclarubicin is an anthracycline drug used in cancer chemotherapy. TMC-310911 is an antiviral drug and Faldaprevir is an experimental drug under clinical trial for the treatment of hepatitis C disease. Further experimental validation of the compounds is recommended.

The outcome of this study gave a comprehensive understanding of the structural factors important for the inhibition of kinases and COVID-19 main protease.

Seketoulie Keretsu Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

PART 1

Brief Overview of Kinases

1. Introduction

Kinases are biological enzymes that catalyze the transfer of phosphates in proteins through a process known as phosphorylation. Kinases are involved in the phosphorylation of up to one-third of the proteome. During phosphorylation, a phosphate from the high energy Adenosine Triphosphate (ATP) is transferred to a specific substrate. The phosphorylation of a protein, in turn, leads to its activation. This mechanism serves as the primary means of signal transduction. Hence, kinases play a very important role in the signaling regulation of various cellular processes such as cell progression, differentiation, apoptosis, and signal transduction [1].

2. Structure

Kinase monomers typically consist of 300 to 500 amino acid residues. Kinases may form a functional unit such as dimers and tetramers. Though kinases may structurally vary, they share a highly conserved kinase domain. The kinase domain consists of an N-terminal domain and a C-terminal domain. The two terminal domains are connected by the hinge. The N-terminal domain is characterized by five-stranded anti-parallel β -sheets (β 1- β 5) and an α C-helix whereas the C-terminal domain is globular consisting of six helices. The cavity at the interface of the N and C domain forms the ATP binding site. This ATP binding site is also commonly referred to as the active binding site or ligand binding site [2].

The ATP binding site is enveloped by residues from the activation loop, catalytic loop, p-loop (glycine-rich loop), α C-Helix, and the hinge region. The residues forming the active site are highly conserved throughout the kinase family and play significant roles in the catalytic processes through the binding of the ATP and other cofactors.

3. Classification of Kinases

The human kinome consists of 518 kinases which can be classified into various families and subfamilies based on the sequence and structural similarities. The human kinase families are AGC (Kinase A, G, and C) family, CMGC (CDK, MAPK, GSK2, CLK) family, CAMK (CMK1 and CMK2) family, CK1 (Casein Kinase 1) family, and others. The ‘other’ kinase family consists of STE (Homologs of STE7, STE11, and STE20), TK (Tyrosine Kinase), TKL (Tyrosine-Like Kinase), RGC, PKL, and Atypical kinases [3].

Based on the residue of phosphorylation, kinases within the human kinome can also be classified as either tyrosine kinases or serine/threonine kinases. Tyrosine kinases are involved in the transfer of phosphate from an ATP to a free hydroxyl group in the tyrosine residue of a target protein. Whereas, serine/threonine kinases catalyze the transfer of phosphates from ATP to serine or tyrosine residue of a target protein [4].

4. Kinase as Drug Target

The first oncogene was identified as a protein kinase in 1978 by Ray Erikson. This was followed by the discovery of other kinases associated with cancer and other diseases. The phosphorylation by kinases mediates important cellular processes hence, their overactivation and dysregulation lead to cancer and the pathogenesis of many diseases such as autoimmune, inflammatory, degenerative, metabolic, and cardiovascular diseases. Hence, kinases are regarded as important drug targets. Survey studies showed that one-third of the small molecules drug candidates reported target the protein kinases [5] [6]. Recent studies of kinases as drug targets have shown the AGC and CMGC were the major kinase families that have been targeted for inhibition. However, there are subfamilies of kinases for which inhibitors have not been developed. Suggesting that not all kinases

received equal attention. This also suggested that the full potential of kinases as drug targets have not be achieved yet.

5. Kinase Inhibitor

The earliest kinase inhibitors were Naphthalene sulphonamides developed by Hiroyoshi Hidaka. These early inhibitors target the cyclic-AMP and cGMP-dependant protein kinases, protein kinase C and cyclic-nucleotide dependent kinase. Most kinase inhibitors are small-molecule kinase inhibitors while a small number of them are nucleotide-based inhibitors. The majority of the kinase inhibitors are ATP complete and its functional activity depends on the ATP concentration. However, non-competitive and uncompetitive inhibitors that target other binding sites are also being considered. These types of inhibitors typically target allosteric bind sites [7].

Due to the various binding mechanisms of the small molecule kinase inhibitors, these inhibitors are classified into various types (Type 1 - Type 4). Type 1 inhibitors target the active conformation of the ATP binding site and are therefore ATP competitive inhibitors. A majority of the existing kinase inhibitors are classified as type 1. Due to the catalytic activity, most kinases share sequence and structural similarity in this region hence type 1 inhibitor is considered to show multi-kinase inhibitory activity. Type 2 inhibitors target the inactive conformation of kinases. While the type 3 and type 4 target binding site outside the catalytic domain. These inhibitors are commonly referred to as allosteric inhibitors. Due to the unique mechanism of binding associated with allosteric inhibitors. This class of inhibitors is considered to be ideal candidates for selective inhibition of specific kinases [8].

PART II

Docking and 3D-QSAR Studies of Hydrazone and Triazole Derivatives for Selective Inhibition of GRK2 over ROCK2.

1. INTRODUCTION

Heart failure has become a global health problem with an estimation of more than 38 million patients worldwide. The future of heart failure treatment looks pessimistic as further projection showed the increase of heart failure patients [9].

During heart failure, the heart fails to produce sufficient myocardial contraction, subsequently reducing the flow of blood throughout the body [10]. The positive inotropic effect needed for myocardial contraction to counter the reduced blood flow is achieved via the beta-adrenergic receptors (β -AR) signaling. These β -ARs are tightly regulated by the feedback enzyme G protein-coupled receptor kinase 2 (GRK2), which promotes beta-arrestin binding at the receptor. Beta-arrestin binding at the β -ARs blocks substrate binding at the receptor and leads to reduced β -AR signaling [11-13]. Therefore, the up-regulation of the β -AR signaling by inhibiting GRK2 is considered as a potential route to mitigate heart failure [10]. Several studies in animal models have also shown the association between overexpression of GRK2 and heart failure [14-16], indicating the importance of GRK2 inhibition as a potential drug target for heart failure treatment.

GRK2 belongs to G protein-coupled receptor kinase (GRK) subfamily under the A, G, and C family (AGC family) of kinases [13]. They share high sequence and structural similarity with other kinases in the AGC family. The AGC family of proteins play a vital role in regulating many physiological processes. Consequently, their mutation and dysregulation contribute to the pathogenesis of many human diseases, including cancer and diabetes [17-19].

Rho-associated coiled-coil containing kinase (ROCK2) is a serine/threonine kinase and belongs to the AGC kinase family. They are known to play an important role in cell migration and invasion, centrosome duplication, cytokinesis, and apoptosis [13, 20]. Due to the high sequence identity between the catalytic domain of GRK2 and ROCK2 (nearly 32 % identity), cross-activity against them has been observed in previous studies [14, 21]. Hence, to avoid cell

Seketoulie Keretsu Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

toxicity due to inhibition of ROCK2 during the inhibition of GRK2, discriminatory inhibition against GRK2 and ROCK2 is crucial.

Over the last few years, several GRK2 inhibitors have been developed for the treatment of heart failure, such as the natural product balanol [22], CMPD101 and CMPD103A developed by Takeda pharmaceuticals [23], paroxetine [24, 25] and GSK180736A [26]. However, none of these GRK2 inhibitors advanced to clinical trials due to a lack of selectivity and poor bioavailability.

Computational drug design methods such as molecular docking and three-dimensional quantitative structure-activity relationship (3D-QSAR) have emerged as powerful tools in assisting the drug discovery process [27, 28]. Though these computational techniques have been used to study the structure-activity relationship (SAR) and protein-ligand interactions of kinase inhibitors [29, 30], its application in the study of GRK2 and its inhibitors is yet to be seen. In this work, we performed molecular docking and 3D-QSAR study on a series of 93 hydrazone and triazole derivatives that showed inhibitory activity towards GRK2 and its closely related ROCK2 [31]. To analyze the binding affinities and interactions formed between the ligand and receptor, docking of the most active compound with receptor was performed for both GRK2 and ROCK2. Using Docking and 3D-QSAR techniques, comparative molecular field analysis (CoMFA) [32] and comparative molecular similarity indices analysis (CoMSIA) [33] models were developed for both GRK2 and ROCK2. Co-analysis of the molecular docking and 3D-QSAR results from both GRK2 and ROCK2 were performed to identify the important structural features and interactions that drive these compounds to prefer GRK2 over ROCK2.

2. METHODOLOGY

2.1. Dataset

A series of 93 hydrazone and triazole derivatives and their activity values against G protein-coupled receptor kinase (GRK2) and Rho-associated coiled-coil

containing kinase (ROCK2) which were reported in previous studies are collected for molecular modeling studies [31]. The inhibitory concentration (IC_{50}) values of the compounds are converted to negative logarithmic pIC_{50} values. The structure of the compounds and their corresponding pIC_{50} values against GRK2 and ROCK2 are given in Table 1. The series of compounds have a wide activity range of 5 to 8.9 for GRK2 and 5 to 8.1 for ROCK2 and is considered to be suitable for 3D-quantitative structure-activity relationship (3D-QSAR) studies [34].

All the compounds are sketched and optimized in Sybyl-X 2.1 for 3D-QSAR studies. The docked pose of the most active compound is used as a template for sketching the other compounds. The structures are energy minimized using Tripos force field, followed by the application of Gasteiger-Hückel charges as partial charges. The dataset was divided into training set of 63 compounds and test set of 30 compounds based on activity value. The training set was used in deriving the 3D-QSAR model and the test set was used to test the predictive ability of the derived model.

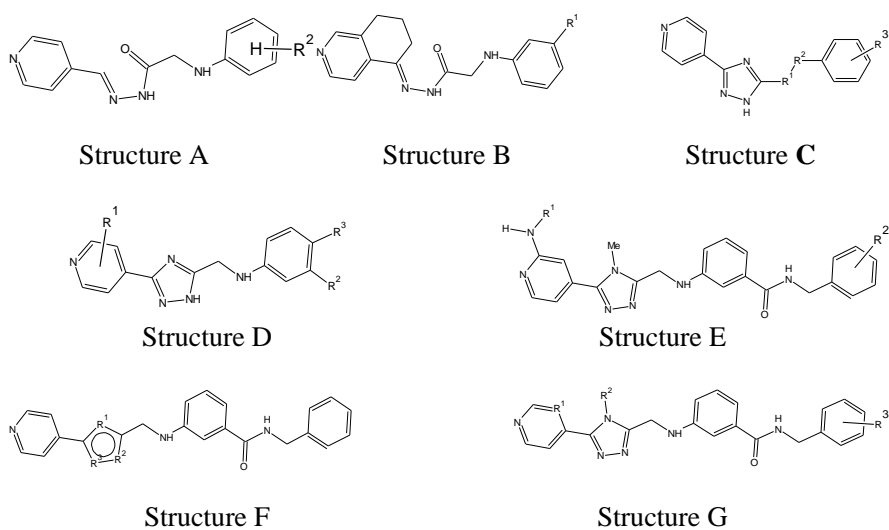
2.2. Molecular Docking

To analyze the ligand-protein interactions, we performed molecular docking of the most active compound with both the receptors, GKR2 and ROCK2 using Autodock 4.2.5.1 [35]. The structure of the ligand-GRK2 complex (PDB code: 5UUU) [31] and ligand-ROCK2 (PDB code: 4QL6) [36] were already reported in previous studies. To validate the docking protocol, we have sketched the ligand structure and minimized it using Tripos force field in Sybyl-X 2.1 outside of the receptor and then docked to the apo-receptor to perform the re-docking. The docked pose of the ligand showed an RMSD value of 0.6.

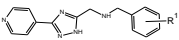
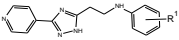
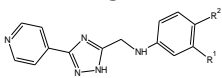
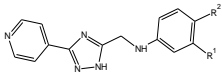
The docking of the most active compound is preceded by the preparation of the ligand and the protein. During the protein preparation, the water molecules are removed and polar hydrogen atoms are added to the protein. Kollman and

Gasteiger charges are calculated for all atoms in the protein. In the ligand preparation, non-polar hydrogen atoms are merged to the ligand and Gasteiger charges are applied. A grid box of dimension 70 X 70 X 70 points was set around the ligand to define the area to be searched during docking. All points on the grid box are separated by 0.375 Å. In the docking process, Lamarckian Genetic Algorithm (LGA) was used to generate 100 conformations with 2500000 evaluations per run. Finally, the outcome of the docking process is analyzed and a pose is selected based on binding energy and interactions reported in earlier studies.

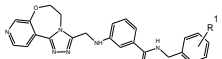
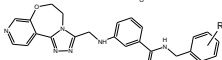
Table 1 Structure of the hydrazone and triazole derivatives and their pIC₅₀ values for GRK2 and ROCK2.



Compounds	Structures	R1	R2	R3	pIC ₅₀	
					GRK2	ROCK2
1	A	H	-	-	5.7	6.1
2	A	2-Cl	-	-	5.0	5.5
3	A	3-Cl	-	-	6.1	6.5
4	A	4-Cl	-	-	5.2	6.2
5	A	2-OMe	-	-	5.0	5.1

6	A	3-OMe	-	-	7.3	7.2
7	A	4-OMe	-	-	5.0	5.4
8	B	H	-	-	5.9	5.7
9	B	OMe	-	-	7.8	6.4
10	B	OEt	-	-	6.9	5.2
11	B	OPh	-	-	7.3	5.2
12	B	OBn	-	-	7.2	5.0
13	B	CO- morpholino	-	-	5.0	5.0
14	B	CONH Bn	-	-	7.6	7.2
15	C	-CH ₂ -	-S-	-Cl	5.9	6.2
16	C	-CH ₂ -	-O-	-Cl	5.3	5.7
17	C	-CH ₂ -	-NH-	-Cl	7.3	7.0
18	C	-S-	-CH ₂ -	-Cl	5.2	5.7
19	C	-NH-	-CH ₂ -	-Cl	6.0	6.5
20	C	-NH-	-CO-	-Cl	5.0	5.0
21		4-H	-	-	5.0	5.0
22		4-Cl	-	-	5.0	5.0
23	C	C	NH	H	7.0	6.2
24	C	C	NH	2-OMe	5.9	5.7
25	C	C	NH	3-OMe	6.7	6.1
26	C	C	NH	4-OMe	6.3	6.5
27	C	C	NH	2-Cl	6.6	5.7
28	C	C	NH	3-Cl	6.8	6.4
29	C	C	NH	3-SO ₂ NH ₂	7.3	6.7
30	C	C	NH	4-SO ₂ NH ₂	5.9	6.4
31	C	C	NH	2-OPh	6.5	5.0
32	C	C	NH	2-OBn	6.4	5.8
33	C	C	NH	2-O(CH ₂) ₂ Ph	6.5	5.5
34	C	C	NH	2-SBn	6.5	5.9
35	C	C	NH	2-SO ₂ NHBn	7.0	7.0
36	C	C	NH	2-SO ₂ Bn	5.7	5.9
37	C	C	NH	2-CONHMe	7.1	6.6
38	C	C	NH	2-CONHPh	7.2	7.5
39	C	C	NH	2-CONHBn	8.2	7.6
40	C	C	NH	2-CONH(CH ₂) ₂ Ph	7.4	7.3
41	C	C	NH	2-CONH(CH ₂) ₂ Ph	7.6	7.0
42	C	C	NH	2-CON(Me)Bn	6.1	5.6
43	C	C	NH	2-NHAc	7.1	7.1
44		NHAc	Cl	-	6.8	7.1
45		NHCO Bn	Cl	-	7.9 208	7.7

46	C	C	NH	2-NHCONHBn	7.7	7.5
47	C	C	NH	2-CH ₂ CONHBn	6.4	6.3
48	D	2-Me	H	-	5	5.1
49	D	3-Me	CONHBn	-	7.3	7.0
50	D	3-CH ₂ OH	CONHBn	-	6.8	6.2
51	D	3-(CH ₂) ₂ OH	H	-	5.4	5.0
52	D	3-CH ₂ OCONH ₂	CONHBn	-	6.2	5.6
53	E	Ac	2-CF ₃	-	6.7	5.0
54	E	ME	2-F-6-F	-	7.2	5.1
55	E	Bn	2-F-6-F	-	7.1	5.0
56	E	(CH ₂) ₂ Ph	2-F-6-F	-	6.6	5.0
57	E	(CH ₂) ₃ Ph	2-F-6-F	-	6.9	5.0
58	F	N	NH	N	7.8	7.2
59	F	CCH ₃	NH	N	8.0	7.6
60	F	N	NH	C	7.6	8.1
61	F	C	N	O	7.6	7.7
62	F	C	O	N	7.3	6.8
63	F	N	C	NH	6.3	6.7
64	F	NCH ₃	N	N	7.3	6.2
65	F	C	NH	C	6.8	7.2
66	F	C	C	C	5.7	6.4
67	F	O	N	N	7.1	6.4
68	F	N	O	N	6.6	6.4
69	F	N	S	C	6.7	7.6
70	F	N	C	S	6.3	6.9
71	F	CPh	NH	N	6.2	6.7
72	G	C	H	4-OMe	7.7	7.5
73	G	C	H	3-OMe	7.9	7.6
74	G	C	H	2-OMe	8.5	7.0
75	G	C	H	2-Cl	8.4	7.3
76	G	C	Me	H	7.3	6.2
77	G	C	Me	4-CF ₃	7.3	6.2
78	G	C	Me	3-CF ₃	7.4	6.2
79	G	C	Me	2-CF ₃	7.7	5.8
80	G	C	Me	2-OMe	7.3	5.7
81	G	C	Me	2-Cl	7.9	5.7
82	G	C	Et	H	8.1	7.1
83	G	C	n-Pr	H	7.4	6.9
84	G	C	i-Pr	H	6.8	6.5
85	G	C	Bn	H	6.4	5.4
86	G	C	(CH ₂) ₂ OMe	H	7.7	6.4

87	G	C	(CH ₂) ₂ OH	2-CF ₃	8.2	6.4
88		2-CF ₃	-	-	8.7	7.3
89		2-F-6-F	-	-	8.7	6.7
90	G	N	Me	2-CF ₃	8.1	6.1
91	G	N	Me	2-F-6-F	8.1	6.1
					6.75	5.7
92	G	N	n-Pr	2-CF ₃	8.7	6.8
93	G	N	n-Pr	2-F-6-F	8.9	6.7

2.3. CoMFA and CoMSIA

The comparative molecular field analysis (CoMFA) [32, 37] and comparative molecular similarity indices analysis (CoMSIA) [33] models were developed for both GRK2 and ROCK2 using Sybyl-X 2.1.

In CoMFA, electrostatic field and steric field exerted by the compounds are calculated at each point on a regularly spaced 3D grid around the compounds. The field exerted at each point on the 3D grid is calculated using a probe atom (sp³ carbon of +1 charge and having a van der Waal radius of 1.52 Å) [32]. The steric and electrostatic fields are contributed by the Lennard- Jones potential and Coulombic potential respectively. Since CoMFA is highly sensitive to the alignment of the dataset compounds, proper alignment of the molecule in 3D Cartesian space is crucial [38]. Here, we considered the docked pose of the most active compound as the bioactive conformation and aligned all compounds using the most active compound as a template. The compounds are aligned by manually superimposing them on the fragment of the template compound which is common in all compounds. Partial least square (PLS) regression analysis is used to linearly correlate the 3D-QSAR descriptor values (independent variables) to the activity values (dependent variables) [39]. The leave-one-out method is used to derive the cross-validated correlation coefficient (q^2) and optimal number of components

(ONC). Based on the ONC, the non-cross-validation was performed to get the conventional correlation coefficient (r^2), standard error of estimation and F-test value (F) [38].

In CoMSIA, five descriptors namely, steric (S), electrostatic (E), H-bond donor (D), H-bond acceptor (A) and hydrophobic (H) descriptors are used. Similar to CoMFA, the descriptor values are calculated by probing the intersection of the 3D cubic lattice around the structure of the compounds by a probe atom (sp^3 carbon atom having a +1 charge). The attenuation factor was set to the default value of 0.3. CoMSIA models based on all different possible combination of the five descriptors were developed and the model giving the best statistical value was selected as the final model.

2.4. Model Validation

To validate the derived CoMFA and CoMSIA models, the models were tested rigorously using both internal and external validation techniques [40].

To access the robustness and statistical confidence of the derived model, bootstrapping analysis was carried out for 1000 runs. Also, to assess the sensitivity of the derived model to chance correlation, we performed leave-five-out cross-validation. In leave-five-out method, a model is derived by omitting five compounds randomly and the derived model is used to predict the activity of the five compounds.

To test the predictive ability of the derived model against external dataset compounds, predictive correlation coefficient (r^2_{pred}) was calculated based on the equation given below [40]:

$$r^2_{pred} = (SD - PRESS)/SD$$

Where SD represents sum of squared deviation between the inhibitory activities of test set compounds and mean activity of the training set compounds and PRESS represents the sum of square deviation between the predicted and actual activity of each compound in the test set.

3. RESULTS AND DISCUSSION

3.1. Molecular Docking

Molecular docking analysis was performed for the most active compound of GRK2 (**93**) and the most active compound of ROCK2 (**60**) to explore the ligand-protein binding mechanisms. The interactions between the most active compounds and the receptors are shown in Figure (**1a**) (GRK2) and Figure (**1b**) (ROCK2).

The crystal structure of GRK2 in complex with one of the dataset compound (**79**) has been reported in a previous study (PDB code **5UUU**) [31]. This crystal structure was used in the docking study of the most active compound (**93**) with GRK2. The binding site of GRK2 consists of residues from adenine subsite (Leu-324, Met-274, Asp-272, Ala-218, Val-205), ribose subsite (Asp-278), phosphate subsite (Lys-220, GLU-239, ARG-199, GLY-200, GLY-201, PHE202, GLY-203), hydrophobic subsite (ALA-236, LEU-235, GLU-239, GLY-337) and catalytic loop (ASP-335) [21]. Docking of the ligand into the receptor was performed with Autodock 4.2.5.1 using the Lamarckian Genetic Algorithm. The docking procedure was first validated by extracting the ligand from the crystal structure and docking it back into the receptor. The resultant docked pose is then compared with the pose of the ligand in the crystal structure. The docking process for the most active compound generated 100 conformations, out of which one conformation was selected based on its binding energy and important interactions which were reported in previous studies.

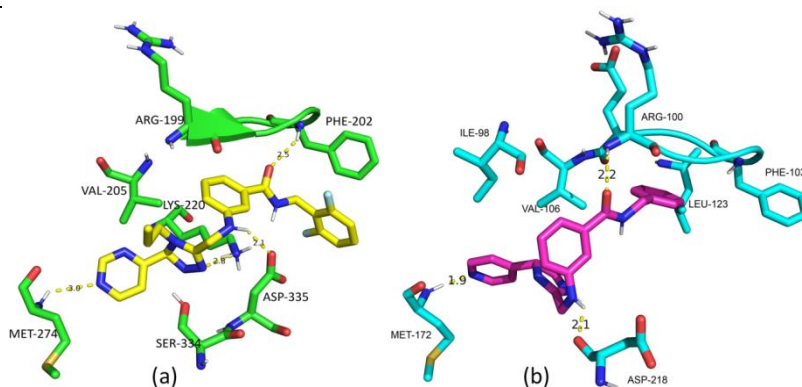


Figure 1 (a) The docked conformation of the most active compound (compound **93**) inside the active site of GRK2. (b) Docked conformation of the most active compound (compound **60**) inside the active site of ROCK2. The hydrogen bond interactions are indicated by dotted lines and bond distances in Angstrom are given (Color figure online).

The most active compound (**93**) of GRK2 forms four hydrogen bonds with the binding site residues as shown in Figure **1a**). One nitrogen atom from the pyrimidine ring of the ligand forms an H-bond with Met-274 at the hinge region. This H-bond was also observed in the crystal structure and is considered to be important as it anchors the ligand inside the binding site. Another H-bond interaction was observed between the nitrogen of ligand triazole ring and the oxygen of ASP-335 of the catalytic loop with the amide nitrogen at the linker between the triazole ring and the aniline ring of the ligand. A fourth interaction was observed between the PHE-202 of the p-loop with the oxygen at the linker between the aniline and the benzyl ring of the ligand. This interaction with the p-loop (PHE-202) was also observed in previous studies [16].

Docking of the most active compounds (**60**) with the ROCK2 was performed using the same procedure used for the docking of compound **93**-GRK2 complex. The crystal structure of ROCK2 (PDB code **4L6Q**) which has previously been

3.2.1. CoMFA and CoMSIA models for GRK2

CoMFA and CoMSIA models were developed for GRK2, followed by the contour map analysis of the selected models to understand the structural properties that are responsible for GRK2 inhibition.

During the 3D-QSAR model development, the docked pose of the most active compound (**93**) was used as the template for aligning the compounds. The fragment of the most active compound (**93**) used in the alignment and the final alignment of all the compounds are shown in Figure (**2a**) and (**2b**) respectively. Two compounds (**21** and **22**) could not be aligned with the template due to structural difference (the presence of an additional atom) and were discarded. Based on the alignment, the CoMFA and CoMSIA models were developed. In CoMFA, the steric and electrostatic descriptors are considered for field calculations. PLS analysis was performed to assess the predictive ability of the models and the results are shown in Table 2. The derived CoMFA model showed a q^2 value of 0.608 and r^2 value of 0.902. The statistical results of the CoMFA model suggested that the model has reasonable predictive ability.

Table 2 Statistical results of the CoMFA and CoMSIA models for GRK2.

Parameters	CoMFA	CoMSIA (SEHD)
q^2	0.608	0.558
ONC	6	5
SEP	0.654	0.689
r^2	0.902	0.877
SEE	0.327	0.363
F value	81.756	77.21
LFO	0.633	0.607

BS r^2	0.928	0.945
BS SD	0.029	0.017
r^2_{pred}	0.574	0.784
Influence of different fields (%)		
S	46.3	14.7
E	53.7	31.8
H	-	19.9
A	-	-
D	-	33.6

q^2 : cross-validated correlation coefficient; ONC: Optimum number of components; SEP: Standard Error of Prediction; r^2 : non-validated correlation coefficient; SEE: Standard Error of Estimation; F value: F-test value; r^2 ; LFO: Leave five out; BS- r^2 : Bootstrapping r^2 mean; BS-SD: Bootstrapping Standard deviation; Q^2 : Progressive scrambling; S: Steric; E: Electrostatic; H: Hydrophobic; A: Acceptor; D: Donor.

Table 3 Statistical results of the CoMFA and CoMSIA models for ROCK2.

Parameters	CoMFA	CoMSIA (SEHD)
q^2	0.621	0.632
ONC	5	5
SEP	0.540	0.53
r^2	0.673	0.867
SEE	0.313	0.32
F value	75.52	71.15
LFO	0.580	0.563
BS r^2	0.913	0.946

BS SD	0.023	0.014
r^2_{pred}	0.511	0.562
Influence of different fields (%)		
S	65.20	17.8
E	34.80	21.6
H	-	20.8
A	-	-
D	-	39.8

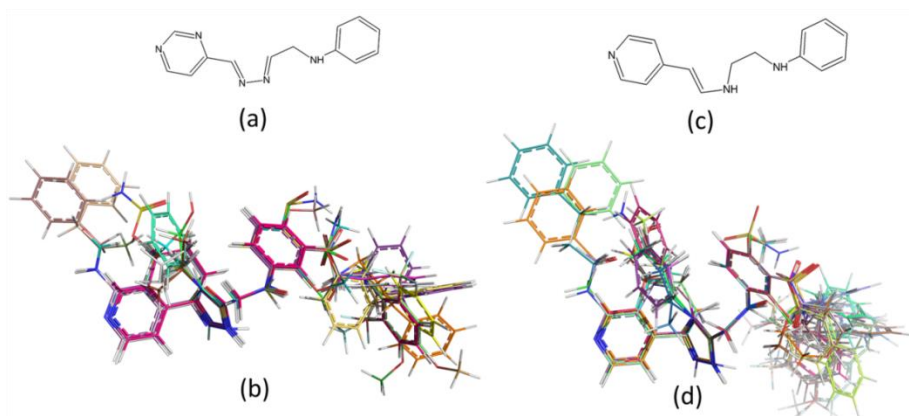


Figure (2) (a) Substructure from template compound (compound **93**) used for the alignment of dataset. (b) Alignment of the dataset compounds used in the GRK2 model development. (c) Substructure from template compound (compound **60**) used for the alignment of dataset. (d) Alignment of the dataset compounds used in the GRK2 model development (Color figure online).

During the CoMSIA model development, all possible combinations of the descriptor fields namely, steric (S), electrostatic (E), hydrophobic (H), H-bond acceptor (A) and H-bond donor (D) were used to develop multiple models. The model based on SEHD descriptors was selected as it resulted in relatively higher

statistical outcome with q^2 , ONC and r^2 value of 0.558, 5 and 0.877 respectively, suggesting that the model has a reasonable predictive ability.

Leave-Five-Out (LFO) method was performed to assess the sensitivity of the models to chance correlation and Bootstrapping (BS) method was used to test the robustness of the derived models. The statistical results obtained from the LFO (0.633) and BS ($BS-r^2 = 0.928$) indicated that the derived CoMFA and CoMSIA models have acceptable accuracy and robustness in its prediction.

To validate the predictive ability of the derived models against external test set, the models were used to predict the activity values of the test set (30 compounds). The CoMFA and CoMSIA models showed acceptable predictive ability for external test set, giving r^2_{pred} value of 0.511 and 0.562 respectively.

3.2.2. CoMFA and CoMSIA models for ROCK2

During the CoMFA and CoMSIA model development for ROCK2, the docked pose of the most active compound (**60**) for ROCK2 was used as the template in aligning the compounds. The fragment of the most active compound (**60**) used in the alignment and the final alignment of all the compounds are shown in Figure (**2c**) and (**2d**) respectively. Two compounds (**20** and **21**) which do not align with the template compound due to structural differences were discarded. Based on the aligned compounds, the models were developed and PLS analysis was performed to assess the predictive ability of the models.

The derived CoMFA model showed a q^2 , ONC and r^2 value of 0.621, 5 and 0.673 respectively, suggesting that the model has a reasonable predictive ability. During the CoMSIA model development the SEHD descriptors were used as it resulted in better statistical value with q^2 , ONC and r^2 value of 0.632, 5 and 0.867 respectively, showing reasonable predictive ability. The statistical results for the ROCK2 CoMFA and CoMSIA models are shown in Table 3.

Both the models developed for ROCK2 were also validated for sensitivity to chance correlation using LFO. The CoMFA and CoMSIA models showed acceptable LFO values of 0.580 and 0.563 respectively. During the test for robustness using bootstrapping (BS) method, the CoMFA model gave a BS- r^2 value of 0.913 with a BS-SD value of 0.023 and the CoMSIA model gave a BS- r^2 value of 0.946 with a BS-SD value of 0.014. These statistical results support that the derived models have reasonable accuracy and robustness in prediction. During the external validation using the test set (30 compounds) the CoMFA and CoMSIA models showed r^2_{pred} value of 0.511 and 0.562 respectively, suggesting that the models can be used for external test set.

The comparison between the predicted and actual pIC_{50} values for both GRK2 CoMFA and CoMSIA models are shown in Table 4. The comparison between the predicted and actual pIC_{50} values for both ROCK2 CoMFA and CoMSIA models are shown in Table 5.

3.3 Contour map analysis

Based on the CoMFA and CoMSIA models derived for GRK2 and ROCK2, contour maps were developed for analysis. The CoMFA and CoMSIA contour maps for GRK2 overlaid with its most active compound **93** for reference are shown in Figure (3) and Figure (4) respectively and the CoMFA and CoMSIA contour maps for ROCK2 overlaid with its most active compound **60** for references are shown in Figure (5) and Figure (6) respectively. In the electrostatic contour maps derived for both GRK2 and ROCK2, blue and red color contours were used to represent electropositive and electronegative substituent favorable regions respectively. Likewise, green and yellow contours were used to represent steric bulk substituent favorable and unfavorable regions respectively. For hydrophobic contour maps derived from the CoMSIA models, yellow contours represent the regions favorable to hydrophobic substituents whereas, and gray contours represent hydrophobic substituent unfavorable regions. In the H-bond

donor contour maps, cyan and purple colors were used to indicate H-bond donor favorable and unfavorable regions respectively.

3.3.1. GRK2 CoMFA contour maps

The steric and electrostatic contour maps developed based on the GRK2 CoMFA model are shown overlaid with the most active compound (**93**) for reference in Figure (3a) and (3b) respectively. In the steric contour map, a green contour was observed near the difluorobenzyl ring suggesting that bulky substituents are favorable in this region and can potentially increase the inhibitory activity toward GRK2. While the yellow contour near the pyrimidine ring and triazole rings suggested that bulky substituents are not favorable in these regions and could lead to a decrease in inhibitory activity of the compounds. This can be observed in compounds **51**, **52**, **56** and **57** (pIC_{50} = 5.420-6.959) which have significantly larger bulky substituent near the pyrimidine ring and showed lower activity values compared to compound **17**, **39** and **54** (pIC_{50} =7.208-8.215) which have relatively smaller substituents.

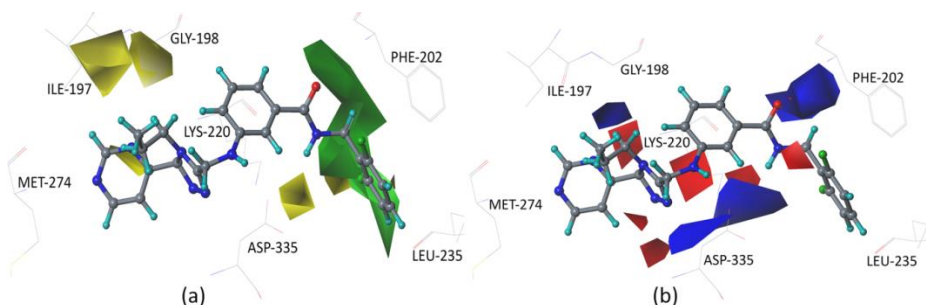


Figure (3) Standard coefficient contour maps obtained from GRK2 CoMFA analysis with the template compound (compound **93**) as reference. (a) Steric contour map. Green contour indicates steric bulk favorable regions and yellow contour indicates steric bulk unfavorable regions. (b) Electrostatic contour map. Blue contour indicates electropositive substituent favorable regions and red contour indicates electronegative substituent favorable regions (Color figure online).

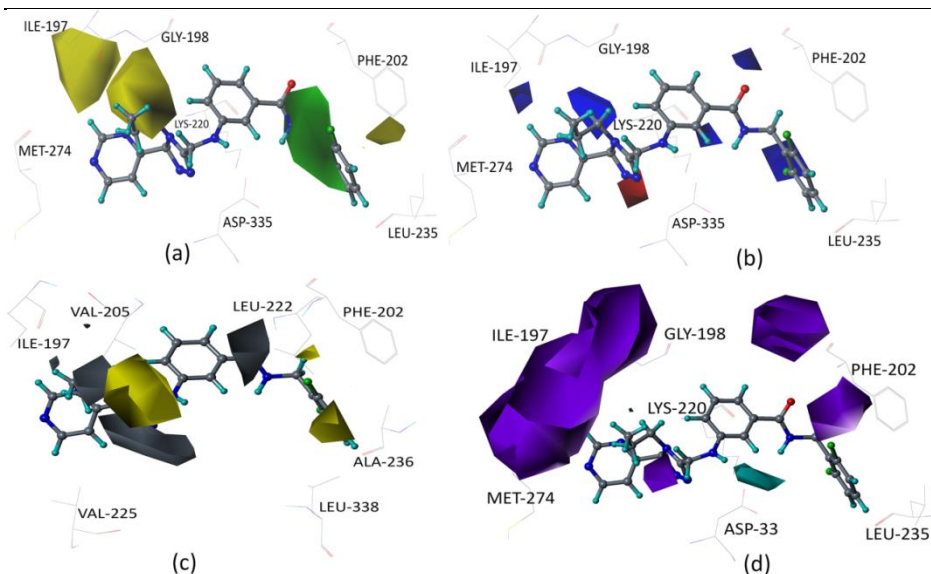


Figure (4) Standard contour maps obtained from GRK2 CoMSIA analysis with the template compound (compound **93**) as reference. **(a)** Electrostatic contour map. Blue contour indicates electropositive substituent favorable regions and red contours indicate electronegative substituent favorable regions. **(b)** Steric contour map. Green contour indicates steric bulk favorable regions and yellow contour indicates steric bulk unfavorable regions. **(c)** Hydrophobic contour map. Yellow and gray contours indicate hydrophobic substituents favorable and unfavorable regions respectively. **(d)** Hydrogen bond donor contour map. Cyan and purple contours indicate hydrogen bond donor substituent favorable and unfavorable regions respectively (Color figure online).

In the electrostatic contour map, blue contours were observed near the linker between the aniline ring and the difluorobenzyl ring and also near the linker between the triazole ring and aniline ring suggesting that electropositive substituents are favorable in these regions. Electropositive substituents at the linker position between the triazole ring and the aniline ring can lead to H-bond interaction with the H-bond acceptor oxygen of ASP-335 at the catalytic loop, as observed in the docking study of the most active compound (**93**) with GRK2 in Figure (1).

Red contour was observed near the triazole ring suggesting that electronegative and H-bond acceptor substituents are favorable at this region. This is observed in

compounds **86** to **93** (pIC_{50} = 7.796-8.921) all of which have H-bond acceptor and electronegative substituents at the triazole ring and showed relatively higher activity value compared to compound 52, 56, 65 and 71 (pIC_{50} = 6.252-6.602) which do not have H-bond acceptor and electronegative substituents. Having electronegative substituents at these regions can form H-bond interactions with donor nitrogen atom of LYS-220 as seen in the docking study of the most active compound (**93**) with GRK2 in Figure (1).

3.3.2. GRK2 CoMSIA contour maps

In the CoMSIA contour map analysis shown in Figure (4a) and (4b), the steric and electrostatic contours were found to closely resemble those of the CoMFA steric and electrostatic contours respectively, hence further analysis of these contours was not done.

In the hydrophobic contour map shown in Figure (4c), the yellow contour near the difluorobenzyl ring and near the linker between the triazole ring and aniline ring suggested hydrophobic substituents are favorable in these regions. This is exemplified by compounds **90**, **91**, **92** and **93** all of which have hydrophobic substituents at these positions and showed relatively higher GRK2 inhibitory activity value in the series (pIC_{50} = 8.108-8.921).

In the H-bond donor contour map in Figure (4d), the only cyan contour was observed near the linker between the triazole ring and aniline ring, indicating the importance of H-bond donor at this region. Having H-bond donor substituents here can lead to H-bond interaction with acceptor oxygen of ASP-335 at the catalytic loop, as observed in the docking study of the most active compound (**93**).

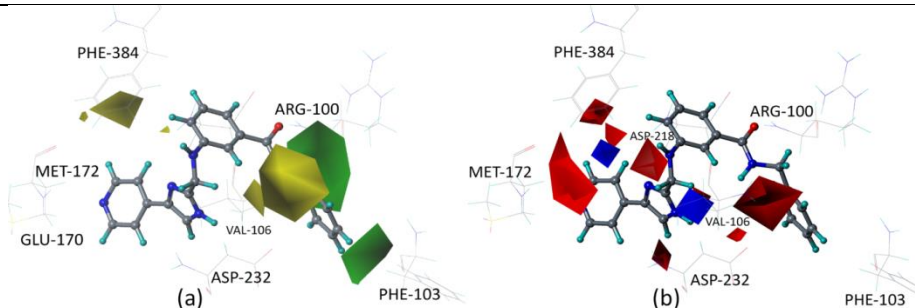


Figure (5) Standard coefficient contour maps obtained from ROCK2 CoMFA analysis with the template compound (compound **60**) as reference. **(a)** Steric contour map. Green contour indicates steric bulk favorable regions and yellow contour indicates steric bulk unfavorable regions. **(b)** Electrostatic contour map. Blue contour indicates electropositive substituent favorable regions and red contour indicates electronegative substituent favorable regions (Color figure online).

3.3.3. ROCK2 CoMFA contour maps

The steric and electrostatic contour map derived from the CoMFA models of ROCK2 is shown in Figure (5a) and Figure (5b). In the steric contour map, two green contours were observed near the benzyl ring suggesting that steric bulk substituents are favorable in these regions. This is exemplified by compounds **31**, **48** and **51** ($pIC_{50}=5.000-5.149$) which does not have the extended benzyl moiety and showed much lower activity value compared to compounds **39**, **45**, **49**, **50** and **60** ($pIC_{50}= 6.268-8.137$) all of which have the extended benzyl moiety. Yellow contours were observed near the pyridine ring and near the linker between the aniline ring and benzyl ring suggesting that bulky substituent in these regions are not favorable. This can be observed in the low active compounds **53**, **54**, **55**, **56** and **57** ($pIC_{50}= 5.000-5.131$) all of which have bulky substituents near the pyridine ring as opposed to compound **39** and **49** ($pIC_{50}= 7.027- 7.620$) which possess smaller substituents. Having bulk substituents at the pyridine ring may cause steric clash with binding site residues such as PHE-384, ILE-98, and TYR-171.

In the electrostatic contour map, a blue contour was observed near the amide linker between the triazole ring and aniline ring, suggesting that electropositive substituents are favorable at this position. Red contours observed near the pyridine ring, near the linker between the pyridine ring and triazole ring, and at the benzyl ring suggested that electronegative substituents are favorable in these regions. Electronegative substituents at the pyridine ring can lead to H-bond interaction with the MET-172 as observed in the docking study of the most active compound (**60**) with ROCK2.

3.3.4. ROCK2 CoMSIA contour maps

The CoMSIA steric and electrostatic contour maps are shown in Figure 6a and 6b. In the electrostatic contour map, as observed in the ROCK2 CoMFA electrostatic contour map, a red contour was observed near the pyridine ring suggesting that electronegative substituents are favorable at this region. A blue contour was observed near the amide linker between the aniline ring and the benzyl ring indicating that electropositive substituents are favorable at this position. As for the steric contour map, it resembles closely to the CoMFA steric contour map and therefore further separate analysis was not done.

In the hydrophobic contour map shown in Figure (**6c**), yellow contour was observed near the benzyl ring, suggesting hydrophobic substituents are favorable in this region. This can be observed in compounds **38**, **39**, **45** and **60** having hydrophobic substituents such as benzyl group and phenyl group in this region and showed higher activity ($pIC_{50}= 7.509-8.137$) compared to compounds **31**, **48** and **51** ($pIC_{50}=5.000-5.149$) which do not have hydrophobic substitution at that region. Yellow contour was also observed near the triazole ring suggesting hydrophobic substituents are favorable at this region whereas, hydrophobic substitutions are not favorable near the aniline ring as indicated by the gray contour.

In the H-bond donor contour map shown in Figure (6d), the purple contour observed near the pyridine ring indicating that H-bond donors are not favorable at that region whereas, a cyan contour was observed near the linker between the aniline and benzyl ring suggesting that H-bond donors are favorable at this region.

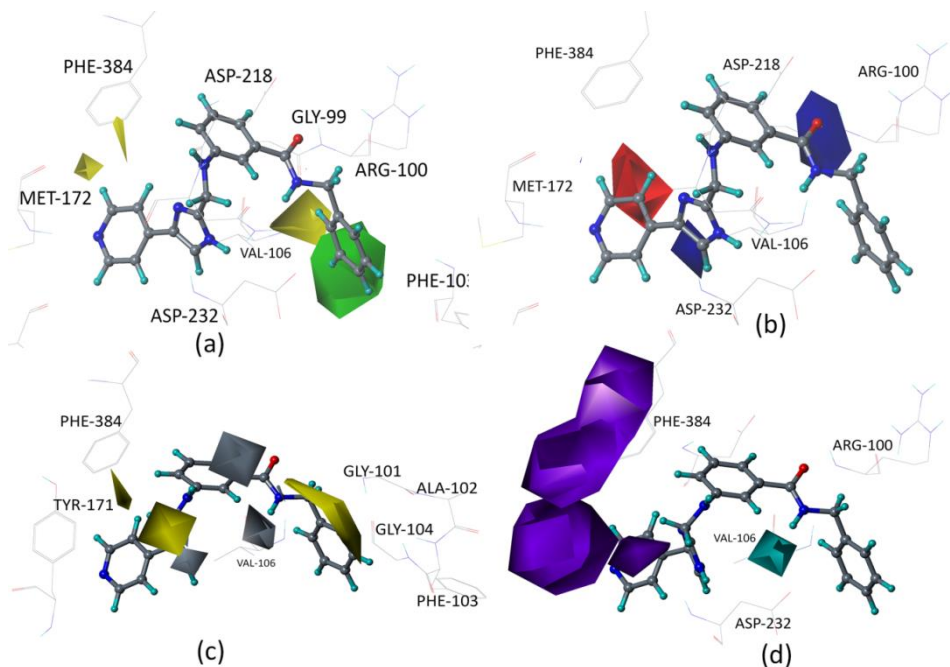


Figure (6) Standard contour maps obtained from ROCK2 CoMSIA analysis with the template compound (compound **60**) as reference. **(a)** Electrostatic contour map. Blue contour indicates electropositive substituent favorable regions and red contours indicate electronegative substituent favorable regions. **(b)** Steric contour map. Green contour indicates steric bulk favorable regions and yellow contour indicates steric bulk unfavorable regions. **(c)** Hydrophobic contour map. Yellow and gray contours indicate hydrophobic substituents favorable and unfavorable regions respectively. **(d)** Hydrogen bond donor contour map. Cyan and purple contours indicate hydrogen bond donor substituent favorable and unfavorable regions respectively (Color figure online).

From the analysis of the CoMFA and CoMSIA contour maps derived from GRK2 and ROCK2, we observed that having non-bulky substituents near the pyridine/pyrimidine ring, steric bulk and hydrophobic substituents at the benzyl/difluorobenzyl ring and electronegative substituents near the linker between the triazole ring and the pyridine/pyrimidine ring of the compounds are favorable in both GRK2 and ROCK2 and can potentially lead to increase in the inhibitory activity of the compounds towards both GRK2 and ROCK2.

Further analysis of the steric contours showed that green contour near the difluorobenzyl ring derived from GRK2 models are larger than the green contour observed near the benzyl ring derived from the ROCK2 models, suggesting that GRK2 has a relatively higher affinity for bulky substitutions in this region compared to ROCK2. This could be a possible reason why compounds having the bulky benzyl ring without additional substitutions are more favorable in the case of ROCK2 inhibition and are among the compounds having the highest inhibitory activity for ROCK2 in the dataset. Adding substitutions to the benzyl ring tend to reduce the inhibitory activity of the compounds toward ROCK2. This can be observed in compounds **79**, **80** and **91** all of which have additional substitutions at the benzyl ring and showed lower activity value (pIC_{50} = 5.721-5.854) compared to compounds **39**, **59**, **60**, **61** and **69** (pIC_{50} = 7.638-8.137) all of which have no substitution on the benzyl ring. Docking analysis of the most active compound (**60**) with ROCK2 showed that the pose adopted by the most active compound place the benzyl ring in close proximity to the residues at the p-loop. Addition of substituents to the benzyl ring could cause steric clash with these residues. This could be a possible explanation of why additional substitutions to the benzyl ring are not favorable and lead to a decrease in inhibitory activity. On the other hand, the addition of substituents like difluoro and trifluoromethyl group on the benzyl ring increases the activity of the compounds towards GRK2 as seen in compounds **90** to **93** (pIC_{50} = 8.108-8.921) which are among the compounds with the highest activity towards GRK2 in the dataset. From docking study of the most active

compound (**93**) with GRK2, it was observed that the compound extends the difluorobenzyl ring into the hydrophobic pocket which can accommodate the added substituents at the benzyl ring.

From the analysis, it was observed that non-bulky substituents near the triazole ring, bulky and hydrophobic substituents near the benzyl/difluorobenzyl ring and electronegative and H-bond acceptor substituents at the triazole ring tends to increase the inhibitory activity of the compounds towards GRK2 over ROCK2. This can be observed in compounds **79**, **80**, **81** and **91** which show high activity preference towards GRK2 ($pIC_{50}=7.357-8.168$) and low activity towards ROCK2 ($pIC_{50}=5.721-5.854$). All these compounds which showed activity preference towards GRK2 have smaller substitutions like methyl group at the pyrimidine ring, electronegative and H-bond acceptor nitrogen at the triazole ring and have bulky and hydrophobic substituents at the benzyl/difluorobenzyl ring.

4. CONCLUSIONS

The selective inhibition of GRK2 is considered to be important for the successful treatment of heart failure. In this study, we performed molecular docking study and 3D-QSAR analysis on a series of triazole derivatives and hydrazone derivatives that showed activity against both GRK2 and ROCK2, to understand the ligand-protein binding mechanisms and the structural properties that drive the inhibitory preference of the compounds toward GRK2 over ROCK2. Docking studies revealed that the most active compound of GRK2 forms four H-bond interactions with residues MET-274, LYS-220, ASP-335 and PHE-202 of the active site. On the other hand, the most active compound of ROCK2 showed three H-bond interactions with MET-172, ASP-218 and ARG-100 residues of the active site. CoMFA and CoMSIA models were developed for both GRK2 and ROCK2 and contour maps were derived based on these models.

Co-analysis of the molecular docking results and 3D-QSAR results derived from GRK2 and ROCK2 revealed that the presence of electronegative substituents near the linker between the pyridine/pyrimidine ring and triazole ring, hydrophobic substituents near the benzyl/difluorobenzyl ring and non-bulky substituents near the pyridine/pyrimidine ring of the compounds increases its inhibitory activity towards both GRK2 and ROCK2. On the other hand, the addition of hydrophobic and bulky substituents near the benzyl/difluorobenzyl ring and the presence of electronegative and H-bond acceptor substituents at the triazole ring tend to selectively increase the activity of the compounds towards GRK2 over ROCK2. The outcome of this study may be used in rational drug design to develop potent GRK2 inhibitors with selectivity over ROCK2.

Table 4. Experimental and predicted pIC_{50} values with their residuals of selected CoMFA and CoMSIA model derived from GRK2.

Compound	Actual	CoMFA		CoMSIA	
	pIC_{50}	Predicted	Residual	Predicted	Residual
1 [*]	5.745	5.753	-0.008	5.665	0.080
2	5.000	5.332	-0.332	5.615	-0.615
3	6.155	5.992	0.162	5.716	0.439
4	5.252	5.588	-0.336	5.518	-0.266
5	5.000	4.972	0.02	5.287	-0.287
6 [*]	7.310	6.325	0.984	5.860	1.450
7 [*]	5.000	5.529	-0.529	5.526	-0.526
8	5.959	6.062	-0.103	6.309	-0.350
9 [*]	7.854	6.455	0.465	7.563	0.291
10 [*]	6.921	7.385	-0.008	6.678	0.243

11	7.377	7.418	-0.142	7.187	0.190
12	7.276	5.753	-0.008	7.020	0.256
13 [*]	5.000	5.229	-0.229	5.707	-0.707
14	7.699	7.643	0.056	7.611	0.088
15 [*]	5.921	5.509	0.411	5.688	0.233
16	5.357	5.611	-0.254	5.601	-0.245
17 [*]	7.387	5.589	1.798	6.008	1.379
18	5.244	5.317	-0.072	5.452	-0.208
19 [*]	6.009	5.614	0.394	5.297	0.712
20	5.000	5.174	-0.174	5.147	-0.147
23	7.000	6.1452	0.854	6.975	0.025
24 [*]	5.921	5.492	0.428	5.794	0.127
25	6.796	6.925	-0.129	6.539	0.257
26	6.357	6.383	-0.026	6.155	0.201
27	6.602	5.703	0.899	6.008	0.594
28 [*]	6.854	5.895	0.958	5.975	0.879
29	7.357	7.037	0.3196	7.388	-0.032
30	5.921	6.124	-0.203	5.889	0.032
31 [*]	6.523	6.296	0.226	6.448	0.075
32	6.444	6.045	0.398	6.111	0.333
33	6.553	6.518	0.034	6.479	0.074
34 [*]	6.538	6.22	0.317	5.977	0.561

35	7.000	7.041	-0.040	7.336	-0.336
36	5.745	5.733	0.011	5.933	-0.188
37	7.187	7.124	0.063	7.171	0.016
38 [*]	7.268	7.105	0.162	7.099	0.169
39 [*]	8.215	7.276	0.938	8.213	0.002
40 [*]	7.420	8.038	-0.617	7.772	-0.352
41	7.638	7.784	-0.145	7.633	0.005
42	6.155	6.799	-0.644	6.330	-0.175
43	7.187	7.69	-0.503	7.752	-0.565
44 [*]	6.886	7.572	-0.685	7.638	-0.752
45	7.921	7.843	0.078	7.708	0.213
46	7.721	7.594	0.126	7.451	0.270
47	6.495	6.793	-0.298	6.534	-0.039
48 [*]	5.000	5.746	-0.746	5.008	-0.008
49	7.377	7.194	0.183	7.398	-0.021
50	6.886	6.808	0.078	6.811	0.075
51 [*]	5.420	5.548	-0.127	5.968	-0.548
52	6.252	6.136	0.115	5.863	0.389
53	6.770	6.837	-0.067	6.774	-0.004
54	7.208	6.974	0.233	7.080	0.128
55	7.174	7.294	-0.120	7.084	0.090
56	6.602	6.543	0.058	6.587	0.015

57 [*]	6.959	6.987	-0.028	6.624	0.335
58	7.854	7.579	0.275	7.731	0.123
59	8.004	7.517	0.487	7.695	0.309
60	7.678	6.706	0.971	7.177	0.501
61	7.620	7.077	0.543	7.112	0.508
62	7.337	7.196	0.141	6.984	0.353
63 [*]	6.337	6.211	0.126	6.203	0.134
64	7.347	7.544	-0.197	7.421	-0.074
65 [*]	6.886	7.135	-0.248	7.471	-0.585
66	5.770	6.684	-0.914	6.453	-0.683
67 [*]	7.174	7.057	0.116	7.035	0.139
68	6.699	7.012	-0.313	6.926	-0.227
69	6.796	6.495	0.301	6.508	0.288
70	6.328	6.42	-0.092	6.017	0.311
71	6.292	6.271	0.021	6.926	-0.634
72	7.770	7.706	0.063	8.154	-0.384
73	7.921	7.996	-0.074	8.285	-0.364
74	8.523	8.53	-0.007	8.532	-0.009
75	8.409	8.167	0.242	8.484	-0.075
76	7.347	7.493	-0.145	7.415	-0.068
77 [*]	7.347	7.516	-0.169	7.789	-0.442
78 [*]	7.409	7.939	-0.530	7.868	-0.459

79	7.745	7.928	-0.183	8.031	-0.286
80*	7.357	8.292	-0.935	7.654	-0.298
81	7.921	7.951	-0.031	7.668	0.253
82*	8.125	7.387	0.738	7.362	0.763
83	7.495	7.704	-0.208	7.322	0.173
84	6.854	6.966	-0.112	7.136	-0.282
85*	6.432	6.991	-0.559	7.176	-0.744
86*	7.796	6.808	0.987	7.089	0.707
87	8.260	7.916	0.343	8.085	0.175
88	8.745	8.681	0.063	8.598	0.147
89*	8.770	8.661	0.108	8.665	0.105
90	8.108	8.457	-0.349	8.585	-0.477
91	8.168	8.461	-0.293	8.618	-0.451
92	8.721	8.708	0.012	8.461	0.260
93	8.921	8.712	0.208	8.494	0.427

* represents the test set compounds.

Table 5. Experimental and predicted pIC_{50} values with their residuals of selected CoMFA and CoMSIA model for ROCK2.

Compound	Actual	CoMFA		CoMSIA	
	pIC_{50}	Predicted	Residual	Predicted	Residual
1*	6.1938	6.235	-0.041	5.933	0.260
2	5.585	5.770	-0.185	5.799	-0.214

3	6.5229	6.330	0.193	5.908	0.614
4*	6.2676	6.273	-0.005	5.681	0.586
5	5.1308	5.268	-0.137	5.654	-0.523
6*	7.2076	5.946	1.261	5.715	1.492
7*	5.432	6.282	-0.850	6.219	-0.787
8	5.745	5.689	0.056	5.653	0.091
9*	6.409	5.652	0.757	5.427	0.981
10	5.244	5.170	0.075	5.305	-0.060
11	5.208	4.919	0.289	5.184	0.023
12*	5.000	5.765	-0.765	5.594	-0.594
13	5.000	4.920	0.080	4.814	0.186
14	7.237	6.795	0.442	6.891	0.345
15*	6.229	5.571	0.658	5.538	0.691
16	5.770	5.593	0.177	5.563	0.206
17*	7.027	5.638	1.389	5.659	1.367
18	5.721	5.534	0.187	5.484	0.237
19*	6.569	5.709	0.860	6.111	0.457
20	5.000	5.078	-0.078	4.913	0.087
23	6.201	6.235	-0.034	6.04	0.160
24	5.745	5.257	0.488	5.753	-0.008
25	6.143	5.953	0.189	5.823	0.319
26	6.553	6.721	-0.168	6.329	0.223

27	5.770	5.765	0.005	5.906	-0.136
28	6.420	6.331	0.089	6.015	0.405
29	6.796	6.807	-0.011	7.063	-0.267
30	6.444	6.789	-0.345	6.86	-0.416
31	5.000	4.949	0.051	5.243	-0.243
32	5.854	5.987	-0.134	5.935	-0.081
33	5.569	5.567	0.001	5.86	-0.291
34*	5.921	6.144	-0.223	5.935	-0.014
35	7.009	7.227	-0.218	6.992	0.016
36	5.959	5.979	-0.020	5.859	0.099
37	6.658	6.597	0.061	5.848	0.809
38	7.509	7.281	0.227	7.2	0.308
39*	7.620	7.358	0.262	6.848	0.771
40*	7.328	7.028	0.300	7.2	0.127
41	7.000	7.410	-0.410	7.243	-0.243
42	5.658	6.447	-0.789	5.744	-0.086
43	7.155	7.304	-0.149	7.415	-0.260
44	7.131	7.238	-0.107	7.155	-0.024
45	7.745	7.362	0.383	7.685	0.059
46*	7.523	6.764	0.759	7.155	0.367
47*	6.319	6.659	-0.340	6.685	-0.366
48	5.149	5.495	-0.346	5.855	-0.706

49	7.027	6.911	0.116	7.126	-0.099
50	6.268	6.628	-0.360	6.682	-0.414
51	5.000	5.151	-0.151	4.89	0.110
52	5.638	5.920	-0.282	5.228	0.410
53	5.000	4.851	0.149	4.759	0.241
54	5.131	5.262	-0.132	5.383	-0.252
55*	5.000	4.633	0.367	4.759	0.241
56	5.000	4.772	0.228	5.06	-0.060
57*	5.000	5.148	-0.148	5.105	-0.105
58*	7.222	7.245	-0.023	7.06	0.161
59	7.638	7.150	0.488	7.210	0.428
60	8.137	7.397	0.740	7.503	0.633
61*	7.745	6.915	0.830	7.21	0.534
62	6.886	6.812	0.074	6.622	0.264
63	6.796	7.115	-0.320	6.619	0.176
64	6.260	6.285	-0.025	6.379	-0.119
65*	7.284	6.839	0.445	7.674	-0.390
66	6.469	6.647	-0.179	6.69	-0.221
67*	6.495	6.802	-0.307	5.674	0.821
68	6.456	6.988	-0.532	6.745	-0.289
69*	7.658	6.875	0.783	7.141	0.516
70	6.921	6.871	0.050	6.633	0.287

71*	6.770	6.998	-0.228	6.985	-0.215
72	7.569	7.319	0.250	7.248	0.320
73	7.699	7.330	0.369	7.291	0.408
74	7.071	7.159	-0.088	7.226	-0.155
75	7.301	6.777	0.525	7.034	0.267
76	6.260	6.285	-0.025	6.379	-0.119
77*	6.208	6.662	-0.454	7.034	-0.826
78*	6.276	6.732	-0.456	6.379	-0.103
79	5.854	6.415	-0.561	6.32	-0.466
80*	5.796	6.371	-0.575	6.425	-0.629
81	5.796	6.318	-0.522	6.32	-0.524
82*	7.119	6.376	0.743	6.452	0.667
83*	6.959	6.437	0.522	6.32	0.638
84	6.585	6.619	-0.034	6.43	0.15
85	5.495	5.570	-0.075	5.914	-0.411
86*	6.495	5.910	0.585	6.43	0.064
87*	6.495	6.395	0.100	5.914	0.580
88*	7.310	6.407	0.903	6.926	0.383
89	6.721	6.283	0.438	6.362	0.359
90	6.143	6.655	-0.513	6.658	-0.515
91*	5.721	6.525	-0.804	6.362	-0.640
92	6.854	6.779	0.075	6.549	0.304

Seketoulie Keretsu Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

93 6.721 6.728 -0.007 6.651 0.070

* **represents test set compounds**

PART III

Computational study of paroxetine-like inhibitors reveals new molecular insight to inhibit GRK2 with selectivity over ROCK1

1. INTRODUCTION

Heart failure is a condition in which the heart fails to produce sufficient myocardial contraction needed to effectively circulate blood throughout the body. The low circulation of blood is mitigated by the release of catecholamines by the sympathetic nervous system [42] [43]. Catecholamines bind to the β -adrenergic receptor (β -AR) on the cell surface and activate the downstream release of cAMP, which induces the positive inotropic needed for myocardial contraction of the heart [44]. The stimulated β -AR is desensitized through the phosphorylation of its serine and threonine residues by G protein-coupled receptor kinase 2 (GRK2) [45-47]. The phosphorylation by GRK2 induces arrestin binding at the β -AR, thereby blocking the pathway responsible for increased myocardial contraction [48-50]. Hence, the desensitization of β -AR by inhibition of GRK2 is considered as a potential route for heart failure treatment [51].

GRK2 is a serine/threonine kinase and is one of the members of A, G, and C family (AGC family) of kinases. AGC kinases play a vital role in cell survival, insulin signalling, regulation of ion transporters and channels, and blood pressure among others and its aberrant activity has been shown to be implicated in several diseases [13, 52]. Due to high sequence and structural similarity at the kinase domain among AGC kinases (~33% identity), the inhibition of GRK2 leads to inhibition of other AGC kinases [53]. The Rho-associated coiled-coil containing kinase 1 (ROCK1) is a member of the AGC kinase family and plays crucial role in several vital cellular functions including gene transcription, proliferation, differentiation, apoptosis and oncogenic transformation [54-56]. Rho-associated coiled-coil containing kinase (ROCK2) is also another member of the AGC kinase family. They are known to play an important role in cell migration and invasion, centrosome duplication, cytokinesis, and apoptosis [20]. Several studies have shown that the inhibition of GRK2 leads to the inhibition of ROCK1 and ROCK2. In addition, cross activity between GRK2 and other AGC kinases such as GRK1, GRK3 and protein kinase A (PKA) have been observed and reported

[21, 26, 53]. Therefore, the selective inhibition of GRK2 is considered to be crucial, to avoid unwanted side effects that may result from the inhibition of other AGC kinases.

The catalytic domain of the AGC kinases is highly conserved and consists of a small lobe (N-lobe) and a large lobe (C-lobe). The active site, where most AGC kinase inhibitors bind, is formed at the intersection between the two lobes [21]. The conserved active site consists of the adenine subsite which is adjacent to the hinge moiety, the ribose subsite, the polyphosphate subsite and the hydrophobic subsite. The hydrophobic subsite is made up of residues from the p-loop, the α C-Helix and the DFG motif [57]. A comparison of the residues at the active site of GRK2, ROCK1 and ROCK2 are shown in Table 1.

Table 1. Comparison of the residues at the adenine subsite, polyphosphate subsite, ribose subsite and hydrophobic subsite for GRK2, ROCK1 and ROCK2

Subsites		GRK2	ROCK1	ROCK2
Adenine Subsite		Met274, Asn275, Gly276, Gly277, Asp278, Leu279, His280	Met156, Pro157, Gly158, Gly159, Asp160, Leu161, Val162	Met172, Pro173, Gly174, Gly175, Asp176, Leu177, Val178.
Polyphosphate Subsite		Tyr217, Ala218, Met219, Lys220, Cys221, Leu222	Tyr102, Ala103, Met104, Lys105, Leu106, Leu107	Tyr118, Ala119, Met120, Lys121, Leu122, Leu123
Ribose Subsite		Gly276, Gly277, Asp278, Leu279, His280, Tyr281, His282, Leu283, Ser284	Gly156, Gly159, Asp160, Leu161, Val162, Asn163, Leu164, Met165	Gly174, Gly175, Asp176, Leu177, Val178, Asn179, Leu180, Met181,
Hydrophobic Subsite	P-loop	Ile197, Gly198, Arg199, Gly200, Gly201, Phe202, Gly203, Glu204, Val205	Ile98, Gly99, Arg100, Gly101, Ala102, Phe103, Gly104, Gly105, Val106	Ile82, Gly83, Arg84, Gly85, Ala86, Phe87, Gly88, Glu89, Val90

	αC-Helix	Thr234, Leu235, Ala236, Leu237, Asn238, Glu239, Arg240, Ile241, Met242, Leu243, Ser244, Leu245, Val246, Ser247	Ser118, Ala119, Phe120, Phe121, Trp122, Glu123, Glu124, Arg125, Asp126, Ile127, Met128, Ala129, Phe130, Ala131	Ser134, Ala135, Phe136, Phe137, Trp138, Glu139, Glu140, Arg141, Asp142, Ile143, Met144, Ala145, Phe146, Ala147
	DFG/DL G motif	Asp335, Leu336, Gly337	Asp216, Phe217, Gly218	Asp232, Phe233, Gly234

Several GRK2 inhibitors have been reported over the last decade. The natural product balanol potently inhibits GRK2 with an IC_{50} of 50 nM (at 3 μ M ATP) but lacks the selectivity against protein kinase A (PKA) and protein kinase C (PKC) [18]. Paroxetine, which is an FDA approved serotonin reuptake inhibitor is modestly potent towards GRK2 with an IC_{50} of 1.1 μ M and selective against other GRKs [25] GSK180736A which was originally developed as an ROCK1 inhibitor is a potent inhibitor of GRK2 (IC_{50} = 0.77 μ M) and is selective against other GRKs. However GSK180736 exhibited limited bioavailability [26, 58] CMPD101 and CMPD103 developed by Takeda pharmaceuticals showed high activity for GRK2 with selectivity over other AGC kinases but are not bioavailable [23, 31]. Bouley et al., developed a series of indazole hybrid compounds that showed high potency for GRK2 but these compounds also showed activity for GRK5, ROCK1 and PKA [59]. Recently, Waldschmidt et al., have reported a series of paroxetine-like compounds that showed high inhibitory activity for GRK2 and selectivity over other AGC kinases [16, 53]. A study of this series of paroxetine-like compounds with the objective to understand the structural factors that drive its potency and selectivity for GRK2 poses an interesting challenge. Therefore, these paroxetine-like compounds were selected for computational study.

In-silico drug design techniques have emerged as powerful methods in assisting drug discovery [60-62]. In this study, we have performed molecular docking, molecular dynamics simulation and molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) free energy calculations on 53 paroxetine-like compounds [16, 53] to gain detailed insight into the binding interactions and binding stability of the inhibitors. Using three dimensional quantitative structure activity relationship (3D-QSAR) studies, CoMFA models were developed for both GRK2 and ROCK1. The contour maps developed from the CoMFA models were analyzed to understand the structural changes favorable for high activity. The contour map results and docking analyses of individual receptors were co-analysed to identify the crucial interactions and structural properties that are important to increase the inhibitory activity for GRK2 and selectivity over ROCK1.

2. METHODOLOGY

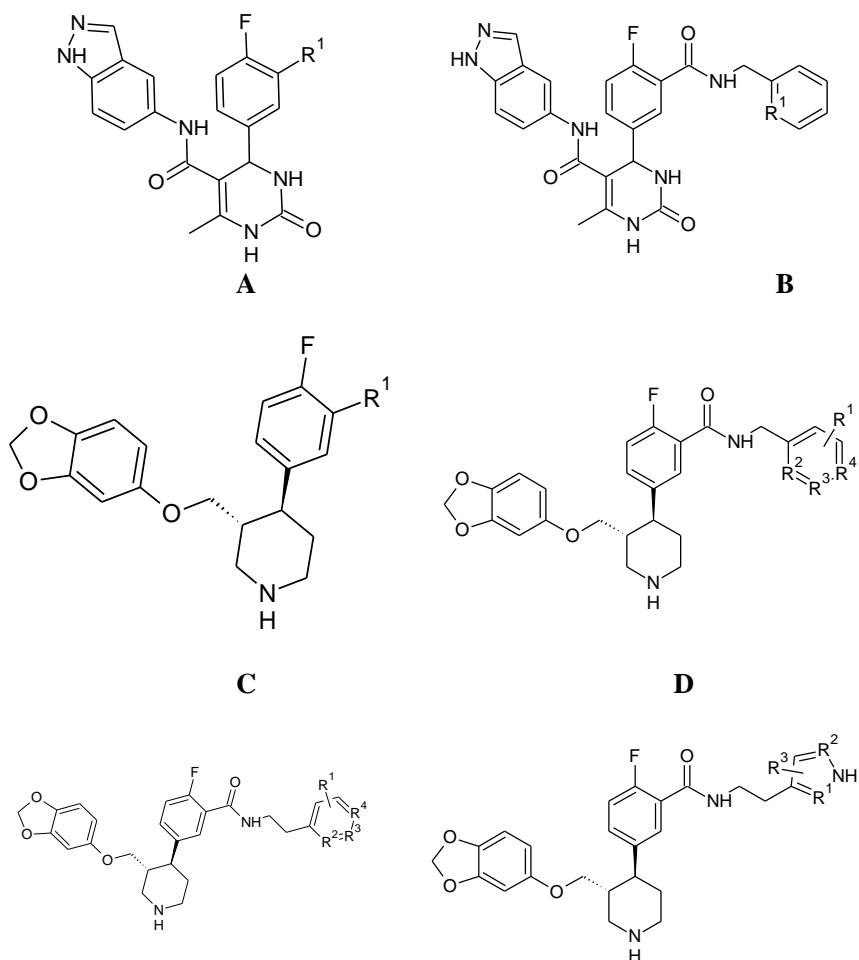
2.1. Dataset:

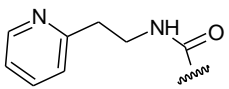
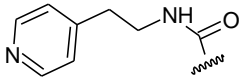
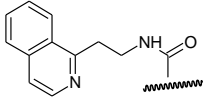
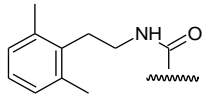
A series of 53 paroxetine-like compounds having activity values for GRK2 and ROCK1 were collected from recent literature [16, 53]. The inhibitory concentration IC_{50} value of the compounds were converted to pIC_{50} ($-\log IC_{50}$) values. The series of compounds showed an activity range of 4.42 to 7.52 for GRK2 and an activity range of 5.17 to 7.96 for ROCK1. The structure of the compounds and their pIC_{50} values for GRK2 and ROCK1 are provided in Table 2.

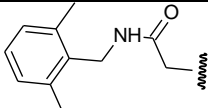
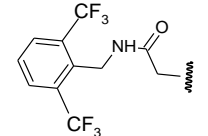
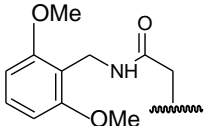
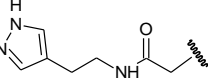
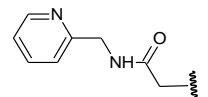
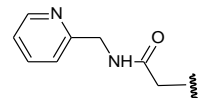
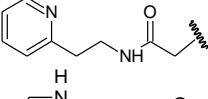
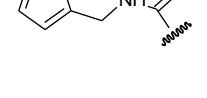
The most active compound for GRK2 (compound **47**) showed a pIC_{50} value of 7.523 for GRK2 and showed more than 230-fold selectivity over the other kinases including ROCK1. The most active compound for ROCK1 (compound **11**) showed pIC_{50} value of 6.824 and 7.959 for GRK2 and ROCK1 respectively. The most selective compound (compound **17**) showed activity pIC_{50} value of 6.886 for GRK2 with more than 700-fold selectivity over other kinases including ROCK1.

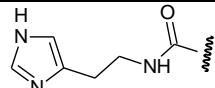
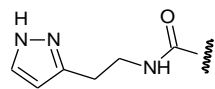
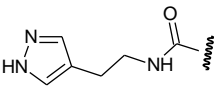
Based on the binding pose of the most active compound, the rest of the compounds were sketched and minimized using the Tripos force field in Sybyl-X 2.1. The dataset was randomly divided into a test set and a training set. A training set of 33 compounds were used to develop the CoMFA model for GRK2 and 19 compounds were used to validate the model. During the CoMFA model development for ROCK1, 21 compounds were used to build the model after removing all compounds that do not have a specified activity value for ROCK1.

Table 2. Structure of the paroxetine-like derivatives and their pIC₅₀ values for GRK2 and ROCK1.



Compound ds	Structures	E				F			
		R ¹	R ²	R ³	R ⁴	GR K2 pIC 50	ROC K1 pIC ₅₀		
1	C	H	-	-	-	5.8	10%*		
2	A	H	-	-	-	6.1	7		
3	A	COOH	-	-	-	4.6	6.7		
4	A	CONHMe	-	-	-	5.3	6.2		
5	B	C	-	-	-	6.1	7.1		
6	B	C	3-F	-	-	6.6	7.6		
7	B	C	2-F-6-F	-	-	6.6	6.9		
8	B	C	2-OMe	-	-	7.2	7.2		
9	B	C	3-OMe	-	-	6.3	7.0		
10	B	C	4-OMe	-	-	6.3	7.3		
11	B	N	-	-	-	6.8	7.9		
12	A		-	-	-	6.5	7.6		
13	A		-	-	-	5.3	7.0		
14	A		-	-	-	6.5	7.9		
15	B	C	2-Cl-6-Cl	-	-	6.8	5.1		
16	B	C	2-Me-6-Me	-	-	7.1	5.2		
17	B	C	2-OMe-6-OMe	-	-	6.8	0%*		
18	B	C	2-CF3-6-CF3	-	-	5.9	22%*		
19	B	C	3-CF3-5-CF4	-	-	5.5	5.7		
20	B	N	2-Me	-	-	5.7	6.8		
21	A		-	-	-	6.6	6.5		

22	A		-	-	-	4.6	6.3
23	A		-	-	-	ND	6.3
24	A		-	-	-	6.3	6.6
25	C		-	-	-	6.2	27%*
26	C	CONHCH3				5.6	11%*
27	D	-	C	C	C	6.1	17%*
28	E	-	C	C	C	5.5	22%*
29	D	2-F-6-F	C	C	C	5.8	34%*
30	E	2-F-6-F	C	C	C	5.6	0%*
31	D	2-CF3	C	C	C	4.9	5%*
32	D	2-Me-6-Me	C	C	C	5.6	0%*
33	D	2-Cl-6-Cl	C	C	C	5.7	0%*
34	D	2-OMe-6-OMe	C	C	C	5.6	13%*
35	D	2-OMe	C	C	C	4.4	1%*
36	C		-	-	-	5.6	18%*
37	D	-	N	C	C	6.2	6%*
38	D	-	C	N	C	5.8	6%*
39	D	-	C	C	N	5.6	8%*
40	E	-	N	C	C	5.4	15%*
41	E	-	C	N	C	5.5	18%*
42	E	-	C	C	N	5.4	21%*
43	C		-	-	-	5.2	0%*
44	C		-	-	-	5.2	0%*
45	C		-	-	-	6.1	18%*

46	C		-	-	-	6.2	11%*
47	F	N	-	-	-	7.5	9%*
48	C		-	-	-	6.1	18%*
49	F	N	-	Me	-	7.5	12%*
50	F	N	-	Me	-	5.9	17%*
51	F	N	-	Me	-	5.6	14%*
52	F	.	N	-	-	6.4	19%*
53	C		-	-	-	4.7	13%*

2.2. Protein preparation

The crystal structures of compound **11** with GRK2 (PDB ID: **5HE0**), compound **17** with GRK2 (PDB ID: **5HE2**) and compound **47** with GRK2 (PDB ID: **5UKM**) of the same dataset compounds were retrieved from the protein databank (<https://www.rcsb.org/>). The crystal structure of ROCK1 (PDB ID: **6E9W**) reported by Hobson et al. was used for docking study of the most active compound (compound **11**) with the binding site of ROCK1 [63]. The alignment of the amino acid sequences in the kinase domains of GRK2 and ROCK1 are shown in Figure 1. The missing residues in protein structures were modeled using the homology modeling program MODELLER v9.21 [64-66]. The final model after refinement was selected based on statistical potentials (GA341) score and Discrete Optimized Protein Energy (DOPE) score [67].

Protein	Residue number	Identities: 89/271 (33 %) Positives: 143/271 (52%)		Residue number																																													
		Sequences																																															
GRK2	191	ES	HR	IGRGG	EL	YGC	KAD	GL	MY	AM	CL	D	KR	KM	QGETL	ALN	E	M	L	S	V	T	G	D	250																								
ROCK1	76	YE	VK	IQE	A	EL	QLV	HKS	R	Y	AM	L	S	F	E	M	I	K	S	D	S	A	F	F	W	E	E	D	I	M	A	F	A	S	-	133													
GRK2	251	C	F	E	C	M	S	A	H	T	P	D	K	S	F	T	D	L	A	N	G	G	D	H	Y	H	Q	H	G	F	S	A	D	M	E	V	A	A	F	I	G	E	H	M	I	N	310		
ROCK1	134	-	W	V	Q	F	A	Q	D	D	R	Y	M	V	M	E	V	P	G	D	V	N	L	M	N	D	E	P	K	W	A	R	E	V	T	A	V	A	D	A	I	S	191						
GRK2	311	R	F	V	V	H	E	K	A	E	E	H	H	V	R	T	S	L	L	A	C	D	F	S	K	R	-	-	K	P	H	A	R	K	G	H	G	M	A	P	E	V	K	G	V	A	367		
ROCK1	192	M	G	F	H	R	V	D	M	L	D	R	S	G	T	L	A	F	F	T	C	M	K	M	K	G	M	V	C	D	T	A	S	C	P	D	I	S	P	E	N	T	R	S	Q	G	G	251	
GRK2	368	-	-	-	D	S	S	A	E	F	E	C	M	E	K	L	R	H	S	R	Q	H	K	T	K	D	K	E	I	D	R	T	L	T	M	A	V	E	L	-	-	S	F	P	L	E	422		
ROCK1	252	D	G	Y	I	G	R	E	C	D	M	W	V	V	F	Y	E	M	V	D	E	Y	A	D	S	L	V	G	T	S	-	K	I	N	H	K	N	E	L	T	F	D	N	D	I	K	A	R	310
GRK2	423	S	L	E	G	L	Q	R	V	N	R	E	C	L	G	R	A	C	V	E	S	P	E	453																									
ROCK1	311	N	E	C	A	F	T	R	E	V	E	-	R	N	V	E	I	R	H	L	338																												

Figure 1. The alignment of the amino acid sequences in the kinase domains of GRK2 and ROCK1. Identical and positive matches in the sequences are highlighted by red and green colours respectively.

2.3. Molecular Docking

Docking study of the most active compound (compound **11**) into the binding site of ROCK1 was done using Autodock 4.2.5.1 [35]. The crystal structure of ROCK1 in complex with a pyridinylbenzamide based inhibitor (PDB ID **6E9W**) was already reported in a previous study [63]. The docking protocol was validated by redocking the co-crystallized ligand. The ligand structure was sketched and minimized with the Tripos force field in Sybyl-X 2.1 outside the receptor and then docked to the apo-receptor to perform the redocking. The docked pose showed a root-mean-square deviation (RMSD) value of 1.07 Å.

The docking of the most active compound to ROCK1 was preceded by the preparation of the ligand and the protein. During the protein structure preparation, polar hydrogen atoms were added to the protein. Gasteiger charges were added as partial charges. A grid box of size of 70x70x70 was created around the ligand to define the area of the receptor to be searched during the docking process. Lamarckian Genetic Algorithm (LGA) was selected to perform the docking. Finally, the docking process was executed to generate 100 docking conformation with 2500000 evaluations per run. The docking results were analyzed using AutoDockTools. Based on its binding energy and important interactions reported

in earlier studies, a docked pose was selected and used as input for molecular dynamics simulation studies.

This docking protocol was also used for docking study of compound **17** and compound **47** with ROCK1.

2.4. Molecular Dynamics (MD) simulation

MD simulations were carried out in Gromacs 2018 [68-72]. The protein topology and structure files were prepared using Amber99SB force field [73]. The ligand topology files were generated with ACYPE package using general AMBER force field (GAFF) [74, 75]. The three-point water model (TIP3 water) was used as the solvent. A dodecahedron box was built around the protein-ligand complex and the system was solvated. Sodium ions (Na^+) were added to the protein-ligand system to neutralize the charge of the system. The system was energy minimized using steepest descent algorithm with the maximum force (F_{Max}) set to 1000 KJ/Mol. The system was subjected to constant Number of particles, Volume, and Temperature (NVT) ensemble equilibration for 100 ps to equilibrate the solvent and ions around the protein at 300 K. The temperature coupling was done using modified Berendsen thermostat [76]. Constant number of particle, pressure, and temperature (NPT) ensemble equilibration was performed for 100 ps to stabilize the pressure. During NPT equilibration, Parrinello-Rahman barostat was used for pressure coupling [77]. LINCS algorithm was used to keep the bonds constrained [78]. During NVT and NPT equilibration, the positions of the protein and the ligand were kept restrained. Production MD simulations were carried out for 40 ns without restraints.

2.5. Free energy calculation

Molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) free energy calculation was performed using the `g_mmpbsa` package [79, 80]. The last 5 ns from the production run of the 40 ns MD simulation were used to calculate

binding energy. Snapshots were extracted every 50 ps. The binding energy consists of three energetic terms (potential energy in vacuum, polar-solvation energy, and non-polar solvation energy) [81, 82]. The vacuum potential energy includes both bonded (angle, bond, and dihedral) and non-bonded (electrostatics and van der Waals) interactions and was calculated based on molecular mechanics force field parameters [73, 83]. Polar solvation energy was calculated by solving the Poisson-Boltzmann equation [80, 84, 85] and non-polar solvation energy was calculated based on the solvent accessible surface area (SASA) model [86, 87]. The binding energy contributed by individual residue was calculated based on the equation given below:

$$\Delta R_x^{BE} = \sum_{i=0}^n (A_i^{bound} - A_i^{free})$$

Where, ΔR_x^{BE} represents the binding energy of the residue x , and A_i^{bound} and A_i^{free} are the energy of i^{th} atom from x residue in bound and unbound forms respectively.

2.6. 3D-QSAR

The comparative molecular field analysis (CoMFA) models were developed for both GRK2 and ROCK1 using Sybyl-X 2.1[32]. In CoMFA model development, the electrostatic field and steric field exerted by the compounds were calculated at each point of a regularly spaced 3D grid around the compounds. A probe atom (sp^3 carbon of +1 charge and having a van der Waal radius of 1.52 Å) was used to calculate the field exerted. The steric fields were contributed by Lennard-Jones potential and the electrostatic fields were contributed by Coulombic potential.

During the CoMFA model development for GRK2, the binding pose of the most active compound (compound **47**) given in the co-crystal structure (PDB ID **5UKM**) was used for aligning the dataset compounds. Since the co-crystalized

structure of ROCK1 with its most active compound (compound **11**) was not available, the average structure of the most active compound extracted from the last 5 ns of the 40 ns MD simulation was used as a template for developing the CoMFA model for ROCK1.

The dataset compounds were aligned by superimposing on the substructure which was common to all compounds using the ‘database align’ method given in Sybyl-X 2.1.

The alignments used for developing the CoMFA models for GRK2 and ROCK1 are shown in Figure 2. Partial least square (PLS) analysis was performed to linearly correlate the 3D-QSAR descriptor values to the activity values. The leave-one-out method was used to derive the cross-validated correlation coefficient (q^2) and optimal number of components (ONC) of the model. The non-cross-validated correlation coefficient (r^2), standard error of estimation and F-test value (F) were evaluated for the CoMFA model based on the ONC value[88].

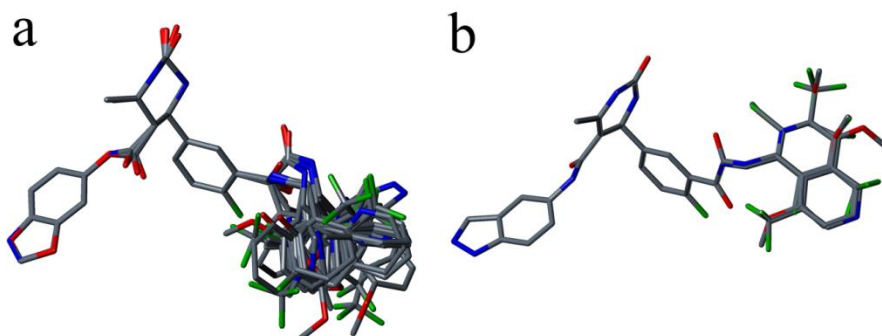


Figure 2. (a) Alignment of the dataset compounds used in the CoMFA model development for GRK2. (b) Alignment of the dataset compounds used in the CoMFA model development for ROCK1.

2.7 Model validation

The CoMFA models were validated for its robustness and statistical confidence using bootstrapping (BS) analysis. Leave-five-out (LFO) analysis was performed to assess the sensitivity of the models to chance correlation [37]. To test the predictive ability of the models against external test set, predictive correlation coefficient (r^2_{pred}) was calculated based on the equation given below [40]:

$$r^2_{pred} = (SD - PRESS)/SD$$

where SD represents the squared deviation between the activity value of the test set compounds and the mean activity value of the training set compounds. PRESS represents the sum of square deviation between the actual activity and the predicted activity of each compound in the test set.

3. RESULTS

3.1. Molecular Docking

The x-ray crystal structure of ROCK1 (PDB ID **6E9W**) in complex with a pyridinylbenzamide derivative reported by Hobson et al. [63] was used for the docking study of compound **11**, **17** and **47**. The docking protocol was validated by redocking the co-crystal ligand into the apo-receptor of ROCK1. The redocked ligand pose showed a root-mean-square deviation (RMSD) value of 1.07 Å.

Docking of the most active compound for ROCK1 (compound **11**) resulted in 100 conformations. The docking results were analyzed and a pose was selected based on low binding energy and H-bond interactions. The binding site of ROCK1 consisted of residues Gly85, Ala86, Phe87, Lys105, Leu106, Met156, Tyr155, Glu154, Ala215, Asp216, Glu124, Phe120, Phe217, and Leu107. Analysis of the non-bonded interactions showed that the compound **11** formed H-bond interactions with the Glu154 and Met156 at the hinge region, Asn203, and Asp216 at the ribose subsite and Lys105 at the phosphate binding site of ROCK1.

The interactions between compound **11** and the binding site residues of ROCK1 are shown in Figure 3

Docking studies of compound **17** (most selective compound) and compound **47** (most active compound for GRK2) were also performed to understand the binding modes of the inhibitors inside ROCK1. In the docking of compound **17** with ROCK1, the compound **17** formed H-bond interactions with Met156 at the hinge region and also with Arg84 and Phe87 at the P-loop. Analysis of the docking results for compound **47** with ROCK1 showed that the benzodioxole, piperidine, and pyrazole of compound **47** formed H-bond interactions with the ROCK1 residues Met156, Asp160, and Gly88 respectively.

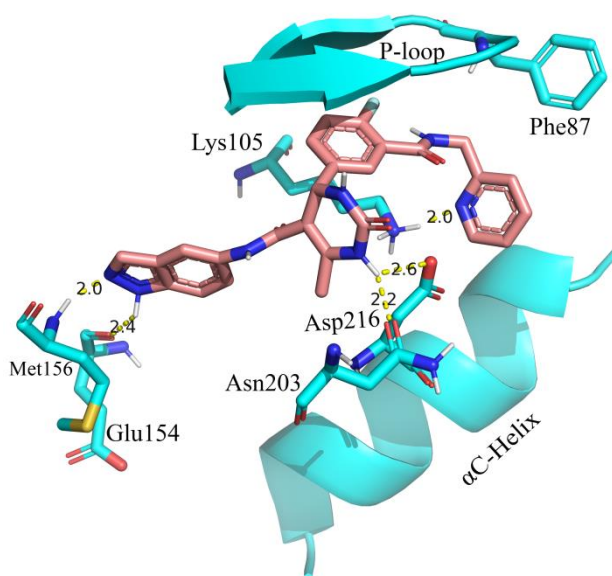


Figure 3. The docked conformation of the most active compound for ROCK1 (compound **11**) inside the active site of ROCK1. H-bond interactions were represented as yellow dotted lines.

From the docking studies, it was observed that compound **11**, **17** and **47** formed H-bond interaction with Met156 at the hinge region of ROCK1. This interaction with the hinge region was considered to be important as it anchors the inhibitor

inside the receptor and also induces significant conformational changes in the kinase domain [89]. The docked structures of compound **11**, **17** and **47** inside ROCK1 were used for molecular dynamics simulations studies to understand the dynamic interactions between the inhibitors and ROCK1.

3.2. Molecular Dynamics (MD) Simulation

During the MD simulation studies of compound **11**, **17** and **47** with GRK2, the crystal structure having PDB ID **5HE0** (compound **11**-GRK2 complex), **5HE2** (compound **17**-GRK2 complex) and **5UKM** (compound **47**-GRK2 complex) were used as initial structures [53]. In the MD studies of compound **11**, **15** and **47** with ROCK1, the inhibitor-protein complex structures obtained from the docking studies were used as starting structures. The root-mean-square deviation (RMSD) values of the inhibitors and proteins for the 40 ns MD simulations are shown in Figure 4. The snapshots of the inhibitor-protein complexes after 40 ns MD simulations were extracted and analyzed to understand the non-bonded interactions between the inhibitors and the receptors. Analysis of the H-bond interactions and hydrophobic interactions were shown in Figure 5 and 6 respectively.

The compound **11** showed H-bond interactions with the GRK2 binding site residues Met274, Asp272, Asn322, Lys 319 and Lys220 as shown in Figure 5a. The interactions with Met274, Asp272, and Asn322 were observed in the crystal structure of compound **11** with GRK2 (PDB ID **5HE0**). Compound **11** also formed hydrophobic interactions with the GRK2 binding site residues Ile197, Gly200, Gly201, Gly203, Val205, Ala218, Asn322 and Leu324 as shown in Figure 6a. In the MD study of compound **11** with ROCK1, H-bond interactions between compound **11** and the binding site residues Glu154, Met156, Ala215, Asp202, and Asp216 were observed. Hydrophobic interactions were also observed between compound **11** and the binding site residues Gly83, Gly85, Gly88, Val90, Phe120, and Leu205. The H-bond interactions and hydrophobic

interactions between compound **11** and the binding site residues of ROCK1 are shown in Figure **5d** and **6d** respectively. From the analysis, it was observed that compound **11** formed stable H-bond interactions and hydrophobic interactions with both GRK2 and ROCK1, which could be a possible reason behind the high activity value of the compound for both GRK2 ($pIC_{50}=6.8$) and ROCK1 ($pIC_{50}=7.9$).

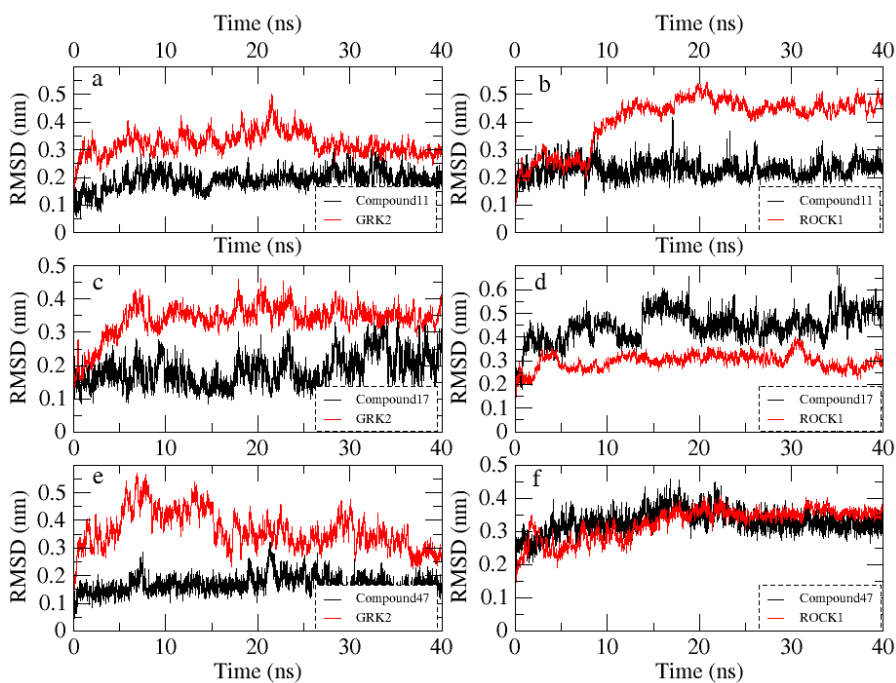


Figure 4. The RMSD diagrams for the 40 ns MD simulation runs. MD production run for each protein-ligand complex was performed once only. (a) Compound **11** with GRK2. (b) Compound **11** with ROCK1. (c) Compound **17** with GRK2. (d) Compound **17** with ROCK1. (e) Compound **47** with GRK2. (f) Compound **47** with ROCK1.

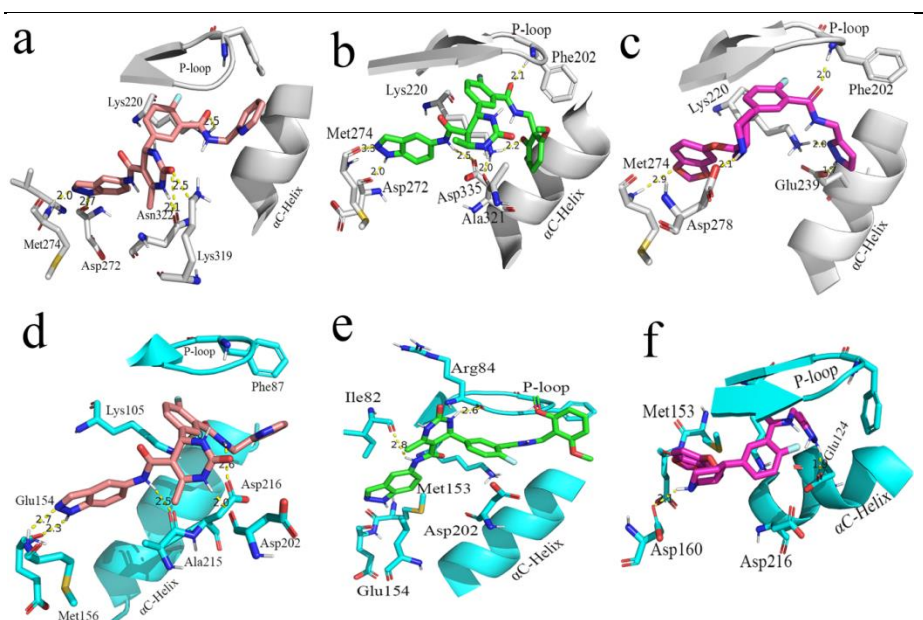


Figure 5. H-bond interactions between the compound **11** (salmon), **17** (green) and **47** (magenta) with GRK2 and ROCK1. Snapshots were collected after 40 ns simulations. The GRK2 and ROCK1 residues are shown in grey and cyan colours respectively. H-bond interactions were represented as yellow dotted lines. **(a)** Compound **11** with GRK2. **(b)** Compound **17** with GRK2. **(c)** Compound **47** with GRK2. **(d)** Compound **11** with ROCK1. **(e)** Compound **17** with ROCK1. **(f)** Compound **47** with ROCK1.

In the MD study of the most selective compound (compound **17**) with GRK2, the compound **17** formed H-bond interactions with the GRK2 binding site residues Met274, Asp272, Asp335, Ala321, Lys220, and Phe202. All these H-bond interactions, except the interaction with Lys220, were also observed in the crystal structure of compound **17** with GRK2 (PDB ID **5HE2**). Compound **17** also formed hydrophobic interactions with residues Arg199, Phe202, Gly203, Val205, Leu235, Leu273 and Leu324 of the GRK2 binding site. The H-bond interactions

and hydrophobic interactions of compound **17** with GRK2 are shown in Figure **5b** and Figure **6b** respectively. In ROCK1, compound **17** formed H-bond interactions with Ile82 and Arg84. The loss of the crucial H-bond interactions with Met156 (hinge region) and Phe87 (P-loop) in ROCK1 indicated that compound **17** was unable to form stable interactions with ROCK1. The H-bond interactions and hydrophobic interactions of compound **17** with ROCK1 are shown in Figure **5e** and Figure **6e** respectively. It was also observed that the dimethoxybenzene ring of compound **17** extended away from the binding site and was unable to form hydrophobic interactions with residues at hydrophobic subsite of ROCK1. Hydrophobic interactions were observed between compound **17** and residues from the adenine subsite and P-loop of ROCK1 such as Ile82, Gly83, Arg84, Gly85, Ala86, and Leu205.

From the analysis of the MD results for compound **47** with GRK2, it was observed that compound **47** formed H-bond interactions with Met274, Asp278, Glu239, Lys220, and Phe202. In addition to the interactions observed in the co-crystal structure of compound **47** with GRK2 (PDB ID **5UKM**), new interactions with the Met274 (hinge region) and Lys220 (phosphate subsite) were observed. Hydrophobic interactions were also observed between compound **47** and GRK2 binding site residues Ile197, Val205, Gly203, Val205, Met274, Leu324, Asp335, and Gly337. The H-bond interactions and hydrophobic interactions are shown in Figure **5c** and Figure **6c** respectively. In ROCK1, the compound **47** formed H-bond interactions with residues Met153 and Glu124 however, failed to form interactions with Met156 (Hinge region) and Gly88 (P-loop) of ROCK1. The inability of the compound **47** to form interactions at the hinge region and the P-loop could be the reason why the inhibitor was unable to form stable binding with ROCK1 and extended out of the binding pocket. Compound **47** also formed hydrophobic interactions with residues Gly85, Gly88, Glu89, Val90, Leu106, and Leu107 at the binding site of ROCK1. The H-bond interactions and hydrophobic interactions of compound **47** with ROCK1 are given in Figure **5f** and Figure **6f**.

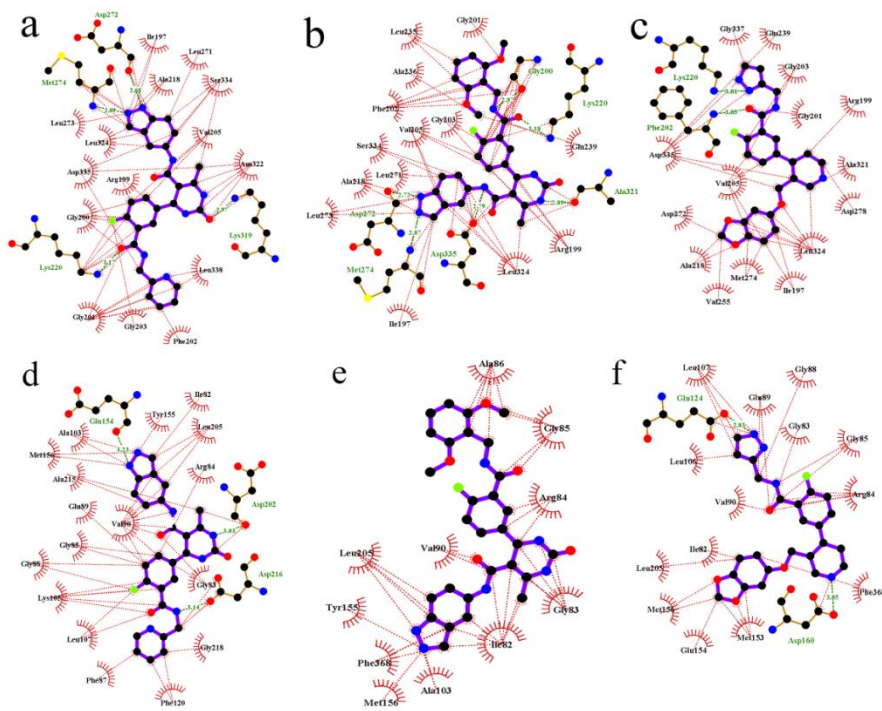


Figure 6. Hydrophobic interactions between compound **11**, **17**, and **47** with GRK2 and ROCK1. Snapshots were collected after 40 ns simulations. Hydrophobic interactions are represented as red dotted line. H-bond interactions are represented as green dotted lines. (a) Compound **11** with GRK2. (b) Compound **17** with ROCK1. (c) Compound **47** with GRK2. (d) Compound **11** with ROCK1. (e) Compound **17** with ROCK1. (f) Compound **47** with ROCK1.

The analysis of the binding interactions from the MD studies showed that the most active compound for GRK2 (compound **47**) and the most selective compound (compound **17**) were able to adopt conformations that allow the pyrazole/pyridine rings to form interactions with the Lys220 at the phosphate binding site of GRK2, which was not observed in the interactions with ROCK1. H-bond interaction was also observed between the pyrazole of the compound **47** and Glu239 (α C-Helix) in GRK2 which was not observed in the other inhibitor-protein interactions. The ability of the compound **47** to form stable interactions

with Lys220 and Glu239 could be vital for stabilizing the dimethoxybenzene ring at the hydrophobic subsite of GRK2.

3.3. MM/PBSA based free energy calculations

The binding energies of the inhibitor-protein interactions were calculated from the last 5 ns of the 40 ns MD production runs. The results of the binding energy calculations are given in Table 3. The total binding energy for each inhibitor-protein complex was contributed by the following energy terms: van der Waals, electrostatic, polar solvation, and non-polar solvation. From the analyses, we observed that Van der Waals energy and electrostatic energy made the major contribution to the total binding energies. In the interaction of compound **11** with ROCK1, the van der Waals and electrostatic energy values were -225.88 kJ/mol and -120.99 kJ/mol respectively suggesting that van der Waals interactions (hydrophobic interactions) were the major forces in the binding of ROCK1 and its most active compound. In the interaction of compound **47** with GRK2, the contribution of van der Waals energy and electrostatic energy to the total binding energy were -254 kJ/mol and -242.27 kJ/mol respectively suggesting that compound **47** can form favorable van der Waals interactions (hydrophobic interactions) and electrostatic interactions (H-bond interactions) with the binding site residues of GRK2.

Table 3. The energy contribution of the various energetic terms (van der Waals energy, electrostatic energy, polar solvation energy, and non-polar solvation energy/SASA) to the total binding energy.

Complexes	Van der Waals (kJ/Mol)	Electrostatics (kJ/Mol)	Polar solvation (kJ/Mol)	SASA (kJ/Mol)	Total Binding Energy (kJ/Mol)
Compound 11-GRK2	-249.62	-85.93	246.63	-24.66	-113.58
Compound 11-ROCK1	-255.88	-120.99	307.92	-24.08	-63.03

Compound 17-GRK2	-262.38	-112.65	275.98	-25.58	- 124.63
Compound 17-ROCK1	-189.84	-76.00	160.00	-20.40	- 126.25
Compound 47-GRK2	-254.96	-242.27	293.74	-23.97	- 227.48
Compound 47-ROCK1	-225.91	-73.84	190.62	-21.19	- 130.34

The energy contributions of the residues to the total binding energies were calculated for each inhibitor-protein complex to understand the residues that made significant contributions in the inhibitor-protein interactions. The binding energy values for the residues at the binding site of GRK2 that made significant contributions to the total binding energy are shown in Figure 7a and the energy values for the corresponding residues in ROCK1 are shown in Figure 7b. In binding of compound **11**, **17** and **47** with GRK2, the residues Gly200, Gly201, Phe202, Val 204 and Lys205 from the P-loop, the residues Leu222 and Glu235 from the phosphate subsite and the residues Leu271, Asp272, Leu273, met274, Asn275 and Asp278 from the adenine subsite made vital contributions to the total binding energies. Whereas in ROCK1, the residues Glu89 and Val90 from the P-loop, Met153, Glu154, Tyr155 and Met156 from the adenine subsite and residues Asp160, Asp 202, Leu205 and Asp216 from the ribose subsite made vital contributions to the total binding energies.

3.4. 3D-QSAR

The CoMFA models for GRK2 and ROCK1 were developed using Sybyl-X 2.1. During the development of the CoMFA model for GRK2, the structure of the most active compound for GRK2 (compound **47**) given in the co-crystallized structure (PDB ID **5UKM**) was used as the template for aligning the dataset compounds. The model was built based on a training set of 30 compounds and the

remaining 23 compounds were used for testing the model. The aligned compounds are shown in Figure 2a. The CoMFA model showed a cross-validated correlation coefficient (q^2) value of 0.67 and non-cross-validated correlation coefficient (r^2) value of 0.92. The statistical results of the CoMFA model are shown in Table 4. The derived CoMFA model showed an LFO value of 0.54 and also showed reasonable BS- r^2 and BS-SD value of 0.96 and 0.03 respectively. During the external validation, the CoMFA model exhibited acceptable predictive ability showing an r^2_{pred} value of 0.61.

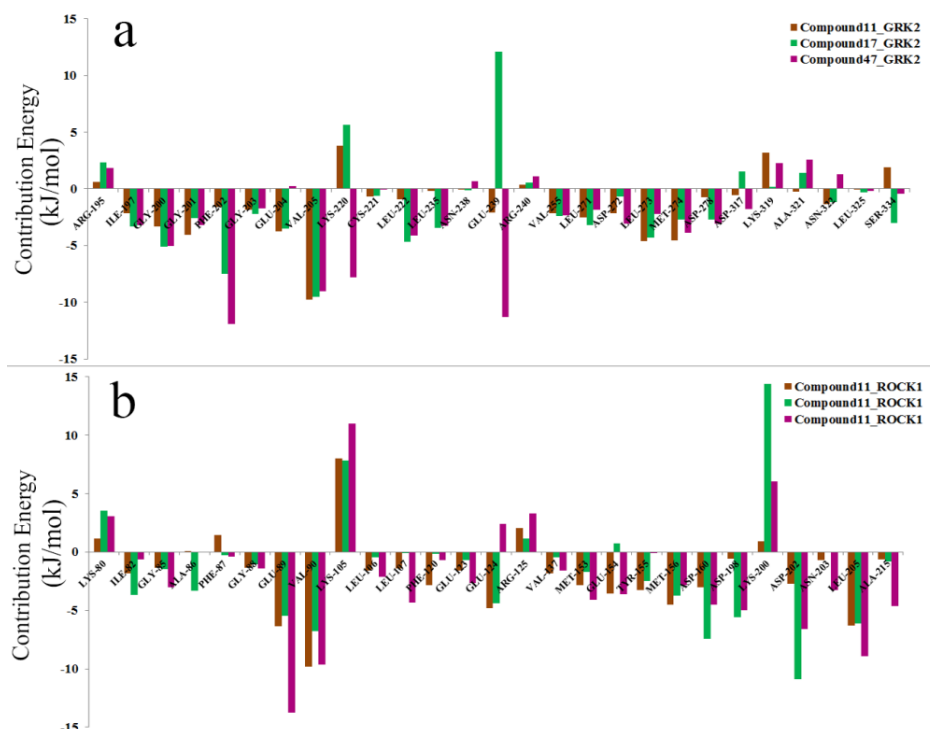


Figure 7. The energy contributions (in kJ/mol) of the key residues to the total binding energy; (a) Interaction of compound 11 (brown), 17 (green) and 47 (magenta) with GRK2. (b) Interaction of compound 11 (brown), 17 (green) and 47 (magenta) with ROCK1.

During the development of the CoMFA model for ROCK1, the average structure of the most active compound for ROCK1 (compound 11) extracted from the last 5

ns of the 40 ns MD simulation (compound **11**-ROCK1 complex) was used as the template for aligning the dataset compounds. The CoMFA model was build using 21 compounds. The aligned compounds for the CoMFA model for ROCK1 are shown in Figure **2b**. The derived model showed a q^2 value of 0.59 and r^2 value of 0.94. During the model validation, the CoMFA model showed an LFO value of 0.62 and showed reasonable BS- r^2 and BS-SD value of 0.98 and 0.01 respectively. The statistical results are shown in Table **4**.

Table 4. Statistical results of the CoMFA models for GRK2 and ROCK1.

Parameters	CoMFA (GRK2)	CoMFA (ROCK1)
q^2	0.67	0.59
ONC	5	6
SEP	0.50	0.57
r^2	0.92	0.94
SEE	0.22	0.2
F value	52.46	35.93
LFO	0.54	0.62
BS r^2	0.96	0.98
BS SD	0.03	0.01
r^2_{pred}	0.61	NA
Influence of different fields (%)		
S	0.49	0.73
E	0.50	0.27

q^2 : cross-validated correlation coefficient; ONC: Optimal number of components; SEP: Standard Error of Prediction; r^2 : non-cross-validated correlation coefficient;

SEE: Standard Error of Estimation; F value: F-test value; r^2 ; LFO: Leave five out; BS- r^2 : Bootstrapping r^2 mean; BS-SD: Bootstrapping Standard deviation; r^2_{pred} : predictive correlation coefficient; S: Steric; E: Electrostatic; ND: Not Determined. The statistical results from the CoMFA models for GRK2 and ROCK1 suggested that the models have acceptable robustness and predictive ability. The comparison of the actual activity values and the predicted activity values for the CoMFA models for GRK2 and ROCK1 are shown in Table 5 and 6.

3.5. Contour Map Analysis

The electrostatic and steric contour maps developed from the CoMFA models for GRK2 and ROCK1 are shown in Figure 8. In the electrostatics contour maps, the regions favorable to electropositive substituents were shown in blue color contours and the electronegative substituents favorable regions were shown in red color contours. In the steric contour maps, the bulky substituents and non-bulky substituents favorable regions were represented in green and yellow contours respectively.

In the CoMFA contour maps for GRK2, compound 47 (most active compound for GRK2) was used as a reference (Figure 8a and 8b). The blue contours observed near the benzodioxole, the piperidine ring and near the linker between the benzodioxole and the piperidine ring suggested that electropositive substituents at these positions are favored. Red color contours were observed near the pyrazole ring, the benzodioxole and near the linker between the benzene ring and the pyrazole ring suggesting that electronegative substituents are favored in these regions. The electronegative and electropositive substituents near the benzodioxole can lead to H-bond interactions with the GRK2 hinge region residues such as Met274 and Asp272 as observed in compound 11, 17 and 47. Electronegative substituents at the pyrazole ring can lead to H-bond interaction with Lys220 as observed in compound 17 and 47. In the steric contour map (Figure 8b), green contours were observed near the benzodioxole and the

pyrazole ring suggesting that bulky substituents are favored in these regions. As the benzodioxole and the pyrazole ring occupied the adenine subsite and the hydrophobic subsite respectively, the presence of bulky substituents may result in favorable non-bonded interactions with residues surrounding the hydrophobic subsite. Yellow contour at the back of the pyrazole ring suggested that extended bulky substituents are not favorable in this region. Extended bulky substituents in this region can cause steric clash with binding site residues. This is exemplified by compound **22**, **30**, **40**, **41** and **42** all of which have relatively lower activity value for GRK2 in the series.

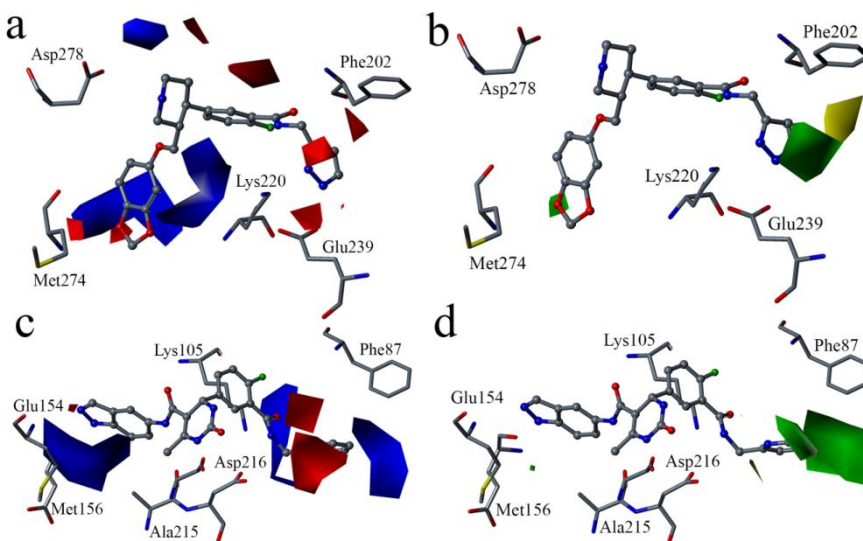


Figure 8. Standard coefficient contour maps obtained from GRK2 and ROCK1 CoMFA analyses. In the electrostatic contour maps, Blue contour indicates electropositive substituent favorable regions and red contour indicates electronegative substituent favorable regions. In the steric contour maps, green contour indicates steric bulk favorable regions and yellow contour indicates steric bulk unfavorable regions. (a) Electrostatic contour map for GRK2 CoMFA model with the template compound (compound **47**) as reference. (b) Steric contour map for GRK2 CoMFA model with the template compound (compound **47**) as reference. (c) Electrostatic contour map for ROCK1 CoMFA model with the template compound (compound **11**) as reference. (d) Steric contour map for ROCK1 CoMFA model with the template compound (compound **11**) as reference.

In the CoMFA contour maps for ROCK1, the compound **11** (most active compound for ROCK1) was used as a reference (Figure **8c** and **8d**). The blue contours near the indazole and near the linker between the benzene ring and the pyridine ring in the electrostatic contour map (Figure **8c**) suggested that electropositive substituents at these positions are favorable and may increase the activity of the compounds for ROCK1. Electropositive substituents at the linker between the benzene ring and the pyridine can lead to H-bond interactions with Ala215 as observed in the interaction of compound **11** with ROCK1 (Figure **5d**). The red contour near the pyridine ring suggested that electronegative substituents are favored in that region. Having electronegative substituents at the pyridine ring can lead to favorable H-bond interactions with Lys105 as observed in the interaction compound **11** with ROCK1. Green contour was observed near the pyridine ring, suggesting that bulky substituents are favorable in that region (Figure **8d**).

From the analysis of the CoMFA contour maps for GRK2 and ROCK1, it was observed that electronegative substituents near the benzodioxole and near the piperidine ring and having bulky substituents near the piperidine ring increased the activity for both GRK2 and ROCK1. Whereas, having electropositive substituents at the piperidine ring, having electropositive and electronegative substituents at the benzodioxole and having electronegative substituents near the amide linker between the benzene ring and the pyrazole ring are favorable for increasing the activity for GRK2 with selectivity over ROCK1.

4. DISCUSSIONS

From the analysis of the inhibitor-protein interactions from the MD simulation results, it was observed that compound **11**, **17** and **47** were able to form stable H-bond interactions with residues from the hinge region (Met274 and Asp272) and the ribose subsite (Asp278, Lys319, Asn322, and Asp335) of GRK2 which anchored the inhibitors at the binding site (Figure **5a**, **5b**, and **5c**). Compound **11**

also formed H-bond interactions with residues from the hinge region (Glu154) and the ribose subsite (Asp216 and Asp202) of ROCK1 (Figure 5d). However, compound **17** and **47** could not form stable interactions with residues at the hinge region and at the ribose subsite of ROCK1 (Figure 5e and 5f), which could be a reason behind the poor activity of these compounds towards ROCK1.

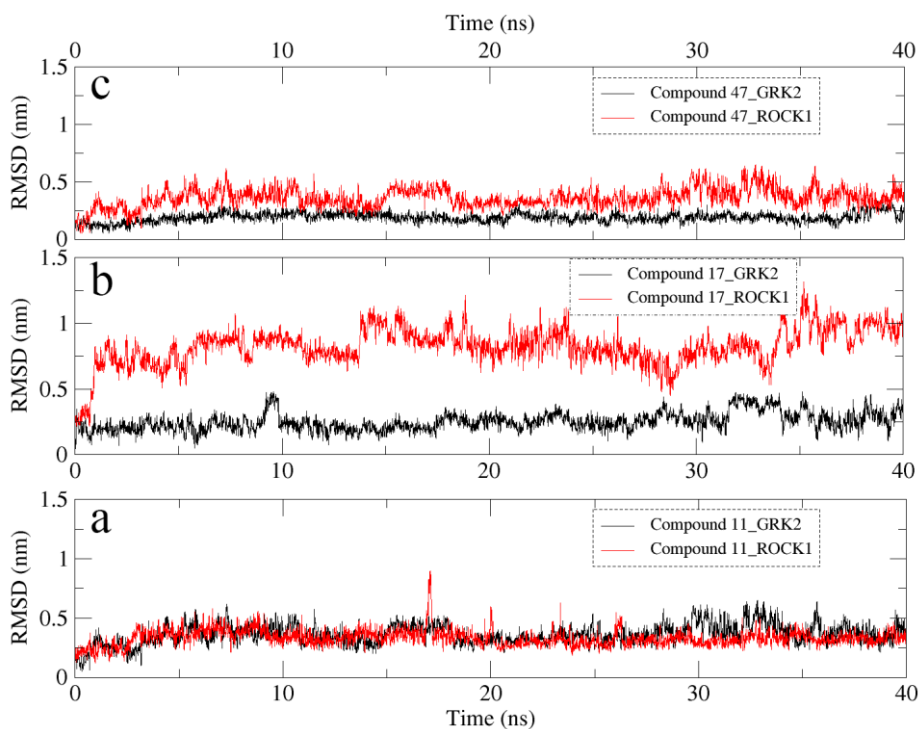


Figure 9. RMSD values of pyridine ring (compound **11**), dimethoxybenzene ring (compound **17**) and pyrazole ring (compound **47**) inside GRK2 and ROCK1 for 40 ns MD simulations. MD production run for each protein-ligand complex was performed once only. (a) RMSD of pyridine ring of compound **11** at the hydrophobic pocket of GRK2 (black) and ROCK1 (red). (b) RMSD of dimethoxybenzene ring of compound **17** at the hydrophobic pocket of GRK2 (black) and ROCK1 (red). (c) RMSD of pyrazole ring of compound **47** at the hydrophobic pocket of GRK2 (black) and ROCK1 (red).

In both compound **17** and compound **47**, the oxygen at the amide linker between the benzene ring and the dimethoxybenzene ring/pyrazole ring formed H-bond

interactions with the nitrogen of Phe202 (P-loop) and extended the dimethoxybenzene/pyrazole ring into the hydrophobic pocket of GRK2 (Figure **5b** and **5c**). This allowed the dimethoxybenzene ring/pyrazole rings to form H-bond interactions with Lys220 (phosphate subsite) of GRK2.

These H-bond interactions with the P-loop (Phe87) and the phosphate subsite (Lys105) were not observed in ROCK1. These interactions with the Phe202 and Lys220 could be crucial for the binding stability of compound **17** and compound **47** inside GRK2. To investigate the influence of the interactions with the Phe202 and Lys220 on the stability of the dimethoxybenzene/pyrazole ring inside the hydrophobic pockets, we calculated the RMSD of the pyridine ring (compound **11**), dimethoxybenzene ring (compound **17**) and pyrazole ring (compound **47**) inside the hydrophobic subsites of GRK2 and ROCK1. MD production run for each protein-ligand complex was performed once only. The outcomes of the RMSD calculations for each protein-ligand complex are shown in Figure **9**. The pyridine ring of Compound **11** showed an average RMSD value of 0.36 Å and 0.33 Å inside GRK2 and ROCK1 respectively (Figure **9a**). Though compound **11** formed multiple H-bond interactions with the binding site residues of GRK2 and ROCK1, it did not form interactions with Lys220 (GRK2) or Lys105 (ROCK1) which could be the reason behind the increased in the average RMSD values of 0.36 Å and 0.33 Å for GRK2 and ROCK1 respectively. The pyrazole of compound **47** showed an average RMSD value of 0.19 Å and 0.37 Å in GRK2 and ROCK1 respectively (Figure **9c**). The low average RMSD value of the pyrazole ring inside GRK2 suggested that the pyrazole of compound **47** was stably locked at the hydrophobic subsite. This stability may be attributed to the H-bond interactions with Lys220 and Glu239 at the hydrophobic subsite of GRK2. The biggest difference in the RMSD value was observed for the dimethoxybenzene ring of compound **17** which showed an average RMSD value of 0.25 Å and 0.85 Å inside GRK2 and ROCK1 respectively (Figure **9b**). The difference in the average RMSD values may be attributed to the fact that

compound **17** was able to form interactions with Phe202 and Lys220 in GRK2 which stabilized the dimethoxybenzene ring inside the hydrophobic subsite, however, in ROCK1, the dimethoxybenzene ring extended out of the hydrophobic subsite and did not form interactions at the hydrophobic subsite.

The H-bond interactions with Phe202 and Lys220 could be the reason behind the stability of compound **17** and **47** in GRK2, resulting in higher activity of the compounds for GRK2. These observations suggested that the H-bond formation with the Phe202 and Lys220 may be crucial for the stability of the inhibitors at the hydrophobic pocket of GRK2 and could potentially lead to selective inhibition of GRK2 over ROCK1.

Table 5. Experimental and predicted pIC_{50} values with their residuals of CoMFA for GRK2

Compound	Actual pIC_{50}	GRK2 CoMFA	
		Predicted pIC_{50}	Residual
1 [*]	5.9	6.0	-0.1
2 [*]	6.1	5.8	0.3
3	4.7	4.9	-0.2
4	5.4	5.4	-0.1
5 [*]	6.2	6.8	-0.7
6	6.7	6.8	-0.1
7	6.7	6.6	0.1
8	7.3	7.2	0.0
9 [*]	6.4	6.1	0.3
10	6.4	6.6	-0.3

11	6.8	6.8	0.0
12	6.6	6.5	0.0
13	5.3	5.2	0.1
14*	6.6	6.2	0.4
15	6.9	7.1	-0.2
16*	7.2	7.5	-0.4
17*	6.9	7.1	-0.2
18	5.9	6.0	0.0
19	5.6	5.4	0.2
20	5.7	5.7	0.0
21*	6.6	6.7	-0.1
22	4.6	4.5	0.1
24*	6.3	5.6	0.8
25	6.2	6.4	-0.2
26	5.7	5.5	0.2
27*	6.1	6.1	0.0
28	5.6	5.6	-0.1
29	5.8	6.0	-0.2
30*	5.7	5.9	-0.2
31*	4.9	4.9	0.0
32	5.7	5.9	-0.3
33	5.7	5.8	0.0

34	5.7	5.7	0.0
35*	4.4	5.1	-0.7
36	5.6	5.6	0.0
37*	6.2	6.6	-0.4
38	5.8	5.9	-0.1
39	5.7	5.7	0.0
40*	5.5	5.9	-0.4
41*	5.5	5.7	-0.2
42	5.5	5.6	-0.1
43	5.2	5.2	0.1
44*	5.2	5.9	-0.7
45	6.1	6.1	0.0
46	6.2	6.2	0.0
47	7.5	7.0	0.6
48	6.1	6.7	-0.6
49	7.5	7.1	0.4
50*	5.9	5.7	0.2
51	5.7	5.4	0.2
52	6.4	6.4	0.0
53*	4.8	5.7	-0.9

* Test set compounds

Table 6. Experimental and predicted pIC_{50} values with their residuals of CoMFA for ROCK1.

Compound	Actual pIC_{50}	ROCK1 CoMFA	
		Predicted pIC_{50}	Residual
2	7.0	7.1	-0.1
3	6.7	6.6	0.1
4	6.3	6.3	0.0
5	7.2	7.1	0.0
6	7.7	7.4	0.3
7	7.0	6.6	0.3
8	7.2	7.3	-0.1
9	7.0	7.0	0.0
10	7.3	7.3	0.0
11	8.0	8.2	-0.3
12	7.6	7.6	0.0
13	7.1	7.1	0.0
14	7.9	7.9	0.0
15	5.2	5.7	-0.5
16	5.2	5.1	0.2
19	5.7	5.2	0.5
20	6.8	6.9	-0.1
21	6.5	6.5	-0.1
22	6.3	6.4	-0.1

23	6.4	6.4	-0.1
24	6.7	6.7	-0.1

5. CONCLUSIONS

In this study, we have used molecular docking, molecular dynamics simulation, free energy calculation and 3D-QSAR methods to study a series of 53 paroxetine-like inhibitors to understand the structural properties that drive the inhibitory preference for GRK2 over ROCK1. The observations from the MD studies suggested that H-bond interactions of the inhibitors with the residues at hinge regions and ribose subsites are crucial for anchoring the inhibitors at the binding site in GRK2 and ROCK1. It was also observed that H-bond interactions with Phe202 and Lys220 increased the stability of the inhibitors at the hydrophobic subsite of GRK2. Hence, H-bond interactions with Phe202 and Lys220 were considered to be vital for the selective inhibition of GRK2. Free energy calculations of the inhibitor-protein interactions suggested that van der Waals and electrostatic energies were the major contributors to the total binding energies in GRK2 and ROCK1. Residue-wise energy decompositions indicated that van der Waals interactions and electrostatic interactions with residues Phe202, Val205, Lys220, and Glu239 were important for the inhibition of GRK2 with selectivity over ROCK1. Analysis of the contour maps from the 3D-QSAR models suggested that having electropositive substituents at the piperidine ring, electronegative and electropositive substituents at the benzodioxole and electronegative substituent near the amide linker between the benzene ring and the pyrazole ring were favorable in GRK2 and may lead to increased inhibitor activity for GRK2 with selectivity over ROCK1.

Seketoulie Keretsu Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

The statistical results and scientific observations reported in this study contributed in understanding the structural properties required for the selective inhibition of GRK2 with selectivity over ROCK1. The outcome of this study could be useful in designing potent GRK2 inhibitors with selectivity over ROCK1 for therapeutic intervention of heart failure diseases.

PART IV

Molecular Modelling Study of c-KIT/PDGFR α Dual Inhibitors for the Treatment of Gastrointestinal Stromal Tumors

1. Introduction

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal tumors of the gastrointestinal tract that arise from interstitial cells of Cajal (ICC) or from stem cells that differentiate towards ICCs [90, 91]. GISTs are commonly originated in the stomach (70%), with rare cases of it observed in the small intestine (20%) or esophagus (10%) [92]. It has an incidence rate of 14.5 per million per year [93].

Stem cell factor receptor (c-KIT) [94] and platelet derived growth factor receptor alpha (PDGFR α) [95] kinases are members of Type 3 transmembrane receptor protein-tyrosine kinase (RPTK) family and play important roles in various cellular signaling processes. Structurally, the member of the RPTK subfamily consists of five extracellular immunoglobulin (Ig) domains, one transmembrane domain, one juxtamembrane helix, and one cytoplasmic kinase domain [96]. The c-KIT kinases are primarily expressed on the hematopoietic stem cell surface and binds to the stem cell factor at the extracellular Ig domain. The binding of the stem cell factor to c-KIT leads to the dimerization of the kinase domains and phosphorylation of specific tyrosine residues in the juxtamembrane regions, which in turn activates downstream signaling cascades that mediate cell survival, proliferation, and differentiation [97, 98]. The PDGFR α kinases are expressed on the surface of several cell types and bind to the platelet-derived growth factor. The binding of the platelet derived growth factor leads to kinase domain dimerization and activation. PDGFR α plays an important role in the regulation of embryonic development, cell proliferation, and cell survival [99, 100]. Experimental studies have shown that the tumorigenesis of GISTs is associated with the gain-of-function mutation in c-KIT and PDGFR α [101-103]. Rammohan et al. (2013) reported that positive c-KIT expression was observed in approximately 90% of GISTs cases [104]. Mutations that lead to constitutive

PDGFR α activation have been found in approximately 10% of GISTs cases [105]. Because of the role of the c-KIT and PDGFR α in the development and progression of GISTs, these kinases are considered to be promising therapeutic targets for the treatment of GISTs.

So far, three non-selective inhibitors namely, Imatinib, Sunitinib, and Regorafenib have been approved by the Food and Drug Administration (FDA) for the treatment of GISTs [106]. Imatinib is a non-selective inhibitor of c-KIT, PDGFR α , and ABL (Abelson) kinases and has been approved for use as a first-line treatment of GISTs. However, one-half of the responding GISTs patients gain Imatinib resistance within 2 years of treatment via mutation at the T670 residue of c-KIT [107]. The result is more optimistic in the case of PDGFR α with only 5%-7% of cases with PDGFR α mutation showing resistance to Imatinib [108]. Sunitinib is another non-selective inhibitor of c-KIT, PDGFR α , vascular endothelial growth factor receptor (VEGFR), and Fms like tyrosine kinase 3 (FLT3) that has been approved for second-line treatment of GISTs [109]. Regorafenib is also a non-selective multi-kinase inhibitor with activity against c-KIT, PDGFR α , VEGFR, rapidly accelerated fibrosarcoma 1 (RAF1), rearranged during transfection (RET), and fibroblast growth factor receptor (FGFR) [110, 111]. However, efficacy and safety study of Regorafenib in patients with advanced GISTs has shown that 15% of the patients experience an exacerbation of cancer-related symptoms [112]. In addition to the approved drugs, several other c-KIT/ PDGFR α inhibitors such as dovitinib [113], masitinib [114], crenolanib [115], and ripretinib [116] are also under investigation for the treatment of GISTs [117]. Though progress has been made in the treatment of GISTs, current therapeutic options have various drawbacks such as low efficacy, clinical resistance, and side effects due to the non-selective property of the existing drugs.

Computer-aided drug discovery (CADD) methods have become popular in the drug discovery process and have been widely used in several drug discovery

studies [29, 118]. Taking advantage of the CADD techniques, our research group has performed several computational studies particularly in the area of kinase inhibitors [119-122]. Given the role of both c-KIT and PDGFR α in tumorigenesis of GISTs, the computational study of the dual inhibitory mechanism of these kinases could provide valuable insight into developing more effective drugs against GISTs. In this spirit, we have performed the computational studies of a series of potent and selective pyrazolopyridine inhibitors to explore the structural features important for dual inhibition of c-KIT/ PDGFR α . Molecular docking and molecular dynamics (MD) simulation were performed to study the inhibitor-protein binding interactions. Comparative molecular field analysis (CoMFA) [32] and comparative molecular similarity indices analysis (CoMSIA) [123] models were developed and the contour maps were analyzed to explore the important structural features. Binding energy evaluation was carried out to predict the binding affinity of the compounds.

2. Methodology

2.1. Data Preparation

The dataset of 48 pyrazolopyridine derivatives and their inhibitory values against c-KIT and PDGFR α were collected for computational study [124]. The half-maximal inhibitory concentration (IC₅₀) values of the compounds were converted to its log (pIC₅₀) values. The compounds were sketched and minimized in Sybyl X 2.1. The compound **14** which showed the highest activity for both c-KIT (pIC₅₀=8.6) and PDGFR α (pIC₅₀=8.1) was selected as a representative compound for the dataset.

The X-ray crystal structure of Imatinib-c-KIT complex (PDB ID **1T46**) and Imatinib- PDGFR α complex (PDB ID **6JOL**) were collected from the protein databank (RCSB.ORG) [94]. The water molecules, ions, and other small

molecules were removed from the protein file. The missing residues of the proteins were modeled using the SWISS-MODEL [125].

2.2. Molecular Docking

The binding interactions of compound **14** with the c-KIT and PDGFR α were studied using Autodock 4.2 [35]. The receptor was prepared by removing the heteroatoms. This was followed by the addition of hydrogen atoms and the application of partial charges to all the atoms. The inhibitor was prepared by assigning partial charges and the number of rotatable bonds. A grid box of 70 \times 70 \times 70 was developed to define the search space inside the receptor using the Autogrid program. The Lamarckian genetic algorithm was selected to perform the docking of the ligand. To validate the docking protocol, the crystal ligand (Imatinib) was minimized outside the receptor and docked into the receptor. The docking results showed that the docked pose closely overlapped with the crystal ligand in both c-KIT and PDGFR α . The result of the docking was analyzed using the autodock tools. This docking protocol was used for all protein-ligand interactions in this study.

2.3. Molecular Dynamics Simulation

The protein-ligand dynamics simulation was performed with the Gromacs 2020 [68, 126, 127]. The protein-ligand complex from the molecular docking study was used as the initial structure for the MD simulation. The protein parameter files were generated with the CHARMM36 all-atom force field (2019) [128]. The ligand topology and parameter files were prepared using the CHARMM General Force Field (CGenFF) [129]. Initially, the system was set up containing the protein-ligand complex inside a dodecahedron box and solvated with TIP3 waters. The charge of the system was neutralized by adding Na⁺ and Cl⁻ counterions. Steepest descent energy minimization step was performed to remove

steric clashes and inappropriate geometries. This was followed by 100 ps isothermal-isochoric ensemble (NVT ensemble) equilibration and 100 ps isothermal-isobaric ensemble (or NPT ensemble) equilibration of the system to stabilize the water around the protein and ligand. The protein was kept restrained during the equilibrations. Temperature and pressure coupling was performed using Berendsen thermostat and Parrinello-Rahman barostat respectively. Long-range electrostatics were treated using the Particle-mesh Ewald method. The thermodynamic properties of the system were collected every 1 ps. The unrestrained MD production run was performed for 100 ns at the temperature and pressure of 300 K and 1 bar respectively.

2.4. Evaluation of Binding Energy

The binding energy between the protein and the ligand was calculated using the *g_mmpbsa* package[79]. Molecular mechanics energies combined with the Poisson–Boltzmann and surface area continuum solvation (MM/PBSA) methods have been successfully used to predict the relative binding energy values of congeneric compounds. In *g_mmpbsa*, the MM potential energy term is calculated based on the molecular mechanics force-field parameters as given in the equation below.

$$E_{MM} = E_{bonded} + E_{nonbonded} = E_{bonded} + (E_{vdw} + E_{elec})$$

Where, E_{bonded} is a bonded energy term consisting of bond, angle, dihedral, and improper interactions. The non-bonded energy term is made up of electrostatics and van der Waals energy terms and was calculated based on Coulomb and Lennard-Jones potential function respectively. The free energy of solvation was calculated based on an implicit solvent model where the electrostatic (G_{polar}) and non-electrostatic ($G_{nonpolar}$) energy terms were calculated as given below.

$$G_{MM} = G_{polar} + G_{nonpolar}$$

The electrostatic term and non-electrostatic terms were calculated based on the

Poisson-Boltzmann equation and solvent accessible surface area (SASA) model respectively.

Before calculating the binding energy, the trajectory was processed to correct the periodicity and center the protein within the unit cell. Water molecules and ions were removed from the trajectory file. Finally, the binding energy was calculated from the converged region of the MD trajectory at an interval of 0.5 ns (100 frames).

2.5. 3D-QSAR

To perform the 3D-QSAR studies, we have selected a binding conformation of compound **14** from the MD simulation trajectories. A confirmation that corresponded to the highest number of H-bond interactions with the receptor was selected. The structure of the compound was relaxed by quick minimization in Sybyl X 2.1. Using compound **14** as a template the other compounds were sketched and minimized. The alignment of the compounds was performed based on the common substructure of the compounds [32, 123]. The aligned compounds were randomly divided into a training set and a test set .

The training set compounds were used to develop various Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) models [9, 130]. In CoMFA, the electrostatic and van der Waal energy terms were calculated for each compound. In CoMSIA the hydrogen bond donor, hydrogen bond acceptor, hydrophobic, steric, and electrostatic energy terms were calculated for each compound [131]. The energy terms were calculated by probing the 3D-grid around the compound by using an sp³ hybridized carbon atom (charge +1). The partial least square (PLS) method was used to establish the relationship between the dependent and the independent variables. Leave-One-Out crossvalidated analysis was performed to determine the predictive ability of the 3D-QSAR models and to determine the optimal number of components

(ONC). Based on the crossvalidated q^2 and the ONC, the non-crossvalidated predictive r^2 value was calculated. Based on the statistical results, a model with high q^2 and r^2 values was selected for further analysis.

Bootstrapping (BS) was performed to estimate the confidence intervals of the parameters predicted by the 3D-QSAR models. A bootstrap sampling size of 100 was used during the validation. In addition, the predictive ability of the derived models against an external test set was also evaluated using the equation given below.

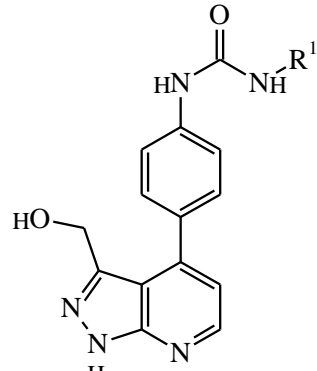
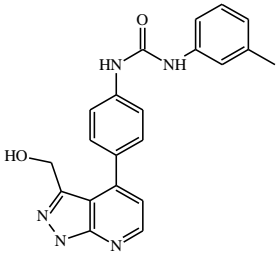
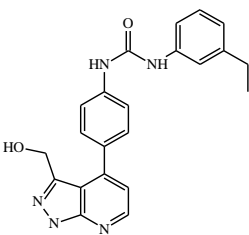
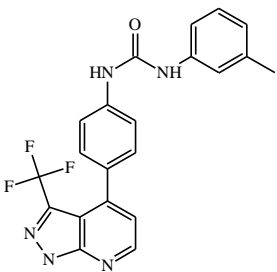
$$r_{pred}^2 = (SD - PRESS) / SD$$

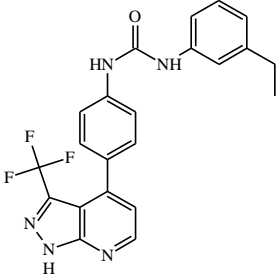
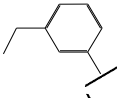
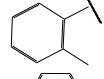

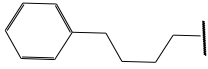
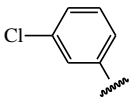
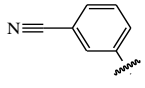
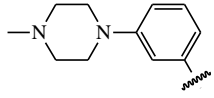
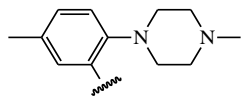
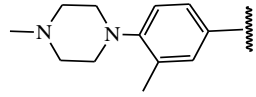
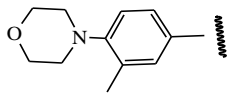
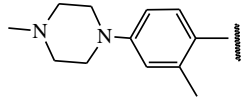
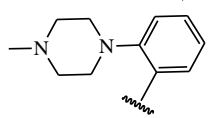
Where, SD represented the standard deviation between the activity value (pIC₅₀) of the test set compounds and the mean activity value of the training set compounds. PRESS represented the sum of the square deviation between the predicted and the actual activity value of each compound in the test set.

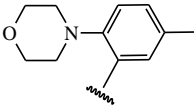
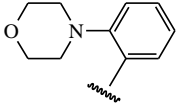
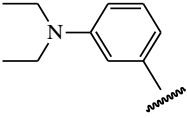
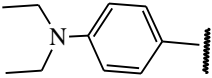
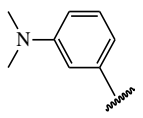
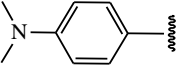
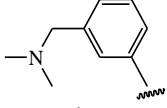
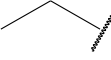
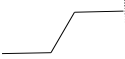
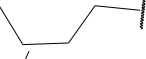
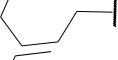

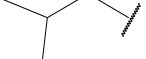

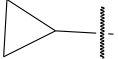
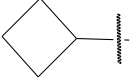
3. Results

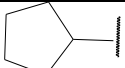
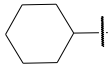
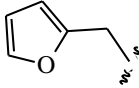
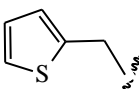
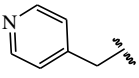
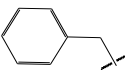
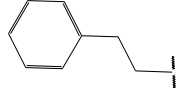
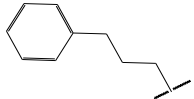
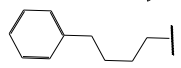
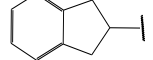
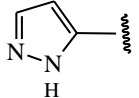
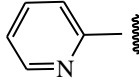
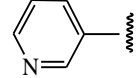
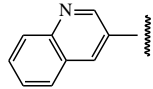
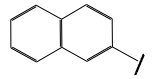
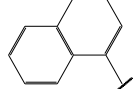
The X-ray crystal structures of Imatinib in complex with the inactive form of c-KIT (PDB ID **1T46**) and PDGFR α (PDB ID **6JOL**) were collected from the protein databank (www.rcsb.org). Imatinib showed an pIC₅₀ value of 7.4 and 8.3 for c-KIT and PDGFR α respectively and was used as a reference compound [124]. The dataset compound 14 which showed high pIC₅₀ values of 8.6 for c-KIT and 8.1 for PDGFR α was used as a representative compound for the dataset. The dataset compounds and their log activity values are shown in Table 1.

Table 1. Structure of the pyrazolopyridine derivatives and their pIC₅₀ values for c-KIT and PDGFR α .

Compounds	Structures	R1	c-KIT (pIC ₅₀)	PDGF R α (pIC ₅₀)
	 <p style="text-align: center;">Structure A</p>			
1		3-Methylphenyl	8.62	7.06
2		4-Ethylphenyl	8.43	7.49
3		3-Methylphenyl	>4.3	>4.3

4			5.2	>4.3
5	A		8.41	7.66
6	A		4.96	6.52
7	A		5.74	7.19
8	A		5.32	6.58
9	A		5.62	6.74
10	A		6.70	4.67
11	A		7.72	6.67
12	A		5.71	5.75
13	A		8.14	6.90
14	A		8.62	8.14
15	A		5.65	5.96
16	A		5.34	6.03

17	A		6.40	4.72
18	A		5.19	5.39
19	A		7.92	6.87
20	A		7.92	6.75
21	A		8.59	7.08
22	A		8.03	6.51
23	A		4.66	5.34
24	A		4.4	>4.3
25	A		4.7	>4.3
26	A		4.69	4.36
27	A		8.46	6.79
28	A		8.85	7.57
29	A		6.49	5.95
30	A		5.85	5.83
31	A		6.26	5.95
32	A		6.50	6.14

33	A		6.84	6.24
34	A		>4.3	>4.3
35	A		7.29	6.41
36	A		7.59	6.80
37	A		7.11	6.33
38	A		>4.3	>4.3
39	A		8.43	7.09
40	A		8.77	7.62
41	A		8.21	6.87
42	A		8.06	7.08
43	A		6.92	6.39
44	A		7.70	6.40
45	A		7.85	6.50
46	A		8.72	7.66
47	A		7.70	6.91
48	A		8.44	7.01

3.1 *Molecular Docking*

Molecular docking was performed using Autodock 4.2. Validation of the docking procedure was performed by docking the crystal ligand (imatinib) into the receptors.. Docking results showed that imatinib formed H-bond interactions with Cys673, Thr670 (Gatekeeper), Glu640, and Ile769 in c-KIT and Cys677, Thr836, Glu644, Val815, and His816 in PDGFR α .

Docking of the compound 14 with c-KIT showed H-bond interactions with Cys673 at the hinge region, Glu640 at the α C-helix, and Asp810 at the DFG motif of the activation loop. The binding interaction of pyrazolopyridine of compound 14 with hinge residue Cys673 was analogous to the interaction of the pyridine of imatinib with Cys673 observed in the X-ray structure (PDB ID **1T46**). This interaction with the hinge region was crucial for anchoring the ligand at the binding site. Compound 14 showed H-bond interaction with PDGFR α residues Cys677 (hinge), Glu644 (α C-helix), and Asp836 (DFG motif). The results suggested that compound 14 was bound to c-KIT and PDGFR α in a similar binding pattern.

3.2 *Molecular Dynamics Simulation*

Classical MD simulations of imatinib and compound 14 with c-KIT and PDGFR α were performed for 100 ns using Gromacs. The interactions of imatinib and compound 14 with the receptors and the pairwise RMSD of the ligands from the MD trajectories were shown in Figure 1. During the simulation of the imatinib-c-KIT complex, the α C-helix and the activation loop formed a narrow pocket around the binding site (Figure 1a). This allowed the Glu640 from the α C-helix to form a stable salt bridge with Lys623 of β 3 and one H-bond with the amide linker of imatinib. Imatinib also formed a stable H-bond with Cys673 at the hinge and

Thr670 of the gate-keeper residue. The methyl piperazine moiety of imatinib also formed a weak H-bond with the Ile789 of the catalytic loop. The overlap between the crystal ligand and the MD binding pose showed an RMSD value of 0.9 Å.

In PDGFR α (Figure 1b), imatinib formed H-bond interaction with Cys677 and Thr674 at the hinge region. The carbonyl linker between the two benzyl rings of imatinib also formed H-bond interaction with the Asp836 at the DFG motif. Additionally, a weak H-bond interaction was also observed between the methyl piperazine of imatinib and the catalytic loop residue Val815. These interactions were also observed in the X-ray structure of the imatinib-PDGFR α (PDB ID 6JOL). The overlap of the crystal ligand and the MD binding pose showed an RMSD value of 0.4 Å.

In the compound 14-c-KIT complex simulation, the pyrazolopyridine of compound 14 occupied the pocket close to the hinge region and formed H-bond interactions with Cys673. Additionally, compound 14 formed H-bond interactions with Glu640 and Asp810 from the α C-helix and the DFG-motif respectively. The morpholine moiety of compound 14 extended into the hydrophobic pocket formed by residues Ile571, Val643, Leu647, Phe782, Leu783, Cys788, and Ile789 from the α C-helix and the catalytic domain. The binding pose of compound 14 inside c-KIT is shown in Figure 1c. In PDGFR α , compound 14 formed H-bond interactions with Cys677 (hinge), Glu644 (α C-helix), and Asp836 (DFG motif). The binding pose of compound 14 inside PDGFR α is given in Figure 1d. These results indicated that compound 14 formed interactions with c-KIT and PDGFR α in a similar pattern.

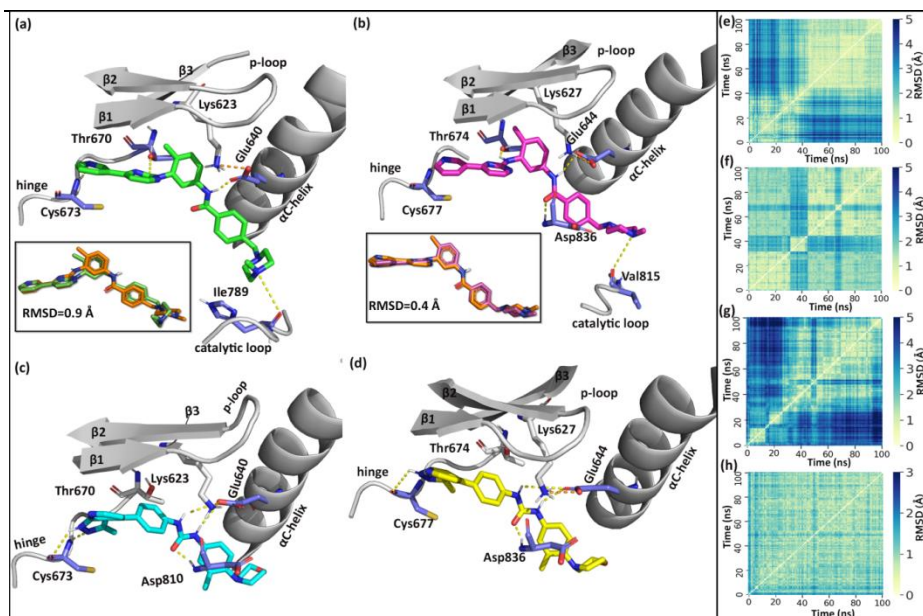


Figure 1. H-bond interactions of imatinib and compound 14 with c-KIT and PDGFR α from the MD simulations. H-bond interactions are represented by yellow dotted lines and residues forming H-bonds are shown in purple color. **(a)** Binding interactions between imatinib and c-KIT. The overlap between the crystal ligand pose (salmon) and the MD binding pose (green) of imatinib at the binding site (RMSD=0.9 Å). **(b)** Binding interactions between imatinib and PDGFR α . The overlap between the crystal ligand pose (salmon) and the MD binding pose (magenta) of imatinib at the binding site (RMSD=0.4 Å). **(c)** Binding interactions between compound 14 and c-KIT. **(d)** Binding interactions between compound 14 and PDGFR α . The pairwise RMSD plots of the ligands from the MD simulation of **(e)** imatinib and c-KIT **(f)** imatinib and PDGFR α **(g)** compound 14 and c-KIT **(h)** compound 14 and PDGFR α .

3.3. Evaluation of Binding Energy

The binding energy (BE) of imatinib and compound 14 with c-KIT and PDGFR α were evaluated using the *g_mmpbsa* package. The contribution of van der Waals,

electrostatic, polar, and non-polar solvation energy terms to the total BE of the protein-ligand complexes were given in Table 2. The total BE of imatinib-c-KIT and imatinib-PDGFR α were -105 kJ/mol and -104 kJ/mol respectively. The van der Waals and electrostatic energy terms made the major contributions to the total BE. In imatinib-c-KIT binding, the van der Waals and electrostatic energies contributed -236 kJ/mol and -70 kJ/mol to the total BE respectively. In imatinib-PDGFR α , the van der Waals and electrostatic energy contributions were -244 kJ/mol and -58 kJ/mol respectively. The total BE values of compound 14-c-KIT and compound 14-PDGFR α were -120 kJ/mol and -117 kJ/mol respectively. In the compound 14-c-KIT interaction, the van der Waals and electrostatic energies contributed -257 kJ/mol and -57 kJ/mol to the total BE respectively. In compound 14-PDGFR α interaction, the van der Waals energy contribution was -251 kJ/mol and the electrostatic energy contribution was -55 kJ/mol.

Table 2. The energy contribution of the various energetic terms (Van der Waals energy, electrostatic energy, polar solvation energy/SASA) to the total binding energy during the binding of imatinib and compound 14 with c-KIT and PDGFR α .

Complexes	Van der Waals (kJ/Mol)	Electrost atics (kJ/Mol)	Polar solvation (kJ/Mol)	SASA (kJ/Mol)	Total Binding Energy (kJ/Mol)
Imatinib-c-KIT	-260	-74	257	-28	-105
Imatinib- PDGFRα	-244	-58	225	-27	-104
Compound14-c- KIT	-257	-57	219	-25	-120
Compound14- PDGFRα	-251	-55	213	-25	-118
Compound31 - c-KIT	-183	-48	185	-19	-65
Compound31- PDGFRα	-180	-50	163	-18	-85

Imatinib-c-KIT/I670	-227	-48	228	-26	-73
Imatinib-PDGFRα/I674	-248	-18	227	-28	-67
Compound 14-c-KIT/I670	-250	-61	205	-25	-131
Compound 14-PDGFRα/I674	-265	-39	203	-25	-126

The residues that made a high contribution to the total BE in compound 14-c-KIT interaction were compared with the corresponding residues in imatinib-c-KIT interaction in Table 3. The results indicated that the hydrophobic residues Val603, Leu644, Val654, Cys809, and Phe811 individually contributed more than -5 kJ/mol to the total BE in compound 14-c-KIT interaction. In imatinib-c-KIT, the hydrophobic residues Leu644, Val654, Tyr672, and Cys809 individually contributed more than -5 kJ/mol to the total BE. The hydrophobic residues Val607, Met648, Val658, Leu825, and Cys835 individually contributed more than -5 kJ/mol to the total BE in both imatinib-PDGFR α and compound 14-PDGFR α interactions. The polar residue Asp836 contributed -5.1 kJ/mol in compound 14-PDGFR α interaction and 1.8 kJ/mol in imatinib-PDGFR α interaction. Similarly, the corresponding c-KIT residue Asp810 contributed -2.1 kJ/mol in compound 14-c-KIT interaction and 9.3 kJ/mol in imatinib-c-KIT interaction. The high energy contribution of Asp836/Asp810 in the interactions with compound 14 may be attributed to the H-bond interaction between the carbonyl oxygen of compound 14 with Asp836/Asp810 in c-KIT/PDGFR α . The PDGFR α residue Phe837 also contributed -4.8 kJ/mol and -7.1 kJ/mol in the interactions with compound 14 and imatinib respectively. The corresponding c-KIT residue Phe811 also contributed -6.8 kJ/mol and -4 kJ/mol in interactions with compound 14 and imatinib respectively through hydrophobic interactions. These hydrophobic interactions with Phe811/Phe837 of the DFG motive were possible as a result of the DFG-in conformation of the inactive form of c-KIT and PDGFR α [94].

Table 3. Residues that showed a high contribution to the total binding energy during the MD simulations of **Compound14-c-KIT**, **Imatinib-c-KIT**, **Compound14-PDGFR α** , and **Imatinib-PDGFR α** . The energy values of the residues are given in kJ/mol.

c-KIT Residues	Compound1 4-c- KIT (kJ/mol)	Imatini b-c-KIT (kJ/mol)	PDGFR α Residues	Compound 14- PDGFR α (kJ/mol)	Imatini b- PDGFR α (kJ/mol)
Asp572	-0.9	-0.7	Glu587	-0.66	-0.42
Leu595	-2.6	-3.4	Leu599	-2.48	-2.87
Val603	-5.4	-3.2	Gly600	-0.91	-0.82
Ala621	-2.7	-3.1	Val607	-5.23	-5.10
Val620	-0.7	-0.8	Val608	-1.10	-0.69
Val622	-1.0	-1.7	Glu609	-1.31	-2.33
Glu635	-0.8	-0.7	Val624	-0.76	-1.10
Val643	-3.1	-3.6	Ala625	-2.20	-1.87
Leu644	-6.7	-6.6	Val626	-1.15	-1.84
Leu647	-2.1	-0.5	Glu637	-0.83	-0.96
Ile653	-3.0	-0.9	Ile647	-4.19	-4.35
Val654	-5.2	-7.6	Met648	-8.93	-7.98
Tyr672	-3.9	-5.6	Leu651	-2.11	-0.96
Cys673	-2.1	-2.5	Ile657	-2.52	-0.76
Gly676	-0.8	-0.4	Val658	-5.46	-5.28
Leu783	-2.9	-0.6	Ile672	-0.83	-2.95
Cys788	-1.3	-1.8	Tyr676	-3.39	-4.99
His790	-3.6	-1.9	Cys677	-2.48	-1.82
Asp792	-1.0	-0.2	Gly680	-0.92	-0.76
Leu799	-4.9	-4.3	Leu809	-2.58	-1.14
Lys807	-1.1	2.2	Cys814	-2.27	-1.81
Ile808	-1.0	0.4	Leu825	-5.59	-5.59
Cys809	-6.4	-6.2	Ile834	-1.21	-0.40
Asp810	-2.1	9.3	Cys835	-6.27	-5.11
Phe811	-6.8	-4.0	Asp836	-5.11	1.89
Asp851	-1.5	-0.5	Phe837	-4.87	-7.16

3.4. 3D-QSAR

The 48 pyrazolopyridine derivatives and their activity values were used to perform the 3D-QSAR study. The specific activity values of compounds 3, 4, 24, 25, 34, and 38 for both receptors were not available and were excluded from the 3D-QSAR study. The compounds were randomly separated into a training set and a test set of 30 compounds and 12 compounds respectively.

In the c-KIT CoMFA and CoMSIA models, the binding pose of compound 14 from the MD simulation with c-KIT was used as a template for the alignment of the compounds. The aligned compounds are shown in Figure 2(g). The developed CoMFA model showed a crossvalidated q^2 value of 0.63 and an optimal number of components (ONC) value of 6. In the non-validated analysis, the model showed an r^2 value of 0.98 and SEE value of 0.2, suggesting that the model has a reasonable predictive ability. The CoMSIA model based on the hydrophobic (H) and steric (S) descriptors gave relatively higher statistical results. Hence, this model was selected for further analysis. The selected CoMSIA model exhibited q^2 and ONC values of 0.6 and 5 respectively. In the non-crossvalidated analysis, the CoMSIA model showed r^2 and SEE values of 0.9 and 0.46. The statistical results of the c-KIT CoMFA and CoMSIA models are shown in Table 4.

Table 4. Statistical results of the CoMFA and CoMSIA models for c-KIT and PDGFR α . q^2 : cross-validated correlation coefficient; ONC: Optimal number of components; r^2 : non-cross-validated correlation coefficient; SEE: Standard Error of Estimation; F value: F-test value; r^2 ; BS- r^2 : Bootstrapping r^2 mean; BS-SD: Bootstrapping Standard deviation; r^2_{pred} : predictive correlation coefficient; S: Steric; E: Electrostatic; H: Hydrophobic.

Parameters	CoMFA (c-KIT)	CoMSIA (c-KIT)	CoMFA (PDGFR α)	CoMSIA (PDGFR α)
q^2	0.63	0.6	0.61	0.62
ONC	6	5	6	3
r^2	0.98	0.9	0.98	0.81

SEE	0.2	0.46	0.12	0.39
F value	204	43	232	46
BS r^2	0.98	0.94	0.98	0.97
BS SD	0.15	0.32	0.1	0.14
r^2_{pred}	0.59	0.58	0.56	0.59
Influence of different fields (%)				
S	59	50	67	42
E	41	-	33	-
H	-	50	-	58

In the PDGFR α CoMFA and CoMSIA models, the compounds were aligned based on the binding pose of compound 14 from the MD simulation with PDGFR α . The aligned compounds are shown in Figure 2(h). The PDGFR α CoMFA model showed a q^2 value of 0.61 and an ONC value of 6. In the non-validated analysis, the model showed r^2 and SEE values of 0.98 and 0.12 respectively. The PDGFR α CoMSIA model was developed based on the HS descriptors. The PDGFR α CoMSIA model showed q^2 and ONC values of 0.62 and 3 respectively. The non-crossvalidated r^2 and SEE values were 0.81 and 0.39 respectively. The statistical results of the PDGFR α CoMFA and CoMSIA models are shown in Table 4.

Internal and external validation of the derived 3D-QSAR models were performed using bootstrapping (BS) and external r^2_{pred} analysis. The c-KIT CoMFA model showed BS- r^2 and BS-SD values of 0.98 and 0.15 respectively. The c-KIT CoMSIA (SH) model showed a BS- r^2 value of 0.94 and a BS-SD value of 0.32. The BS analysis suggested that the c-KIT CoMFA and CoMSIA models have reasonable robustness. The PDGFR α CoMFA model showed a BS- r^2 value of 0.98 and a BS-SD value of 0.1. The BS- r^2 and BS-SD values for the CoMSIA model were 0.97 and 0.14 respectively. These results suggested that the derived CoMFA and CoMSIA models have reasonable robustness. In the external validation, c-KIT CoMFA and CoMSIA models showed r^2_{pred} values of 0.59 and 0.58 respectively. The PDGFR α CoMFA and CoMSIA models showed r^2_{pred}

values of 0.56 and 0.59 respectively. The external validation results suggested that the derived models have reasonable predictive ability against an external dataset. The predicted activity values of the compounds for c-KIT and PDGFR α are given in Table 7 and 8.

3.5. Analysis of Contour Map

In the CoMFA and CoMSIA contour maps, compound 14 was used as a reference. The contour maps are shown in Figure 2. In the electrostatic contour map, the red and blue contours represented electronegative and electropositive substituents favorable and unfavorable regions respectively. The Green and yellow color contours in the steric contour map represented bulky substituents favorable and unfavorable regions respectively. In the hydrophobic contour map, cyan and purple color contours represented hydrophobic favorable and unfavorable regions respectively.

In the c-KIT CoMFA electrostatic contour map (Figure 2a), a blue contour was observed near the meta position of the methylbenzene ring suggesting that electropositive substituents were favored at that position. In the steric contour map (Figure 2b), a green contour was observed near the meta position. The yellow contour near the ortho position of methylbenzene indicated that bulky substituents were not favored in that position. Bulky substituents at the ortho position of methylbenzene could lead to a steric clash with binding site residues. In the CoMSIA hydrophobic contour map (Figure 2c), a cyan contour was seen near the para position and meta position of methylbenzene suggesting that hydrophobic substituents were favorable in these regions. CoMSIA steric contour map was similar to that of the CoMFA steric contour map and was not included in the analysis.

In the PDGFR α CoMFA electrostatic contour map (Figure 2d), a blue contour was observed near the benzene and the para position of the methylbenzene

suggesting that electropositive substituents were favored in these regions. In the steric contour map (Figure 2e), the green contour was observed near the meta-position of methylbenzene suggesting that bulky substituents were favored in that region. The yellow contour near the ortho position suggested that bulky substituents were not favored in that region and could lead to decreased activity for both c-KIT and PDGFR α . In the CoMSIA hydrophobic contour map (Figure 2f), a cyan contour was observed near the ortho position of the methylbenzene and purple contours were observed near the pyrazolopyridine and the meta position of the methylbenzene suggesting that hydrophobic substituents are not favored in these regions.

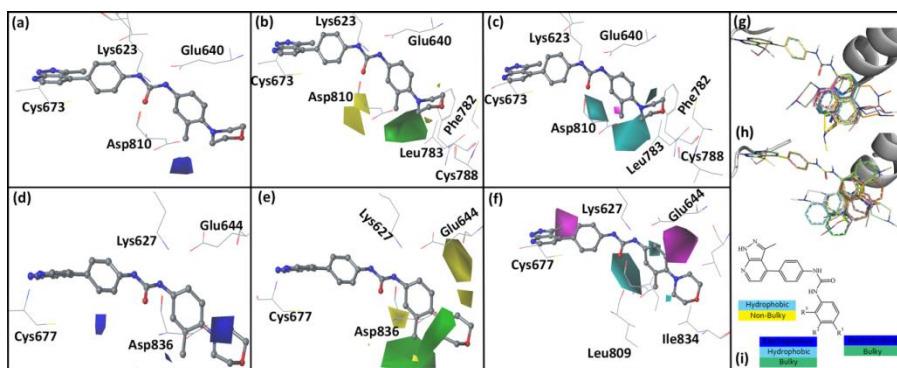


Figure 2. Contour maps generated based on the CoMFA and CoMSIA models for c-KIT and PDGFR α with Compound 14 used as a reference. Blue and red contours indicate electropositive and electronegative substituents favorable regions respectively. Green and yellow contours indicate steric bulk substituents favorable and unfavorable regions respectively. Cyan and purple colors contours represent hydrophobic favorable and unfavorable regions. (a) Electrostatic contour map for the c-KIT CoMFA model. (b) Steric contour map for c-KIT CoMFA model (c) Hydrophobic contour map for c-KIT CoMSIA model. (d) Electrostatic contour map for the PDGFR α CoMFA model. (e) Steric contour map for PDGFR α CoMFA model (f) Hydrophobic contour map for PDGFR α CoMSIA model. Alignments used for the development of the 3D-QSAR models.

(g) Alignment of the compounds inside c-KIT (h) Alignment of the compounds inside PDGFR α . (i) Scheme developed based on the 3D-QSAR models for designing new compounds.

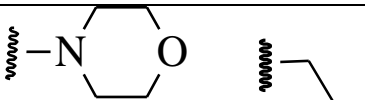
3.6. Designed Compounds

Based on the 3D-QSAR contour maps, a design scheme was developed as shown in Figure 2(g). Following the scheme, 50 compounds were designed and the activity values for c-KIT and PDGFR α were predicted using the derived CoMSIA (SH) models. Base on the predicted pIC₅₀ values, eight compounds that showed higher activity values than the compound 14 were selected for further evaluations. The predicted activity values of the designed compounds for both the receptors are given in Table 5.

MD simulation of the eight designed compounds with the receptors was performed for 70 ns. The binding interactions of the designed compounds with the receptors are shown in Figure 3. The results showed that the designed compounds were able to form stable interactions with both c-KIT and PDGFR α throughout the simulation. In the designed compounds, the pyrazolopyridine moiety was anchored near the hinge region of c-KIT and PDGFR α through H-bond interactions with Cys673 and Cys677. The eight designed compounds also showed H-bond interaction with DFG motif residues Asp810/Asp836 (c-KIT/PDGFR α). Except for compound D23, D25, and D44 in PDGFR α , the designed compounds also formed weak H-bond interactions with the Lys623/Lyss627 (c-KIT/ PDGFR α) and Glu640/Glu644 (c-KIT/ PDGFR α).

Table 5. The chemical structures and the predicted pIC₅₀ values of the newly designed compounds for c-KIT and PDGFR α .

Compounds	R1	R2	R3	Predicted Activity (pIC ₅₀)	
				c-KIT	PDGFR α
Compound D18			H	10.4	8.3
Compound D23			CH3	10.1	8.2
Compound D25			CH3	10.5	8.1
Compound D28			CH3	9.6	8.4
Compound D32			CH3	9.1	8.3
Compound D39			H	10.3	8.1
Compound D44			H	10.2	8.3

Compound D45		H	9.3	8.1
------------------------	---	---	-----	-----

The BE values of the designed compounds with both c-KIT and PDGFR α are given in Table 6. The evaluation of the BE values showed that the designed compounds possessed higher predicted binding affinity than imatinib and compound 14 for both receptors. Among the eight designed compounds, the D39 showed the highest binding affinity against both the receptors. In c-KIT, D39 formed H-bond interactions with Cys673, Glu640, and Asp810. In PDGFR α , D39 formed H-bond interactions with Cys677, Asp836, and Lys627. In addition, the benzyl and butyl substituents at the R1 and R2 positions of the methylbenzene extended into the hydrophobic pocket. This allowed the formation of hydrophobic interactions with residues Leu647, Val643, His790, Cys809, and Ile808 in c-KIT and Ile657, Met648, Val815, Leu809, and Ile834 in PDGFR α . The benzyl and butyl substituents were unique in D39 which suggested that having hydrophobic substituents in the R1 and R2 positions may increase the binding affinity towards both c-KIT and PDGFR α . The hydrophobic interactions of D39 with the receptors are given in Figure 4.

Table 6. The energy contributions of the various energetic terms (Van der Waals energy, electrostatic energy, polar solvation energy, and non-polar solvation energy/SASA) to the total binding energies are shown for the designed compounds.

Complexes (Designed Compounds- Receptor)	Van der Waals (kJ/Mol)	Electrostat ics (kJ/Mol)	Polar solvation (kJ/Mol)	SASA (kJ/Mol)	Total Binding Energy (kJ/Mol)
D18-c-KIT	-272	-57	218	-28	-139
D18 - PDGFRα	-282	-57	224	-27	-142
D23-c-KIT	-287	-48	242	-29	-122
D23-	-286	-36	212	-28	-138

PDGFRα					
D28-c-KIT	-278	-62	242	-27	-126
D28- PDGFRα	-289	-38	227	-29	-129
D32-c-KIT	-271	-53	229	-27	-122
D32- PDGFRα	-283	-44	224	-28	-130
D39-c-KIT	-268	-65	212	-26	-148
D39- PDGFRα	-274	-48	200	-26	-150
D44-c-KIT	-262	-56	216	-25	-129
D44- PDGFRα	-261	-42	209	-26	-120
D45-c-KIT	-262	-76	244	-26	-121
D45- PDGFRα	-274	-58	215	-26	-143

The synthetic accessibility of the designed compounds was evaluated with SwissADMET (<http://www.swissadme.ch/>) and the results were given in Table 9. The synthetic accessibilities of the compounds were scored within the range of 1 to 10 where a synthetic accessibility score of 1 indicates easy synthesis and a score of 10 indicates difficult synthesis. The designed compounds showed a reasonable synthetic accessibility score of less than 5. The absorption (A), distribution (D), metabolism (M), excretion (E), and toxicity (T) properties of the designed compounds were also evaluated using the pkCSM online server (<http://biosig.unimelb.edu.au/pkcsm/>)[132]. Steady-state volume of distribution (VD_{ss}) of a compound represents the degree to which the compound will likely get distributed in the body rather than the plasma. VD_{ss} score is considered low if below -0.15 log L/kg. The designed compounds showed low to moderate VD_{ss} scores, suggesting a reasonable distribution rate. The designed compounds also showed a positive outcome for the cytochrome P450 substrate test suggesting that the compounds are likely to be metabolized by cytochrome P450. Except for compound D32 and D39, the designed compounds showed a total clearance rate

of at least 0.8. The compound D32 and D39 showed a clearance rate of 0.6 and 0.5 respectively. The toxicity prediction showed that, except compound D39, the designed compounds tested negative for mutagenic potential.

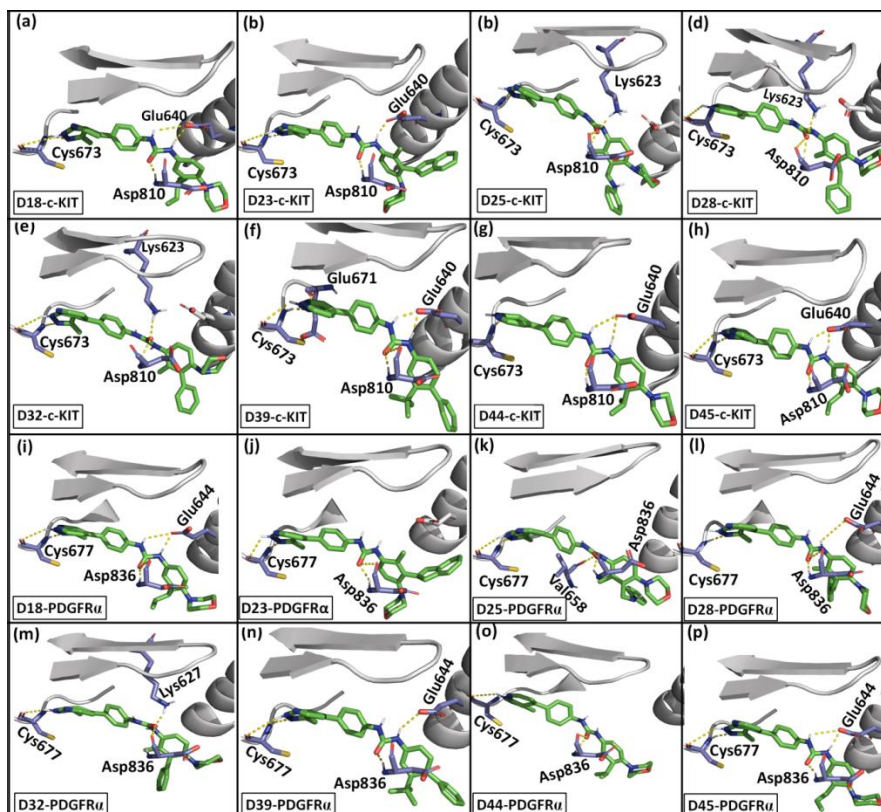


Figure 3. H-bond interactions of the designed compounds with c-KIT and PDGFR α from the MD simulations. H-bond interactions are represented by yellow dotted lines and residues forming H-bonds are shown in purple color. (a) D18-c-KIT (b) D23-c-KIT (c) D25-c-KIT (d) D28-c-KIT (e) D32-c-KIT (f) D39-c-KIT (g) D44-c-KIT (h) D45-c-KIT (i) D18-PDGFR α (j) D23-PDGFR α (k) D25-PDGFR α (l) D28-PDGFR α (m) D32-PDGFR α (n) D39-PDGFR α (o) D44-PDGFR α (p) D45-PDGFR α

4. Discussion

Molecular docking and molecular dynamics simulation of compound 14 showed H-bond interactions with Cys673, Glu640, and Asp810 in c-KIT and Cys677, Glu644, and Asp836 in PDGFR α . For comparative study, MD simulations of the

compound 31 which showed low activity values for c-KIT ($pIC_{50} = 6.2$) and PDGFR α ($pIC_{50} = 5.9$) was performed. In c-KIT, Compound 31 form H-bond interactions with Cys673, Glu640, and Asp810 which were also observed in compound 14-c-KIT interaction. However, compound 31 did not have the extended methylbenzene and morpholine moiety present in compound 14 and lost the hydrophobic interactions with Leu644, Ile768, Leu783, Leu647, Val643, and Ile808 which were observed in compound 14-c-KIT interaction. The hydrophobic and H-bond interactions of compound 31 with c-KIT are shown in Figure 4 and Figure 5 respectively. Similarly, compound 31 formed H-bond interactions with PDGFR α residues Cys677, Lys627, and Asp836 which were observed in compound 31-PDGFR α interaction. However, compound 31 did not form hydrophobic interactions at the catalytic loop and the α C-Helix due to the absence of the extended methylbenzene moiety. The loss of the hydrophobic interactions at the catalytic loop and the α C-Helix could be the reason why compound 31 showed lower activity value against both c-KIT and PDGFR α .

Contour map analysis suggested that positive, bulky, and hydrophobic substituents were favored near the meta position of the methylbenzene of compound 14 and could increase activity for c-KIT and PDGFR α . The presence of bulky hydrophobic substituents at the meta position may lead to the formation of crucial hydrophobic interaction with residues from the α C-helix and the catalytic loop (Figure 5). The result is also supported by the BE evaluation which showed that hydrophobic residues Leu644, Val643, Leu647 from the α C-helix, and Leu783 and Cys788 from the catalytic loop made key contributions to the total BE in c-KIT (Table 3). Further analysis also showed that hydrophobic residues Val603, Leu644, Val654, Cys809, and Phe811 contributed more than -5 kJ/mol to the total BE in compound 14-c-KIT. Whereas, the hydrophobic residues Val607, Met648, Val658, Leu825, and Cys835 contributed more than -5 kJ/mol to the total BE in compound 14-PDGFR α . These results suggested that hydrophobic interactions were dominant in the binding of compound 14 with both

the receptors. The eight designed compounds showed a higher binding affinity with both c-KIT and PDGFR α compared to compound 14 and imatinib. The higher binding affinity could be attributed to the hydrophobic substituents in the designed compounds which were able to form interaction with hydrophobic residues from the catalytic loop and the α C-Helix (Figure 5a).

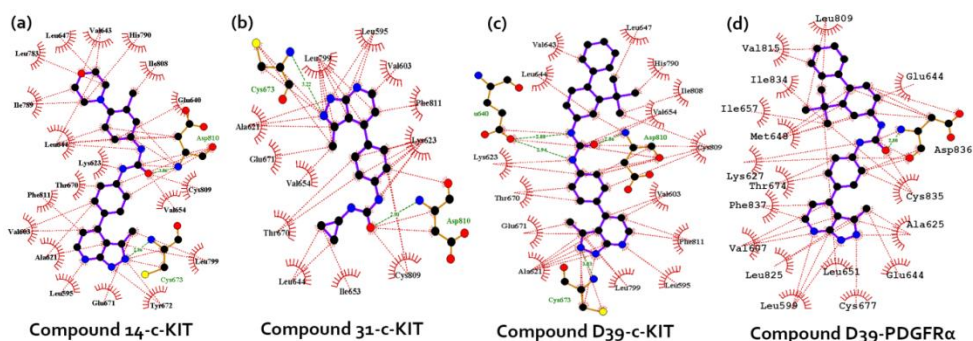


Figure 4. Showing the hydrophobic interactions of the inhibitors with c-KIT and PDGFR α . Hydrophobic interactions were represented in red dotted lines. Residues that showed Hydrophobic and H-bond interactions were given in green and black label respectively. (a) **Compound 14-c-KIT** (b) **Compound 31-c-KIT** (c) **D39-c-KIT** (d) **compound D39- PDGFR α** .

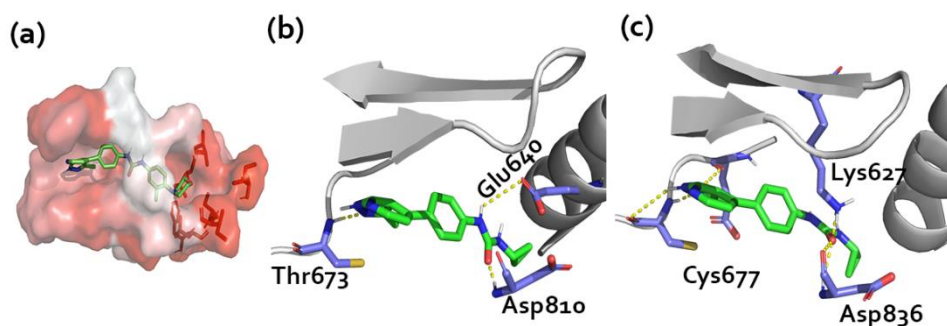


Figure 5 (a) The hydrophobic surface of the c-KIT binding site residues; Hydrophobic residues from the α C-helix and the catalytic loops are shown in stick representation **(b)** Binding interactions between **compound 31** and c-KIT. **(c)** Binding interactions between **compound 31** and PDGFR α .

Earlier studies have shown that imatinib resistance is achieved via T670I substitution in c-KIT and T674I substitution in PDGFR α [108, 124]. However, the molecular mechanisms underlying the drug resistance remained unclear. We have performed MD simulations of the imatinib with c-KIT/I670 and PDGFR α /I674 mutants to study the effect of the T670I/T674I substitutions on the binding interactions. The results showed that imatinib formed H-bond interaction with Cys673 and Glu640 in c-KIT/I670. In imatinib-PDGFR α /I674 complex, imatinib formed only one H-bond interaction with Glu644 and the pyridine of imatinib moved out of the hinge region. This outward movement could be attributed to the loss of H-bond interaction with Cys677 as a result of the T674I substitution. The binding interaction of imatinib and compound 14 with c-KIT/I670 and PDGFR α /I674 are shown in Figure 4. In c-KIT/I670, compound 14 formed H-bond interactions with Cys673, Glu640, Asn810, Ile789, and His790. Whereas, compound 14 formed H-bond interactions with Cys677, Glu675, Glu644, and Asp836 in PDGFR α /I674. Following the T670I substitution, the total BE of imatinib was reduced from -105 kJ/mol (wild-type) to -72 kJ/mol (c-KIT/I670). Similarly, the total BE value of imatinib reduced from -104 kJ/mol (wild-type) to -67 kJ/mol (PDGFR α /I674) after T674I substitution. On the other hand, compound 14 showed total BE values of -131 kJ/mol with c-KIT/I670 and -122 kJ/mol with PDGFR/I674. These results suggested that the T670I/T674I substitutions disrupted the interaction of imatinib with c-KIT/I670 and PDGFR/I674 which consequently reduced the binding affinity against the receptors. Whereas, compound 14 was able to retain the interactions with c-KIT/I670 and PDGFR α /I674 resulting in high binding affinity against both receptors. In contrast to the H-bond interaction between the amide linker of

imatinib and T670/T674 in c-KIT/PDGFR α , compound 14 formed hydrophobic interactions with T670/T674. Hence, the substitution of the hydrophilic threonine residue with the hydrophobic isoleucine residue could have led to the loss of H-bond interaction with imatinib while increased the binding affinity for compound 14 through hydrophobic interactions.

Residue contact map analysis was calculated for the Ile670 (c-KIT/I670) and Ile674 (PDGFR α /I674) to study the effect of the T670I/T674I substitutions on the residue interactions at the binding sites. The residue contact map shows how often a residue of interest interacted with its surrounding residues throughout the simulation[133]. The I670/I674 contact maps were generated from the MD trajectories of imatinib and compound 14 with c-KIT, c-KIT/I670, PDGFR α , and PDGFR α /I674. The contact maps are shown in Figure 6. Comparison of the Thr670 (imatinib-c-KIT) and the Ile670 (imatinib-c-KIT/I670) contact maps showed that the T670I substitution led to the loss of interactions with the Val620 and Val668 (Figure 6e & 4f). Similarly, the T674I substitution also led to the loss of interactions with Val672 and Lys627 in imatinib-PDGFR α /I674 (Figure 6i & 4j). The comparison of the Thr670 (compound 14-c-KIT) and Ile670 (compound 14-c-KIT/I670) contact maps showed that the substitution led to the loss of interactions with Val668, Val620, and Lys623. However, the substitution has also led to more interactions with Asn655. Comparison of the contact maps for Thr674 and Ile674 in compound 14-PDGFR α and compound 14-PDGFR α /ILE674 showed that that the substitutions led to the loss of interaction with Val624, Val626, and Lys627. However, T674I substitution also led to more interaction with Asn659, Ile672, and Met648.

These results suggested that the T670I substitution in c-KIT induced conformation changes at the binding site which led to reduced interaction with Val620, Val668, and Lys623 while increased the interactions with Asn655. Similarly, T674I substitution in PDGFR α led to the loss of interaction with Val624 and Lys627. The loss of interaction with Val620/Val624, and

Lys623/Lys627 in c-KIT/PDGFR α upon T670I/T674I substitution was characterized by the movement of the I670/I674 at the DFG motif away from the β 3 and moving closer to the α C-helix. These results complemented earlier claims that the T670I/T674I substitutions modified the binding pocket of c-KIT/PDGFR α [108, 134].

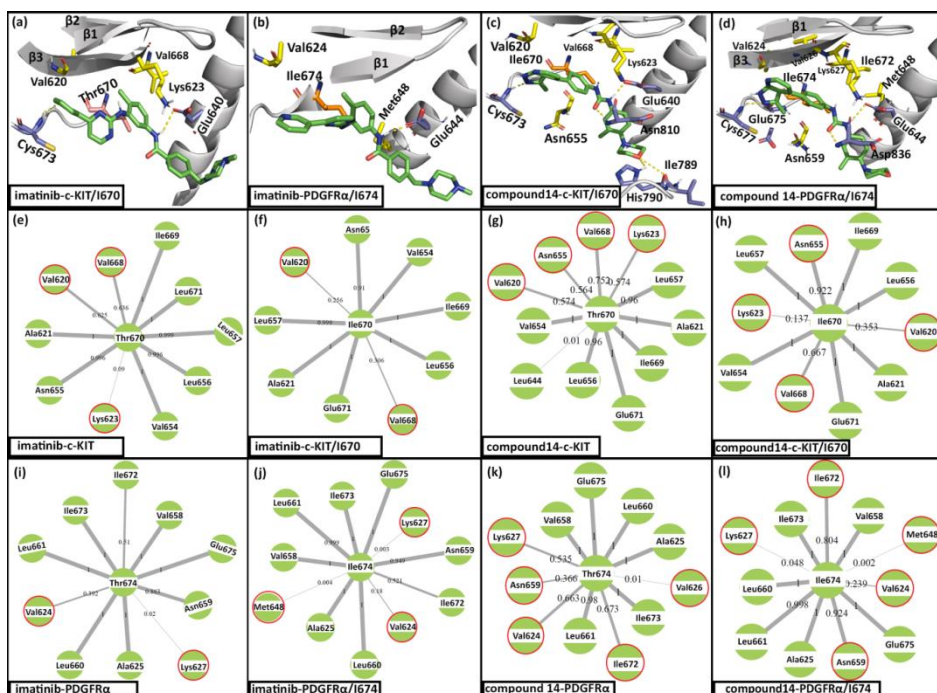


Figure 6. H-bond interactions and residue contact map from the MD simulation of imatinib and compound 14. H-bond interactions observed in (a) imatinib-c-KIT/I670 (b) imatinib-PDGFR α /I674 (c) compound 14-c-KIT/I670 (d) compound 14-PDGFR α /I674. Residue contact map for (e) Thr670 in imatinib-c-KIT (f) Ile670 in imatinib-c-KIT/I670 (g) Thr670 in compound 14-c-KIT (h) Ile670 in compound 14-c-KIT/I670 (i) Thr674 in imatinib-PDGFR α (j) Ile674 in imatinib-PDGFR α /I674 (k) Thr674 in compound 14-PDGFR α (l) Ile674 in compound 14-PDGFR α /I674. Weight on the edge between two residues represented how often the interaction existed between the residues during the simulation. Residues that showed a significant increase or decrease in interactions with the residue of interest were highlighted in red circles.

Table 7. The experimental/actual and predicted pIC₅₀ values with their residuals for the CoMFA and CoMSIA for c-KIT.

Compound	Experimental pIC ₅₀	c-KIT CoMFA		c-KIT CoMSIA	
		Predicted pIC ₅₀	Residual	Predicted pIC ₅₀	Residual
1	8.62	8.27	-0.35	8.27	0.35
2	8.43	8.61	0.17	9.10	-0.67
5	8.41	8.58	0.17	8.13	0.28
6*	4.96	5.80	0.84	5.96	-0.99
7	5.74	5.91	0.17	6.73	-0.99
8	5.32	5.26	-0.06	5.86	-0.54
9	5.62	5.51	-0.11	5.51	0.11
10*	6.70	7.21	-0.52	7.33	-0.63
11	7.72	7.62	-0.10	7.57	0.15
12	5.71	5.87	0.16	5.88	-0.17
13	8.14	8.16	0.02	8.18	-0.04
14	8.62	8.68	0.06	8.27	0.35
15*	5.65	6.55	-0.90	6.60	-0.95
16	5.34	5.17	-0.17	5.16	0.18
17	6.40	6.12	-0.28	6.07	0.33
18	5.19	5.50	0.31	5.35	-0.16
19	7.92	7.97	0.05	8.26	-0.34
20*	7.92	7.62	0.30	8.90	-0.98
21	8.59	8.71	0.13	8.41	0.17
22	8.03	7.86	-0.17	7.78	0.25
23*	4.66	5.58	-0.93	5.60	-0.94
26	4.69	4.73	0.04	5.31	-0.62
27	8.46	8.37	-0.09	8.51	-0.05
28*	8.85	8.09	0.77	8.44	0.42
29	6.49	6.41	-0.08	5.88	0.62
30	5.85	6.10	0.25	5.82	0.03
31	6.26	6.18	-0.08	5.93	0.32
32*	6.50	6.46	0.03	6.30	0.20
33	6.84	6.94	0.10	7.08	-0.24
35*	7.29	8.10	-0.81	7.92	-0.62
36	7.59	7.48	-0.10	7.45	0.14
37*	7.11	6.11	1.00	6.47	0.64
39	8.43	8.57	0.14	8.37	0.06
40	8.77	9.02	0.25	8.88	-0.11
41	8.21	8.32	0.11	8.71	-0.51
42*	8.06	8.91	-0.85	8.68	-0.62
43	6.92	6.98	0.06	7.15	-0.23
44	7.70	7.75	0.05	7.75	-0.05
45*	7.85	6.92	0.94	7.62	0.23
46	8.72	8.37	-0.35	8.45	0.28
47*	7.70	8.42	-0.72	8.49	-0.79

48	8.44	8.14	-0.30	7.52	0.92
-----------	------	------	-------	------	------

* Test set compounds

Table 8. The experimental/actual and predicted pIC₅₀ values with their residuals for the CoMFA and CoMSIA for PDGFR α .

Compound	Experimental pIC ₅₀	PDGFR α CoMFA		PDGFR α CoMSIA	
		Predicted pIC ₅₀	Residual	Predicted pIC ₅₀	Residual
1	7.06	7.14	-0.08	7.32	-0.26
2	7.49	7.49	0.01	7.45	0.04
5	7.66	7.54	0.12	7.23	0.43
6*	6.52	6.11	0.40	6.78	-0.27
7	7.19	6.98	0.21	6.99	0.20
8	6.58	6.55	0.03	6.81	-0.22
9	6.74	6.87	-0.13	7.05	-0.31
10*	4.67	3.85	0.82	5.56	-0.89
11	6.67	6.70	-0.03	6.57	0.09
12	5.75	5.57	0.18	5.25	0.50
13	6.90	7.00	-0.11	6.96	-0.06
14	8.14	7.98	0.16	7.96	0.18
15*	5.96	6.34	-0.38	6.65	-0.69
16	6.03	6.25	-0.22	5.50	0.53
17	4.72	4.66	0.06	5.30	-0.57
18	5.39	5.35	0.04	5.56	-0.17
19	6.87	6.89	-0.03	7.05	-0.19
20*	6.75	6.81	-0.06	7.04	-0.29
21	7.08	7.04	0.04	6.97	0.11
22	6.51	6.54	-0.03	6.86	-0.35
23*	5.34	5.99	-0.66	6.17	-0.83
26	4.36	4.33	0.03	4.59	-0.23
27	6.79	6.86	-0.07	6.80	-0.01
28*	7.57	6.94	0.63	6.90	0.67
29	5.95	5.96	-0.01	5.78	0.17
30	5.83	6.09	-0.25	5.90	-0.07
31	5.95	5.91	0.04	6.00	-0.06
32*	6.14	5.88	0.26	5.80	0.34
33	6.24	6.20	0.04	5.73	0.51
35*	6.41	6.88	-0.46	6.18	0.23
36	6.80	6.80	0.00	6.47	0.34
37*	6.33	6.86	-0.53	6.27	0.05
39	7.09	7.08	0.01	7.45	-0.36
40	7.62	7.66	-0.04	7.33	0.29
41	6.87	7.05	-0.18	6.90	-0.03
42*	7.08	7.54	-0.46	6.97	0.11

43	6.39	6.29	0.10	6.16	0.23
44	6.40	6.41	-0.01	6.43	-0.03
45*	6.50	6.29	0.22	6.03	0.47
46	7.66	7.57	0.09	7.50	0.16
47*	6.91	6.21	0.70	6.61	0.30
48	7.01	6.99	0.03	6.43	0.58

* Test set compounds

Table 9: The predicted ADMET values and synthetic accessibility values for the 8 designed compounds.

Compounds	Absorption	Distribution	Metabolism							Excretion	Toxicity	Synthetic Accessibility
	Intestinal absorption (human)	VD _{ss} (human)	CYP							Total Clearance	AMES toxicity	
			2D6	3A4	1A2	2C19	2C9	2D6	3A4			
			substrate			inhibitor						
Numeric (% Absorbed)	Numeric (log L/kg)	Categorical (Yes/No)							Numeric (log ml/min/kg)	Categorical (Yes/No)		
D18	89	0.2	Yes	Yes	No	Yes	Yes	No	Yes	0.9	No	3.8
D23	98	-0.5	Yes	Yes	Yes	Yes	No	Yes	No	0.8	No	4.4
D25	87	-0.4	Yes	Yes	No	No	No	No	No	0.8	No	4.0
D28	92	-0.7	Yes	Yes	Yes	Yes	Yes	No	Yes	0.8	No	3.9
D32	93	-0.8	Yes	Yes	Yes	Yes	Yes	No	No	0.6	No	3.8
D39	88	-1.2	Yes	Yes	Yes	Yes	No	No	No	0.5	Yes	3.5
D44	87	0.1	Yes	Yes	Yes	Yes	Yes	No	No	0.8	No	3.5
D45	86	0.2	Yes	Yes	Yes	Yes	Yes	No	Yes	0.9	No	3.6

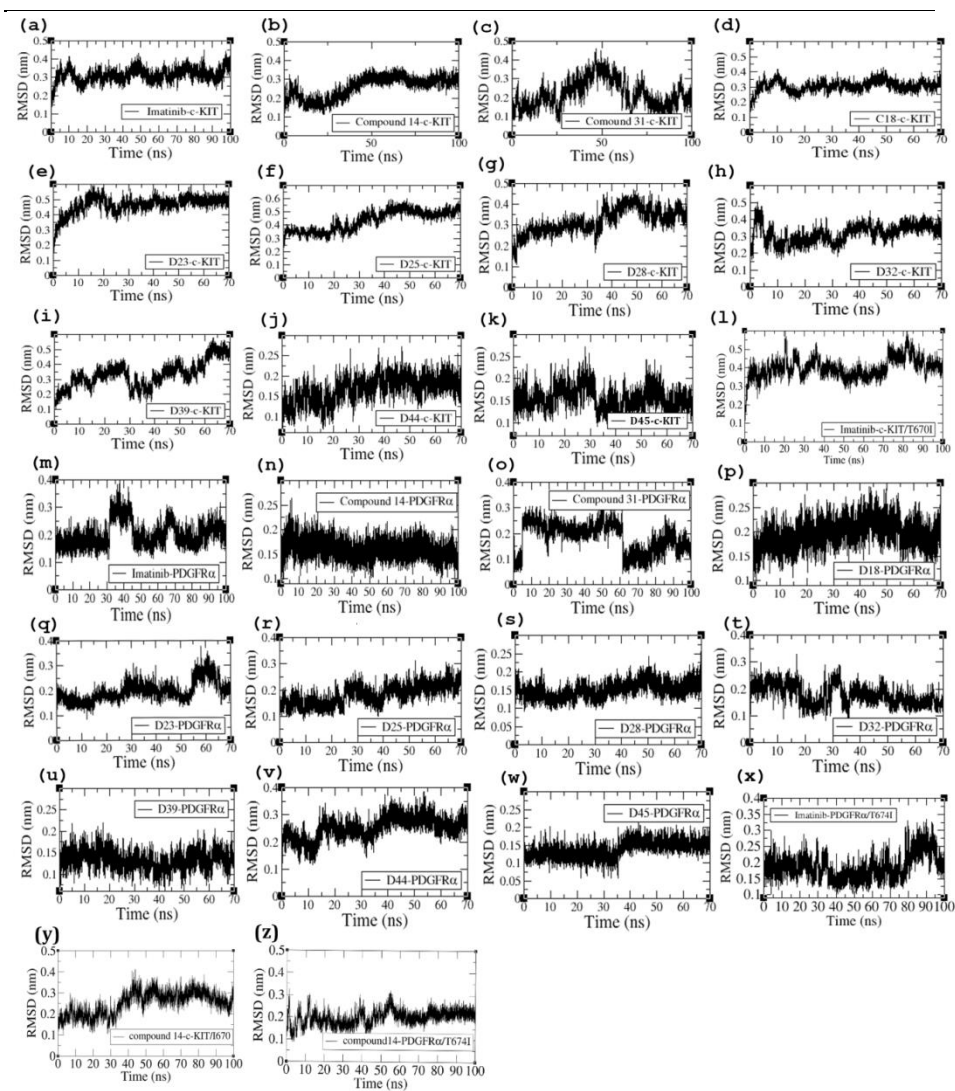


Figure 7. The ligand RMSD plots from the MD simulations. (a) Imatinib-c-KIT (b) Compound 14-c-KIT (c) Compound 31-c-KIT (d) D18-c-KIT (e) D23-c-KIT (f) D25-c-KIT (g) D28-c-KIT (h) D32-c-KIT (i) D39-c-KIT (j) D44-c-KIT (k) D45-c-KIT (l) Imatinib-c-KIT/T670I (m) Imatinib-PDGFR α (n) Compound 14-PDGFR α (o) Compound 31-PDGFR α (p) D18-PDGFR α (q) D23-PDGFR α (r) D25-PDGFR α (s) D28-PDGFR α (t) D32-PDGFR α (u) D39-PDGFR α (v) D44-PDGFR α (w) D45-PDGFR α (x) Imatinib-PDGFR α /T674I (y) Compound 14-c-KIT α /T670I (z) Compound 14-PDGFR α /T674I

Abbreviations

S	Steric
E	Electrostatic
H	Hydrophobic
ADMET	Absorption, distribution, metabolism, excretion and toxicity
BE	Binding Energy
BS SD	Bootstrap standard deviation
CADD	Computer-aided drug discovery
c-KIT	Stem cell factor receptor
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity indices analysis
FDA	Food and Drug Administration
FGFR	Fibroblast growth factor receptor
FLT3	Fms like tyrosine kinase 3
GISTs	Gastrointestinal stromal tumors
ICC	Interstitial cells of Cajal
MD	Molecular dynamics
MM/PBSA	Molecular mechanics energies combined with the Poisson–Boltzmann and surface area continuum solvation
ONC	Optimal number of components
PDB	Protein data bank
PDGFR α	Platelet derived growth factor receptor alpha
RAF1	Rapidly accelerated fibrosarcoma 1
RET	Rearranged during transfection
RMSD	Root mean square deviation

SASA	Solvent accessible surface area
SEE	Standard error of estimation
VEGFR	Vascular endothelial growth factor receptor
3D-QSAR	Three-dimensional quantitative structure-activity relationship

5. Conclusion

We performed modelling study of pyrazolopyridine derivatives which showed inhibitory activity for both c-KIT and PDGFR α . 3D-QSAR study was performed to understand the structural properties important for the dual inhibition of c-KIT and PDGFR α . Contour maps analysis showed that positive and bulky substituents are favorable near the meta and para position of compound 14 and may lead to an increase in activity against c-KIT and PDGFR α . Whereas, bulky substituents near the ortho position of reference compound 14 were not favored and could lead to steric clashes with binding site residues in both receptors. Comparative study of compound 14 and compound 31 (low activity value for c-KIT/PDGFR α) showed that the compound 14 was able to form additional interactions at the hydrophobic pocket formed by residues from the catalytic loop and the α C-helix. These interactions were not observed in the interactions of compound 31 with c-KIT/PDGFR α . The results suggested that possessing substituents that extended into the hydrophobic pocket could be crucial to increase the activity against c-KIT and PDGFR α . Based on the predicted activity values from the 3D-QSAR models, eight compounds were selected as potential c-KIT/ PDGFR α inhibitors. The eight designed compounds showed higher BE values against c-KIT/ PDGFR α than imatinib. Residue contact map analysis indicated that the T670I/T674I substitution in c-KIT/ PDGFR α led to conformational changes at the binding sites. BE calculation showed that, following the T670I/T674I substitution, the

activity values of imatinib against c-KIT/I670 and PDGFR α /I674 reduced to -72 kJ/mol and -67 kJ/mol respectively. Whereas, the activity value of compound 14 against c-KIT and PDGFR α were -131 kJ/mol and -126 kJ/mol suggesting that compound 14 was able to retain the activity values against both receptors after the T670I/T674I substitutions. The high activity values of compound 14 against both wild-type and mutant c-KIT/PDGFR α showed the potential of the pyrazolopyridine derivatives as c-KIT/PDGFR α inhibitors for the treatment of imatinib-resistant GIST. The outcome of this study could provide valuable insight into designing more potent dual c-KIT and PDGFR α inhibitors.

PART V

Rational Approach toward COVID-19 Main Protease Inhibitors via Molecular Docking, Molecular Dynamics Simulation and Free Energy Calculation

1. Introduction

The coronavirus disease (COVID-19) is an acute respiratory tract disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and was first reported in December 2019 in Wuhan, China [135]. The disease was declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [136]. Since then, it has spread to 218 countries and has infected more than 25 million people and claimed the lives of 852,000 people until September 2, 2020. The common symptoms observed in COVID-19 patients include fever, cough, fatigue, shortness of breath, and loss of smell [137].

Epidemiological analyses have shown that the SARS-CoV-2 has a lower fatality rate (5%) but a higher transmissibility rate (2-2.5%) than those of the previously known coronaviruses Middle East respiratory syndrome (MERS) coronavirus (34.4% fatality, <1% transmissibility) and severe acute respiratory syndrome (SARS) coronavirus (9.5 % fatality, 1.7-1.9 % transmissibility)[138, 139]. Serial viral load analyses in COVID-19 patients using reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) indicated that peak viral load was observed during the first week of symptom onset, with a median viral shedding period of 20 days [140, 141]. Antibody production starts approximately 10 days after symptom onset. Cohort studies of COVID-19 patients associated old age, multiple organ dysfunction, and high blood coagulation activity on admission with increased odds of death.. Sepsis, respiratory failure, acute respiratory distress syndrome (ARDS), heart failure, and septic shock were the commonly observed complications among the cohorts[141, 142]. Chakraborty et al. have provided a detailed review regarding the diagnostic and proposed therapeutic options for COVID-19 treatment [143-146].

Following the outbreak of the COVID-19 pandemic, several drug candidates from the repository of existing drugs have been tested for activity against SARS-CoV-2 [147-149]. The Food and Drug Administration (FDA) has also created a special

emergency program, the Coronavirus Treatment Accelerated Program (CTAP), that has reviewed 270 trials and is currently monitoring more than 570 drug development programs in the planning stage [150]. A review of the currently available literature shows that several existing antiviral drugs that target the viral replicating mechanism are under investigation for the treatment of COVID-19. The list of antiviral drugs being tested for COVID-19 includes remdesivir, hydroxychloroquine, chloroquine, lopinavir, darunavir, baloxavir, imatinib, and Favipiravir [151]. Immunomodulating drugs that reduce inflammatory responses such as corticosteroids, tocilizumab, ruxolitinib, infliximab, acalabrutinib, and azithromycin are also under clinical investigation [152-155]. Various adjunctive drugs such as vitamins C and D and antithrombotics are also being considered for COVID-19 treatment [156-158].

The ritonavir-lopinavir drug combination (Kaletra) has been used for the treatment of hospitalized patients in China and its benefits have been noted by the WHO [159]. Phase 3 clinical trials are underway to evaluate the performance and safety of the influenza drug Favipiravir [160, 161]. Remdesivir, an RNA-dependent RNA polymerase inhibitor, has been identified as a potential therapeutic agent for COVID-19 based on in vitro studies of SARS-CoV-2 clinical isolates [162, 163]. The FDA issued an Emergency Use Authorization (EUA) for the emergency use of the drug following promising results from a placebo-controlled randomized clinical trial of remdesivir for COVID-19 treatment (<https://www.fda.gov/>). In a clinical study, remdesivir showed effectiveness in reducing the recovery time in COVID-19 patients. However, the drug did not contribute to significant improvement in survival rates, and the efficiency of the drug in reducing viral load in patients remained unclear. The FDA has also authorized the emergency use of the anti-malarial drugs chloroquine and hydroxychloroquine for COVID-19 treatment [164, 165]. However, the clinical efficacy of these drugs remains inconclusive.

Viruses that cause diseases in humans are known to encode one or more proteases that play important roles in the viral life cycle. Proteases are ideal drug targets for viral diseases as they are responsible for cleaving the viral polyprotein, thus continuing the viral replication process [166, 167]. Protease inhibitors have been used in combination drug therapy in diseases where the virus developed resistance by mutation. This strategy of using combination therapy to combat drug resistance has been successfully used in the treatment of viral diseases such as acquired immunodeficiency syndrome, in which protease inhibitors were used in combination with nucleoside reverse transcriptase inhibitors [168].

The SARS-CoV-2 replicase enzyme encodes two polyproteins, pp1a and pp1ab, that produce all functional polypeptide units responsible for replication and transcription. Polypeptides are released by the catalytic cleavage activity of 3CL^{pro} at various subsites of the polyproteins. This cleavage process is known to be conserved in 3CL^{pro} for all coronaviruses [169, 170]. Due to the important role of 3CL^{pro} in the viral replication process and the absence of a close homolog in humans, this protease has been regarded as a promising therapeutic target for COVID-19 treatment [171]. However, despite its potential, the quest for 3CL^{pro} inhibitors feasible for therapeutic use against COVID-19 has been unsuccessful so far.

Computer-aided drug discovery (CADD) methodologies have emerged as powerful tools in the drug discovery process and have been used over the last decade to identify protein inhibitors and to study protein-drug interactions and protein-protein interactions [119, 149, 172, 173]. Since the development of a candidate drug into an approved drug is a long and costly process, a combination of computational methodologies such as virtual screening, docking, molecular dynamics (MD) simulation, and binding free energy evaluation, serves as a promising alternative for identifying potential drug candidates from compound libraries [174]. Cava et al. studied the mechanism of the angiotensin-converting enzyme 2 (ACE2) and its co-expressed genes using gene expression profiles in

silico and suggested several interesting potential drug candidates for COVID-19 [175]. Wang et al. performed virtual screening of the approved drugs and of those that are in clinical trials and identified several existing drug candidates that showed high binding affinity against 3CL^{pro} [176]. Zhang et. al (2020) used in silico screening to identify potential SARS-CoV-2 inhibitors from a repository of traditional Chinese medicines [177]. Liang et al. performed MD simulation to demonstrate the binding stability of an α -ketoamide inhibitor inside the SARS-CoV-2 main protease [178].

In this rapidly evolving pandemic, repurposing existing drugs and evaluating commercially available inhibitors against the druggable targets of SARS-CoV-2 should be an effective strategy to accelerate the drug discovery process. Consequently, taking advantage of the availability of the X-ray crystal structure of 3CL^{pro} in complex with the inhibitor N3 (PDB code 6LU7)[170], we performed a docking-based virtual screening of the protease inhibitor database MEROPS[179] (<http://www.ebi.ac.uk/merops/>) to identify potential 3CL^{pro} inhibitors.

Molecular docking and dynamic simulations were carried out to study the binding interactions of the inhibitor compounds with 3CL^{pro} [180]. Binding energy calculations were performed using the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method to evaluate the binding affinity of the compounds and to identify residues important for binding with 3CL^{pro} [180][79]. The results of the modeling study were carefully analyzed to identify commercially available potential 3CL^{pro} inhibitors.

2. Methods

2.1. Data preparation

The X-ray crystal structure of 3CL^{Pro} in complex with the inhibitor N3 (PDB ID 6LU7) prepared by Jin et al. was used as the receptor for our study [170]. The heteroatoms and water molecules were removed from the protein file for further study.

A total of 2700 protease inhibitors were collected from the MEROPS database. MEROPS is a database of proteases and their inhibitors [179]. The two-dimensional (2D) structures provided in the simplified molecular-input line-entry system (SMILES) format and the PubChem IDs of the compounds were collected from the PubChem website (<https://pubchem.ncbi.nlm.nih.gov/>) using an in-house script. The 2D structures were converted to three-dimensional (3D) structures using the concord module in Sybyl-X 2.1 [181].

2.2. Virtual Screening

Docking-based virtual screening was performed using the Surflex-Dock module[182] in Sybyl-X 2.1 and the autodock vina [183] program to identify potential 3CL^{Pro} inhibitors. Since Surflex-Dock and autodock vina use different approaches in scoring the binding affinity of the compounds, using both methods increased the credibility of the virtual screening results.

2.2.1. Surflex-Dock

The protein structure was prepared by adding hydrogen atoms and assigning Amber 7FF99 atom types, followed by brief energy minimization. During ligand preparation, a general cleanup process was carried out by filling valences and removing duplicates and compounds that are not drug-like. A computational representation of the binding site, called the protomol, was generated based on the crystal ligand coordinate, as shown in Figure 1. The protomol was used to direct the initial placement of the ligands during docking. Virtual screening was carried out via the Surflex-Dock program using a molecular similarity-based search

engine. Binding interactions were evaluated using an empirical scoring function based on hydrophobic, polar, repulsive, entropic, and solvation energy terms.

2.2.2. Autodock vina

3D coordinates of the compounds bearing partial charges were generated and saved in the pdbqt format. The receptor coordinates and grid parameters were generated using autodock tools [184]. The virtual screening process and the analysis of the results were performed using in-house scripts that incorporated the autodock vina program. The binding energies of the compounds were analyzed and used to rank the compounds.

2.3. Molecular docking

Molecular docking was performed to evaluate the binding energy and to provide initial coordinates and topology parameters for the MD simulations. The docking procedure was validated by extracting the irreversible inhibitor N3 [170] (PDB ID 6LU7) and the α -ketoamide inhibitor 13b [185] (PDB ID 6Y2F) from the crystal structures and docking them back into the receptor.

During the molecular docking of the compounds, the binding pose of the selected compounds from the virtual screening was used as the input. Polar atoms were added to the protein and Kollman charges were added as partial charges. A grid box with dimensions 60×60×60 centered at the coordinates X=-10, Y=13, and Z=70 was used to represent the search area. The Lamarckian genetic algorithm (LGA) was used to perform the docking process, generating 100 conformations for each compound. Based on the binding energy and binding interactions with the receptor, a representative binding pose for the ligands was selected.

2.4 Molecular dynamics simulation

Classical MD simulations were carried out on selected compounds, using GROMACS 2019 [72], to evaluate their binding interactions with 3CL^{pro}. The protein-ligand complexes from the docking study were used for the MD simulation study. The ligands were parameterized with the general amber force field (GAFF) [74] using the Acypype program [75]. Protein topology and coordinate files were generated using the Amber99SB force field provided in GROMACS. The protein-ligand complex was contained in a dodecahedron and solvated with TIP3P water. Counter ions were added to neutralize the solvated system followed by quick energy minimization with the steepest descent minimization algorithm. This was followed by a restrained constant number of particles, volume, and temperature (NVT) ensemble equilibration for 500 ps and a constant number of particles, pressure, and temperature (NPT) ensemble for 1 ns equilibration. Thermodynamic properties such as pressure, density, potential energy, and temperature of the systems were monitored to ensure adequate equilibration before the production run. The particle mesh Ewald method was used to calculate the long-range electrostatics. Modified Berendsen thermostat and Parrinello-Rahman barostat were used for temperature and pressure coupling, respectively. Finally, unrestrained 50 ns production simulations were carried out for the systems at 310 K and 1 bar atmospheric pressure. The MD simulation procedure used here has been used in several protein-ligand interaction studies by our group and others [119, 173].

2.5. Calculation of binding free energy

The *g_mmpbsa* package developed by Kumari et al. (2014) was used to calculate the binding free energy or simply, the binding energy (BE) of the protein-ligand complexes. The *g_mmpbsa* program used subroutines sourced from GROMACS and APBS packages to integrate high-throughput molecular dynamics simulation with binding energy calculations [79].

The vacuum potential energy was calculated from the bonded and non-bonded interactions based on the molecular mechanics (MM) force field. The electrostatic and van der Waals (E_{vdw}) energy contributions were calculated based on Coulomb potential and Lennard-Jones potential functions, respectively. During the evaluation of the free energy of solvation, the polar contribution was calculated by solving the Poisson-Boltzmann equation. The non-polar contribution was calculated based on the assumption that the non-electrostatic solvation energy is linearly related to the solvent-accessible surface area (SASA). The non-polar energy term ($G_{nonpolar}$) includes both repulsive and attractive forces between the solute and solvent developed due to cavity (G_{cavity}) formation as well as the van der Waals interaction (G_{vdw}). This can be represented by the equation below [173, 186].

$$G_{nonpolar} = G_{cavity} + G_{vdw}$$

During the calculation of the BE, snapshots were generated from the equilibrated region of the MD trajectory. Energy components were evaluated for 51 snapshots extracted every 0.1 ns from the trajectory. The decomposition of the energy term to individual residues was carried out using the `MmPbsaDecomp.py` script provided with the *g-mmpbsa* package. The default parameters set by Kumari et al. were used for all the calculations.

3. Results

The X-ray structures of the irreversible inhibitor N3 [170] and the α -ketoamide inhibitor 13b [185] in complex with 3CL^{pro} were retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) database. In the cell-based study by Jin et al. [170], N3 showed inhibitory activity against SARS-CoV-2 with a half-maximal effective concentration (EC_{50}) value of 16.77 μ M. However, N3 covalently binds to 3CL^{pro} as an irreversible inhibitor, and its half-maximal inhibitory concentration (IC_{50}) value could not be determined. The α -ketoamide

inhibitor 13b showed an IC_{50} value of $0.67 \mu\text{M}$ for purified recombinant SARS-CoV-2 main protease $3CL^{\text{pro}}$ and also showed inhibitory activity against COVID-19 with an EC_{50} value of 4 to $5 \mu\text{M}$ in human Calu-3 cells infected with SARS-CoV-2[185]. The moderate inhibitory activity values of the existing inhibitors necessitate the development of high-affinity $3CL^{\text{pro}}$ inhibitors.

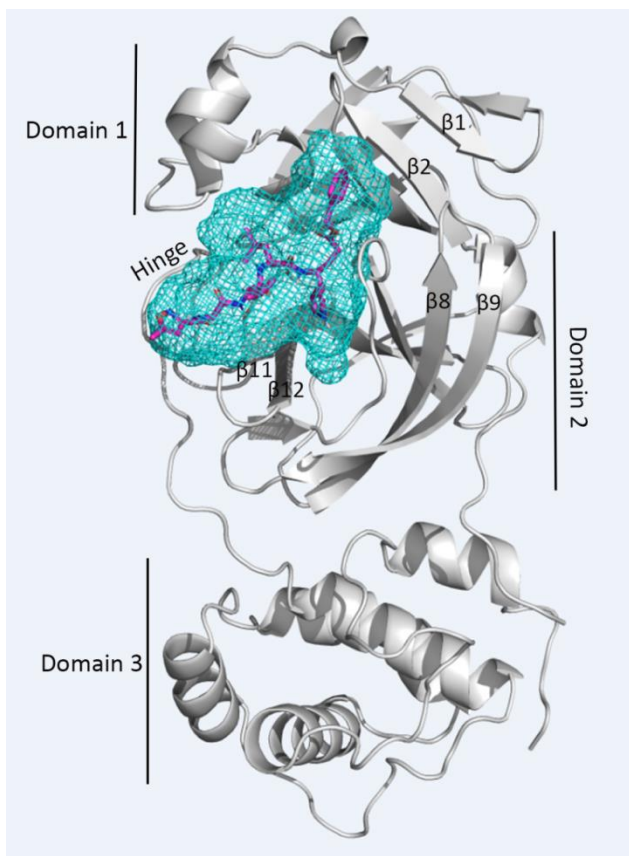


Figure 1. Structure of the inhibitor **N3** in complex with $3CL^{\text{pro}}$ as given in **6LU7**. The protein and ligand were shown in gray and magenta color. The mesh (cyan) representation between domain 1 and domain 2 represents the area to be searched by surflex dock (protomol).

The 3CL^{PRO} protease consists of three domains: domain 1 (residues 3-99), domain 2 (residues 100-182), and domain 3 (residues 199-307), as shown in Figure 1 [185]. Domains 1 and 2 comprise six-stranded antiparallel β -barrels with the substrate binding site at the intersection of the two domains. As shown in Figure 2 (a), the binding site is made up of subsites S1, S2, S3, S4, and S1', which are represented based on the binding position of the substrate polyprotein [170]. Domains 2 and 3 are connected by a hinge region (residues 182-198), which contributes to the formation of the S3 and S4 subsites. Domain 3 consists of five α -helices arranged in a globular cluster and regulates the dimerization of 3CL^{PRO}. The tight dimerization of 3CL^{PRO} is necessary for its catalytic activity, as it leads to crucial conformational changes at the S1 subsite and subsequent binding of the substrate.

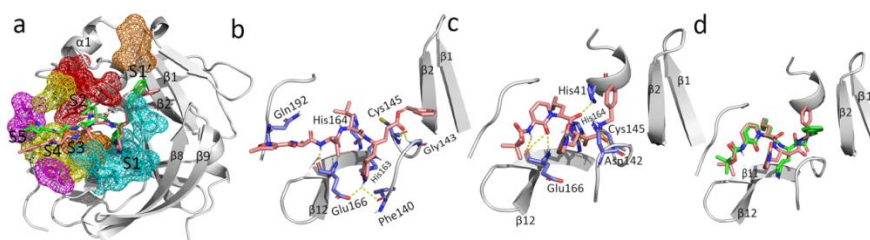


Figure 2. Showing the H-bond interactions of the inhibitors **N3** and **13b** with 3CL^{PRO} from the MD simulation studies. H-bond interactions were represented by yellow dotted lines and residues forming H-bonds were shown in purple color. **(a)** The overlap between the crystal ligand pose (green) and the MD binding pose (salmon) of **N3** at the binding site. The subsites S1 (cyan), S2 (red), S3 (yellow), S4 (magenta) and S1' (salmon) were shown in mesh representation. **(b)** Binding interactions between **N3** and 3CL^{PRO}. **(c)** Binding interactions between **13b** and 3CL^{PRO}. **(d)** The overlap between the crystal ligand pose (green) and the MD binding pose (salmon) of **13b** at the binding site of 3CL^{PRO}.

3.1. Virtual screening

The protease inhibitor dataset consisting of 2700 compounds was retrieved from the MEROPS database [179]. Using an in-house script, we retrieved the PubChem IDs and simplified molecular-input line-entry system (SMILES) structures of the compounds from the PubChem website. 3D structures were generated using the concord module in Sybyl-X 2.1 [181] and the Open Babel package [187]. Virtual screening of the protease inhibitor dataset was performed with the Surflex-Dock [182] program in Sybyl-X 2.1 and autodock vina. The Surflex-Dock program in Sybyl-X 2.1 uses a scoring function that includes hydrophobic, polar, repulsive, entropic, and solvation energy terms, whereas the autodock vina uses a scoring function based on steric, hydrophobic, and hydrogen bonding energy terms [182, 183]. The protein file was prepared by stripping the water molecules and other heteroatoms present in it and then converting the file to pdbqt file format. The methods and parameters used for virtual screening were validated by redocking the crystal ligands N3 (PDB ID 6LU7) and 13b (PDB ID 6Y2F) into the receptor.

The total binding score from Surflex-Dock and the binding energy from autodock vina were collected and used to rank the compounds. Based on the total score and binding energy, 32 compounds were selected and further studied using molecular docking, MD simulation, and free energy calculation methods. The Surflex-Dock binding scores and the autodock vina binding energies for the 32 compounds, including the reference compounds N3 and 13b, are presented in Table 1.

Table 1. The PubChem IDs, total scores (Surflex-Dock), autodock vina and autodock binding energies (kcal/mol), and MM-PBSA based binding energies (kJ/mol) of compounds N3, 13b, and the 32 selected compounds.

Sl. No	PubChem ID	Total Score (Surflex Dock)	Binding Energies (kcal/mol)		MM-PBSA Score (kJ/mol)
			Autodock Vina	Autodock	

1	53361968	7.1	-8.5	-11.0	-151
2	451415	7.0	-8.7	-8.6	-150
3	134815261	7.1	-8.6	-9.7	-133
4	15942730	10.9	-8.3	-8.8	-129
5	644196	10.1	-7.1	-7.7	-129
6	441243	8.6	-9.1	-9.4	-125
7	46178275	9.2	-8.1	-9.5	-123
8	9828551	7.1	-8.3	-10.2	-120
9	446837	7.9	-8.8	-10.8	-115
10	132531950	9.0	-7.8	-9.3	-114
11	102285029	10.3	-7.8	-8.8	-111
12	11962092	8.3	-8.7	-9.5	-108
13	446918	7.0	-8.5	-8.3	-108
14	92727	9.7	-8.4	-8.8	-104
15	45358152	9.9	-8.1	-7.4	-102
16	443119	9.1	-8.1	-6.4	-98
17	134691740	7.0	-8.9	-10.2	-97
18	5492607	10.5	-8.5	-8.0	-96
19	121304016	6.3	-7.9	-7.9	-94
20	447216	7.1	-8.7	-9.4	-92
21	103535	10.5	-8.3	-9.2	-90
22	6918046	10.7	-8.1	-7.4	-90
23	6324659	9.3	-8	-8.7	-88
24	213039	7.5	-8	-9.1	-81
25	5464035	9.3	-8.2	-9.6	-79
26	132585244	7.2	-8.6	-10.7	-77
27	134823859	7.0	-8.7	-10.0	-73
28	102207029	7.0	-8.9	-9.8	-63
29	21881944	7.0	-8.6	-8.1	3

30	4322	9.9	-8.3	-8.7	12
31	100997107	8.4	-8.4	-8.5	24
32	3451	7.0	-8.3	-9.8	4
Inhibitors from crystal structures					
33	N3 (6LU7)	10	-7.8	-6.3	-150
34	13b (6Y2F)	7.8	-9.7	-9.7	-99

3.2. Molecular docking

Molecular docking of the selected 32 compounds was performed to study the binding interactions and to provide initial coordinates of the protein-ligand complexes for subsequent MD simulation studies. The X-ray crystal structure of 3CL^{pro} (PDB ID **6LU7**) provided by Jin et al. (2020) was used as the receptor for this study. The docking protocol was validated by redocking the crystal ligands N3 (PDB ID **6LU7**) and 13b (PDB ID **6Y2F**) into the receptor. The docking showed that N3 formed H-bond interactions with residues His41, Asn142, Glu166, and Gln189 of 3CL^{pro}. Compound 13b showed interactions with Asn142, Gly143, Ser144, His163, and Glu166 with the binding site residues of 3CL^{pro}. The binding interactions of the inhibitors with 3CL^{pro} are shown in Figure 3.

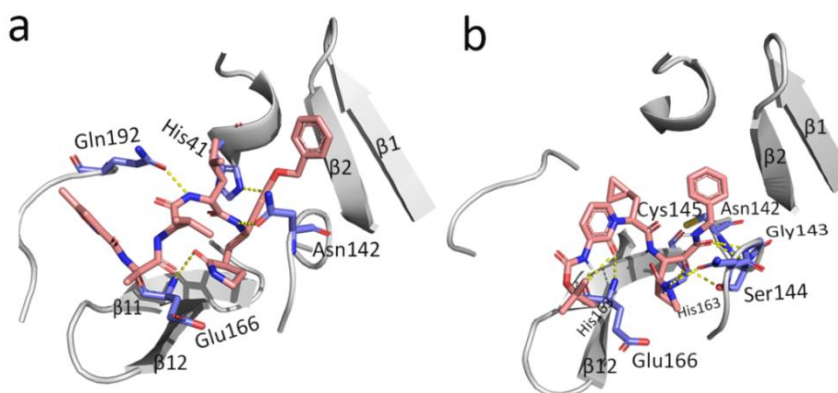


Figure 3. Showing the H-bond interactions of the inhibitors **N3** and **13b** with 3CL^{PRO} from the molecular docking studies. H-bond interactions were represented by yellow dotted lines and residues forming H-bonds were shown in purple color. **(a)** Binding interactions between **N3** and 3CL^{PRO}. **(b)** Binding interactions between **13b** and 3CL^{PRO}.

Both the X-ray structure and docked structure overlapped within a similar space inside the receptor. The docked pose of N3 overlapped with the pose in the X-ray crystal structure (PDB ID **6LU7**) at a root mean square deviation (RMSD) value of 2.6 Å, whereas the docked pose of 13b and X-ray structure (PDB ID **6Y2F**) showed an RMSD value of 1.8 Å.

The docking protocol used in docking the 3CL^{PRO} inhibitors was used to dock the selected 32 compounds. The resultant binding energy values of the 32 compounds are presented in Table 1. Binding conformations of the compounds were carefully selected based on the binding energy values and also based on important non-bonded interactions observed with 3CL^{PRO}. The protein-ligand complexes from the docking study were used as initial coordinates in the MD simulations.

3.3. Molecular Dynamics Simulation

GROMACS 2019 [72] was used to perform classical MD simulations of the selected 32 protein-ligand complexes to study the dynamic binding interactions of the compounds with 3CL^{PRO}.

For a comparative study, we also performed MD simulations of the N3-3CL^{PRO} complex (PDB ID **6LU7**) and the 13b-3CL^{PRO} complex (PDB ID **6Y2F**). The observed H-bond interactions and hydrophobic interactions are shown in Figure 1 and Figure 2, respectively. The X-ray structure of N3-3CL^{PRO} showed H-bond interactions with Phe140, Gly143, His160, Glu166, Glu189, and Thr190 [170]. The ligand N3 also formed a covalent bond with Cys145. However, this covalent bond with Cys145 was not observed in the MD simulation result because the

standard force field (AMBER99SB) cannot account for the formation of covalent bonds. The inhibitor N3 formed H-bond interactions with Phe140, Gly143, Cys145, His163, and His164 at the S1 subsite of 3CL^{pro}. H-bond interactions were also observed between N3 and the hinge residue Glu192 near the S4 subsite. Isopropyl moieties of N3 were seen at the adjacent hydrophobic subsites S2 and S3. The benzene moiety of N3 was observed near the S1' subsite, as seen in the crystal structure (PDB ID 6LU7). The binding interactions of N3 with 3CL^{pro} are shown in Figure 2 (b). Similar binding patterns were also observed in the α -ketoamide compound 13b, where the moiety between the benzene ring and the pyridone ring formed H-bond interactions with His41, Asn142, Cys145, and His164 at the S1 subsite. The benzene ring of 13b also formed hydrophobic interactions with residues at the S1' subsite. The pyridone ring formed H-bond interactions with Glu166 of β 11 and the cyclopropyl moiety extended into the small hydrophobic subsite S2. The tert-butyloxycarbonyl protecting (Boc) group of 13b did not fully extend into the S4 subsite to form interactions with the hinge residues, as observed in the N3-3CL^{pro} complex, and instead formed hydrophobic interactions with Leu167 and Pro168 at β 11. The binding interactions observed in the MD simulation of 13b-3CL^{pro} are shown in Figure 2 (c). The least-square fit RMSD of the ligands N3 and α -ketoamide 13b during the 50 ns simulations are shown in Figure 7. When compared with the respective crystal ligand poses, compounds N3 (Figure 2a) and 13b (Figure 2d) showed RMSD values of 1.5 Å and 2 Å, respectively. Similarly, the dynamic binding interactions of the 32 compounds with 3CL^{pro} were also studied. The trajectories from the MD study were used to evaluate the free energy of binding for the selected 32 compounds. The hydrophobic interactions observed from the MD simulations of N3 and 13b are shown in Figure 4.

3.4. Calculation of binding free energy

MM-PBSA based binding energy (BE) calculations were performed for the selected 32 protein-ligand complexes, followed by evaluation of the energy

contribution of the individual residues. For a comparative study, we also calculated the BE and the BE distribution for both, N3-3CL^{PRO} and 13b-3CL^{PRO} complexes. Compounds N3 and 13b showed BE values of -150 kJ/mol and -99 kJ/mol, respectively. Calculation of the BE distribution identified residues that contributed highly to the total BE, as shown in Figure 5. The binding site residue Met165 from the S2 subsite showed the highest BE contribution, that may be attributed to the hydrophobic interaction observed with the compounds N3 and 13b. Pro168 at β 11 also showed a high BE contribution, that may be attributed to the hydrophobic interaction with the methylisoxazole of N3 and the Boc group of 13b.

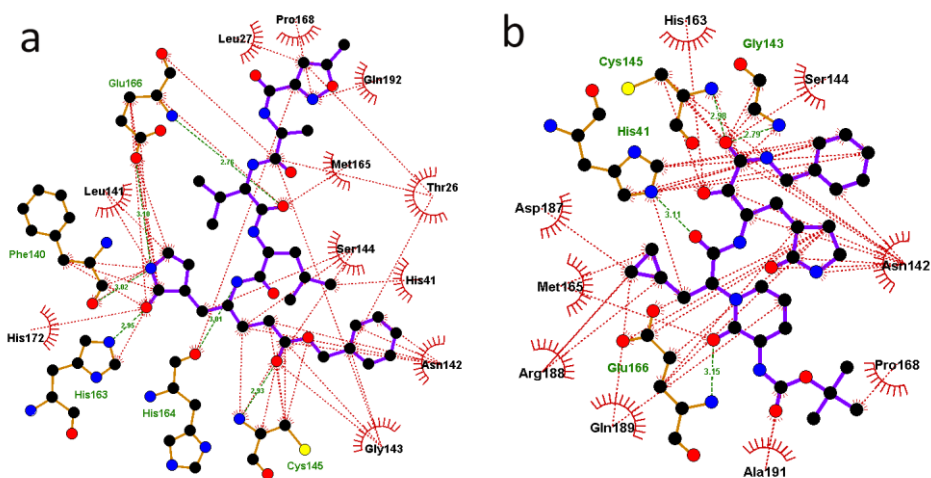


Figure 4. Showing the hydrophobic interactions of the inhibitors **N3** and **13b** with 3CL^{PRO}. Hydrophobic interactions were represented in red dotted lines. Residues that showed Hydrophobic and H-bond interactions were given in green and black label respectively. Carbon, oxygen, nitrogen and sulfur atoms were shown in black, red, blue and yellow color respectively. **(a)** N3-3CL^{PRO} complex **(b)** 13b-3CL^{PRO} complex.

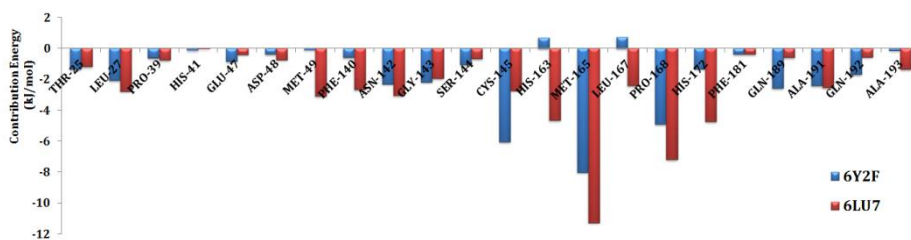


Figure 5. Residues that showed high contribution to the total binding energy during the MD simulation of **N3-3CL^{pro}** (**6LU7**) and **13b-3CL^{pro}** (**6Y2F**) complex. The residues from the **N3-3CL^{pro}** complex and **13b-3CL^{pro}** complex simulations were shown in red and blue color respectively.

The residues from the S1 subsite, namely Phe140, Asn142, Gly143, Ser144, and Cys145, and the residues from β 11, namely His163, His164, Met165, Leu167, and Pro168, were involved in both H-bond interaction and hydrophobic interactions in N3-3CL^{pro} as well as 13b-3CL^{pro} complexes. Consequently, these residues showed relatively high BE contributions (Figure 4). Trajectory analyses also showed that the binding interactions with S1, S2, and β 11 were stable throughout the simulations. However, interactions at the S1' and S4 subsites were transient, resulting in flexible movement as indicated by the flipping of the benzene ring at the S1' subsite and the movement of Boc and methylisoxazole at the S4 subsite.

Following the calculation of the BE for the 32 compounds, 16 compounds showed total BE values higher than the BE of compound 13b (-99 kJ/mol), suggesting potential inhibitory activity for 3CL^{pro}. Compounds 53361968 (-151 kJ/mol) and 451415 (-150 kJ/mol) showed higher BE values than the potent inhibitor N3. We also observed that 12 compounds showed BE values in the range of -98 kJ/mol and -63kJ/mol, suggesting a moderate binding affinity with 3CL^{pro}. Compounds 21881944, 4322, 100997107, and 3451 showed positive BE values, possibly due to non-converging simulations. The total BE values of the compounds are presented in Table 1. Based on the MM-PBSA based BE evaluations, the residue energy contributions of 10 protein-ligand complexes with

high binding affinity were analyzed, as shown in Table 3. Analysis of the energy decomposition results for the selected 10 compounds showed that residues Thr25, Leu27, His41, Asp48, Met49, Leu50, Leu141, Cys145, His164, Met167, Pro168, Asp187, Gln189, and Ala191 play important roles in the binding of the compounds with 3CL^{PRO}. The interactions with these residues were dominated by electrostatic and hydrophobic interactions (Table 2).

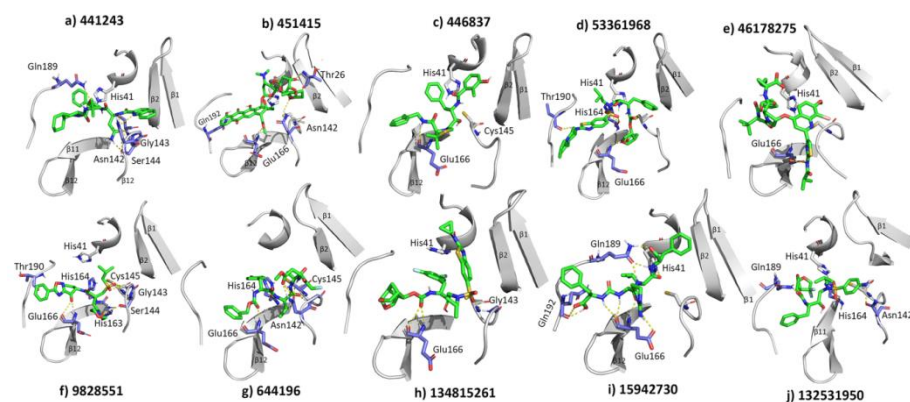


Figure 6. Showing the H-bond interactions of the inhibitors with 3CL^{PRO}. The protein and ligand and ligand were shown in gray and green color respectively. H-bond interactions were represented by yellow dotted lines and residues forming H-bonds were shown in purple color. **(a) 441243-3CL^{PRO} (b) 451415-3CL^{PRO} (c) 446837-3CL^{PRO} (d) 53361968-3CL^{PRO} (e) 46178275-3CL^{PRO} (f) 9828551-3CL^{PRO} (g) 644196 -3CL^{PRO} (h) 134815261-3CL^{PRO} (i) 15942730-3CL^{PRO} and (j) 132531950-3CL^{PRO}.**

4. Discussion

We selected 10 compounds that showed high potential for 3CL^{PRO} inhibition based on the total binding free energy to analyze the structural features critical for binding with 3CL^{PRO}. The structures of the selected compounds are presented in Table 4. The binding interactions with 3CL^{PRO} and the RMSD values for the 10 compounds are shown in Figure 6 and Figure 7, respectively.

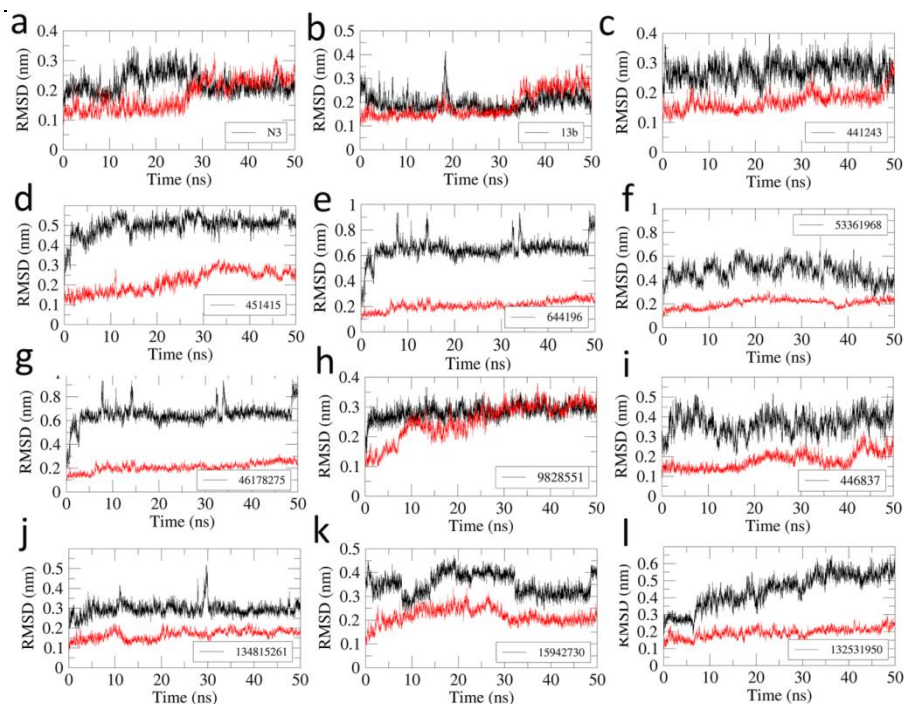


Figure 7. Least Square Fit Root Mean Square Deviation (RMSD) of the protein (red) and ligand (black) from the 50 ns MD simulations. (a) N3-3CL^{pro} (b) 13b-3CL^{pro} (c) 441243-3CL^{pro} (d) 451415-3CL^{pro} (e) 644196-3CL^{pro} (f) 53361968-3CL^{pro} (g) 46178275-3CL^{pro} (h) 9828551-3CL^{pro} (i) 446837-3CL^{pro} (j) 134815261-3CL^{pro} (k) 15942730-3CL^{pro} and (l) 132531950-3CL^{pro}.

Compound 441243 formed H-bond interactions with Gln189 and His41, while forming multiple interactions at the S1 subsite with Asn142, Ser144, and Gly143. Compounds 451415 (-150 kJ/mol) and 53361968 (-151 kJ/mol), which showed relatively high binding energy values (Table 2), had a relatively less number of H-bond interactions. However, further analysis showed that these compounds had relatively high hydrophobic energy contributions, resulting in higher total binding energy values. The donor nitrogen atoms of compound 446837 formed two H-bond interactions: with His41 and Glu166. Compound 46178275 showed only one stable H-bond interaction with Glu166. However, the oxygen and donor nitrogen atoms of 46178275 near the hinge region could form transient interactions with Gln189 and Thr190. Compound 15942730 showed the highest

number of H-bonds, forming multiple interactions with His41, Glu166, Gln189, and Gln192. In the binding energy analysis of the 15942730-3CL^{pro} complex (Table 2), the contribution of the electrostatic component to the total binding energy was -91 kJ/mol, which was higher than that of the other selected compounds. Compounds 9828551 and 644196 formed several H-bond interactions with residues from β 11 and S1. Consequently, these two compounds showed high electrostatic energy terms in the binding energy calculations. Compound 134815261 formed H-bond interactions with Glu166 and Gly143. Compound 13231950 formed H-bond interactions with His41, Asn142, His164, and Gln189. The analyses suggested that compounds showing higher binding affinities with 3CL^{pro} were able to form H-bond interactions with residues from multiple subsites and also showed higher number of hydrophobic interactions.

Table 2. Energy contributions of the various energetic terms to the total binding energies of the inhibitors with 3CL^{pro}.

Complexes	Van der Waals (kJ/mol)	Electrostatics (kJ/mol)	Polar solvation (kJ/mol)	Non-polar (kJ/mol)	Total Binding Energy (kJ/mol)
441243-3CL^{pro}	-256	-81	238	-26	-125
451415-3CL^{pro}	-293	-44	217	-29	-150
446837-3CL^{pro}	-217	-60	185	-23	-115
53361968-3CL^{pro}	-276	-33	184	-26	-151
46178275-3CL^{pro}	-213	-38	152	-24	-123
9828551-3CL^{pro}	-215	-72	188	-21	-120

644196-3CL^{pro}	-231	-88	214	-24	-129
134815261-3CL^{pro}	-262	-54	210	-27	-133
15942730-3CL^{pro}	-239	-91	225	-24	-129
132531950-3CL^{pro}	-254	-53	219	-26	-114
N3-3CL^{pro}	-300	-108	286	-27	-150
13b-3CL^{pro}	-229	-78	232	-23	-99

Table 3. Residues with a high contribution to the total binding energy during the MD simulations of the complexes 441243-3CL^{pro}, 451415-3CL^{pro}, 446837-3CL^{pro}, 53361968-3CL^{pro}, 46178275-3CL^{pro}, 9828551-3CL^{pro}, 644196-3CL^{pro}, 134815261-3CL^{pro}, 15942730-3CL^{pro}, and 132531950-3CL^{pro}. The energy values of the residues are in kJ/mol.

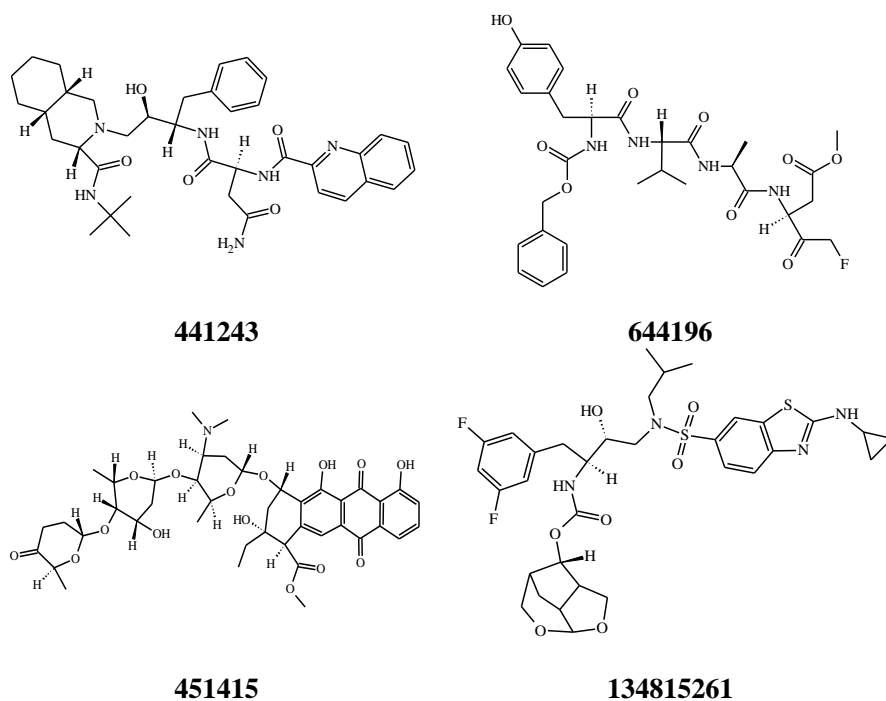
Residues	441243	451415	446837	134815261	53361968	15942730	46178275	9828551	132531950	644196
THR25	-1.16	-4.25	-0.89	-2.02	-1.56	-2.07	-3.03	-0.16	-1.94	-0.38
Leu27	-2.43	-3.88	-0.55	-0.93	-2.43	-1.05	-2.41	-1.79	-3.16	-3.69
His41	-1.31	0.11	-4.13	-0.36	-8.03	-4.21	1.78	-0.79	-4.73	-1.39
Cys44	-0.23	-0.04	0.39	0.22	-1.35	2.46	-0.08	-0.10	-0.31	-0.20
Asp48	-0.28	-0.92	-0.56	-2.22	-1.63	-1.45	-2.28	-1.81	-1.68	-0.97
Met49	-5.81	-6.99	-5.29	-6.89	-6.68	-6.14	-5.97	-2.97	-7.65	-2.37
Leu50	-0.51	-0.49	-0.63	-0.43	-1.24	-0.69	-0.40	-0.29	-0.75	-0.12
Leu141	-2.01	-0.67	-0.49	-0.52	-1.76	-0.36	-3.69	-2.67	-0.30	-2.68
Cys145	-6.26	-3.84	-1.70	-2.14	-4.28	-4.19	-1.78	-2.51	-4.35	-6.67
His164	-5.05	-2.89	0.57	4.14	-0.62	1.41	-0.27	-1.25	2.90	-5.98
Met165	-9.01	-	-9.62	-	-	-11.13	-2.59	-6.64	-	-
		11.29		12.06	10.39				10.07	11.23
Leu167	-2.28	-3.61	-3.40	-4.74	-1.49	-2.89	-0.61	-0.85	-2.00	-0.56
Pro168	-1.54	-2.88	-5.76	-4.07	-3.42	-7.68	-0.33	-0.36	-1.11	-1.04

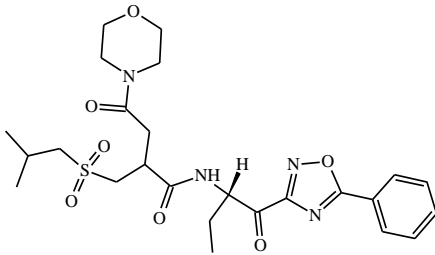
Asp187	0.02	-2.55	-0.41	0.09	-2.61	-2.03	-2.04	-2.37	-3.70	-2.14
Gln189	-7.24	-0.09	-3.47	-2.68	-6.20	-3.55	-3.45	-1.92	-8.39	1.01
Thr190	2.46	-0.28	2.41	0.31	-3.98	-2.89	-1.10	-1.74	2.29	0.27
Ala191	-0.98	-0.81	-0.95	-1.12	-3.09	-2.07	0.00	-3.92	-1.04	-2.02

From the analyses of the binding interactions, we observed that the interactions of the compounds with the S1 subsite residues such as His41, Asn142, Gly143, Ser144, and the Glu166 residue of β 11 were crucial for stable interaction with 3CL^{pro}. These interactions with the S1 and β 11 residues were also observed in experimental studies [170, 185]. Interactions of the compounds at the S2 subsite were predominantly hydrophobic. Since the S2 subsite is a small hydrophobic pocket, compounds with substituents such as isopropyl and cyclopropyl, which can fit into the hydrophobic pocket, could be promising 3CL^{pro} inhibitors, as in the cases of inhibitor 13b (Figure 2b), compound 451415/aclarubicin, and 53361968/TMC-310911 [188]. It was also observed that compounds with substituents that extend into the S4 subsite tend to show higher binding energies. The compounds 451415/aclarubicin [189] and 53361968, which formed H-bonds (Figure 6b and 6d) and hydrophobic interactions at the S4 subsite showed relatively high total BE of -150 kJ/mol and -151 kJ/mol, respectively, suggesting the importance of these interactions in 3CL^{pro} inhibition. This observation was also made in the experimental study of inhibitor 13b, wherein removing substituents that extended into the S4 subsite reduced the activity value [185]. Having substituents that extend into the S4 subsite induced conformational changes at the hinge region between subunits 1 and 2, as noted by Zhang et al. (2020) [185]. However, the exact mechanism behind the conformational change leading to increased affinity for 3CL^{pro} remains unclear. From interaction studies, it was observed that having substituents that form hydrophobic interactions at the S1' subsite was important for binding with 3CL^{pro}. These interactions at S1' were dominated by hydrophobic interactions, as observed in the interactions of N3 and 13b (Figure 4) and also in the cases of compounds 446837/KNI-764 [190] (-115 kJ/mol), 53361968 [188] (-150 kJ/mol), and 15942730/chemostatin [191] (-129 kJ/mol). However, in the case of compound 451415, that lacks an extended

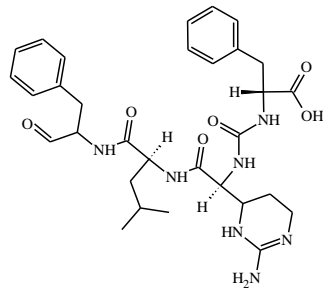
hydrophobic benzyl substituent, H-bond interaction was observed with Thr26 at the S1' subsite (Figure 6b). From these observations, we speculate that having substituents that can form hydrophobic and H-bond interactions with S1' residues may increase the binding affinity since having both hydrophobic and H-bond interactions with 3CL^{pro} closely emulates the substrate-binding pattern [192]. Binding energy decomposition for individual residues identified His41, Met49, Met165, and Glu189 as key locations. These residues were also identified as important hotspot residues by Wang et al. [176] in a recent study. Additionally, analysis of BE decomposition also revealed that the residues Thr25, Leu27, Asp48, Leu50, Leu141, Cys145, His164, Leu167, Pro168, Asp187, and Ala191 were significant for the binding of the inhibitors with 3CL^{pro}.

Table 4. Structures of the 10 compounds selected on basis of binding energy (MM-PBSA)

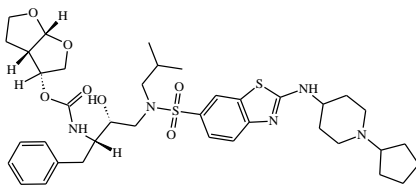




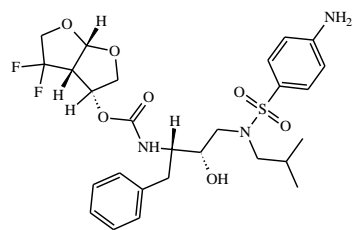
9828551



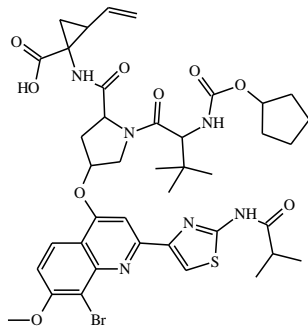
15942730



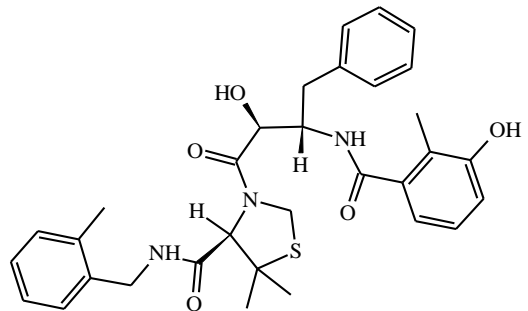
53361968



132531950



46178275



446837

Table 5: The predicted ADMET values for the 10 selected compounds.

Compounds	Absorption	Distribution	Metabolism							Excretion	Toxicity
	Intestinal absorption (human)	VDss (human)	CYP							Total Clearance	AMES toxicity
			2D6	3A4	1A2	2C19	2C9	2D6	3A4		
	Numeric (% Absorbed)	Numeric (log L/kg)	Categorical (Yes/No)							Numeric (log ml/min/kg)	Categorical (Yes/No)
substrate			inhibitor								
441243	57	1.0	No	Yes	No	No	No	No	Yes	0.3	No
451415	84	1.4	No	Yes	No	No	No	No	Yes	0.9	No
446837	63	0.7	No	Yes	No	Yes	Yes	No	Yes	-0.1	No
53361968	82	1.9	No	Yes	No	No	No	No	Yes	0.2	No
46178275	53	-0.6	No	Yes	No	No	No	No	No	0.3	No
9828551	66	0.25	No	Yes	No	No	No	No	Yes	0.5	No
644196	39	0.3	No	Yes	No	No	No	No	Yes	1.0	No

134815261	92	0.85	No	Yes	No	No	No	No	Yes	0.4	No
15942730	17	-1.3	No	No	No	No	No	No	Yes	0.03	No
132531950	75	0.5	No	Yes	No	No	No	No	No	0.5	No

The absorption, distribution, metabolism, and excretion (ADMET) properties of the compounds were also evaluated using the pkCSM server[132] and the results are presented in Table 5. In the ADMET analyses, compounds that showed an intestinal absorption value of less than 30% were considered to have poor absorption rates. Except for compound 15942730, all the selected compounds showed reasonable intestinal absorption rates. Steady-state volume of distribution (VD_{ss}) represents the degree to which the compounds are distributed in the body rather than the plasma, and compounds with log (VD_{ss}) values greater than -0.15 are considered to have a reasonable distribution rate. All the compounds in Table 5 except 46178275 and 15942730 showed VD_{ss} values greater than -0.15, indicating that the compounds have satisfactory distribution rates. Analyses of the metabolism results suggest that the compounds are poor cytochrome P450 inhibitors. Compounds with positive results for the CYP3A4 substrate test suggest that they can be metabolized by cytochrome P450. The selected compounds also showed a reasonable total clearance rate from the body, except compound 446837. The negative Ames toxicity test results suggest that the compounds have poor mutagenic potential.

Being in the middle of the COVID-19 pandemic, availability of these compounds is crucial for in vivo and in vitro experimental studies. Hence, we checked the availability of the selected compounds in commercial libraries by referencing vendor data through the PubChem website and the ZINC database[193]. The details regarding the ZINC ID and the distributor (vendor) of the compounds are

provided in Table 6. Compound 441243/saquinavir [194] is an antiretroviral protease inhibitor approved by the FDA for the treatment of human immunodeficiency virus (HIV) infection. Recently, several computational studies have also reported the encouraging binding ability of saquinavir with 3CL^{pro} [159, 195]. The high binding affinity of saquinavir observed in our study, as well as other by independent research groups, indicates its potential as a 3CL^{pro} inhibitor. The registry of clinical trials maintained by the United States National Library of Medicine under the National Institute of Health (NIH) showed that compound 451415/aclarubicin is an anthracycline drug and has been under evaluation (phase 2 clinical trial) for combination therapy against acute myeloid leukemia (AML). Compound 53361968, which showed the highest BE value (-151 kJ/mol), is an investigational protease inhibitor that is currently being studied for HIV-1 infection treatment [196]. Compound 46178275/faldaprevir is a hepatitis C virus protease inhibitor currently being studied for the treatment of hepatitis C [197]. Additionally, compounds 132531950 (-114 kJ/mol), 102285029 (-111 kJ/mol), 11962092 (-108 kJ/mol), 446918 (-108 kJ/mol), 92727 (-104 kJ/mol), and 45358152 (-102 kJ/mol) also showed BE values greater than that of the inhibitor in the X-ray structure (α -ketoamide with a BE value of -99 kJ/mol), suggesting that these compounds may have higher inhibitory activity against 3CL^{pro}.

Table 6. The binding energies, ZINC compound IDs, and the distributor/vendor names and vendor compound IDs of the 10 compounds selected on basis of high MM-PBSA-based binding energy evaluation.

PubChem ID	Binding Energy (kJ/mol)	ZINC Compound ID	Vendor	Vendor Compound ID
441243 (saquinavir)	-125	<u>ZINC3914596</u>	Molport	MolPort-000-883-824
451415 (aclarubicin)	-150	<u>ZINC8101053</u>	Molport SC Economica I	MolPort-004-845-383
446837	-115	ZINC3941126	eMolecules	92333721

(KNI-764)				
53361968 (TMC-310911)	-151	<u>ZINC98208561</u>	Synblock Inc.	SB17102
46178275 (faldaprevir)	-123	<u>ZINC15033914</u> 5	Compound Cloud	42601552, 56594927
9828551	-120	<u>ZINC13729397</u> 8	eNovation Chemicals	D676701
644196	-129	ZINC78938888	Ambinter	Amb1993041 1
134815261	-133	NA	NA	NA
15942730 (chemostatatin)	-129	<u>ZINC3947583</u>	NA	NA
132531950	-114	<u>ZINC22469939</u> 9	NA	NA

*NA represents not available.

5. Conclusion

In this ongoing COVID-19 pandemic, CADD methodologies can be used effectively to accelerate the process of developing therapeutic agents for the treatment of this disease. In this study, we used docking-based virtual screening to search the protease inhibitor database (MEROPS) to identify potential inhibitors of the SARS-CoV-2 main protease 3CL^{pro}. Molecular docking and dynamics simulations were carried out to study the binding interactions. Binding free energy calculations were performed to identify potential 3CL^{pro} inhibitors. The study identified saquinavir, which is an approved drug for HIV-1 treatment, and several other investigational drugs, such as aclarubicin, TMC-310911, and faldaprevir. We also assessed the commercial availability of the compounds, which could be useful for experimental researchers. Analysis of the binding interactions revealed that electrostatic interactions with residues from the S1 subsite and the β –strand (β 11) were important for the inhibition of 3CL^{pro}. Compounds possessing substituents that extend into the S4 subsite induced conformational changes at the hinge between subunit 1 and subunit 2 and showed

higher binding affinity. Compounds with high binding energies showed either hydrophobic or electrostatic interactions at the S1' subsite. These structural features may be harnessed to design potent 3CL^{Pto} inhibitors. Using CADD methods, we identified 15 compounds with a binding affinity greater than that of the inhibitor inside 3CL^{Pto} in the X-ray structure (α -ketoamide). We suggest further experimental investigation of these compounds.

Seketoulie Keretsu Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

PART VI

Conclusion

Conclusion

Protein kinases play an important role in mediating various signaling processes and their over-activation has been linked to the pathogenesis of various cancers and diseases. This has made kinases interesting targets for drug development. To this end, several kinases have been identified to be associated with the pathogenesis of various cancer types and efforts have been made by drug developers to design potent inhibitors for these kinases. Consequently, several small compounds and peptide-based inhibitors have been designed and tested for inhibition of kinases which resulted in the approval of 52 kinase inhibitors. However, the targeting kinases for pharmaceutical use are still limited due to the lack of selectivity of the inhibitors. Most inhibitors showed activity for multiple kinases, resulting in off-target effects. Hence, the design and development of selective and potent inhibitors constitute an interesting field in drug design.

We have performed a molecular modeling study of several dataset compounds that were reported in recent studies as kinase inhibitors. In the first study, we have used molecular docking, molecular dynamics simulation, and 3D-QSAR to understand the structural factors important for selective inhibition of GRK2 kinases. GRK2 is a protein kinase implicated in heart diseases, making it a target for drug design. MM/PBSA based method was used to calculate the binding energy between the ligand and the receptor. The study resulted in the identification of several structural elements that were important for the inhibition of GRK2. Further analysis also showed that compounds with substituents that extended into the hydrophobic pocket tend to show higher activity and selectivity for GRK2. Similarly, we have also performed a computational study of a series of c-KIT/PDGFR α inhibitors. The study showed that the inhibitors are bound with c-KIT and PDGFR α in a similar pattern. The study also suggested several structural guidelines that were important for dual inhibition of the two kinases.

Finally, virtual screening of the protease inhibitor database MEROPS for potential SARS-COV-2 main protease (3CL^{pro}) inhibitor was performed. The study resulted in the identification of several known protease inhibitors. Based on the binding affinity with 3CL^{pro} we selected 15 protease inhibitors as potential inhibitors of 3CL^{pro} inhibitors. The selected inhibitors include Saquinavir (an approved drug for HIV-1 treatment) and three other investigational drugs namely aclarubicin, TMC-310911, and Faldaprevir. Aclarubicin is an anthracycline drug used in cancer chemotherapy. TMC-310911 is an antiviral drug and Faldaprevir is an experimental drug under clinical trial for the treatment of hepatitis C disease. Further in vivo and in vitro experiment could provide promising insight into the pharmaceutical potential of the compounds for COVID-19 treatment.

Taken together, the computational modeling study provided various important structural clues for the development of potent kinases inhibitors. The studies also lead to the design of several compounds that showed promising predicted activity values against specific kinases. Further validation of the activity of the compounds is recommended.

REFERENCES

1. Pearson RB, Kemp BE: **[3] Protein kinase phosphorylation site sequences and consensus specificity motifs: Tabulations.** *Methods in enzymology* 1991, **200**:62-81.
2. Hubbard SR, Till JH: **Protein tyrosine kinase structure and function.** *Annual review of biochemistry* 2000, **69**(1):373-398.
3. Hanks SK, Quinn AM, Hunter T: **The protein kinase family: conserved features and deduced phylogeny of the catalytic domains.** *Science* 1988, **241**(4861):42-52.
4. Bull SC, Doig AJ: **Properties of protein drug target classes.** *PloS one* 2015, **10**(3):e0117955.
5. Cohen P: **Protein kinases—the major drug targets of the twenty-first century?** *Nature reviews Drug discovery* 2002, **1**(4):309-315.
6. Rask-Andersen M, Masuram S, Schiöth HB: **The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication.** *Annual review of pharmacology and toxicology* 2014, **54**:9-26.
7. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP: **Comprehensive analysis of kinase inhibitor selectivity.** *Nature biotechnology* 2011, **29**(11):1046-1051.
8. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR: **Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity.** *Nature biotechnology* 2011, **29**(11):1039-1045.
9. Bang SJ, Cho SJ: **Comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) study of mutagen X.** *BULLETIN-KOREAN CHEMICAL SOCIETY* 2004, **25**(10):1525-1530.
10. Lefkowitz RJ, Stadel JM, Caron MG: **Adenylate cyclase-coupled beta-adrenergic receptors: structure and mechanisms of activation and desensitization.** *Annual review of biochemistry* 1983, **52**(1):159-186.
11. Sutherland EW, Robison GA, Butcher RW: **Some aspects of the biological role of adenosine 3', 5'-monophosphate (cyclic AMP).** *Circulation* 1968, **37**(2):279-306.
12. Lymperopoulos A, Rengo G, Funakoshi H, Eckhart AD, Koch WJ: **Adrenal GRK2 upregulation mediates sympathetic overdrive in heart failure.** *Nature medicine* 2007, **13**(3):315.
13. Pearce LR, Komander D, Alessi DR: **The nuts and bolts of AGC protein kinases.** *Nature reviews Molecular cell biology* 2010, **11**(1):9.
14. Bouley R, Waldschmidt HV, Cato MC, Cannavo A, Song J, Cheung JY, Yao X-Q, Koch WJ, Larsen SD, Tesmer JJ: **Structural Determinants Influencing the Potency and Selectivity of Indazole-Paroxetine**

-
- Hybrid G Protein-Coupled Receptor Kinase 2 Inhibitors.** *Molecular pharmacology* 2017, **92**(6):mol. 117.110130.
15. Matkovich SJ, Diwan A, Klanke JL, Hammer DJ, Marreez Y, Odley AM, Brunskill EW, Koch WJ, Schwartz RJ, Dorn GW: **Cardiac-specific ablation of G-protein receptor kinase 2 redefines its roles in heart development and β -adrenergic signaling.** *Circulation research* 2006, **99**(9):996-1003.
 16. Waldschmidt HV, Homan KT, Cato MC, Cruz-Rodríguez O, Cannavo A, Wilson MW, Song J, Cheung JY, Koch WJ, Tesmer JJ: **Structure-based design of highly selective and potent G protein-coupled receptor kinase 2 inhibitors based on paroxetine.** *Journal of medicinal chemistry* 2017, **60**(7):3052-3069.
 17. Dzimiri N, Muiya P, Andres E, Al-Halees Z: **Differential functional expression of human myocardial G protein receptor kinases in left ventricular cardiac diseases.** *European journal of pharmacology* 2004, **489**(3):167-177.
 18. Guccione M, Ettari R, Taliani S, Da Settimo F, Zappalà M, Grasso S: **G-protein-coupled receptor kinase 2 (GRK2) inhibitors: current trends and future perspectives.** *Journal of medicinal chemistry* 2016, **59**(20):9277-9294.
 19. Montó F, Oliver E, Vicente D, Rueda J, Agüero J, Almenar L, Ivorra MD, Baretino D, D'Ocon P: **Different expression of adrenoceptors and GRKs in the human myocardium depends on heart failure ethiology and correlates to clinical variables.** *American Journal of Physiology-Heart and Circulatory Physiology* 2012, **303**(3):H368-H376.
 20. Boureux A, Vignal E, Faure S, Fort P: **Evolution of the Rho family of ras-like GTPases in eukaryotes.** *Molecular biology and evolution* 2006, **24**(1):203-216.
 21. Homan KT, Tesmer JJ: **Molecular basis for small molecule inhibition of G protein-coupled receptor kinases.** *ACS chemical biology* 2014, **10**(1):246-256.
 22. Kulanthaivel P, Hallock YF, Boros C, Hamilton SM, Janzen WP, Ballas LM, Loomis CR, Jiang JB, Katz B: **Balanol: a novel and potent inhibitor of protein kinase C from the fungus *Verticillium balanoides*.** *Journal of the American Chemical Society* 1993, **115**(14):6452-6453.
 23. Ikeda S, Keneko M, Fujiwara S: **Cardiotonic agent comprising GRK inhibitor.** *US Patent* 2007.
 24. Mayer G, Wulffen B, Huber C, Brockmann J, Flicke B, Neumann L, Hafenbradl D, Klebl BM, Lohse MJ, Krasel C: **An RNA molecule that specifically inhibits G-protein-coupled receptor kinase 2 in vitro.** *Rna* 2008, **14**(3):524-534.
 25. Thal DM, Homan KT, Chen J, Wu EK, Hinkle PM, Huang ZM, Chuprun JK, Song J, Gao E, Cheung JY: **Paroxetine is a direct inhibitor of g protein-coupled receptor kinase 2 and increases myocardial contractility.** *ACS chemical biology* 2012, **7**(11):1830-1839.

26. Homan KT, Larimore KM, Elkins JM, Szklarz M, Knapp S, Tesmer JJ: **Identification and structure–function analysis of subfamily selective g protein-coupled receptor kinase inhibitors.** *ACS chemical biology* 2014, **10**(1):310-319.
27. Yuriev E, Ramsland PA: **Latest developments in molecular docking: 2010–2011 in review.** *Journal of Molecular Recognition* 2013, **26**(5):215-239.
28. Kubinyi H: **QSAR and 3D QSAR in drug design Part 1: methodology.** *Drug discovery today* 1997, **2**(11):457-467.
29. Keretsu S, Balasubramanian PK, Bhujbal SP, Cho SJ: **Receptor-guided 3D-Quantitative Structure–Activity Relationship and Docking Studies of 6-Substituted 2-Arylamino-purines as CDK2 Kinase Inhibitors.** *Bulletin of the Korean Chemical Society* 2017, **38**(11):1275-1284.
30. Bhujbal SP, Balasubramanian PK, Cho SJ: **In silico studies on 2-substituted phenol quinazoline derivatives as RET receptor tyrosine kinase antagonists.** *Medicinal Chemistry Research* 2017, **26**(12):3228-3239.
31. Okawa T, Aramaki Y, Yamamoto M, Kobayashi T, Fukumoto S, Toyoda Y, Henta T, Hata A, Ikeda S, Kaneko M: **Design, synthesis, and evaluation of the highly selective and potent G-protein-coupled receptor kinase 2 (GRK2) inhibitor for the potential treatment of heart failure.** *Journal of medicinal chemistry* 2017, **60**(16):6942-6990.
32. Cramer RD, Patterson DE, Bunce JD: **Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins.** *Journal of the American Chemical Society* 1988, **110**(18):5959-5967.
33. Klebe G, Abraham U, Mietzner T: **Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity.** *Journal of medicinal chemistry* 1994, **37**(24):4130-4146.
34. Li Y-P, Weng X, Ning F-X, Ou J-B, Hou J-Q, Luo H-B, Li D, Huang Z-S, Huang S-L, Gu L-Q: **3D-QSAR studies of azaoxoisoaporphine, oxoaporphine, and oxoisoaporphine derivatives as anti-AChE and anti-AD agents by the CoMFA method.** *Journal of Molecular Graphics and Modelling* 2013, **41**:61-67.
35. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ: **AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.** *Journal of computational chemistry* 2009, **30**(16):2785-2791.
36. Akama T, Dong C, Virtucio C, Sullivan D, Zhou Y, Zhang Y-K, Rock F, Freund Y, Liu L, Bu W: **Linking phenotype to kinase: identification of a novel benzoxaborole hinge-binding motif for kinase inhibition and development of high-potency rho kinase inhibitors.** *Journal of Pharmacology and Experimental Therapeutics* 2013, **347**(3):615-625.

-
37. Cramer III RD, Bunce JD, Patterson DE, Frank IE: **Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies.** *Quantitative Structure-Activity Relationships* 1988, **7**(1):18-25.
 38. Green SM, Marshall GR: **3D-QSAR: a current perspective.** *Trends in pharmacological sciences* 1995, **16**(9):285-291.
 39. Kubinyi H, Martin YC, Folkers G: **3D QSAR in drug design: volume 1: theory methods and applications**, vol. 1: Springer Science & Business Media; 1993.
 40. Gramatica P: **Principles of QSAR models validation: internal and external.** *QSAR & combinatorial science* 2007, **26**(5):694-701.
 41. Tesmer JJ, Tesmer VM, Lodowski DT, Steinhagen H, Huber J: **Structure of human G protein-coupled receptor kinase 2 in complex with the kinase inhibitor balanol.** *Journal of medicinal chemistry* 2010, **53**(4):1867-1870.
 42. Fukuta H, Little WC: **The cardiac cycle and the physiologic basis of left ventricular contraction, ejection, relaxation, and filling.** *Heart failure clinics* 2008, **4**(1):1-11.
 43. Cohn JN, Levine TB, Olivari MT, Garberg V, Lura D, Francis GS, Simon AB, Rector T: **Plasma norepinephrine as a guide to prognosis in patients with chronic congestive heart failure.** *New England journal of medicine* 1984, **311**(13):819-823.
 44. Lefkowitz RJ, Sharp GW, Haber E: **Specific binding of β -adrenergic catecholamines to a subcellular fraction from cardiac muscle.** *Journal of Biological Chemistry* 1973, **248**(1):342-349.
 45. Choi D-J, Koch WJ, Hunter JJ, Rockman HA: **Mechanism of β -adrenergic receptor desensitization in cardiac hypertrophy is increased β -adrenergic receptor kinase.** *Journal of Biological Chemistry* 1997, **272**(27):17223-17229.
 46. Kelly E, Bailey CP, Henderson G: **Agonist-selective mechanisms of GPCR desensitization.** *British journal of pharmacology* 2008, **153**(S1):S379-S388.
 47. Rapacciuolo A, Suvarna S, Barki-Harrington L, Luttrell LM, Cong M, Lefkowitz RJ, Rockman HA: **Protein kinase A and G protein-coupled receptor kinase phosphorylation mediates β -1 adrenergic receptor endocytosis through different pathways.** *Journal of Biological Chemistry* 2003, **278**(37):35403-35411.
 48. Pippig S, Andexinger S, Daniel K, Puzicha M, Caron M, Lefkowitz R, Lohse M: **Overexpression of beta-arrestin and beta-adrenergic receptor kinase augment desensitization of beta 2-adrenergic receptors.** *Journal of Biological Chemistry* 1993, **268**(5):3201-3208.
 49. Pitcher J, Lohse MJ, Codina J, Caron MG, Lefkowitz RJ: **Desensitization of the isolated. beta. 2-adrenergic receptor by. beta.-adrenergic receptor kinase, cAMP-dependent protein kinase, and protein kinase C occurs via distinct molecular mechanisms.**

-
- Biochemistry* 1992, **31**(12):3193-3197.
50. Feldman DS, Carnes CA, Abraham WT, Bristow MR: **Mechanisms of disease: β -adrenergic receptors—alterations in signal transduction and pharmacogenomics in heart failure.** *Nature Reviews Cardiology* 2005, **2**(9):475.
 51. Najafi A, Sequeira V, Kuster DW, van der Velden J: **β -adrenergic receptor signalling and its functional consequences in the diseased heart.** *European journal of clinical investigation* 2016, **46**(4):362-374.
 52. Rengo G, Lymperopoulos A, Leosco D, Koch WJ: **GRK2 as a novel gene therapy target in heart failure.** *Journal of molecular and cellular cardiology* 2011, **50**(5):785-792.
 53. Waldschmidt HV, Homan KT, Cruz-Rodríguez O, Cato MC, Waninger-Saroni J, Larimore KM, Cannavo A, Song J, Cheung JY, Kirchoff PD: **Structure-based design, synthesis, and biological evaluation of highly selective and potent G protein-coupled receptor kinase 2 inhibitors.** *Journal of medicinal chemistry* 2016, **59**(8):3793-3807.
 54. Nakagawa O, Fujisawa K, Ishizaki T, Saito Y, Nakao K, Narumiya S: **ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein serine/threonine kinase in mice.** *FEBS letters* 1996, **392**(2):189-193.
 55. Shimokawa H, Rashid M: **Development of Rho-kinase inhibitors for cardiovascular medicine.** *Trends in pharmacological sciences* 2007, **28**(6):296-302.
 56. Zhou L, Xu Z, Ren X, Chen K, Xin S: **MicroRNA-124 (MiR-124) inhibits cell proliferation, metastasis and invasion in colorectal cancer by downregulating Rho-associated protein kinase 1 (ROCK1).** *Cellular Physiology and Biochemistry* 2016, **38**(5):1785-1795.
 57. Johnson LN: **Protein kinase inhibitors: contributions from structure to clinical compounds.** *Quarterly reviews of biophysics* 2009, **42**(1):1-40.
 58. Sehon CA, Wang GZ, Viet AQ, Goodman KB, Dowdell SE, Elkins PA, Semus SF, Evans C, Jolivet LJ, Kirkpatrick RB: **Potent, selective and orally bioavailable dihydropyrimidine inhibitors of Rho kinase (ROCK1) as potential therapeutic agents for cardiovascular diseases.** *Journal of medicinal chemistry* 2008, **51**(21):6631-6634.
 59. Bouley R, Waldschmidt HV, Cato MC, Cannavo A, Song J, Cheung JY, Yao X-Q, Koch WJ, Larsen SD, Tesmer JJ: **Structural Determinants Influencing the Potency and Selectivity of Indazole-Paroxetine Hybrid G Protein-Coupled Receptor Kinase 2 Inhibitors.** *Molecular pharmacology* 2017, **92**(6):707-717.
 60. Anderson AC: **The process of structure-based drug design.** *Chemistry & biology* 2003, **10**(9):787-797.
 61. Adcock SA, McCammon JA: **Molecular dynamics: survey of methods for simulating the activity of proteins.** *Chemical reviews* 2006, **106**(5):1589-1615.

-
62. Reddy MR, Erion MD: **Free energy calculations in rational drug design**: Springer Science & Business Media; 2001.
 63. Hobson AD, Judge RA, Aguirre AL, Brown BS, Cui Y, Ding P, Dominguez E, DiGiammarino E, Egan DA, Freiberg GM: **Identification of Selective Dual ROCK1 and ROCK2 Inhibitors Using Structure-Based Drug Design**. *Journal of medicinal chemistry* 2018, **61**(24):11074-11100.
 64. Eswar N, Webb B, Marti-Renom MA, Madhusudhan M, Eramian D, Shen My, Pieper U, Sali A: **Comparative protein structure modeling using MODELLER**. *Current protocols in protein science* 2007, **50**(1):2.9. 1-2.9. 31.
 65. Marti-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A: **Comparative protein structure modeling of genes and genomes**. *Annual review of biophysics and biomolecular structure* 2000, **29**(1):291-325.
 66. Fiser A, Do RKG: **Modeling of loops in protein structures**. *Protein science* 2000, **9**(9):1753-1773.
 67. Shen My, Sali A: **Statistical potential for assessment and prediction of protein structures**. *Protein science* 2006, **15**(11):2507-2524.
 68. Berendsen HJ, van der Spoel D, van Drunen R: **GROMACS: a message-passing parallel molecular dynamics implementation**. *Computer physics communications* 1995, **91**(1-3):43-56.
 69. Hess B, Kutzner C, Van Der Spoel D, Lindahl E: **GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation**. *Journal of chemical theory and computation* 2008, **4**(3):435-447.
 70. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D: **GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit**. *Bioinformatics* 2013, **29**(7):845-854.
 71. Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E: **Tackling exascale software challenges in molecular dynamics simulations with GROMACS**. In: *International Conference on Exascale Applications and Software: 2014*. Springer: 3-27.
 72. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E: **GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers**. *SoftwareX* 2015, **1**:19-25.
 73. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C: **Comparison of multiple Amber force fields and development of improved protein backbone parameters**. *Proteins: Structure, Function, and Bioinformatics* 2006, **65**(3):712-725.
 74. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA: **Development and testing of a general amber force field**. *Journal of computational chemistry* 2004, **25**(9):1157-1174.

-
75. Da Silva AWS, Vranken WF: **ACPYPE-Antechamber python parser interface**. *BMC research notes* 2012, **5**(1):367.
 76. Berendsen HJ, Postma Jv, van Gunsteren WF, DiNola A, Haak J: **Molecular dynamics with coupling to an external bath**. *The Journal of chemical physics* 1984, **81**(8):3684-3690.
 77. Parrinello M, Rahman A: **Polymorphic transitions in single crystals: A new molecular dynamics method**. *Journal of Applied physics* 1981, **52**(12):7182-7190.
 78. Hess B: **P-LINCS: A parallel linear constraint solver for molecular simulation**. *Journal of Chemical Theory and Computation* 2008, **4**(1):116-122.
 79. Kumari R, Kumar R, Consortium OSDD, Lynn A: **g_mmpbsa□ A GROMACS tool for high-throughput MM-PBSA calculations**. *Journal of chemical information and modeling* 2014, **54**(7):1951-1962.
 80. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA: **Electrostatics of nanosystems: application to microtubules and the ribosome**. *Proceedings of the National Academy of Sciences* 2001, **98**(18):10037-10041.
 81. Gohlke H, Kiel C, Case DA: **Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes**. *Journal of molecular biology* 2003, **330**(4):891-913.
 82. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W: **Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models**. *Accounts of chemical research* 2000, **33**(12):889-897.
 83. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE: **Improved side-chain torsion potentials for the Amber ff99SB protein force field**. *Proteins: Structure, Function, and Bioinformatics* 2010, **78**(8):1950-1958.
 84. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA: **Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate– DNA helices**. *Journal of the American Chemical Society* 1998, **120**(37):9401-9409.
 85. Honig B, Nicholls A: **Classical electrostatics in biology and chemistry**. *Science* 1995, **268**(5214):1144-1149.
 86. Sitkoff D, Sharp KA, Honig B: **Accurate calculation of hydration free energies using macroscopic solvent models**. *The Journal of Physical Chemistry* 1994, **98**(7):1978-1988.
 87. Still WC, Tempczyk A, Hawley RC, Hendrickson T: **Semianalytical treatment of solvation for molecular mechanics and dynamics**. *Journal of the American Chemical Society* 1990, **112**(16):6127-6129.
 88. Wold S, Ruhe A, Wold H, Dunn I, WJ: **The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses**. *SIAM Journal on Scientific and Statistical*

-
- Computing* 1984, **5**(3):735-743.
89. Yao X-Q, Cato MC, Labudde E, Beyett TS, Tesmer JJ, Grant BJ: **Navigating the conformational landscape of G protein-coupled receptor kinases during allosteric activation.** *Journal of Biological Chemistry* 2017, **292**(39):16032-16043.
 90. Yang Z, Wang F, Liu S, Guan W: **Comparative clinical features and short-term outcomes of gastric and small intestinal gastrointestinal stromal tumours: a retrospective study.** *Scientific Reports* 2019, **9**(1):1-8.
 91. Kindblom L-G, Remotti HE, Aldenborg F, Meis-Kindblom JM: **Gastrointestinal pacemaker cell tumor (GIPACT): gastrointestinal stromal tumors show phenotypic characteristics of the interstitial cells of Cajal.** *The American journal of pathology* 1998, **152**(5):1259.
 92. Miettinen M, Lasota J: **Gastrointestinal stromal tumors—definition, clinical, histological, immunohistochemical, and molecular genetic features and differential diagnosis.** *Virchows archiv* 2001, **438**(1):1-12.
 93. Nilsson B, Bümbling P, Meis-Kindblom JM, Odén A, Dortok A, Gustavsson B, Sablinska K, Kindblom LG: **Gastrointestinal stromal tumors: the incidence, prevalence, clinical course, and prognostication in the preimatinib mesylate era: a population-based study in western Sweden.** *Cancer* 2005, **103**(4):821-829.
 94. Mol CD, Dougan DR, Schneider TR, Skene RJ, Kraus ML, Scheibe DN, Snell GP, Zou H, Sang B-C, Wilson KP: **Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase.** *Journal of Biological Chemistry* 2004, **279**(30):31655-31663.
 95. Liang L, Yan X-E, Yin Y, Yun C-H: **Structural and biochemical studies of the PDGFRA kinase domain.** *Biochemical and Biophysical Research Communications* 2016, **477**(4):667-672.
 96. Rosnet O, Birnbaum D: **Hematopoietic receptors of class III receptor-type tyrosine kinases.** *Critical reviews in oncogenesis* 1993, **4**(6):595.
 97. Heldin C-H: **Dimerization of cell surface receptors in signal transduction.** *Cell* 1995, **80**(2):213-223.
 98. Hubbard SR, Mohammadi M, Schlessinger J: **Autoregulatory mechanisms in protein-tyrosine kinases.** *Journal of Biological Chemistry* 1998, **273**(20):11987-11990.
 99. Rosenkranz S, Ikuno Y, Leong FL, Klinghoffer RA, Miyake S, Band H, Kazlauskas A: **Src family kinases negatively regulate platelet-derived growth factor α receptor-dependent signaling and disease progression.** *Journal of Biological Chemistry* 2000, **275**(13):9620-9627.
 100. Kelly JD, Haldeman B, Grant FJ, Murray MJ, Seifert R, Bowen-Pope DF, Cooper J, Kazlauskas A: **Platelet-derived growth factor (PDGF) stimulates PDGF receptor subunit dimerization and intersubunit trans-phosphorylation.** *Journal of Biological Chemistry* 1991, **266**(14):8987-8992.
 101. Yin Y-q, Liu C-j, Zhang B, Wen Y, Yin Y: **Association between CT**

-
- imaging features and KIT mutations in small intestinal gastrointestinal stromal tumors.** *Scientific reports* 2019, **9**(1):1-7.
102. Xu Z, Huo X, Tang C, Ye H, Nandakumar V, Lou F, Zhang D, Jiang S, Sun H, Dong H: **Frequent KIT mutations in human gastrointestinal stromal tumors.** *Scientific reports* 2014, **4**:5907.
103. Rubin BP, Heinrich MC, Corless CL: **Gastrointestinal stromal tumour.** *The Lancet* 2007, **369**(9574):1731-1741.
104. Rammohan A, Sathyanesan J, Rajendran K, Pitchaimuthu A, Perumal S-K, Srinivasan U, Ramasamy R, Palaniappan R, Govindan M: **A gist of gastrointestinal stromal tumors: A review.** *World journal of gastrointestinal oncology* 2013, **5**(6):102.
105. Belinsky MG, Cai KQ, Zhou Y, Luo B, Pei J, Rink L, von Mehren M: **Succinate dehydrogenase deficiency in a PDGFRA mutated GIST.** *BMC cancer* 2017, **17**(1):1-11.
106. Call JW, Wang Y, Montoya D, Scherzer NJ, Heinrich MC: **Survival in advanced GIST has improved over time and correlates with increased access to post-imatinib tyrosine kinase inhibitors: results from Life Raft Group Registry.** *Clinical sarcoma research* 2019, **9**(1):4.
107. Zhao R, Wang Y, Huang Y, Cui Y, Xia L, Chen Y, Zhuang W, Zhou Y, Wu X: **Adjuvant imatinib for patients with high-risk gastrointestinal stromal tumors: a retrospective cohort study.** *Scientific Reports* 2017, **7**(1):1-7.
108. Corless CL, Schroeder A, Griffith D, Town A, McGreevey L, Harrell P, Shiraga S, Bainbridge T, Morich J, Heinrich MC: **PDGFRA mutations in gastrointestinal stromal tumors: frequency, spectrum and in vitro sensitivity to imatinib.** *Journal of clinical oncology* 2005, **23**(23):5357-5364.
109. Janeway KA, Albritton KH, Van Den Abbeele AD, D'Amato GZ, Pedrazzoli P, Siena S, Picus J, Butrynski JE, Schlemmer M, Heinrich MC: **Sunitinib treatment in pediatric patients with advanced GIST following failure of imatinib.** *Pediatric blood & cancer* 2009, **52**(7):767-771.
110. Waddell T, Cunningham D: **Evaluation of regorafenib in colorectal cancer and GIST.** *The Lancet* 2013, **381**(9863):273-275.
111. Serrano C, Mariño-Enríquez A, Tao DL, Ketzer J, Eilers G, Zhu M, Yu C, Mannan AM, Rubin BP, Demetri GD: **Complementary activity of tyrosine kinase inhibitors against secondary kit mutations in imatinib-resistant gastrointestinal stromal tumours.** *British journal of cancer* 2019, **120**(6):612-620.
112. Son MK, Ryu M-H, Park JO, Im S-A, Kim T-Y, Lee SJ, Ryoo B-Y, Park SR, Kang Y-K: **Efficacy and safety of regorafenib in Korean patients with advanced gastrointestinal stromal tumor after failure of imatinib and sunitinib: a multicenter study based on the management access program.** *Cancer research and treatment: official journal of Korean Cancer Association* 2017, **49**(2):350.

-
113. Kang Y, Yoo C, Ryoo B, Lee J, Tan E, Park I, Park J, Choi Y, Jo J, Ryu J: **Phase II study of dovitinib in patients with metastatic and/or unresectable gastrointestinal stromal tumours after failure of imatinib and sunitinib.** *British journal of cancer* 2013, **109**(9):2309-2315.
114. Le Cesne A, Blay J-Y, Bui BN, Bouché O, Adenis A, Domont J, Cioffi A, Ray-Coquard I, Lassau N, Bonvalot S: **Phase II study of oral masitinib mesilate in imatinib-naïve patients with locally advanced or metastatic gastro-intestinal stromal tumour (GIST).** *European Journal of Cancer* 2010, **46**(8):1344-1351.
115. von Mehren M, Tetzlaff ED, Macaraeg M, Davis J, Agarwal V, Ramachandran A, Heinrich MC: **Dose escalating study of crenolanib besylate in advanced GIST patients with PDGFRA D842V activating mutations.** In.: American Society of Clinical Oncology; 2016.
116. Smith BD, Kaufman MD, Lu W-P, Gupta A, Leary CB, Wise SC, Rutkoski TJ, Ahn YM, Al-Ani G, Bulfer SL: **Ripretinib (DCC-2618) is a switch control kinase inhibitor of a broad spectrum of oncogenic and drug-resistant KIT and PDGFRA variants.** *Cancer Cell* 2019, **35**(5):738-751. e739.
117. Ravegnini G, Nannini M, Sammarini G, Astolfi A, Biasco G, Pantaleo MA, Hrelia P, Angelini S: **Personalized medicine in gastrointestinal stromal tumor (GIST): clinical implications of the somatic and germline DNA analysis.** *International journal of molecular sciences* 2015, **16**(7):15592-15608.
118. Macalino SJY, Gosu V, Hong S, Choi S: **Role of computer-aided drug design in modern drug discovery.** *Archives of pharmacal research* 2015, **38**(9):1686-1701.
119. Keretsu S, Bhujbal SP, Cho SJ: **Computational study of paroxetine-like inhibitors reveals new molecular insight to inhibit GRK2 with selectivity over ROCK1.** *Scientific reports* 2019, **9**(1):1-14.
120. Keretsu S, Bhujbal SP, Cho SJ: **Molecular modeling studies of pyrrolo [2, 3-d] pyrimidin-4-amine derivatives as JAK1 inhibitors based on 3D-QSAR, molecular docking, molecular dynamics (MD) and MM-PBSA calculations.** *Journal of Biomolecular Structure and Dynamics* 2020:1-13.
121. Keretsu S, Bhujbal SP, Cho SJ: **Docking and 3D-QSAR Studies of Hydrazone and Triazole Derivatives for Selective Inhibition of GRK2 over ROCK2.** *Letters in Drug Design & Discovery* 2020, **17**(5):618-632.
122. Keretsu S, Bhujbal SP, Cho SJ: **Computational Study of Pyrimidin-2-Aminopyrazol-Hydroxamate-based JAK2 Inhibitors for the Treatment of Myeloproliferative Neoplasms.** *Bulletin of the Korean Chemical Society* 2020, **41**(5):542-551.
123. Klebe G: **Comparative molecular similarity indices analysis: CoMSIA.** In: *3D QSAR in drug design.* Springer; 1998: 87-104.
124. Lu Y, Mao F, Li X, Zheng X, Wang M, Xu Q, Zhu J, Li J: **Discovery of**

-
- Potent, Selective Stem Cell Factor Receptor/Platelet Derived Growth Factor Receptor Alpha (c-KIT/PDGFR α) Dual Inhibitor for the Treatment of Imatinib-Resistant Gastrointestinal Stromal Tumors (GISTs).** *Journal of Medicinal Chemistry* 2017, **60**(12):5099-5119.
125. Schwede T, Kopp J, Guex N, Peitsch MC: **SWISS-MODEL: an automated protein homology-modeling server.** *Nucleic acids research* 2003, **31**(13):3381-3385.
126. Lindahl E, Hess B, Van Der Spoel D: **GROMACS 3.0: a package for molecular simulation and trajectory analysis.** *Molecular modeling annual* 2001, **7**(8):306-317.
127. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ: **GROMACS: fast, flexible, and free.** *Journal of computational chemistry* 2005, **26**(16):1701-1718.
128. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD: **CHARMM36m: an improved force field for folded and intrinsically disordered proteins.** *Nature methods* 2017, **14**(1):71-73.
129. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I: **CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.** *Journal of computational chemistry* 2010, **31**(4):671-690.
130. Gadhe CG, Madhavan T, Kothandan G, Cho SJ: **In silico quantitative structure-activity relationship studies on P-gp modulators of tetrahydroisoquinoline-ethyl-phenylamine series.** *BMC structural biology* 2011, **11**(1):5.
131. Gadhe CG, Kothandan G, Cho SJ: **Large variation in electrostatic contours upon addition of steric parameters and the effect of charge calculation schemes in CoMFA on mutagenicity of MX analogues.** *Molecular Simulation* 2012, **38**(11):861-871.
132. Pires DE, Blundell TL, Ascher DB: **pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures.** *Journal of medicinal chemistry* 2015, **58**(9):4066-4072.
133. Ross C, Nizami B, Glenister M, Sheik Amamuddy O, Atilgan AR, Atilgan C, Tastan Bishop Ö: **MODE-TASK: large-scale protein motion tools.** *Bioinformatics* 2018, **34**(21):3759-3763.
134. Tamborini E, Pricl S, Negri T, Lagonigro M, Miselli F, Greco A, Gronchi A, Casali P, Ferrone M, Fermeiglia M: **Functional analyses and molecular modeling of two c-Kit mutations responsible for imatinib secondary resistance in GIST patients.** *Oncogene* 2006, **25**(45):6140-6146.
135. Surveillances V: **The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020.** *China CDC Weekly* 2020, **2**(8):113-122.
136. Cucinotta D, Vanelli M: **WHO declares COVID-19 a pandemic.** *Acta*

-
- bio-medica: Atenei Parmensis* 2020, **91**(1):157-160.
137. Chen Y, Li L: **SARS-CoV-2: virus dynamics and host response**. *The Lancet Infectious Diseases* 2020.
138. Mahase E: **Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate**. In.: British Medical Journal Publishing Group; 2020.
139. Li Lq, Huang T, Wang Yq, Wang Zp, Liang Y, Huang Tb, Zhang Hy, Sun W, Wang Y: **COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis**. *Journal of medical virology* 2020, **92**(6):577-583.
140. Pan Y, Zhang D, Yang P, Poon LL, Wang Q: **Viral load of SARS-CoV-2 in clinical samples**. *The Lancet infectious diseases* 2020, **20**(4):411-412.
141. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X: **Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study**. *The lancet* 2020.
142. To KK-W, Tsang OT-Y, Leung W-S, Tam AR, Wu T-C, Lung DC, Yip CC-Y, Cai J-P, Chan JM-C, Chik TS-H: **Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study**. *The Lancet Infectious Diseases* 2020.
143. Chakraborty C, Sharma A, Sharma G, Bhattacharya M, Lee S: **SARS-CoV-2 causing pneumonia-associated respiratory disorder (COVID-19): diagnostic and proposed therapeutic options**. *Eur Rev Med Pharmacol Sci* 2020, **24**(7):4016-4026.
144. Saha A, Sharma AR, Bhattacharya M, Sharma G, Lee S-S, Chakraborty C: **Probable Molecular Mechanism of Remdesivir for the Treatment of COVID-19: Need to Know More**. *Archives of Medical Research* 2020.
145. Chakraborty C, Sharma AR, Bhattacharya M, Sharma G, Lee SS, Agoramoorthy G: **Consider TLR5 for new therapeutic development against COVID-19**. *Journal of Medical Virology* 2020.
146. Chakraborty C, Sharma AR, Sharma G, Bhattacharya M, Saha RP, Lee S-S: **Extensive Partnership, Collaboration, and Teamwork is Required to Stop the COVID-19 Outbreak**. *Archives of Medical Research* 2020.
147. Touret F, Gilles M, Barral K, Nougairède A, van Helden J, Decroly E, de Lamballerie X, Coutard B: **In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication**. *Scientific Reports* 2020, **10**(1):1-8.
148. Jácome R, Campillo-Balderas JA, de León SP, Becerra A, Lazcano A: **Sofosbuvir as a potential alternative to treat the SARS-CoV-2 epidemic**. *Scientific Reports* 2020, **10**(1):1-5.
149. Trezza A, Iovinelli D, Prischi F, Santucci A, Spiga O: **An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors**. 2020.

-
150. D'Acquarica I, Agranat I: **Chiral switches of chloroquine and hydroxychloroquine: potential drugs to treat COVID-19.** *Drug Discovery Today* 2020.
 151. Javorac D, Grahovac L, Manić L, Stojilković N, Anđelković M, Bulat Z, Đukić-Čosić D, Curčić M, Djordjević AB: **An overview of safety assessment of the medicines currently used in the treatment of COVID-19 disease.** *Food and Chemical Toxicology* 2020:111639.
 152. Perez-Moreiras JV, Gomez-Reino JJ, Maneiro JR, Perez-Pampin E, Lopez AR, Alvarez FMR, Laguarda JMC, del Estad Cabello A, Sorroche MG, Gregori EE: **Efficacy of tocilizumab in patients with moderate-to-severe corticosteroid-resistant Graves orbitopathy: a randomized clinical trial.** *American Journal of Ophthalmology* 2018, **195**:181-190.
 153. Jagasia M, Perales M-A, Schroeder MA, Ali H, Shah NN, Chen Y-B, Fazal S, Dawkins FW, Arbushites MC, Tian C: **Ruxolitinib for the treatment of steroid-refractory acute GVHD (REACH1): a multicenter, open-label phase 2 trial.** *Blood, The Journal of the American Society of Hematology* 2020, **135**(20):1739-1749.
 154. Byrd JC, Wierda WG, Schuh A, Devereux S, Chaves JM, Brown JR, Hillmen P, Martin P, Awan FT, Stephens DM: **Acalabrutinib monotherapy in patients with relapsed/refractory chronic lymphocytic leukemia: updated phase 2 results.** *Blood, The Journal of the American Society of Hematology* 2020, **135**(15):1204-1213.
 155. Saha A, Sharma AR, Bhattacharya M, Sharma G, Lee S-S, Chakraborty C: **Tocilizumab: A therapeutic option for the treatment of cytokine storm syndrome in COVID-19.** *Archives of Medical Research* 2020.
 156. Carr AC: **A new clinical trial to test high-dose vitamin C in patients with COVID-19.** *Critical Care* 2020, **24**(1):1-2.
 157. Grant WB, Lahore H, McDonnell SL, Baggerly CA, French CB, Aliano JL, Bhattoa HP: **Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths.** *Nutrients* 2020, **12**(4):988.
 158. Bikdeli B, Madhavan MV, Jimenez D, Chuich T, Dreyfus I, Driggin E, Der Nigoghossian C, Agano W, Madjid M, Guo Y: **COVID-19 and Thrombotic or Thromboembolic Disease: Implications for Prevention, Antithrombotic Therapy, and Follow-Up: JACC State-of-the-Art Review.** *Journal of the American College of Cardiology* 2020, **75**(23):2950-2973.
 159. Harrison C: **Coronavirus puts drug repurposing on the fast track.** *Nature biotechnology* 2020, **38**(4):379-381.
 160. Cai Q, Yang M, Liu D, Chen J, Shu D, Xia J, Liao X, Gu Y, Cai Q, Yang Y: **Experimental treatment with favipiravir for COVID-19: an open-label control study.** *Engineering* 2020.
 161. Du YX, Chen XP: **Favipiravir: pharmacokinetics and concerns about clinical trials for 2019-nCoV infection.** *Clinical Pharmacology & Therapeutics* 2020.

-
162. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G: **Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro.** *Cell research* 2020, **30**(3):269-271.
163. Cao Y-c, Deng Q-x, Dai S-x: **Remdesivir for severe acute respiratory syndrome coronavirus 2 causing COVID-19: An evaluation of the evidence.** *Travel Medicine and Infectious Disease* 2020:101647.
164. Colson P, Rolain J-M, Lagier J-C, Brouqui P, Raoult D: **Chloroquine and hydroxychloroquine as available weapons to fight COVID-19.** *Int J Antimicrob Agents* 2020, **105932**(10.1016).
165. Jaffe S: **Regulators split on antimalarials for COVID-19.** *The Lancet* 2020, **395**(10231):1179.
166. Kräusslich H-G, Wimmer E: **Viral proteinases.** *Annual review of biochemistry* 1988, **57**(1):701-754.
167. Tong L: **Viral proteases.** *Chemical Reviews* 2002, **102**(12):4609-4626.
168. Yilmaz NK, Swanstrom R, Schiffer CA: **Improving viral protease inhibitors to counter drug resistance.** *Trends in microbiology* 2016, **24**(7):547-557.
169. Hilgenfeld R: **From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design.** *The FEBS journal* 2014, **281**(18):4085-4096.
170. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C: **Structure of Mpro from COVID-19 virus and discovery of its inhibitors.** *bioRxiv* 2020.
171. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R: **Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs.** *Science* 2003, **300**(5626):1763-1767.
172. Sliwoski G, Kothiwale S, Meiler J, Lowe EW: **Computational methods in drug discovery.** *Pharmacological reviews* 2014, **66**(1):334-395.
173. Keretsu S, Bhujbal SP, Cho SJ: **Molecular modeling studies of pyrrolo [2, 3-d] pyrimidin-4-amine derivatives as JAK1 inhibitors based on 3D-QSAR, molecular docking, molecular dynamics (MD) and MM-PBSA calculations.** *Journal of Biomolecular Structure and Dynamics* 2020(just-accepted):1-26.
174. Elmezayen AD, Al-Obaidi A, Şahin AT, Yelekçi K: **Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes.** *Journal of Biomolecular Structure and Dynamics* 2020(just-accepted):1-12.
175. Cava C, Bertoli G, Castiglioni I: **In silico discovery of candidate drugs against covid-19.** *Viruses* 2020, **12**(4):404.
176. Wang J: **Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) Through Computational Drug Repurposing Study.** *Journal of Chemical Information and Modeling* 2020.
177. Zhang D-h, Wu K-l, Zhang X, Deng S-q, Peng B: **In silico screening of**

-
- Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus.** *Journal of integrative medicine* 2020, **18**(2):152-158.
178. Liang J, Pitsillou E, Karagiannis C, Darmawan KK, Ng K, Hung A, Karagiannis TC: **Interaction of the prototypical α -ketoamide inhibitor with the SARS-CoV-2 main protease active site in silico: Molecular dynamic simulations highlight the stability of the ligand-protein complex.** *Computational Biology and Chemistry* 2020:107292.
179. Rawlings ND, Barrett AJ, Bateman A: **MEROPS: the peptidase database.** *Nucleic acids research* 2010, **38**(suppl_1):D227-D233.
180. Chang MW, Ayeni C, Breuer S, Torbett BE: **Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina.** *PloS one* 2010, **5**(8).
181. SYBYL-X V: **1.2, Tripos International, St. Louis, 2010.** In.; 2017.
182. Spitzer R, Jain AN: **Surflex-Dock: Docking benchmarks and real-world application.** *Journal of computer-aided molecular design* 2012, **26**(6):687-699.
183. Trott O, Olson AJ: **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.** *Journal of computational chemistry* 2010, **31**(2):455-461.
184. Huey R, Morris GM: **Using AutoDock 4 with AutoDocktools: a tutorial.** *The Scripps Research Institute, USA* 2008:54-56.
185. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R: **Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors.** *Science* 2020, **368**(6489):409-412.
186. Levy RM, Zhang LY, Gallicchio E, Felts AK: **On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute– solvent interaction energy.** *Journal of the American Chemical Society* 2003, **125**(31):9523-9530.
187. Howser G: **Babel.** In: *Computer Networks and the Internet.* Springer; 2020: 321-327.
188. Dierynck I, Van Marck H, Van Ginderen M, Jonckers TH, Nalam MN, Schiffer CA, Raouf A, Kraus G, Picchio G: **TMC310911, a novel human immunodeficiency virus type 1 protease inhibitor, shows in vitro an improved resistance profile and higher genetic barrier to resistance compared with current protease inhibitors.** *Antimicrobial agents and chemotherapy* 2011, **55**(12):5723-5731.
189. Jensen PB, Jensen PS, Demant EJ, Friche E, Sørensen BS, Sehested M, Wassermann K, Vindeløv L, Westergaard O, Hansen HH: **Antagonistic effect of aclarubicin on daunorubicin-induced cytotoxicity in human small cell lung cancer cells: relationship to DNA integrity and topoisomerase II.** *Cancer research* 1991, **51**(19):5093-5099.
190. Mimoto T, Kato R, Takaku H, Misawa S, Fukazawa T, Nojima S, Terashima K, Sato H, Shintani M, Kiso Y: **KNI-764, A novel dipeptide-**

-
- based HIV protease inhibitor containing allophenylnorstatine.** In: *Peptide Science—Present and Future*. Springer; 1999: 652-653.
191. Galpin I, Wilby A, Place G, Beynon R: **Synthetic analogues of the proteinase inhibitor: chymostatin.** *International journal of peptide and protein research* 1984, **23**(5):477-486.
192. Xue X, Yu H, Yang H, Xue F, Wu Z, Shen W, Li J, Zhou Z, Ding Y, Zhao Q: **Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design.** *Journal of virology* 2008, **82**(5):2515-2527.
193. Irwin JJ, Shoichet BK: **ZINC– a free database of commercially available compounds for virtual screening.** *Journal of chemical information and modeling* 2005, **45**(1):177-182.
194. Noble S, Faulds D: **Saquinavir.** *Drugs* 1996, **52**(1):93-112.
195. Hall Jr DC, Ji H-F: **A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease.** *Travel medicine and infectious disease* 2020:101646.
196. Geronikaki A, Eleftheriou P, Poroikov V: **Anti-HIV Agents: Current status and recent trends.** In: *Communicable Diseases of the Developing World*. Springer; 2016: 37-95.
197. Zeuzem S, Soriano V, Asselah T, Bronowicki J-P, Lohse AW, Müllhaupt B, Schuchmann M, Bourlière M, Buti M, Roberts SK: **Faldaprevir and deleobuvir for HCV genotype 1 infection.** *New England Journal of Medicine* 2013, **369**(7):630-639.

APPENDIX A

List of Publications

1. **Keretsu, S.**, Bhujbal, S. P., & Cho, S. J. (2019) Computational study of paroxetine-like inhibitors reveals new molecular insight to inhibit GRK2 with selectivity over ROCK1. *Scientific reports*, 9(1), 1-14.
2. **Keretsu, S.**, Bhujbal, S. P., & Cho, S. J. (2019). Docking and 3D-QSAR Studies of Hydrazone and Triazole Derivatives for Selective Inhibition of GRK2 over ROCK2. *Letter in Drug Design and Discovery*, 17(5), 618-632.
3. **Keretsu, S.**, Bhujbal, S. P., & Cho, S. J. (2020). Molecular modeling studies of pyrrolo[2,3-d]pyrimidin-4-amine derivatives as JAK1 inhibitors based on 3D-QSAR, molecular docking, molecular dynamics (MD) and MM-PBSA calculations. *Journal of Biomolecular Structure and Dynamics*, 1-26.
4. **Keretsu, S.**, Bhujbal, S. P., & Cho, S. J. (2020). Rational Approach toward Covid-19 Main Protease Inhibitors via Molecular Docking, Molecular Dynamics Simulation, and Free Energy Calculation. *Scientific reports (accepted)*
5. **Keretsu, S.**, Balasubramanian, P. K., Bhujbal, S. P., & Cho, S. J. (2017). Receptor-guided 3D-Quantitative Structure–Activity Relationship and Docking Studies of 6-Substituted 2-Arylamino purines as CDK2 Kinase Inhibitors. *Bulletin of the Korean Chemical Society*, 38(11), 1275-1284.
6. Bhujbal, S. P., **Keretsu, S.**, & Cho, S. J. (2019). Receptor-guided 3D-QSAR Study of Anilinoquinazolines as RET Receptor Tyrosine Kinase Antagonists. *Bulletin of the Korean Chemical Society*, 40(3), 207-213.

7. Bhujbal, S. P., **Keretsu, S.**, & Cho, S. J. (2019). Macrocyclic effect on inhibitory activity: a modeling study on MerTK inhibitors. *Medicinal Chemistry Research*, 1-16.
8. Bhujbal, S. P., **Keretsu, S.**, & Cho, S. J. (2019). A combined molecular docking and 3D-QSAR studies on tetrahydropteridin derivatives as PLK2 antagonists. *Bulletin of the Korean Chemical Society*, 40(8), 796-802.
9. Bhujbal, S. P., **Keretsu, S.**, & Cho, S. J. (2019). Design of New Therapeutic Agents Targeting FLT3 Receptor Tyrosine Kinase Using Molecular Docking and 3D-QSAR Approach. *Letter in Drug Design and Discovery*, 17(5), 585-596.
10. **Keretsu, S.**, Bhujbal, S. P., & Cho, S. J. (2020). Computational study of pyrimidin-2-aminopyrazol-hydroxamate based JAK2 inhibitors for the treatment of Myeloproliferative Neoplasms (MNPs). *Bulletin of the Korean Chemical Society*, 41(5), 542-551.
11. **Keretsu, S.**, Ghosh, S., & Cho, S. J. (2020). Molecular Modelling Study of c-KIT/PDGFR α Dual Inhibitors for the Treatment of Gastrointestinal Stromal Tumors, 21(21), 8232.

APPENDIX B

Acknowledgment

I would like to extend my gratitude to my supervisor Professor Seung Joo Cho for his constant support and guidance throughout my PhD career.

I thank my lab mates Dr. Swapnil Bhujbal and Suparna Ghosh for the memorable months and years.

I am very thankful to my family members and friends who have been by my side through thick and thin. Their support has been the source of my inspiration all these years.