



저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 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 2023
Ph.D. Dissertation

**Kinase Drug Design based on
Structural Bioinformatics: 3D-
QSAR, Molecular Mechanics, Free
Energy Calculation (MM-PBSA,
Umbrella Sampling)**

Graduate School of Chosun University

Department of Biomedical Sciences

Suparna Ghosh

Kinase Drug Design based on Structural Bioinformatics: 3D- QSAR, Molecular Mechanics, Free Energy Calculation (MM-PBSA, Umbrella Sampling)

생물정보학을 이용한 카이네이즈 신약설계:
3 차원 QSAR, 자유에너지 계산(MM-PBSA,
우산샘플링)

24th February, 2023

Graduate School of Chosun University

Department of Biomedical Sciences

Suparna Ghosh

**Kinase Drug Design based on
Structural Bioinformatics: 3D-
QSAR, Molecular Mechanics, Free
Energy Calculation (MM-PBSA,
Umbrella Sampling)**

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 2022

Graduate School of Chosun University

Department of Biomedical Sciences

Suparna Ghosh

Ph.D. dissertation of Suparna Ghosh 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. Sung Cho

Chosun Univ.: Prof. Seung Joo Cho

January 2023

Graduate School of Chosun University

CONTENTS

CONTENTS.....	i
ABBREVIATIONS	iv
LIST OF TABLES	vi
LIST OF FIGURES	viii
ABSTRACT (KOREAN).....	x
ABSTRACT (ENGLISH).....	xi

PART I

Molecular modeling studies of *N*-phenylpyrimidine-4-amine derivatives for inhibiting FMS-like tyrosine kinase-3

01-38

1. Introduction	01
2. Methodology	04
2.1. Protein structure preparation and molecular docking.....	04
2.2. Molecular dynamics	05
2.3. MM-PB/GBSA and LIE.....	06
2.4. Dataset building, Molecular alignment, and CoMFA-CoMSIA (3D-QSAR) study	07
2.5. Contour maps analysis	09
2.6. Designing of the new compounds	09
3. Results and discussion	10
3.1. Molecular docking analysis.....	10
3.2. MD analysis	12
3.3. MM-PB/GBSA and LIE estimation	14
3.4. Dataset building, 3D-QSAR model development, and Model validation.....	18

3.5. CoMFA and CoMSIA contour map analysis	27
3.6. Designing of new compounds	29
4. Conclusion	33
References	34

PART II

Binding studies and lead generation of Pteridin-7(8H)-one derivatives targeting FLT3

1. Introduction	40
2. Methodology	43
2.1. Structure preparation and molecular docking	43
2.2. MD simulation	44
2.3. MM-PB/GBSA binding energy calculation	44
2.4. FPL simulation	45
2.5. US simulation	46
2.6. Free energy perturbation	46
2.7. Dataset building and molecular alignment	47
2.8. CoMFA and CoMSIA studies	49
2.9. Contour map analysis and design of new compounds	50
3. Results and discussion	50
4. Conclusion	66
References	67

PART III

Comparative binding affinity analysis of dual CDK2/FLT3

inhibitors	71-92
1. Introduction	72
2. Methodology	73
2.1. Dataset preparation.....	73
2.2. CoMSIA model development and contour map analysis	74
2.3. Structure preparation and molecular docking	75
2.4. MD simulation and binding energy calculation	75
2.5. US simulation study	76
3. Results and discussion	77
3.1. Statistical analysis of CoMSIA models.....	77
3.2. Contour map analysis and SAR study.....	82
3.3. Docking, MD simulation, and binding energy calculation	83
3.4. US simulation analysis.....	85
4. Conclusion	89
References	90

APPENDIX

A. List of publications.....	93
B. Acknowledgement.....	95

ABBREVIATIONS

ATP	Adenosine Triphosphate
FLT3	FMS-like tyrosine kinase-3
CDK2	Cyclin dependent kinase 2
DFG	Asp-Phe-Gly motif
FDA	The Food and Drug Administration
AML	Acute Myeloid Leukemia
ITD	Internal Tandem Duplication
TKD	Tyrosine Kinase Domain
RMSD	Root Mean Square Deviation
CoMFA	Comparative Molecular Field Analysis
RF-CoMFA	Region-Focused CoMFA
CoMSIA	Comparative Molecular Similarity Indices Analysis
3D-QSAR	Three Dimensional- Quantitative Structure Activity Relationship
SEP	Standard Deviation of Prediction
SEE	Standard Error of Estimate
BS- r^2	Bootstrapping r^2 mean
BS-SD	Bootstrapping Standard deviation
DOPE	Discrete Optimized Protein Energy
GAFF	General AMBER Force Field
LOO	Leave Out One
PLS	Partial Least Square
AD	Applicability Domain

MD	Molecular Dynamics
US	Umbrella Sampling
SMD	Steered Molecular Dynamics
BE	Binding Energy
PMF	Potential Mean Force
FPL	Fast Pulling Ligand
MM-GBSA	Molecular Mechanics-generalized Born Surface Area
LIE	Linear Interaction Energy
FEP	Free Energy Perturbation
WHAM	Weighted Histogram

LIST OF TABLES

PART I

Table 1. MM-PB/GBSA and LIE energy estimation of the FLT3 inhibitor complexes.....	16
Table 2. Per-residue MM-PB/GBSA binding free energy decomposition in kcal/mol.....	16
Table 3. Structure and activity (pIC ₅₀) values of N-methylpyrimidine-4-amine based FLT3 inhibitors	26
Table 4. Statistics of CoMFA and CoMSIA models	19
Table 5. Progressive scrambling results from the CoMFA and CoMSIA (SEHD) training set compounds	26
Table 6. Designed new compounds with higher predicted pIC ₅₀ values than the most active compound M01	30
Table 7. MM-PB/GBSA binding energy estimation between FLT3 and designed compounds	31

PART II

Table 1. Dataset compounds and their inhibitory activities. The test set compounds are shown by the (*) sign.....	48
Table 2. Experimental binding energies and calculated binding energies from MM-PB/GBSA, FPL, LIE, and FEP calculations. Except for the rupture force (F _{MAX}), all binding energy terms are expressed in kcal/mol.....	57
Table 3. Statistical results and validation of CoMFA and CoMSIA models.	59
Table 4. SAR and MD assisted designed compounds in SMILES format.	63

PART II

Table 1. Dataset compounds and their inhibitory activities against CDK2 and FLT3	77
Table 2. Statistical result and validation of CDK2 and FLT3 CoMSIA models .	82
Table 3. Binding free energy calculation from MM-GBSA and US simulation in kcal/mol.....	88

LIST OF FIGURES

PART I

Figure 1. Molecular docking analysis of compound M01	11
Figure 2. RMSD graphs of protein-ligand complexes.....	13
Figure 3. H-bond interaction distance analysis	13
Figure 4. Average MD pose of the ligands in the FLT3 active site	14
Figure 5. MM-PB/GBSA binding free energy comparison between M01 and FLT3 variants	15
Figure 6. Common substructure-based molecular alignment.....	22
Figure 7. PLS scatter plot, Williams plot, and CoMFA-CoMSIA contour maps analysis.....	24
Figure 8. The three-dimensional structure–activity relationship by taking compound M01 as a reference.....	29
Figure 9. Comparison of the final Δ TOTAL binding energy terms of the newly designed compounds	32

PART II

Figure 1. Structure, function, and inhibitory mechanisms of FLT3	41
Figure 2. Molecular docking, MD simulation, and MM-PB/GBSA binding energy calculation studies	52
Figure 3. FPL and US simulation analysis	55
Figure 4. Correlation plots, applicability domain, and contour maps analysis....	61
Figure 5. SAR study and development of new compounds	61
Figure 6. Comparison of the binding free energy of the designed compounds...	66

PART III

Figure 1. CoMSIA analysis, RMSD plots, and residue-specific binding energy contribution to the ligand 84

Figure 2. Caver predicted unbinding direction of the ligand..... 86

Figure 3. Overview of the unbinding direction and estimation of the PMF curve in the US simulation process 87

Figure 4. Correlation plots between experimental binding energies and computed binding energies..... 89

초록

생물정보학을 이용한 카이네이즈 신약설계: 3 차원 QSAR, 자유에너지 계산(MM-PBSA, 우산샘플링)

수팔나 고스

지도교수: 조 승 주, Ph.D.

의과학과

조선대학교 대학원

결합자리에서 기존 화합물의 화학 그룹을 대체하는 리간드 변형은 카이네이즈 약물 개발의 중요한 전략이다. 이를 위해서는 수용체와 소분자 사이의 결합 친화도에 대한 정확한 평가와 순위 결정이 필요하며, 이는 약물 발견 연구의 개발 비용을 크게 줄인다. 두 생체 분자 간의 결합 친화도 측정을 위한 신뢰할 수 있는 방법을 확립하는 것은 현재 이 분야에서 각광을 받고 있다. 이를 위하여 우리는 FMS 유사 티로신 키나아제-3(FLT3) 및 사이클린 의존성 키나아제 2(CDK2)를 포함한 다양한 키나아제 표적에 대한 생물물리학 및 구조 모델링 기술을 사용했다. FLT3의 과발현 및 빈번한 돌연변이는 중증 급성 골수성 백혈병(AML) 질환에서 다중 세포내 신호 전달 경로의 조기 활성화를 유도한다. 경쟁적인 ATP 억제제를 사용한 FLT3의 선택적 억제는 지난 수십 년 동안 인기를 얻었다. 그러나 약물 내성 및 표적 외 효과가 여전히 우려 사항으로 남아 있어 지속적인 약물 개발 프로세스가 필요하다. 도킹, 분자 동역학, Poisson-Boltzmann/일반화된 Born 표면적(MM-PB/GBSA), FPL(빠른 폴링 리간드) 시뮬레이션, 선형 상호 작용 에너지(LIE), 우산 샘플링(US) 및 자유 에너지 섭동(FEP) 등의 시뮬레이션 연구와, 비교 분자장 분석(CoMFA) 및 비교 분자 유사성 지수 분석(CoMSIA)을 사용한 구조-활성 관계(SAR) 연구를 수행고, 그 결과를 SAR 방식으로 강조했다. 계산된 결합 에너지는 여러 계산 방식에서 실험적 결합 에너지와 잘 상관되어 계산 방식이 신뢰할 수 있고 성공적으로 활용될 수 있음을 나타낸다. CoMFA

및 CoMSIA 모델 모두에서 관찰된 억제 활성과 예측된 억제 활성 사이에 만족스러운 상관 관계 통계를 얻었다. CoMFA 및 CoMSIA 모델에서 계산된 등고선 지도는 화합물의 억제 활성을 증가 또는 감소시킬 수 있는 화학 그룹 치환의 유리한 위치와 불리한 위치에 대한 귀중한 정보를 제공한다. 앞서 언급한 접근 방식은 몇 가지 새로운 화합물을 설계하기 위해 공동으로 활용되었으며 결합 친화도 및 억제 가능성이 결정되었다. 이 연구들은 AML 치료의 합리적인 리드 발견을 위한 포괄적인 이론적인 방향성을 제공할 수 있다.

Abstract

Kinase Drug Design based on Structural Bioinformatics: 3D-QSAR, Molecular Mechanics, Free Energy Calculation (MM-PBSA, Umbrella Sampling)

Suparna Ghosh

Advisor: Prof. Cho Seung Joo, Ph.D.

Department of Biomedical Sciences

Graduate School of Chosun University

Ligand modification by substituting chemical groups of the existing lead compounds within the binding pocket is a popular strategy for kinase drug development. This necessitates the accurate assessment or ranking of the binding affinity between the receptor and small molecules, which significantly reduces the development cost in drug discovery research. Establishing the reliable methods for binding affinity determination between two biomolecules is of great interest. In this work, we employed combined molecular modeling techniques for biophysical and structural studies of inhibitors that target FMS-like tyrosine kinase-3 (FLT3) and cyclin-dependent kinase 2 (CDK2). Overexpression and frequent mutations in FLT3 induce premature activation of multiple intracellular signaling pathways in severe acute myeloid leukemia (AML) disease. Selective inhibition of FLT3 using competitive ATP inhibitors has gained popularity over the past few decades. However, drug resistance and off-target effects remain a concern, necessitating the continuous drug development process. We determined protein-ligand binding affinity by employing physics-based docking, molecular dynamics (MD), molecular mechanics, Poisson-Boltzmann/generalized Born surface area (MM-PB/GBSA), fast pulling ligand (FPL) simulation, linear interaction energy (LIE), umbrella sampling (US), and free energy perturbation (FEP) scoring functions. The structure-activity relationship

(SAR) study was conducted using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), and the results were emphasized as an SAR scheme. In both the CoMFA and CoMSIA models, satisfactory correlation statistics were obtained between observed and predicted inhibitory activity. The contour maps derived from the CoMFA and CoMSIA models could explain valuable information on the favorable and unfavorable positions for chemical group substitution, which can increase or decrease the inhibitory activity of the compounds. The aforementioned approaches were co-utilized to design several new compounds, and the binding affinity and inhibitory potential were determined. The overall study could provide valuable theoretical guidance in the rational lead discovery process targeting FLT3.

Suparna Ghosh Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

PART I

Molecular modeling studies of *N*-phenylpyrimidine-4-amine derivatives for inhibiting FMS-like tyrosine kinase-3

1. Introduction

FLT3 belongs to the type III tyrosine kinase receptor together with KIT, FMS, and platelet-derived growth factor receptor (PDGFR), which are involved in the differentiation, proliferation, and survival of hematopoietic progenitor cells [1]. It is expressed by the stromal cells of bone marrow, placenta, and CD34⁺ cells, as well as the myeloid and B lymphoid cell lineage [2]. The FLT3 structure comprises five extracellular domains similar to immunoglobulin (Ig) at the N-terminal, followed by a single transmembrane (TM) domain, a cytoplasmic juxtamembrane domain (JMD), and a tyrosine kinase domain (TKD) separated by a kinase insert (KI). An intracellular domain was located at the C-terminal end. FL is the endogenous ligand of FLT3, which is also expressed in bone marrow stromal cells, exists in a soluble form, or is bound to the membrane. Predominantly, FLT3 is found in an unbound form as an unphosphorylated monomer coordinated by an inactive kinase domain. The JM domain interacts with the KD to block the ATP binding to the active site. After binding to the ligand FL in the extracellular Ig domains, FLT3 undergoes dimerization and exposes its dimerized domain [6–8]. This event triggers the activation of tyrosine kinase, followed by the phosphorylation of the FLT3 intracellular domain at its various sites. The phosphorylation of FLT3 propagates multiple intracellular signaling pathways, which are essential for phospholipid metabolism, transcription, differentiation, proliferation, cell survival, and apoptosis [9].

FLT3 is overexpressed in patients with AML, which is considered an aggressive hematologic malignancy. Active mutations in FLT3 have been reported in ~30% of total AML cases. These mutations can be subdivided into internal tandem duplicates (FLT3-ITD) and point mutations in the tyrosine kinase domain (FLT3-TKD), which are approximately 25% and 5% of the total AML, respectively [10,11]. Therefore, inhibition of FLT3 is an ideal therapeutic choice. Many inhibitors have been subjected to preclinical and clinical trials and have shown promising results. The tyrosine kinase inhibitors such as Tandutinib, Sunitinib, Midostaurin, Lestaurtinib, and Sorafenib were used as first-generation FLT3 inhibitors [12]. Due to the lack of sensitivity and

selectivity of first-generation inhibitors, second-generation inhibitors, such as Gilteritinib, Quizartinib, and Crenolanib, have been approved [13,14]. Although, their poor bioavailability and off-target effects often increase drug toxicity in patients, which remains a concern. Patients undergoing AML treatment have often developed resistance to lead compounds through on-target secondary mutations in the kinase domain [13]. Two of the most common mutations have been found in the gatekeeper residue F691L and the activation loop D835Y, as shown in Figure 1c. Mutations in residues I836, D839, and Y842 are also found in the FLT3 kinase domain in patients with AML [14].

Inevitably, the rational development of new FLT3 inhibitors based on existing lead compounds is an ideal choice for achieving therapeutic efficacy in AML. Computer-aided drug design (CADD) has emerged as a promising tool for discovering new FLT3 inhibitors [15]. In a prior study, Bensinger et al. [16] used virtual screening and docking to identify lead compounds that were covalently bound to the DGF-in conformer of FLT3 receptor. These lead compounds were chemically modified to increase cytotoxicity and inhibitory efficacy against wild-type and mutant (D835Y) FLT3, suggesting that they could be a suitable starting point for discovering irreversible inhibitors [16]. In another study by Smith et al. [17], residue D835 plays a critical role in maintaining the DFG-out configuration by acting as an amino terminal capping residue for the α C-helix and serving as an essential space for type II inhibitor binding. However, the mutations in the D835 remain sensitive to type I inhibitors.

Herein, we performed computational modeling studies such as docking, molecular dynamics (MD), free energy calculation, and three-dimensional structure-activity relationship (3D-QSAR) to a series of 40 pyrimidine-4,6-diamine derivatives, which are reported as type II-like inhibitors of FLT3 by Bharate et al. [18]. In their docking study, the most active compound 13a bound to the hinge loop by forming two H-bond interactions with C694. The compounds exhibited a diverse inhibitory activity range (IC_{50} 13.9 nM-15111 nM) against FLT3. The molecular docking and molecular dynamics revealed the critical interactions with the inhibitors inside the binding

pocket. We calculated the MM-PB/GBSA and LIE binding energy terms to evaluate the protein–ligand binding affinity. We manually induced the F691L and D835Y mutations in FLT3 and performed MD simulations to understand the mutagenic effect in terms of binding affinity to the inhibitor compound. Finally, we conducted CoMFA and CoMSIA studies of the 40 compounds to produce the 3D-QSAR contour map and established the structure–activity relationship. The CoMFA and CoMSIA contour maps described how modification of chemical groups could enhance the inhibitory activity of the compounds.

2. Methodology

2.1. Protein structure preparation and molecular docking

The FLT3 protein structure with a resolution of 2.20 Å was retrieved from the Protein Databank (PDB: 6JQR) [8]. The water molecules and ligands from the crystallographic solution were removed from the protein crystal. The missing residues were remodeled using Modeler-10.1 (University of San Francisco, San Francisco, CA, USA) in UCSF Chimera-1.14 (RBVI, UCSF, San Francisco, CA, USA), and the final model was endorsed in a DFG-out configuration. The model with the lowest DOPE score was selected and the entire protein structure was verified using the PROCHECK (DOE-MBI service, UCLA, Los Angeles, CA, USA) server.

Compound M01 was the most active compound in the dataset and was thus chosen for the docking study, as described in the previous study [35–37]. Compounds M03, M17, M20, M24, and M34, with different subgroups at their positions R₁ and R₂ exhibiting a diverse range of activity, were also included for the docking study. Briefly, the protein was prepared by assigning polar hydrogens and Kollman charges in AutoDockTools (AutoDock 4.2, Scripps Research, La Jolla, CA, USA). The ligand was sketched and minimized, and hydrogens were added in SYBYL-2.1, followed by the addition of Gasteiger charges. The number of rotatable bonds was fixed at six to the ligands to avoid conformational explosion. We assigned the active site according to the X-ray structure of Quizertinib-bound FLT3 (PDB: 4XUF). The grid box size

was set to $50 \times 60 \times 50$ in the X, Y, and Z directions, respectively, with a grid spacing of 0.375 and a grid center of $X = -34$, $Y = -10$, and $Z = -25$ to compute the grid parameters using AutoGrid. For the conformational search, the Lamarckian Genetic Algorithm (LGA) was employed. Finally, AutoDock-4.2 was used to perform 100 docking runs. This protocol was repeated for the remaining compounds. AutoDockTools, Pymol (Schrodinger, Inc., New York, NY, USA), and Maestro (Schrodinger, Inc., New York, NY, USA) were used to analyze the docking results. All protein–ligand docked complexes were taken for the MD simulation study.

2.2. Molecular dynamics

The MD simulation was carried out with GROMACS 2019.5 [38] using the Amber14SB [39] force field. The topology and parameters of the small molecules were generated using ACPYPE [40]. The system was prepared by placing the protein–ligand complex in a cubic periodic box. The complex was then wrapped with the TIP3P water model in such a way so that the minimum thickness of the water wall was maintained at 10 Å from the protein atoms. An adequate amount of Na^+ and Cl^- ions were added to neutralize the system.

The steepest descent algorithm was applied for 10,000 steps to minimize the system, followed by 200 ps of constant substance, volume, and temperature (NVT) simulation using the modified Berendsen thermostat (V-rescale) to attain a temperature of 300 K. Next, a 400 ps constant substance, pressure, and temperature (NPT) simulation was executed using the modified Berendsen barostat (V-rescale) to achieve 1 bar of pressure. The backbone and heavy atoms of the ligands were kept restrained during the minimization, NVT, and NPT simulation steps. Finally, the system was subjected to the 100 ns MD simulation run by un-restraining the backbone and heavy atoms. The Particle Mesh Ewald (PME) scheme was employed to maintain the electrostatic interactions, and the SHAKE algorithm was employed to deal with the bond length constraints. The cut-off distance was set at 12 Å to calculate the coulombic and van der Waals (vdW) interactions, respectively. The protocol was followed for the rest of

the docked complexes as described in the previous study [41]. RMSD and H-bond distances were calculated by the in-built ‘*gmx rms*’ and ‘*gmx distance*’ functions in gromacs.

2.3. MM-PB/GBSA and LIE

The MM-PB/GBSA is a useful technique for computing the end-state binding free energy between the protein and the ligand. In our study, the binding free energy of MM-PB/GBSA was calculated using the *gmx_MMPBSA* [42] package based on *MMPBSA.py* [43]. The protein–ligand binding free energy of MM-PB/GBSA can be expressed by Equation (1),

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{protein}} - \Delta G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S \quad (2)$$

$$\Delta E_{\text{MM}} = (\Delta E_{\text{vdW}} + \Delta E_{\text{ele}}) \quad (3)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{GB}} + \Delta G_{\text{SA}} \quad (4)$$

where $\Delta G_{\text{complex}}$ stands for the total binding free energy between the protein–ligand complexes. The total free energy of the protein and ligand in the solvent was expressed by $\Delta G_{\text{protein}}$ and ΔG_{ligand} . In Equation (2), the ΔE_{MM} stands for the interaction energy between the protein–ligand complex under the gas-phase condition, which was estimated by calculating the van der Waals (ΔE_{vdW}) and electrostatic (ΔE_{ele}) energies (Equation (3)). The ΔG_{sol} stands for the free energy solvation, which was derived by calculating the polar solvation ΔG_{GB} and non-polar solvation ΔG_{SA} energy in Equation (4). The entropy contribution of the system is represented by $T\Delta S$. The entropy ($-T\Delta S$) calculation through *nmode* or Quasi-harmonic (QH) approximation is a computationally cumbersome process; therefore, the entropy term, ($-T\Delta S$) was not considered in this study.

The final 2 ns trajectories from each system were taken for the CDJPB/GBSA total binding energy and per-residue binding energy decomposition analysis.

For the end-state LIE calculation, each of the docked ligands was simulated in an unbound state explicitly by keeping the same parameters as described in the MD simulation method. In the LIE method, the binding affinity (ΔG_{bind}) can be written as:

$$\Delta G_{\text{bind}} = \alpha(\langle V_{\text{lig-surr}}^{\text{vdW}} \rangle_{\text{bound}} - \langle V_{\text{lig-surr}}^{\text{vdW}} \rangle_{\text{unbound}}) + \beta(\langle V_{\text{lig-surr}}^{\text{ele}} \rangle_{\text{bound}} - \langle V_{\text{lig-surr}}^{\text{ele}} \rangle_{\text{unbound}}) \quad (5)$$

The Coulomb interaction ($V_{\text{lig-surr}}^{\text{vdW}}$) and electrostatic interaction ($V_{\text{lig-surr}}^{\text{ele}}$) values in unbound form were computed by the ‘*gmx enemat*’ function. These values were supplied in the final LIE calculation using the ‘*gmx lie*’ function. As described in these studies [44–46], we used the scaling factors $\alpha = 0.181$ and $\beta = 0.43$ by assigning the neutral charge to the ligands.

2.4. Dataset building, molecular alignment, and CoMFA-CoMSIA (3D-QSAR) study

The dataset of 40 pyrimidine-4,6-diamine-based compounds reported to be FLT3 inhibitors was taken for this study. We selected the last 1 ns MD average structure of compound M01 as a template structure. Based on the template structure, the rest of the compounds were sketched, and partial charges were assigned by the Gasteiger method in SYBYL-X2.1 (Tripos, Inc., St. Louis, MO, USA). All compounds were then subjected to minimization by tripos forcefield using the 0.05 kcal/mol convergence criterion with a maximum iteration of 2000 runs. Molecular alignment is an essential step towards the development of the 3D-QSAR model. We used the ‘common substructure-based alignment’ and ‘database alignment’ functions available in SYBYL-X2.1 to align compounds over the template structure as described in earlier studies [47,48].

For the 3D-QSAR study, the IC_{50} values of the compounds were converted into logarithmic IC_{50} (pIC_{50}) values. The dataset was divided into the training set of 30 compounds to develop the model and a test set of 10 compounds to evaluate the external predictive capability. CoMFA and CoMSIA are the two widely established 3D-QSAR approaches, which were used to establish the structure–activity relationship of the compounds in the dataset. In CoMFA, the steric fields (S) were calculated using

the Lennard–Jones potential function, whereas the electrostatic fields (E) were calculated using the Coulombic potential function. The compounds were placed one after another in a spatial grid box with a grid spacing of 2.0 by maintaining an energy tolerance of 30 kcal/mol. The sp^3 carbon atom was assigned as a probe by setting the van der Waals radii at 1.52 Å, with a net charge of +1.0. The other parameters were accepted by default in SYBYL-X2.1.

In the CoMSIA model, besides the steric and electrostatic fields, three additional descriptors, such as hydrophobic (H), H-bond acceptor (A), and H-bond donor (D) fields, were also adopted. The Gaussian-type functions were used to distinguish the distance between the probe atoms and the molecule's atoms for all grid points. The rest of the parameters were kept similar to the CoMFA parameters. All descriptors were used in different combinations to get the best possible CoMSIA model.

The partial least squares (PLS) method was adopted to analyze the internal validation of CoMFA and CoMSIA models. The leave-one-out (LOO) method was applied to obtain the cross-validation coefficient (q^2) and the optimal number of components (ONC) with a column-filtering value of 2.0 kcal/mol. Subsequently, the non-cross-validated correlation coefficient (r^2), Fisher's statistics (F value), and the standard error of estimation (SEE) were calculated, and finally, the pIC_{50} values were predicted for each compound of the test set. The value of q^2 and r^2 greater than 0.5 and 0.6, respectively, as well as $r^2 - q^2$ not exceeding 0.3 well indicated the internal validity of the QSAR models.

The fitness of the CoMFA and CoMSIA models were assessed by estimating the chi-squared (χ^2) and root-mean-squared error (RMSE) [29]:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{y_i^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 1}} \quad (7)$$

where y_i and \hat{y}_i are the observed and predicted activity, respectively, and n is the number of compounds being measured. The large χ^2 (≥ 0.5) and RMSE (≥ 1.0) values reflect the poor predictive accuracy of the QSAR models.

The external validation of the CoMFA and CoMSIA by test set compounds was a crucial step in determining the true predictive power of any QSAR models. The following criteria proposed by Roy et al. [30], Gramatica et al. [31], and Todeschini et al. [32] were used to externally validate the models. The progressive scrambling study was conducted as described in this study [49].

The applicability domain (AD) analysis of the QSAR models was done using the leverage approach as described in earlier studies [34]. The standardized residuals from the activity values of the training set and test set compounds were plotted against their leverage values in the Williams plot. The leverage value of any compound exceeding the warning leverage (h^*), as shown with the red dotted line, was denoted as outliers and influenced the model quality.

2.5. Contour maps analysis

The contour maps were generated from CoMFA and CoMSIA to explain the structure–activity relationship of the existing compounds. The most active compound M01 was placed in the center as a reference, and the contour maps were shown as 3D StDev*Coeff to elucidate the field effects of the descriptors. The green and yellow contours signify the favorable and unfavorable substitutions for steric groups [50,51]. The blue and red contours convey the favorable and unfavorable substitutions for an electropositive group in CoMFA and CoMSIA. Similarly, the favorable and unfavorable substitutions for the hydrophobic, H-bond acceptor, and H-bond donor are represented by orange-gray, magenta-slate, and cyan-purple color schemes.

2.6. Designing of the new compounds

Based on the SAR study, we designed 30 new compounds and predicted their activity by the CoMSIA (SEHD) model. Sixteen compounds out of them were predicted to be a higher pIC₅₀ value than the most active compound M01. Thereafter, we analyzed

physicochemical properties, the synthetic accessibility (SA) score by the SWISSADME [52] webserver, and the absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties by the pkCSM [53] webserver of these compounds. Finally, the designed compounds were subjected to MM-PB/GBSA binding free energy calculation.

3. Results and discussion

3.1. Molecular docking analysis

As we conducted the MD simulation and 3D-QSAR study based on the docking pose, the verification of docking reliability is an important step. There were no compounds from the dataset available with the co-crystallized FLT3 form in the PDB database. Thus, we selected the *N*-phenylpyrimidine-4-amine substructure of the docked compounds and compared it with the existing FLT3 conjugated crystal ligand FF-10101 (PDB ID 5X02) [19], and Quizartinib (PDB ID 4XUF) [20], and AWO with C-kit (PDB ID 6ITT) [21] by overlapping them in the ligRMSD server. Since FLT3 and C-kit both are members of the tyrosine kinase family, we opted for the AWO bound C-kit to compare the ligand interactions. The FF-10101 formed the critical H-bond interaction with residue C694 by the amine(-NH₂) group of *N*-phenylpyrazine-2-amine. For the ligand Quizartinib, π - π interactions were found between its aniline ring and residue F691 and residue F830. We found the critical H-bond interaction between C694 and -NH₂ of the *N*-(Pyrimidin-4-yl)thiazol-2-amine moiety, which anchored the ligand to the hinge loop. Similar interactions were observed with the docked compounds M01, M03, and M17. Compounds M20, M24, and M34 were anchored to the hinge loop by forming H-bond interactions between the residue C694 and the *N*-phenylpyrimidine-4-amine-NH₂ group and the π - π stacking with phenylalanine (F691). The final docked pose was selected from the lowest binding energy cluster, which also shared the lowest RMSD with the crystal substructure according to the ECIDALs norms [22,23]. The acceptable RMSD values suggest the overall docking reliability of the selected compounds. The BEs from the docking study were estimated

to be -11.31 kcal/mol, -11.68 kcal/mol, -9.88 kcal/mol, -10.54 kcal/mol, -9.68 kcal/mol, and -10.09 kcal/mol for compounds M01, M03, M17, M20, M24, and M34, respectively. The 2D protein–ligand interaction is illustrated in Figure 1a. The surrounding residues within 3.5 Å are shown using a color scheme, which is based on the chemical properties of the amino acids. Figure 1b depicts the H-bond, π – π , and π –anion interactions between M01 and residues inside the ATP pocket of FLT3. With residue C694, two H-bond interactions were found, while a third H-bond interaction was observed with catalytic lysine K644.

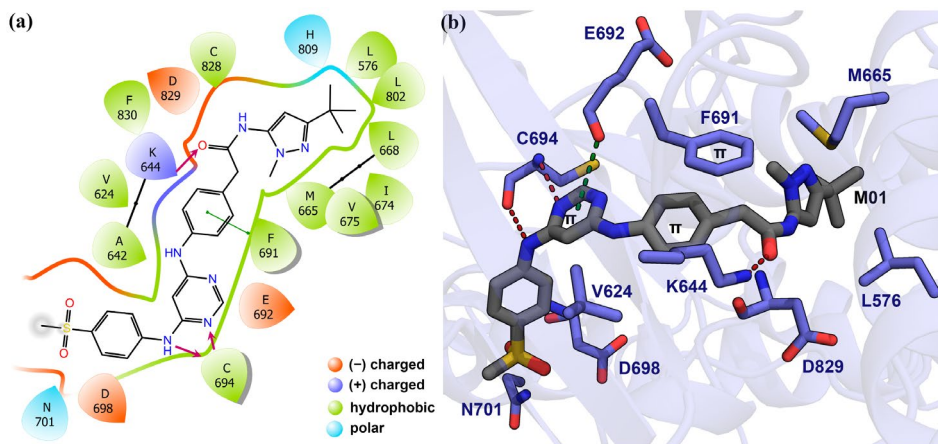


Figure 1. Molecular docking analysis of compound M01 and FLT3. (a) Compound M01 surrounded by the active site residues. The residues are illustrated based on their chemical properties in 2D representation. H-bonds are shown by magenta arrows. (b) Docking interaction of M01 with FLT3 residues. The H-bond interactions are shown by red dashed lines. The π –anion interaction is shown by green dashed lines.

The gatekeeper residue E692 formed the π –anion interaction with the pyrimidine ring. Another gatekeeper residue F691 accomplished the π – π stacking with the phenyl ring of M01. Other residues, such as L616, M665, V675, L856, L576, and I801, participated in hydrophobic interaction inside the hydrophobic cleft between the N-lobe and the C-lobe. The docked complexes of compounds M03, M17, M20, M24, and M34, which also comply with the ECIDALs norms during the docking process. The best performing protein–ligand single-docked coordinate of each complex was taken for further MD simulation study.

3.2. MD analysis

The MD simulations were conducted to evaluate the overall stability of the protein–ligand complexes. During the 100 ns MD run, the protein–ligand complexes achieved convergence within the first 10 ns. The RMSDs of proteins were found in the range of 1.0–4.0 Å, and the RMSDs of ligands were found in the range of 0.5–4 Å, as shown in Figure 2. We observed a low RMSD for compounds M01, M03, and M20, while compounds M17, M24, and M34 had a higher RMSD. In particular, the RMSD of the compounds M24 and M34 reached 4 Å during the simulation. Since the two H-bond interactions with residues K644 and C694 played a critical role in the binding of the ligand inside the ATP pocket, we compared these two interacting H-bond distances for 100 ns in Figure 3. For compound M34, the H-bond distance with K644 was found to be unstable and a major fluctuation was observed, in contrast to the other compounds. Overall, this specific H-bond distance was found to be within the range of 2–3.8 Å in Figure 3a. When comparing the H-bond distance between the ligands and N atom of residue C694, compounds M17 and M34 showed a slightly higher H-bond distance to ~3.7 Å in Figure 3b. However, this H-bond interaction was intact for the remaining compounds. Following that, we compared the average MD structure of protein–ligand complexes from the last 1 ns to their final docked position in Figure 5. Figure 5a,b,e shows that *N*-phenylpyrimidine-4-amine substructures of M01, M03, and M24 closely retained their docking pose and molecular interactions until the end of the MD run. For M17, we observed that the dimethylamine with the pyrimidine ring was displaced from its initial docked position, which might increase the critical distance for H-bond formation as shown in Figure 5c. Compound M20 formed the H-bond interactions to the C694 with its N atoms of the pyridine ring, while another H-bond interaction remained intact with K644 (Figure 4d). The piperazine ring in M34, on the other hand, shifted forward from its initial docking position inside the ATP pocket (Figure 4f), which could be a reason for the extension of the distance from residue C694. We also observed the rotational and translational displacement of the

piperidine ring and the pyrrolidine ring of compounds M03 and M20 by comparing their respective MD trajectories and docking positions.

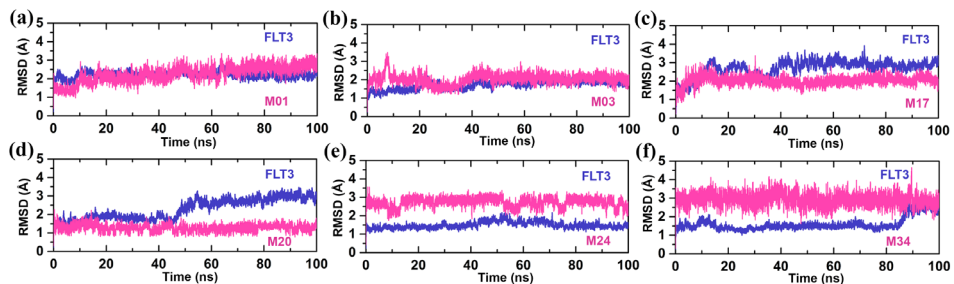


Figure 2. RMSD graphs of protein–ligand complexes. The protein and ligand RMSDs of (a) FLT3-M01, (b) FLT3-M03, (c) FLT3-M17, (d) FLT3-M20, (e) FLT3-M24, and (f) FLT3-M34 are shown in slate and pink colors, respectively.

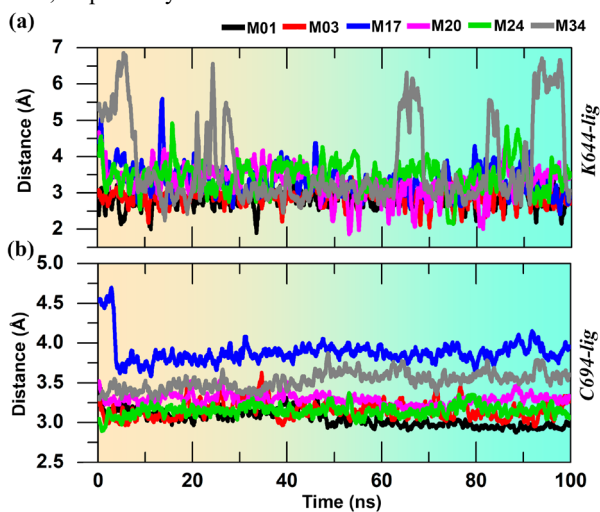


Figure 3. H-bond interaction distance analysis from (a) residue K644 and (b) C694 to the ligands during the production simulation.

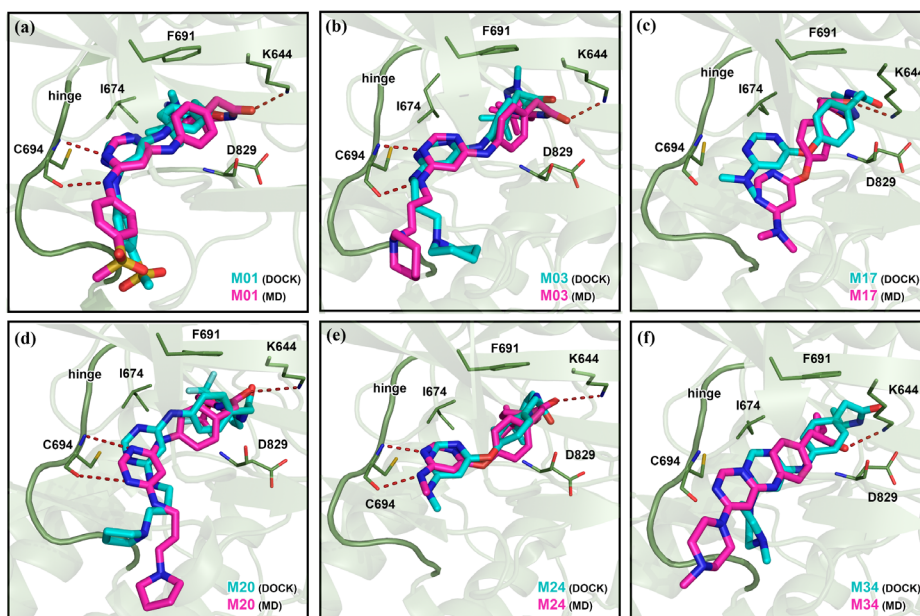


Figure 4. Average MD pose of the ligands in the FLT3 active site. The last 1 ns average MD poses of (a) M01, (b) M03, (c) M17, (d) M20, (e) M24, and (f) M34 inside the binding pocket are shown in pink color. The corresponding docked pose is shown in cyan color. The H-bond interactions are shown by red dashed lines.

3.3. MM-PB/GBSA and LIE estimation

We implemented the MM-PB/GBSA and LIE estimation to investigate the binding affinity of FLT3 to selected ligands. To calculate the MM-PB/GBSA and LIE binding energy, we used the final 2 ns or 200 frames of each protein–ligand MD trajectory. The final ΔT_{TOTAL} BEs were found to be -62.80 , -60.27 , -47.32 , -60.68 , -59.56 , and -49.31 kcal/mol for compounds M01, M03, M17, M20, M24, and M34, respectively. The final ΔG_{bind} from the LIE estimation was found to be -18.39 , -13.43 , -12.56 , -21.37 , -9.75 , and -8.14 kcal/mol for compounds M01, M03, M17, M20, M24, and M34, respectively. The detailed MM-PB/GBSA energy and LIE terms are shown in Table 1. We estimated the per-residue BE decomposition by selecting residues within 4.0 \AA from the ligand atoms. The common residues with positive or negative BE decomposition are compared in Table 2. We found that residues L616, V624, A642, K644, M665, I674, V675, F691, Y693, C694, G697, L818, C828, D829,

and F830 were the key interacting common residues, contributing the critical BE to the ligands shown in Figure 6a. We acquired the substantially equilibrated 80th ns wild-type (WT) FLT3-M01 complex and manually mutated D835Y, F691L, and both (D835Y and F691L) at the same time. The newly prepared three different FLT3-M01 complexes were subjected to standard 20 ns MD simulations, followed by MM-PB/GBSA and LIE calculations from the last 2 ns trajectory. The Δ TOTAL binding energy of MM-PB/GBSA was found to be -63.55 , -62.03 , and -62.79 kcal/mol for the D835Y, F691L, and both mutants of FLT3 variants to compound M01, respectively. However, we were unable to observe any significant changes in Δ TOTAL BE terms with M01.

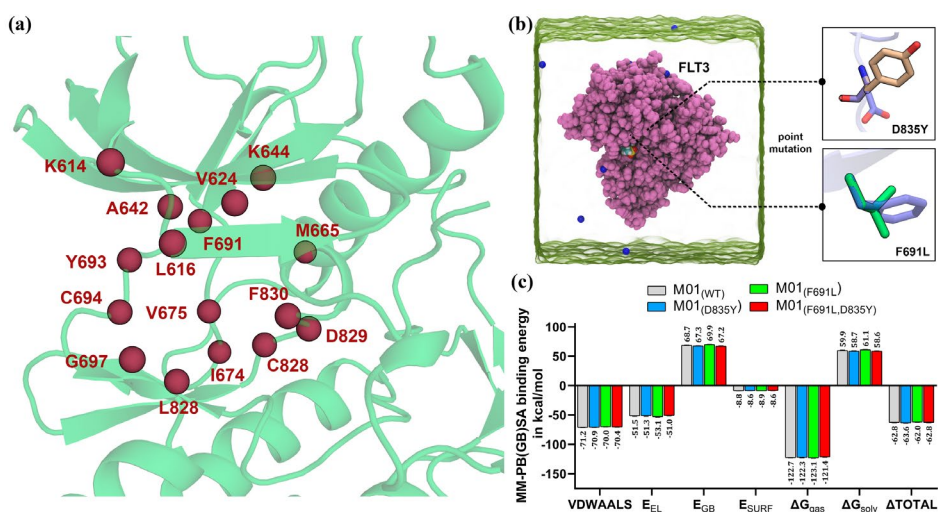


Figure 5. MM-PB/GBSA binding free energy comparison between M01 and FLT3 variants. (a) Common residues that contributed to the Δ TOTAL BE to the ligands are shown in the sphere representation. (b) Point mutations were manually performed at D835 and F691 in Pymol to compare the MM-PB/GBSA binding energy. (c) Comparison of the MM-PB/GBSA binding energy terms between WT, D835Y, F691Y, and both (F691L, D835Y) mutant FLT3 and M01.

Table 1. MM-PB/GBSA and LIE energy estimation of the FLT3 inhibitor complexes

Complexes	MM-PB/GBSA binding energy terms in kcal/mol							LIE (kcal/mol)
	VDWAALS	E _{EEL}	E _{PB/GB}	E _{SURF}	ΔG_{gas}	ΔG_{solv}	ΔT_{TOTAL}	ΔG_{bind}
FLT3-M01	-71.17	-51.49	68.69	-8.82	-122.66	59.86	-62.80	-18.39
FLT3-M03	-72.23	-275.18	295.32	-8.72	-347.42	286.60	-60.27	-13.43
FLT3-M17	-53.77	-21.27	34.20	-6.48	-75.04	27.72	-47.32	-12.56
FLT3-M20	-69.48	-222.37	240.20	-9.02	-291.85	231.17	-60.68	-19.35
FLT3-M24	-62.38	-28.98	39.64	-7.83	-91.36	31.80	-59.56	-9.75
FLT3-M34	-56.90	-248.23	262.59	-6.76	-305.14	255.82	-49.31	-8.14
FLT3 _(D835Y) -M01	-70.94	-51.33	67.31	-8.58	-122.28	58.72	-63.55	-18.39
FLT3 _(F691L) -M01	-69.99	-53.10	69.92	-8.86	-123.09	61.05	-62.03	-17.95
FLT3 _(D835Y,F691L) - M01	-70.43	-50.96	67.16	-8.56	-121.39	58.60	-62.79	-17.95

VDWAALS: van der Waals contribution from MM; E_{EEL}: electrostatic energy as calculated by the MM force field; E_{PB/GB}: electrostatic contribution to the solvation free energy; E_{SURF}: non-polar solvation free energy; ΔG_{gas} : ΔG in gas phase; ΔG_{solv} : ΔG in solvation state; ΔT_{TOTAL} : total binding free energy from MM-PB/GBSA; ΔG_{bind} : final LIE binding energy; LIE: linear interaction energy.

Table 2. Per-residue MM-PB/GBSA binding free energy decomposition in kcal/mol

Residues	Compounds								
	M01	M03	M17	M20	M24	M34	M01 (F691L)	M01 (D835Y)	M01 (F691L, D835Y)
K614	NA	NA	NA	NA	NA	NA	-1.92	-0.66	-0.93
L616	-2.53	-1.90	-2.20	-2.20	-1.13	-2.44	-2.35	-2.38	-2.63
V624	NA	-1.61	-0.60	-1.50	-1.54	NA	-1.68	-1.60	-1.48
A642	-1.25	-1.26	NA	-1.04	-1.00	-1.05	-1.24	-1.30	-1.38

K644	-2.15	-2.26	0.36	-1.43	-0.75	-0.38	-1.84	-1.24	-2.24
M665	-1.48	-1.63	-0.96	-0.99	-1.39	-0.46	-1.46	-1.40	-1.42
I674	-1.03	-1.26	NA	-0.98	NA	NA	-0.92	-1.29	-0.91
V675	-1.81	NA	NA	-0.98	-1.64	-0.93	-1.69	-1.67	-1.62
F691	-2.44	-2.80	-3.63	-2.54	-2.76	-2.15	-1.40 *	-2.63	-1.09 *
Y693	-1.99	-2.04	-0.68	-1.73	-2.04	-1.38	-1.77	-1.98	-1.98
C694	-2.82	-2.24	-0.13	-0.74	-2.41	-0.77	-2.77	-2.72	-2.81
G697	-1.39	-0.77	-1.47	-1.29	-0.39	-1.72	-1.33	-1.41	-1.40
L818	-1.76	-1.67	-1.89	-1.67	-1.56	-1.70	-1.78	-1.68	-1.64
C828	-2.81	-2.86	NA	NA	-4.17	-0.96	-2.79	-3.27	-2.42
D829	-1.22	-1.41	-2.21	-1.12	-0.94	-2.08	-1.19	-1.29	-1.73
F830	-0.78	-2.35	-1.63	-1.71	-1.17	-1.61	NA	NA	NA

NA: Distance of the residues that are more than 4 Å from the compounds or contributed negligible binding energy to the ligand; (*): BE decomposition from the mutated residue.

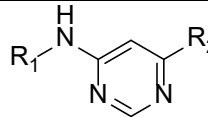
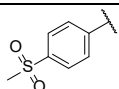
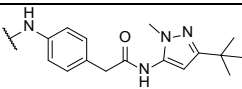

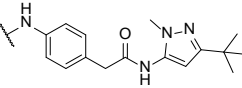
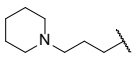
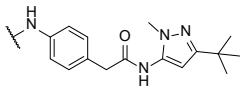
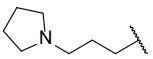
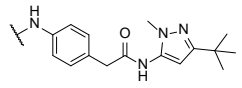
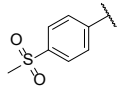
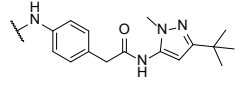
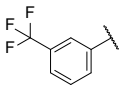
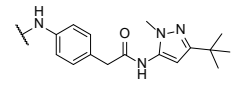
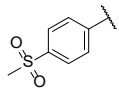
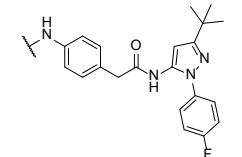
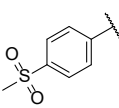
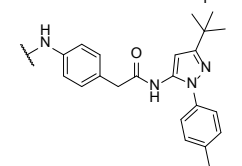
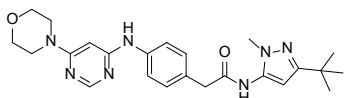
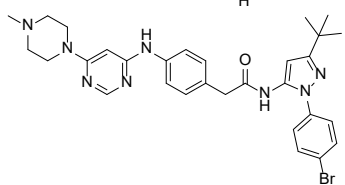
During the calculation process, we restricted the analysis of per-residue MM-PB/GBSA decomposition within the 4.0 Å distance from the ligand atoms. The per-residue BE decomposition between the F830 of FLT3 mutants and compound M01 was not generated, suggesting that the distance between M01 and F830 increased to more than 4.0 Å, whereas the appearance of BE decompositions from residue K614 signify its proximity to compound M01. However, rather lower free energy values were observed in the VDWAALS and E_{EL} energy terms with the FLT3_(F691L)-M01 complex, as shown in Figure 6b, c. Minor energy differences could occur due to the substitution of the bulky hydrophobic residue F691 with a non-bulky leucine residue, as F691 is the key gatekeeper residue and had a strong π - π interaction with the phenyl ring of M01. The LIE values were found to be -18.39, -17.75, and -17.95 kcal/mol for F691L, and both mutant variants of FLT3 to compound M01, respectively. Residue D835Y mutation in the activation loop was 14.1 Å away from the active site. This mutation in the DFG-out FLT3 did not influence the final BE to the type II-like M01 in the estimation of MM-PB/GBSA or LIE. Therefore, we assumed that the F691L mutation could partially affect selectivity and ligand interaction by changing the

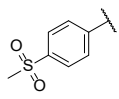
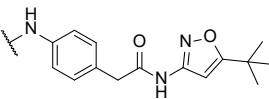
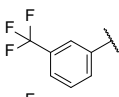
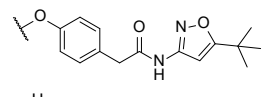
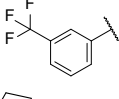
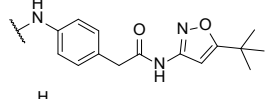
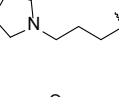
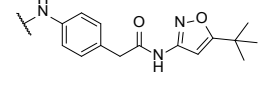
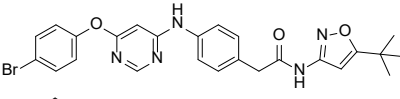
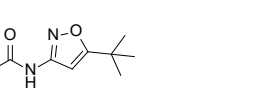
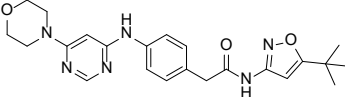

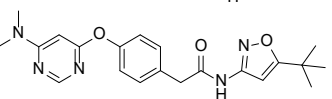
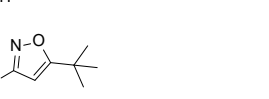

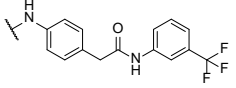
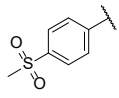
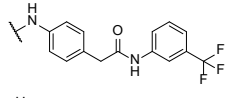
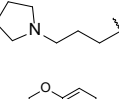
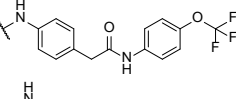

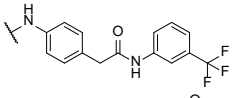

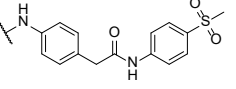
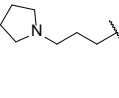
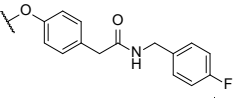

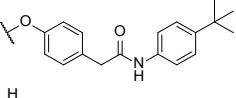
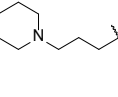
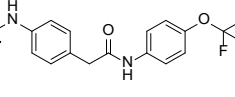
hydrophobic property of the nucleotide-binding pocket. This could also decrease the relative competitiveness of M01 against ATP and induce drug resistance [24,25]. In contrast, the D835Y mutation in the activation loop is strongly associated with the resistance mechanism of the active conformation (DFG-in) tyrosine kinase inhibitors (TKI). A pathway of the allosteric network has been proposed between the activation loop and the DFG motif [26], which plays an important role in stabilizing the adenosine triphosphate (ATP) molecule by chelating the Mg^{2+} ions at the catalytic site. The D835Y mutation could trigger an alteration of the allosteric mechanism in FLT3 by forming a dead mutant of kinases or disfavoring inhibitor binding by increasing entropy, ultimately influencing the entry of the ligand into the ATP-binding pocket [27].


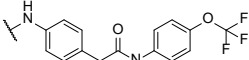
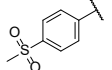
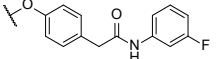
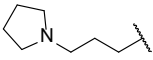
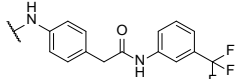
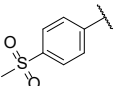
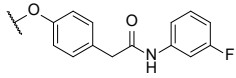

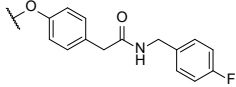
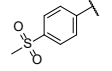
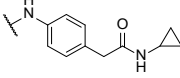
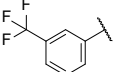
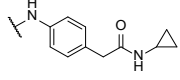
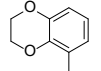
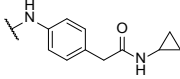
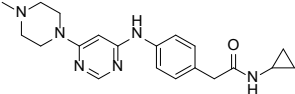

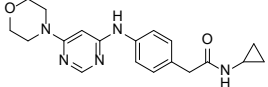

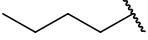
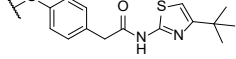
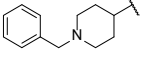
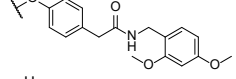
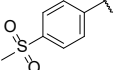
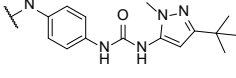

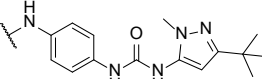
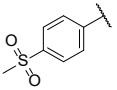
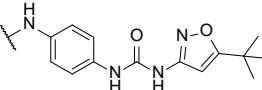
3.4. Dataset building, 3D-QSAR model development, and model validation

Coherence selection of the training set and test set compounds is a vital step toward the development of a 3D-QSAR statistical model. The chemical structures and their respective pIC_{50} values, which were well spanned over three log units, are illustrated in Table 3. Initially, we classified the compounds into high, medium, and low activity groups. Next, we chose compounds at random from each group to form the test set while maintaining structural diversity. We developed two sets of 3D-QSAR models, and their statistical validation is summarized in Table 4. The final 1 ns MD average structure of compound M01 was regarded as a 3D bioactive conformer and selected as a template molecule, as described here [28]. The rest of the compounds were modeled based on the template molecule and aligned with the common skeleton of *N*-methylpyrimidine-4-amine. The maximum common substructure (MCS) functionality was used to find the common core. The alignment of the dataset compounds inside the hydrophobic cleft and their unique chemical core are shown in Figure 6a,b.

Table 3. Structure and activity (pIC₅₀) values of *N*-methylpyrimidine-4-amine based FLT3 inhibitors

#Cpd	Structure A		pIC ₅₀	
				
	Structure	R ₁	R ₂	
01	A			7.85
02	A			7.42
03	A			6.88
04	A			6.42
05	A			6.57
06	A			6.01
07	A			5.49
08	A			5.31
09				6.36
10				5.22

11*	A			7.20
12	A			6.24
13	A			5.92
14	A			5.98
15				4.82
16				6.40
17*				5.05
18*	A	H_3C 		6.38
19	A			6.06
20*	A			6.04
21*	A			5.87
22	A	H_3C 		5.67
23*	A			5.34
24*	A	H_3C 		5.29
25	A			5.18

26*	A			5.13
27	A			5.10
28	A			5.03
29*	A			4.93
30	A			4.89
31	A			5.25
32	A			5.09
33	A			4.97
34				4.86
35				4.82
36	A			5.12
37	A			4.86
38	A			7.38
39	A			7.41
40*	A			7.53

(*) Test set compounds. #Cpd.: Compounds

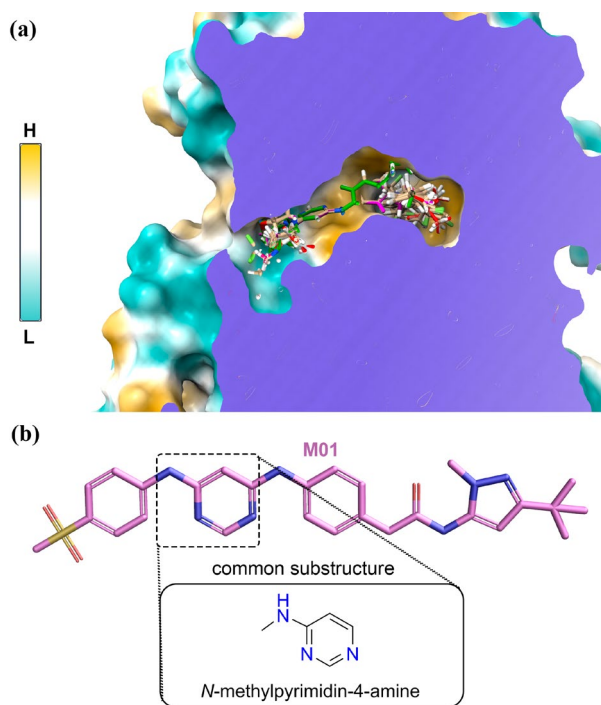


Figure 6. Common substructure-based molecular alignment. **(a)** Alignment of the compounds is shown inside the hydrophobic binding pocket of FLT3 by Z-plane clipping. Cyan to gold-yellow color bar indicate low (L) to high (H) hydrophobicity. **(b)** Common substructure of the compounds in the dataset.

The first model was built by taking 40 compounds altogether. We strictly adhered to the statistical parameters listed in the threshold values column, as well as those described in the Methodology Section when validating each model. In the first 3D-QSAR model, we obtained the best q^2 and r^2 values of 0.735 and 0.956, respectively, at an ONC of 6 in the CoMFA scheme. The SEHA produced the best statistical CoMSIA model with q^2 and r^2 values of 0.725 and 0.912, respectively at an ONC of 5 among the different combinations of descriptor fields. The q^2 and r^2 values of both CoMFA and CoMSIA appeared to be greater than the accepted threshold value ($q^2 > 0.5$ and $r^2 > 0.6$), which indicated a good agreement for the internal validation of both models. The χ^2 and RMSE values were found to be 0.052 and 0.219, respectively. In CoMSIA, the χ^2 and RMSE values were found to be 0.078 and 0.265, respectively.

The χ^2 and RMSE values are within the acceptable threshold value, which signifies the fitness and accuracy of the models. However, any QSAR model would be uncertain without being validated externally. Therefore, the second model was developed using the training set of 30 compounds. The remaining 10 compounds were used as a test set to analyze the external predictivity. In the second 3D-QSAR model, the training set compounds generated the best q^2 and r^2 values of 0.802 and 0.960, respectively, at an ONC of 6 in the CoMFA scheme. The χ^2 ($\chi^2 = 0.012$) and RMSE (RMSE = 0.119) values were determined within the acceptable range of <0.3 and <0.5 [29], indicating for a good internal validation and model's fitness. For external validation, we calculated multiple statistical parameters described in these studies [30–32]. The k and k' were found to be 1.033 and 0.962, and the $|r_0^2 - r'^2|$, $(r^2 - r_0^2) / r^2$, r_m^2 , and Δr_m^2 were predicted to be 0.109, 0.041, 0.724, and 0.059, respectively. In addition, we acquired the Q_{F3}^2 and Q_{CC}^2 metrics determination to evaluate the predictive ability of the CoMFA model. The Q_{F3}^2 value was found to be 0.698, which greater than the acceptable range (>0.6), and a high Q_{CC}^2 value of 0.821 signifies the true predictive power of the model. To generate the best possible CoMSIA model, we applied five molecular descriptor fields i.e., steric (S), electrostatic (E), hydrophobic (H), H-bond acceptor (A), and H-bond donor (D) in different combinations. The combination of SHD produced the best statistical q^2 values ($q^2 = 0.730$) at an ONC of 5. However, the best q^2 value does not necessarily indicate a good predictivity. Instead, the best CoMSIA model was selected based on the highest Q_{F3}^2 metrics value. We found that the SEHD combination produced q^2 and r^2 values of 0.725 and 0.965, respectively, at an ONC of 5. The model had the Q_{F3}^2 value of 0.665, the highest among all. Despite this, the model successfully passed all other statistical parameters. The k and k' were found to be 1.033 and 0.961, and the $|r_0^2 - r'^2|$, $(r^2 - r_0^2) / r^2$, r_m^2 , and Δr_m^2 were predicted to be 0.109, 0.041, 0.724, and 0.059, respectively. Thereafter, the SEHD combination was selected as the final CoMSIA model. Overall, the second QSAR model, including CoMFA and CoMSIA (SEHD), produced the highest statistically significant values; therefore, we selected these models for contour map generation and SAR analysis. The

PLS regression graph of CoMFA and CoMSIA is shown in Figure 7a,c, which described that our 3D-QSAR model could adequately predict the pIC_{50} values of the test set compounds. The contribution of the steric and electrostatic field was found to be 54.4% and 45.6% in the CoMFA model, respectively, while the contributions of the steric, electrostatic, hydrophobic, and H-bond donor were found to be 20.5%, 29.9%, 32.8%, and 16.8%, respectively, in the CoMSIA model. Both models established good agreement between experimental activity and predicted activity by showing acceptable criteria ($q^2 > 0.6$, $r^2 > 0.8$) while having a satisfactory predictive performance ($Q^2_{F3} > 0.5$).

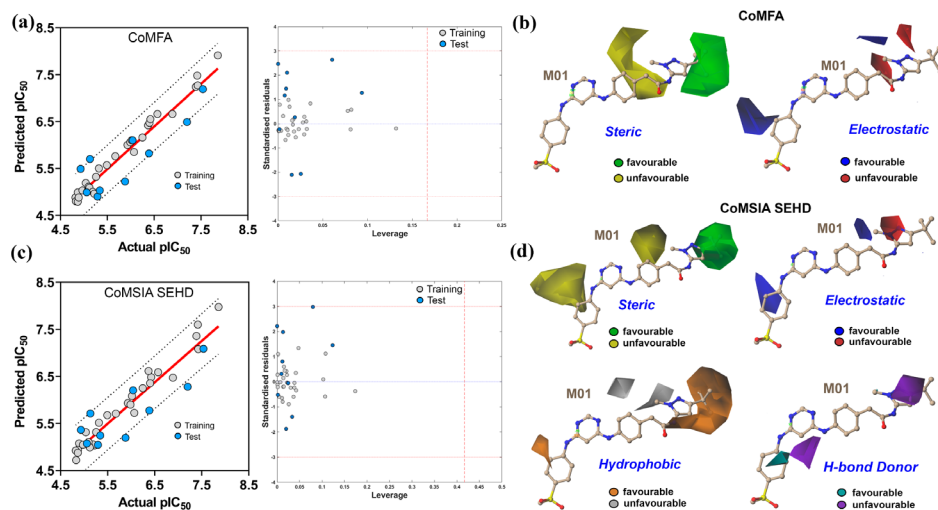


Figure 7. PLS scatter plot, Williams plot, and CoMFA-CoMSIA contour maps analysis. (a) Scatter plot of actual vs. predicted pIC_{50} values from the CoMFA and the corresponding Williams plot for AD analysis. (b) The green and yellow contours from CoMFA indicate the favorable and unfavorable substitution for the steric groups. The blue and red contours indicate the favorable and unfavorable positions for the electropositive groups. (c) Scatter plot of actual vs. predicted pIC_{50} values from the CoMSIA and the corresponding William's plot for AD analysis. (d) From CoMSIA, green and yellow contours signify the favorable and unfavorable positions for the steric group, whereas the blue and red contours show the favorable and unfavorable substitutions for electropositive groups. The orange and gray contours show the favorable and unfavorable positions for the hydrophobic groups. The cyan and purple colors signify the favorable and unfavorable positions for the H-bond donor groups.

Table 4. Statistics of the CoMFA and CoMSIA models

Statistical Parameters	3D-QSAR (All Compounds)		3D-QSAR (Training Set Compounds)					Threshold Values
	CoMFA	CoMSIA (SEHA)	CoMFA	CoMSIA (SHD)	CoMSIA (SEHA)	CoMSIA (SEHD)	CoMSIA (SEHAD)	
q^2	0.735	0.725	0.802	0.730	0.726	0.725	0.721	>0.5
ONC	6	5	6	5	5	5	5	
SEP	0.502	0.503	0.452	0.517	0.521	0.522	0.525	
r^2	0.956	0.912	0.983	0.960	0.962	0.965	0.956	>0.6
SEE	0.204	0.284	0.134	0.199	0.194	0.186	0.209	<<1
F-value	119.97	70.84	216.62	114.54	121.34	131.90	104.16	>100
χ^2	0.052	0.078	0.012	0.028	0.027	0.023	0.028	<0.5
RMSE	0.219	0.265	0.119	0.181	0.176	0.169	0.189	<0.3
MAE	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	≈ 0
RSS	1.873	2.744	0.412	0.953	0.904	0.834	1.046	
k	NA	NA	1.033	1.022	1.032	1.033	1.037	$0.85 \leq k \leq 1.15$
k'	NA	NA	0.962	0.969	0.962	0.961	0.958	$0.85 \leq k' \leq 1.15$
$ r_0^2 - r'^2_0 $	NA	NA	0.109	0.307	0.250	0.224	0.192	<0.3
$(r^2 - r_0^2) / r^2$	NA	NA	0.041	0.002	0.012	0.006	0.117	<0.1
r_m^2	NA	NA	0.724	0.603	0.635	0.649	0.664	>0.5
$\overline{r_m^2}$	NA	NA	0.694	0.449	0.492	0.511	0.623	>0.5
Δr_m^2	NA	NA	0.059	0.307	0.286	0.274	0.083	
Q_{F3}^2	NA	NA	0.698	0.604	0.656	0.668	0.660	>0.6
Q_{ccc}^2	NA	NA	0.821	0.746	0.778	0.787	0.787	

S (%)	52.9	20.2	54.4	28.9	18.5	20.5	16.2
E (%)	47.1	29.9	45.6	NA	28.8	29.9	24.5
H%	NA	37.0	NA	47.3	33.9	32.8	28.3
A%	NA	13.0	NA	NA	18.8	NA	16.0
D%	NA	NA	NA	23.8	NA	16.8	15.1

q^2 : squared cross-validated correlation coefficient; **ONC**: optimal number of components; **SEP**: standard error of prediction; r^2 : squared correlation coefficient; **SEE**: standard error of estimation; **F-value**: F-test value; **RMSE**: root-mean-squared error; **MAE**: mean absolute error; **RSS**: residual sum of error; **k**: slope of the predicted vs. observed activity at zero intercept; **k'**: slope of the observed vs. predicted activity at zero intercept; r_0^2 : squared correlation coefficient between predicted and observed activity; r^2_0 : squared correlation coefficient between predicted and observed activity; Q^2_{F3} : Q^2 metrics for external test set validation; Q^2_{ccc} : concordance correlation coefficient; **S**: steric; **E**: electrostatic; **H**: hydrophobic; **A**: H-bond acceptor; **D**: H-bond donor; **NA**: Not applicable.

Table 5. Progressive scrambling results from the CoMFA and CoMSIA (SEHD) training set compounds

Components	CoMFA			CoMSIA (SEHD)		
	Q^2	cSDEP	$dq^2/dr^2_{yy'}$	Q^2	cSDEP	$dq^2/dr^2_{yy'}$
1	0.191	0.828	0.175	0.230	0.807	0.306
2	0.358	0.750	0.731	0.429	0.707	0.996
3	0.477	0.689	0.855	0.491	0.662	1.052
4	0.489	0.696	1.480	0.550	0.652	1.221
5	0.479	0.711	1.821	0.518	0.770	0.982
6	0.502	0.713	1.198	0.410	0.709	1.513
7	0.518	0.718	1.663	0.400	0.796	2.285

The progressive scrambling method was applied to evaluate the additional stability of the CoMFA and CoMSIA models, as shown in Table 5. We ran 100 independent scrambles with minimum and maximum bin sizes of 2 and 10. At component number 6, scrambling Q^2 and cSDEP were measured to be 0.502 and 0.713, respectively. The $dq^2/dr^2_{yy'}$ did not exceed the value 1.2 ($dq^2/dr^2_{yy'} = 1.198$) in the CoMFA. In the CoMSIA, the scrambling Q^2 , cSDEP, and $dq^2/dr^2_{yy'}$ were found to be 0.518, 0.709, and

0.982, respectively. Progressive scrambling helps to identify the optimal number of components of the model. In addition, it detects the model's sensitivity over a small perturbation when applied to the data.

The QSAR models are generated by a limited number of compounds and can predict inhibitory activity for an unknown chemical that has a very similar chemical constitution. To assess the reliability of the CoMFA and CoMSIA models and specify the outliers, we used the applicability domain analysis by distance-based Williams plot as described in previous work [33]. The CoMFA and CoMSIA PLS plots, and the corresponding Williams plots, portrayed the standardized residuals of the training set and test set compounds against their leverage values. Compounds with a high leverage (h_i) value greater than the warning leverage (h^*) can be detected as outliers and have a substantial impact on the fitness of the model. We used the *Applicability Domain toolbox*, a MATLAB package developed by Sahigara et al. [34] to perform the AD analysis. In our study, all the compounds fell within the ± 3 standardized residuals and the estimated warning leverages in CoMFA ($h^* = 0.16$) and CoMSIA ($h^* = 0.42$), confirming the overall predictive reliability of the models.

3.5. CoMFA and CoMSIA contour map analysis

We generated 3D contour maps around the most active compound M01 in the SYBYL-X2.1 suit to elucidate the structure–activity relationships. The information collected from the contour map could be useful to improve the inhibitory potency of small molecules by altering the chemical groups. Figure 7b displayed the contour maps of the CoMFA analysis. The green and yellow contours around the pyrazole ring represent the favorable and unfavorable substitutions for the bulky and steric chemical groups. Specifically, the location of the green contour arrived near the residues M665 and V675, with which M01 formed the hydrophobic interaction in the docking study, and these residues contributed a BE of -1.48 and -1.81 kcal/mol, respectively. Compounds M01–M06, M11, and M38–M40 had steric groups in their R_2 position at the green contour, which could explain why its inhibitory activity is higher than that of compounds M31–M35.

In contrast, compounds M07, M08, and M10 had the steric phenyl ring facing the yellow contour at the R₂ position, which could decrease their inhibitory potency due to the probability of the steric hindrance effect. In Figure 8b, the blue and red contours show the favorable positions for the electropositive and electronegative groups. Compounds M01–M05, M38–M39, and M40 had the N atom in their pyrazole ring in the R₂ position and exhibited a higher inhibitory activity. Besides, a blue contour near the red contours at the R₂ position was found to show an unfavorable position for an electronegative group. Compounds M12–M16, M36, and M37 bearing the oxygen atom on their isoxazole ring and compounds M23, M27, M29, and M30 bearing the electronegative fluorine atom in the position R₂ exhibited a lower inhibitory activity than M01–M05. For the same reason, compound M40 had less inhibitory potency than compound M01. Another blue contour was present near the phenyl ring at the R₁ position, showing the favorable position of an electropositive chemical group. For compounds M15, M22, M24, and M35, an electropositive group was absent and showed lower inhibitory activity than compound M16.

Figure 7d illustrates the contour maps from the CoMSIA model analysis. We found that the steric and electrostatic descriptors yielded contour maps similar to those produced by CoMFA model. As a result, they are not discussed further. It also provided additional information on the hydrophobic and H-bond donor descriptors fields, making it preferred for SAR study. The favorable and unfavorable substitutions for hydrophobic chemical groups are shown by an orange-gray color scheme. A large orange contour at position R₂ around the pyrazole ring indicated that a hydrophobic chemical group would be beneficial in increasing inhibitory activity. In compounds M18–M21, M25, and M28, the methyl groups were replaced by fluorine atoms, and the presence of the methylsulfonyl group in M22 may affect the inhibitory activity of these compounds.

Similarly, the favorable and unfavorable positions for the H-bond donor groups are shown by cyan and purple contours. For this reason, compound M36 had an N atom in the R₂ position toward the purple contour and exhibited less inhibitory activity.

From the analysis of the CoMFA and CoMSIA contour maps, we summarized an ideal SAR scheme in Figure 8. A positive charge group and an H-bond donor group on the methyl phenyl sulfone ring at position R₁ could enhance the potency of M01. At that position, residue C694 formed two H-bond interactions with the M01 and contributed -2.82 kcal/mol MM-PB/GBSA Δ TOTAL binding free energy. As a result, a hydrophobic group next to this residue would most likely hinder the formation of an H-bond, lowering the binding free energy decomposition.

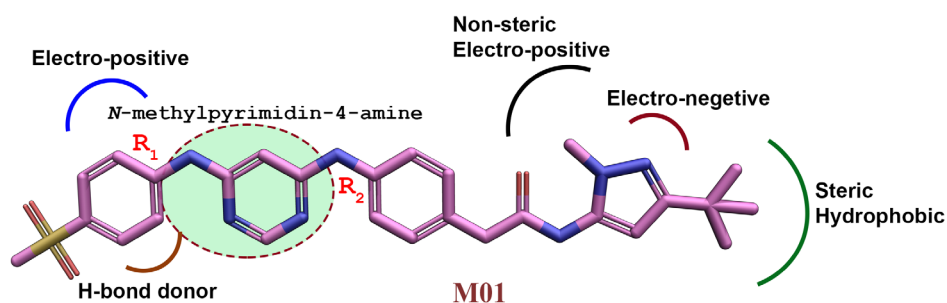


Figure 8. The three-dimensional structure–activity relationship by taking compound M01 as a reference.

Besides, the substitution of a smaller electron-donating group and the addition of a bulky hydrophobic group on the methylpyrazole ring in R₂ could improve the bioactivity of the compounds. However, the docking and MD study suggested that the steric substitution at the R₂ position was not infinite. A larger hydrophobic substitution could cause steric clashes with the hydrophobic residues in the α C-helix. Thereby, a smaller chemical group with steric and hydrophobic properties would be the preferred alteration strategy for an efficient interaction.

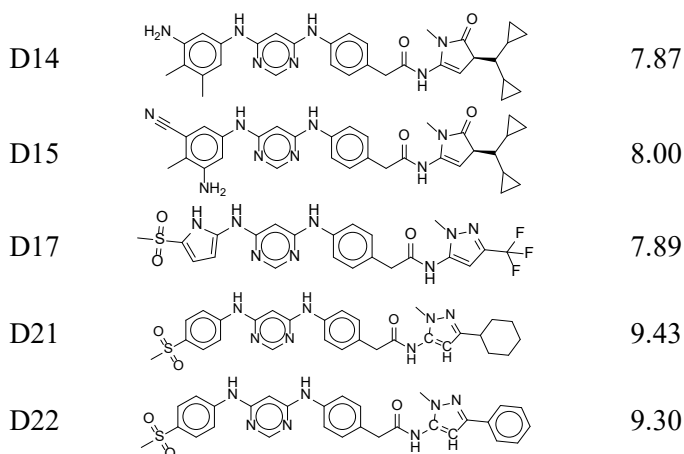
3.6. Designing of new compounds

In the context of SAR analysis, we designed 30 new compounds using M01 as a template and evaluated their predictive activity in the CoMSIA-SHED model. Of these, 16 designed compounds were predicted to have higher inhibitory activity (pIC₅₀) than the most active compound M01 (Table 6). We re-docked these

compounds against FLT3 and determined their physicochemical property and SA score. An SA score predicts the difficulty level in synthesizing the unknown chemical compounds. An SA score of 1 indicates the easy synthesis route, while an SA score of 10 signifies the difficult synthesis route. The SA score of the newly proposed compounds ranged from 3 to 5, suggesting that the compounds would be simple to moderately challenging to synthesize.

Table 6. Designed new compounds with higher predicted pIC₅₀ values than the most active compound M01

Compounds	Structure	Predicted pIC ₅₀
D01		7.90
D02		7.93
D03		7.92
D04		7.93
D05		7.86
D07		7.84
D08		7.84
D09		7.91
D10		8.10
D11		7.80
D12		7.86



D: Designed

The designed compounds with FLT3-bound complexes converged well within the 50 ns of the MD simulation; therefore, they were not extended further. We calculated the MM-PB/GBSA binding free energy of the last 2 ns of each trajectory (Table 7 and Figure 9). The designed compounds, D02–D08, D12, D15, and D22 showed a higher binding free energy compared to the most active compound. The results strongly suggest that SAR-assisted newly designed compounds could have a better binding affinity for FLT3.

Table 7. MM-PB/GBSA binding energy estimation between FLT3 and designed compounds

Complexes	MM-PB/GBSA binding energy terms in kcal/mol						
	VDWAALS	E _{EEL}	E _{PB/GB}	E _{SURF}	ΔG _{gas}	ΔG _{solv}	ΔTOTAL
FLT3-D01	-71.01	-45.14	66.81	-8.84	-116.16	57.96	-58.19
FLT3-D02	-70.82	-32.90	49.71	-9.17	-103.73	40.53	-63.19
FLT3-D03	-71.38	-55.60	71.34	-9.04	-126.98	62.30	-64.68
FLT3-D04	-72.37	-54.24	71.68	-9.30	-126.62	62.38	-64.24
FLT3-D05	-76.94	-63.43	79.81	-9.90	-140.38	69.90	-70.47

FLT3-D07	-73.59	-48.56	60.95	-9.45	-122.15	51.49	-70.66
FLT3-D08	-72.79	-33.30	50.01	-8.72	-106.10	41.29	-64.81
FLT3-D09	-74.03	-54.80	78.07	-9.34	-128.84	68.73	-60.10
FLT3-D10	-69.67	-49.27	71.01	-9.41	-118.94	61.59	-57.35
FLT3-D11	-80.40	-37.05	71.92	-10.24	-117.45	61.68	-55.77
FLT3-D12	-76.69	-27.84	46.29	-9.39	-104.53	36.90	-67.63
FLT3-D14	-80.42	-28.40	56.25	-10.07	-108.82	46.17	-62.65
FLT3-D15	-78.07	-57.04	70.96	-9.45	-135.11	61.50	-73.61
FLT3-D17	-66.50	-64.74	81.42	-8.66	-131.24	72.75	-58.49
FLT3-D21	-72.32	-17.78	46.93	-9.08	-90.11	37.84	-52.27
FLT3-D22	-73.23	-25.85	45.39	-9.44	-79.08	35.94	-63.14

VDWAALS: van der Waals contribution from MM; E_{EL} : electrostatic energy as calculated by the MM force field; $E_{PB/GB}$: electrostatic contribution to the solvation free energy; E_{SURF} : non-polar solvation free energy; ΔG_{gas} : ΔG in gas phase; ΔG_{sol} : ΔG in solvation state; $\Delta TOTAL$: total binding free energy from MM-PB/GBSA.

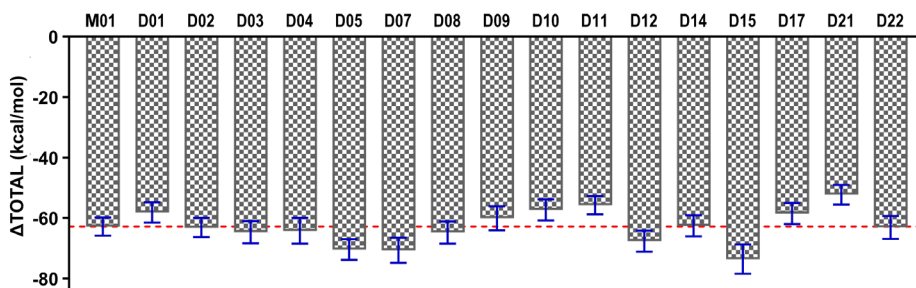


Figure 9. Comparison of the final $\Delta TOTAL$ binding energy terms of the newly designed compounds. Compounds D02–D08, D12, D15, and D22 were estimated to have higher binding free energy compare to M01 (red dashed line) and the standard deviation sign is shown in blue.

4. Conclusions

The development of FLT3 inhibitors is a promising strategy for achieving the therapeutic goal of the treatment of AML. In this work, we used molecular docking and MD simulations to better understand the crucial interaction and stability of the ligands inside the binding pocket. The protein–ligand affinity was estimated by computing the MM-PB/GBSA and LIE-binding energy, including the mutant FLT3 complexes. In a previously conducted docking study, residues Q575, L576, Q577, K644, F691, C694, L818, and L832 alongside the DFG-motif (D829, F830, and G831) formed the fundamental docking interface for a group of diverse inhibitors. In our study, the combination of docking and MD study exposed several key residues, such as L616, M665, F691, E692, N701, L818, and F830 inside the ATP-binding pocket, which could be responsible for the affinity and selectivity of the type II inhibitors. K644 and C694 are the two important residues that form H-bond interactions. Subsequently, we established the CoMFA ($q^2 = 0.802$, $r^2 = 0.983$) and CoMSIA ($q^2 = 0.725$, $r^2 = 0.965$) models, both of which show a reasonable statistical correlation between actual and predictive activity and internal verification ability. The models also showed satisfactory stability and sensitivity in progressive scrambling analysis. The Q_{F3}^2 metrics from CoMFA and CoMSIA were found to be 0.698 and 0.668, respectively, indicating the external predictive power. The models could predict the activity values of new compounds having a similar scaffold. Following that, we developed the CoMFA and CoMSIA contour maps for the SAR study. Contour maps could provide crucial information regarding how chemical group substitution may improve the inhibitory activity of chemical compounds, which may be explained further by their docking and MD poses. Finally, we designed 30 new compounds, 16 of which had a higher predictive pIC_{50} than compound M01. In addition, free energy calculation of the selected designed compounds revealed a greater binding affinity to the FLT3. Compounds D02–D08, D12, D15, and D22, in particular, might offer potential inhibitory activity against FLT3.

In this study, several effective computational approaches and reliable statistics were used, and they may provide several key mechanistic interpretations at the molecular level. The overall findings of this work may be beneficial in delivering theoretical guidance in the future development and synthesis of *N*-phenylpyrimidine-4-amine-based FLT3 inhibitors.

References

1. Wang, Z.; Cai, J.; Cheng, J.; Yang, W.; Zhu, Y.; Li, H.; Lu, T.; Chen, Y.; Lu, S. FLT3 Inhibitors in Acute Myeloid Leukemia: Challenges and Recent Developments in Overcoming Resistance. *J. Med. Chem.* **2021**, *64*, 2878–2900, doi:10.1021/acs.jmedchem.0c01851.
2. Cueto, F.; Sancho, D. The Flt3L/Flt3 Axis in Dendritic Cell Biology and Cancer Immunotherapy. *Cancers* **2021**, *13*, 1525, doi:10.3390/cancers13071525.
3. Pannecoucke, E.; Raes, L.; Savvides, S.N. Engineering and crystal structure of a monomeric FLT3 ligand variant. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2021**, *77*, 121–127, doi:10.1107/s2053230x21003289.
4. Yen, S.-C.; Chen, L.-C.; Huang, H.-L.; Ngo, S.-T.; Wu, Y.-W.; Lin, T.E.; Sung, T.-Y.; Lien, S.-T.; Tseng, H.-J.; Pan, S.-L.; et al. Investigation of Selected Flavonoid Derivatives as Potent FLT3 Inhibitors for the Potential Treatment of Acute Myeloid Leukemia. *J. Nat. Prod.* **2021**, *84*, 1–10, doi:10.1021/acs.jnatprod.0c00589.
5. Wilson, K.R.; Villadangos, J.A.; Mintern, J.D. Dendritic cell Flt3—regulation, roles and repercussions for immunotherapy. *Immunol. Cell Biol.* **2021**, *99*, 962–971, doi:10.1111/imcb.12484.
6. Almatani, M.F.; Ali, A.; Onyemaechi, S.; Zhao, Y.; Gutierrez, L.; Vaikari, V.P.; Alachkar, H. Strategies targeting FLT3 beyond the kinase inhibitors. *Pharmacol. Ther.* **2021**, *225*, 107844, doi:10.1016/j.pharmthera.2021.107844.
7. Hasegawa, H.; Taniguchi, H.; Nakamura, Y.; Kato, T.; Fujii, S.; Ebi, H.; Shiozawa, M.; Yuki, S.; Masuishi, T.; Kato, K.; et al. FMS-like tyrosine kinase 3 (FLT3) amplification in patients with metastatic colorectal cancer. *Cancer Sci.* **2021**, *112*, 314, doi:10.1111/cas.14693.
8. Kawase, T.; Nakazawa, T.; Eguchi, T.; Tsuzuki, H.; Ueno, Y.; Amano, Y.; Mori, M.; Yoshida, T. Effect of Fms-like tyrosine kinase 3 (FLT3) ligand (FL) on antitumor activity of gilteritinib, a FLT3 inhibitor, in mice xenografted with FL-overexpressing cells. *Oncotarget* **2019**, *10*, 6111, doi:10.18632/oncotarget.27222.
9. Roskoski, R., Jr. The role of small molecule Flt3 receptor protein-tyrosine kinase inhibitors in the treatment of Flt3-positive acute myelogenous leukemias. *Pharmacol. Res.* **2020**, *155*, 104725.
10. Bjelosevic, S.; Gruber, E.; Newbold, A.; Shembrey, C.; Devlin, J.R.; Hogg, S.J.; Kats, L.; Todorovski, I.; Fan, Z.; Abrehart, T.C.; et al. Serine Biosynthesis Is a

-
- Metabolic Vulnerability in FLT3-ITD–Driven Acute Myeloid Leukemia. *Cancer Discov.* **2021**, *11*, 1582–1599, doi:10.1158/2159-8290.cd-20-0738.
11. Qiao, X.; Ma, J.; Knight, T.; Su, Y.; Edwards, H.; Polin, L.; Li, J.; Kushner, J.; Dzinic, S.H.; White, K.; et al. The combination of CUDC-907 and gilteritinib shows promising in vitro and in vivo antileukemic activity against FLT3-ITD AML. *Blood Cancer J.* **2021**, *11*, 1–17, doi:10.1038/s41408-021-00502-7.
 12. Scholl, S.; Fleischmann, M.; Schnetzke, U.; Heidel, F.H. Molecular Mechanisms of Resistance to FLT3 Inhibitors in Acute Myeloid Leukemia: Ongoing Challenges and Future Treatments. *Cells* **2020**, *9*, 2493, doi:10.3390/cells9112493.
 13. JA, D.A.; Pérez, G.; MM, P.E.; JL, B.L.; Martinelli, G.; Cerchione, C. FLT3 inhibitors in the treatment of Acute Myeloid Leukemia: Current status and future perspectives. *Minerva Med.* **2020**, *111*, 427–442.
 14. Bertoli, S.; Dumas, P.-Y.; Bérard, E.; Largeaud, L.; Bidet, A.; Delabesse, E.; Tavitian, S.; Gadaud, N.; Leguay, T.; Leroy, H.; et al. Outcome of Relapsed or Refractory FLT3-Mutated Acute Myeloid Leukemia before Second-Generation FLT3 Tyrosine Kinase Inhibitors: A Toulouse–Bordeaux DATAML Registry Study. *Cancers* **2020**, *12*, 773, doi:10.3390/cancers12040773.
 15. Bhujbal, S.P.; Keretsu, S.; Cho, S.J. Design of New Therapeutic Agents Targeting FLT3 Receptor Tyrosine Kinase Using Molecular Docking and 3D-QSAR Approach. *Let. Drug Des. Discov.* **2020**, *17*, 585–596, doi:10.2174/1570180816666190618104632.
 16. Bensinger, D.; Stubba, D.; Cremer, A.; Kohl, V.; Waßmer, T.; Stuckert, J.; Engemann, V.; Stegmaier, K.; Schmitz, K.; Schmidt, B. Virtual Screening Identifies Irreversible FMS-like Tyrosine Kinase 3 Inhibitors with Activity toward Resistance-Confering Mutations. *J. Med. Chem.* **2019**, *62*, 2428–2446, doi:10.1021/acs.jmedchem.8b01714.
 17. Smith, C.C.; Lin, K.; Stecula, A.; Sali, A.; Shah, N.P. FLT3 D835 mutations confer differential resistance to type II FLT3 inhibitors. *Leukemia* **2015**, *29*, 2390–2392, doi:10.1038/leu.2015.165.
 18. Bharate, J.B.; McConnell, N.; Naresh, G.; Zhang, L.; Lakkaniga, N.R.; Ding, L.; Shah, N.P.; Frett, B.; Li, H.-Y. Rational Design, Synthesis and Biological Evaluation of Pyrimidine-4,6-diamine derivatives as Type-II inhibitors of FLT3 Selective Against c-KIT. *Sci. Rep.* **2018**, *8*, 3722, doi:10.1038/s41598-018-21839-3.
 19. Yamaura, T.; Nakatani, T.; Uda, K.; Ogura, H.; Shin, W.; Kurokawa, N.; Saito, K.; Fujikawa, N.; Date, T.; Takasaki, M.; et al. A novel irreversible FLT3 inhibitor, FF-10101, shows excellent efficacy against AML cells with FLT3 mutations. *Blood* **2018**, *131*, 426–438, doi:10.1182/blood-2017-05-786657.
 20. Wu, T.-S.; Lin, W.-H.; Tsai, H.-J.; Hsueh, C.-C.; Hsu, T.; Wang, P.-C.; Lin, H.-Y.; Peng, Y.-H.; Lu, C.-T.; Lee, L.-C.; et al. Discovery of Conformational Control Inhibitors Switching off the Activated c-KIT and Targeting a Broad

- Range of Clinically Relevant c-KIT Mutants. *J. Med. Chem.* **2019**, *62*, 3940–3957, doi:10.1021/acs.jmedchem.8b01845.
21. Zorn, J.A.; Wang, Q.; Fujimura, E.; Barros, T.; Kuriyan, J. Crystal Structure of the FLT3 Kinase Domain Bound to the Inhibitor Quizartinib (AC220). *PLoS ONE* **2015**, *10*, e0121177, doi:10.1371/journal.pone.0121177.
 22. Gutiérrez, C.M.; Cáceres-Rojas, D.; Adasme-Carreño, F.; Palomo, I.; Fuentes, E.; Caballero, J. Docking and quantitative structure–activity relationship of bi-cyclic heteroaromatic pyridazinone and pyrazolone derivatives as phosphodiesterase 3A (PDE3A) inhibitors. *PLoS ONE* **2017**, *12*, e0189213, doi:10.1371/journal.pone.0189213.
 23. Caballero, J.; Morales-Bayuelo, A.; Navarro-Retamal, C. Mycobacterium tuberculosis serine/threonine protein kinases: Structural information for the design of their specific ATP-competitive inhibitors. *J. Comput. Mol. Des.* **2018**, *32*, 1315–1336, doi:10.1007/s10822-018-0173-3.
 24. Nakaoku, T.; Kohno, T.; Araki, M.; Niho, S.; Chauhan, R.; Knowles, P.P.; Tsuchihara, K.; Matsumoto, S.; Shimada, Y.; Mimaki, S.; et al. A secondary RET mutation in the activation loop conferring resistance to vandetanib. *Nat. Commun.* **2018**, *9*, 1–9, doi:10.1038/s41467-018-02994-7.
 25. Mondal, J.; Tiwary, P.; Berne, B.J. How a Kinase Inhibitor Withstands Gatekeeper Residue Mutations. *J. Am. Chem. Soc.* **2016**, *138*, 4608–4615, doi:10.1021/jacs.6b01232.
 26. Ahuja, L.G.; Aoto, P.; Kornev, A.P.; Veglia, G.; Taylor, S.S. Dynamic allostery-based molecular workings of kinase:peptide complexes. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15052–15061, doi:10.1073/pnas.1900163116.
 27. Barouch-Bentov, R.; Sauer, K. Mechanisms of drug resistance in kinases. *Expert Opin. Investig. Drugs* **2011**, *20*, 153–208, doi:10.1517/13543784.2011.546344.
 28. Zhang, L.; Tsai, K.-C.; Du, L.; Fang, H.; Li, M.; Xu, W. How to generate reliable and predictive CoMFA models. *Curr. Med. Chem.* **2011**, *18*, 923–930, doi:10.2174/092986711794927702.
 29. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C.P.; Agrawal, R.K. Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov.* **2011**, *3*, 511–519.
 30. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P.K.; Kar, S.; Das, R.N. Some case studies on application of “rm2” metrics for judging quality of quantitative structure-activity relationship predictions: Emphasis on scaling of response data. *J. Comput. Chem.* **2013**, *34*, 1071–1082, doi:10.1002/jcc.23231.
 31. Gramatica, P.; Sangion, A. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. *J. Chem. Inf. Model.* **2016**, *56*, 1127–1131, doi:10.1021/acs.jcim.6b00088.
 32. Todeschini, R.; Ballabio, D.; Grisoni, F. Beware of Unreliable Q2! A Comparative Study of Regression Metrics for Predictivity Assessment of QSAR

- Models. *J. Chem. Inf. Model.* **2016**, *56*, 1905–1913, doi:10.1021/acs.jcim.6b00277.
33. Abdizadeh, R.; Hadizadeh, F.; Abdizadeh, T. QSAR analysis of coumarin-based benzamides as histone deacetylase inhibitors using CoMFA, CoMSIA and HQSAM methods. *J. Mol. Struct.* **2020**, *1199*, 126961.
 34. Sahigara, F.; Ballabio, D.; Todeschini, R.; Consonni, V. Assessing the validity of QSARs for ready biodegradability of chemicals: An applicability domain perspective. *Curr. Comput. Drug Des.* **2014**, *10*, 137–147, doi:10.2174/1573409910666140410110241.
 35. Ghosh, S.; Keretsu, S.; Cho, S.J. 3D-QSAR, Docking and Molecular Dynamics Simulation Study of C-Glycosylflavones as GSK-3 β Inhibitors. *J. Chosun Natural Sci.* **2020**, *13*, 170–180.
 36. Ghosh, S.; Keretsu, S.; Cho, S.J. Computational Modeling of Novel Phosphoinositol-3-kinase γ Inhibitors Using Molecular Docking, Molecular Dynamics, and 3D-QSAR. *Bull. Korean Chem. Soc.* **2021**, *42*, 1093–1111.
 37. Ghosh, S.; Keretsu, S.; Cho, S.J. Designing of the N-ethyl-4-(pyridin-4-yl)benzamide based potent ROCK1 inhibitors using docking, molecular dynamics, and 3D-QSAR. *PeerJ* **2021**, *9*, e11951, doi:10.7717/peerj.11951.
 38. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718, doi:10.1002/jcc.20291.
 39. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713, doi:10.1021/acs.jctc.5b00255.
 40. Da Silva, A.W.S.; Vranken, W.F. ACPYPE-Antechamber python parser interface. *BMC Res. Notes* **2012**, *5*, 1–8.
 41. Keretsu, S.; Ghosh, S.; Cho, S. Molecular Modeling Study of c-KIT/PDGFR α Dual Inhibitors for the Treatment of Gastrointestinal Stromal Tumors. *Int. J. Mol. Sci.* **2020**, *21*, 8232, doi:10.3390/ijms21218232.
 42. Tresanco, M.S.V.; Valdes-Tresanco, M.E.; Valiente, P.A.; Frías, E.M. gmx_MMPBSA (Version v1.4.3). *Zenodo* **2021**, doi:10.5281/zenodo.4814044.
 43. Miller, B.R., III; McGee, T.D., Jr.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
 44. Rifai, E.A.; van Dijk, M.; Geerke, D.P. Recent Developments in Linear Interaction Energy Based Binding Free Energy Calculations. *Front. Mol. Biosci.* **2020**, *7*, 7, doi:10.3389/fmolb.2020.00114.
 45. Ngo, S.T.; Hung, H.M.; Nguyen, M.T. Fast and accurate determination of the relative binding affinities of small compounds to HIV-1 protease using non-equilibrium work. *J. Comput. Chem.* **2016**, *37*, 2734–2742, doi:10.1002/jcc.24502.

-
46. Rifai, E.A.; Van Dijk, M.; Vermeulen, N.P.E.; Yanuar, A.; Geerke, D.P. A Comparative Linear Interaction Energy and MM/PBSA Study on SIRT1–Ligand Binding Free Energy Calculation. *J. Chem. Inf. Model.* **2019**, *59*, 4018–4033, doi:10.1021/acs.jcim.9b00609.
 47. Bang, S.; Cho, S. Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Index Analysis (CoMSIA) Study of Mutagen X. *Bull. Korean Chem. Soc.* **2004**, *25*, 1525–1530, doi:10.5012/bkcs.2004.25.10.1525.
 48. Gadhe, C.G.; Madhavan, T.; Kothandan, G.; Cho, S.J. In Silico Quantitative Structure–Activity Relationship Studies on P-gp Modulators of Tetrahydroisoquinoline-Ethyl-Phenylamine Series. *BMC Struct. Biol.* **2011**, *11*, 5, doi:10.1186/1472-6807-11-5.
 49. Zięba, A.; Laitinen, T.; Patel, J.; Poso, A.; Kaczor, A. Docking-Based 3D-QSAR Studies for 1,3,4-oxadiazol-2-one Derivatives as FAAH Inhibitors. *Int. J. Mol. Sci.* **2021**, *22*, 6108, doi:10.3390/ijms22116108.
 50. Gadhe, C.G.; Kothandan, G.; Cho, S.J. Large variation in electrostatic contours upon addition of steric parameters and the effect of charge calculation schemes in CoMFA on mutagenicity of MX analogues. *Mol. Simul.* **2012**, *38*, 861–871, doi:10.1080/08927022.2012.659182.
 51. San Juan, A.A.; Cho, S.J. 3D-QSAR study of microsomal prostaglandin E 2 synthase (mPGES-1) inhibitors. *J. Mol. Model.* **2007**, *13*, 601–610.
 52. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 1–13, doi:10.1038/srep42717.
 53. Pires, D.E.V.; Blundell, T.L.; Ascher, D.B. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **2015**, *58*, 4066–4072, doi:10.1021/acs.jmedchem.5b00104.

Suparna Ghosh Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

PART II

Binding studies and lead generation of pteridine-7(8H)-one derivatives targeting FLT3

1. Introduction

FLT3 is best described for its pivotal role in the constitutive activation and development of acute myeloid leukemia (AML) in humans[1]. It is primarily expressed in murine and hematopoietic stem cells and is responsible for the natural development of the immune system. Structurally, FLT3 consists of five immunoglobulin (Ig)-like extracellular domains, a single transmembrane (TM) domain, a juxtamembrane (JM) domain inside the cytoplasm, a cytoplasmic tyrosine kinase domain (TKD) separated by a kinase insert (KI), and a C-terminal intracellular domain (Figure 1a)[2, 3]. In the inactive state, FLT3 exists in its unbound, monomeric, and unphosphorylated forms. Upon binding to the indigenous ligand FL, FLT3 undergoes a conformational change. This conformational change occurs by unfolding the receptor and subsequent receptor-receptor homodimerization, bringing the kinase domain in proximity to the intracellular module, allowing phosphorylation of the tyrosine residues (Y589, Y591, and Y599) in the JM domain. This leads to a cascade of phosphorylation and activation of secondary mediators, including STAT5, PI3K/Akt/mTOR, and Ras/Raf/MAPK oncogenic signal transduction (Figure 1b)[4, 5]. Premature activation of transcription factors triggers cell proliferation and impedes cell differentiation and apoptosis in leukemia cells. As simplified in Figure 1c, the autoinhibited kinase domain (KD) consists of an N and C bi-lobal structure with an activation loop and a JM domain (PDB ID 6JQR)[6]. The interaction between the KD and JM domains prevented ATP binding. The N-lobe has an α -helix (α C-helix) and five antiparallel β -sheets, namely, β 1- β 5. The C-lobe, on the other hand, has seven α -helices and three β -sheets, namely, α D - α I and β 6 to β 8. The activation loop comprised two twisted β -sheets (β 10 and β 11). An additional β -sheet is present in the JM domain, termed as β J2. The N and C lobes are connected by a polypeptide stretch, called the hinge loop, which allows the rotational movement of the two lobes relative to each other.

Active and inactive FLT3 kinase domains can be distinguished by their characteristic 'DFG-in' and 'DFG-out' configurations. The phenylalanine residue of the DFG motif

flipped 180° from its active configuration to an inactive conformation facing the active site. This creates an additional hydrophobic pocket for type II inhibitors with an elongated geometry to interact with the residues in the α C-helix[7]. Several clinical and preclinical studies have found that mutations and overexpression of FLT3 are associated with a poor prognosis of AML. One of the most common mutations, D835Y, has been detected in the kinase activation loop. Nevertheless, mutations in residues I836, D839, Y842, and the gatekeeper residue F691 adjacent to the active site pocket, were also found in patients with a lower frequency[8, 9].

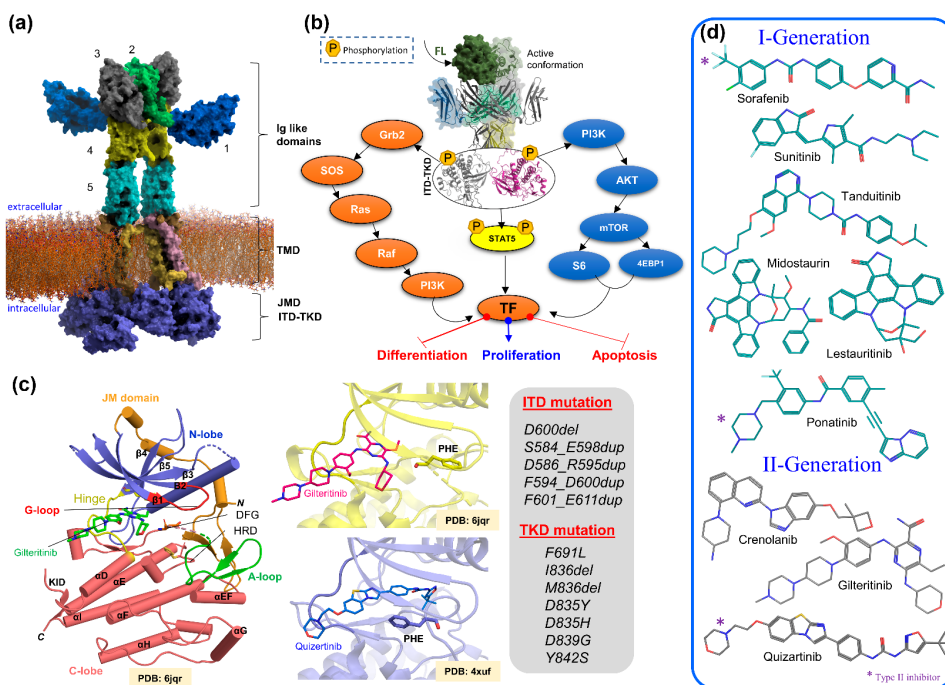


Figure 1: Structure, function, and inhibitory mechanisms of FLT3. (a) A model representation of the FLT3 receptor that is bound to the membrane, consists of five immunoglobulin-like domains (PDB ID 3QS9), a transmembrane domain (PDB ID 410U), a juxtamembrane domain, and two kinase domains (PDB ID 6JQR). (b) Down-signaling of FLT3 by Ras/ Raf/MAPK, STAT5, and Akt/mTOR pathways upon binding to the FL ligand. (c) X-ray structure of the gilteritinib (type I) bound FLT3 kinase domain (PDB ID 6JQR). The binding modes of type I and type II (quizartinib) in the active (DFG-in) and inactive (DFG-out) configuration. The frequently found ITD and TKD mutations are listed in the inset box. (d) First generation and second-generation type I and type II (* sign) FLT3 inhibitors displayed in green and grey C-atoms.

Mutations in TKD consecutively activate tyrosine kinase, which phosphorylates the intracellular domain at various sites and recruits many cytoplasmic adapter proteins for protein-protein interactions with the FLT3 receptor. As FLT3 involvement becomes more prevalent in oncogenic conditions, many small molecules that target FLT3 tyrosine kinase have been discovered. Midostaurin, sorafenib, and lestaurtinib were used as multikinase C inhibitors to improve clinical outcomes in patients with AML[10]. However, their antileukemic activities were limited when used as monotherapy, and adverse cytotoxicity was observed. Therefore, researchers are working to develop next-generation inhibitors that selectively target the FLT3 receptor[11]. Many of them are currently being evaluated in clinical trials and have higher potencies than multikinase inhibitors, such as gilteritinib, quizartinib, and crenolanib. Crenolanib and gilteritinib fall into the category of type I inhibitors and target both active and inactive kinase states, whereas quizartinib is a type II inhibitor specific to inactive-state conformations (Figure 1d)[12]. On the other hand, type I inhibitors have an identical set of chemical interactions in the ATP pocket, forming one to three H-bond interactions similar to that of the adenine moiety of ATP molecules. In addition, they occupy the proximal A-loop or allosteric site, front pocket, and P loop, providing an additional selectivity property for the type I inhibitors compared to that of the type II inhibitors.

Computational drug design is a popular choice for discovering small molecules targeting kinase receptors. In our previous study[13], we performed molecular modeling of pyrimidine4,6-diamine derivatives against inactive FLT3 as type II inhibitors. Here, we conducted the modeling study of 35 pteridin-7(8H)-one compounds as type I inhibitors targeting the active FLT3 conformer. The compounds exhibited a wide range of inhibitory activity (pIC_{50} 5.26-8.80) against FLT3, which was studied by Sun et al. [14]. The molecular docking and MD simulation studies of the most active compound C31 were conducted together with other structurally diversified compounds from the dataset to study the critical interactions and binding stability of the complexes. We conducted MM-PB/GBSA, FPL, LIE, and FEP

calculations to evaluate protein-ligand binding affinity and build the scoring models. We also applied US methods to estimate the effective binding free energy of C31 in complexes with wild-type and mutant receptors through the unbinding pathway. The last 1 ns average MD structure of C31 was retrieved to develop the CoMFA and CoMSIA models for the SAR study. Several new compounds were designed, which were further investigated for inhibitory activity prediction using the CoMSIA model. The binding affinities of the newly designed compounds were evaluated by MM-PB/GBSA, LIE, and FEP calculations.

2. Methodology

2.1. Structure preparation and molecular docking

The Surflex-Dock module in Sybyl X 2.1 (Tripos Inc., St. Louis, MO, USA) was used to perform the molecular docking study. Before the docking experiment, the FLT3 crystal (PDB ID 6JQR) with a resolution of 2.20 Å was retrieved from the RCSB protein databank. Water molecules, solvent ligands, and co-crystallized ligands were removed from the protein structure. Missing residues K634-G636, K649-A650, and G831-I836 were modeled using the web version of the MODELLER webserver (University of San Francisco, San Francisco, CA, USA) in Chimera-1.14 (RBVI, UCSF, San Francisco, CA, USA). Residue S711-L780 or KI domain was excluded during model development. The final model was selected based on the lowest DOPE score and Ramachandran plot from the PROCHECK (DOE-MBI service, UCLA) analysis. Compound C31 was the most active compound and was chosen as a representative candidate for the docking study. To prepare the 3D structure of C31, Sybyl X 2.1 was used to sketch, minimize, and assign gasteiger charges as described in earlier studies[15, 16]. The receptor was prepared using the structure preparation tool performed with the Amber7 99 force field. The docking cavity was defined using protomol generation, in which the gilteritinib-bound position was used as the reference. Finally, the docking score between the receptor and C31 was calculated using the empirical Hammerhead scoring function. Several parameters, such as the

polar, hydrophobic, repulsive, solvation, and entropy terms, were considered during the scoring assignment in Surflex-Dock. The final docking score was expressed in terms of $-\log K_d$ units; where K_d stands for the dissociation constant of the ligand. The docking protocol was repeated for the compounds C01, C03, C06, C17, C22, and C28.

2.2. MD simulation

MD simulations were performed in GROMACS 2019.5 using the Amber14sb force field according to our previous studies[17, 18]. The topology and parameter files of the ligands were generated using ACEPYPE (or AnteChamber PYthon Parser interfacE)[19] with gasteiger charges. The complexes were placed in a cubic periodic box and solvated using the TIP3P water model. The minimum thickness of the water wall was maintained at ~ 10 Å from the protein atoms. Adequate amounts of Na^+ and Cl^- were added to neutralize the system and bring the salt concentration to 150 mM. Next, each system was energy minimized by the steepest descent integrator followed by a 200 ps constant volume ensemble (NVT) to achieve a temperature of 300 K and 400 ps constant pressure ensemble (NPT) to achieve a pressure of 1 bar using the modified Berendsen (V -rescale) thermostat and barostat algorithms. During the NVT and NPT runs, the protein backbone and the heavy atoms of the ligand were restrained. Thereafter, the systems were subjected to 100 ns of MD production run by removing the backbone restraint. Particle mesh Ewald (PME) and LINCS algorithms with a cut-off value of 12 Å were employed to control the electrostatic interaction and bond length constraints. Protein and ligand RMSDs were calculated using the built-in '*gmx rms*' function in the gromacs.

2.3. MM-PB/GBSA binding energy calculation

According to our previous research[13], the *gmx_MMPBSA* package[20] was used to calculate the various MM-PB/GBSA terms. The binding energy ($\Delta G_{\text{MM-PB/GBSA}}$) of MM-PB/GBSA can be expressed by the following equation (eq-1):

$$\Delta G_{MM-PB/GBSA} = \Delta G_{COM} - \Delta G_{PROT} - \Delta G_{LIG} = \Delta E_{MM} + \Delta E_{sol} - T\Delta S = \Delta E_{vdW} + \Delta E_{ELE} + \Delta E_{GB} + \Delta E_{SA} - T\Delta S \quad (1)$$

ΔG_{COM} , ΔG_{PROT} and ΔG_{LIG} stands for the free energy estimation from the protein-ligand complex, protein, and ligand, separately, in the solvent condition. The ΔE_{MM} expresses the interaction energy between the protein and ligand in the gas phase, which can be calculated using van der Waals (ΔE_{vdW}) and electrostatic interactions (ΔE_{ELE}). The ΔG_{SOL} represents the solvation free energy which was obtained by calculating the polar solvation (ΔE_{GB}) and non-polar solvation (ΔE_{SA}) free energy. The $-T\Delta S$ represents the entropy term, which was computed as a more rigorous and concise interaction entropy (IE) proposed by Duan et al.[21] using the same GMX_MMPBSA package.

2.4. FPL simulation

The unbinding pathway was determined using Caver 3.0.3 analysis[22]. The last 1 ns average protein-ligand complex from the MD trajectory was retrieved as the initial structure for the SMD simulation. The protein-ligand complex was placed in a periodic box of $12 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$. The TIP3P water model was used to solvate the system, neutralize with Na^+ and Cl^- counterions, and gradually increase the ion concentration to 0.15 M. The system was then minimized using the steepest descent integrator for 10,000 steps, followed by 200 ps of NPT simulation. The ligand was then pulled from the binding pocket to a distance of about 5 nm. A harmonic force constant of $250 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ in the X-axis direction was used in the FPL simulation for 500 ps. The pulling speed was maintained at 0.010 nm/ps and the ligand displacement was recorded through the unbinding pathway every 0.1 ps. Each FPL simulation was performed three times to guarantee sufficient sampling. The LIE approximation (ΔG_{LIE}) from FPL simulation was calculated according to this study[23]:

$$\Delta G_{LIE} = \frac{1}{2} \Delta E^{cou} + \frac{1}{2} \Delta E^{vdW} \quad (2)$$

where, $\Delta E^{\text{cou}} = E_{\text{unbound}}^{\text{cou}} - E_{\text{bound}}^{\text{cou}}$, and $\Delta E^{\text{cou}} = E_{\text{unbound}}^{\text{cou}} - E_{\text{bound}}^{\text{cou}}$ were calculated from the vdW and electrostatic interaction of the *bound* and *unbound* states of the ligand, respectively.

2.5. US simulation

The US process was divided into two stages. In the first stage, the unbinding procedure was performed by SMD/FPL simulation. In the second stage, the initial structures for US simulation were extracted from the SMD trajectories with a spacing distance of 0.2 nm[24]. However, four additional coordinates were assigned for the first 0.8 nm distance, resulting in the first eight windows having a spacing distance of 0.1 nm. A total of 25 protein-ligand conformations from bound to unbound process were collected as reaction coordinates. After that, each conformation was first equilibrated with a 200 ps NPT ensemble, followed by 2.5 ns of US simulation. The built-in ‘*gmx wham*’ function was used to analyze the potential mean force (PMF) along with its reaction coordinates using the weighted histogram (WHAM) method. Finally, the binding free energy (ΔG_{US}) was calculated by subtracting the lowest and highest values from the PMF curve. The computational error estimation was carried out by 100 bootstrapping runs.

2.6. Free energy perturbation


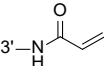
The last 1 ns average MD structures of the receptor-ligand complexes were taken as the initial structure for the FEP simulation study, according to the previous literature[25, 26]. In this method, the ligand interaction was turned on over the two-coupling process in the receptor-bound form and isolated form in the solvent. The ligands were transitioned from non-interaction (0) to the full-interaction state (1) by turning on vdW and coulombic interactions with the surrounding by changing the coupling parameter (λ). Nine λ values: 0.00, 0.10, 0.25, 0.35, 0.50, 0.65, 0.75, 0.90 and 1.00 were used to change the vdW and coulombic interactions, respectively, and a total of seventeen alter- λ simulations of each 2 ns were performed. The total energy


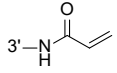
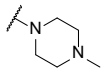
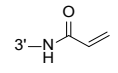

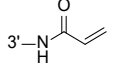
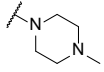
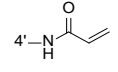

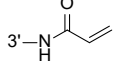
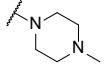

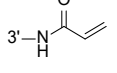
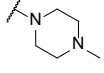

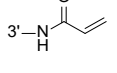
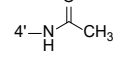
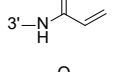
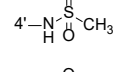
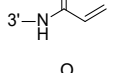
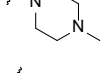
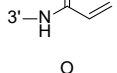
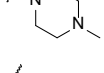
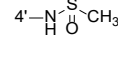

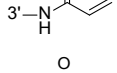
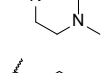

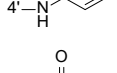
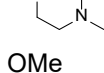
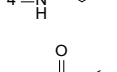
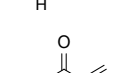
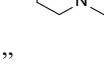
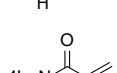

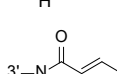
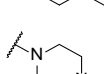
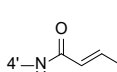
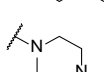


change in the FEP process through the λ alteration was deduced using Bennet's acceptance ratio (BAR) method.

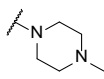
2.7. Dataset building and molecular alignment

The 35 pteridin-7(8*H*)-one-based compounds that were reported to be FLT3 inhibitors were taken as a dataset for this study (Table 1). The last 1 ns average structure of the C31 from the MD simulation was selected as a representative structure from the dataset. The pteridin-7(8*H*)-one chemical entity was chosen as a common skeleton. Based on the common skeleton, the rest of the compounds were sketched and minimized with a convergence force of 0.05 kcal/mol at the maximum iteration of 2000 run by the tripos force field and added Gasteiger-Hückel partial charges, in the Sybyl suit. The biological activities (IC_{50}) of these compounds were converted into logarithmic IC_{50} (pIC_{50}) values. The pIC_{50} values were well distributed across the three log units ($pIC_{50} = 5.26$ to 8.80). The entire dataset was categorized into low, medium, and high activity segments, as described here[27, 28]. The 9 compounds were randomly selected from each segment as the test set compounds in such a way that the compounds would cover different activity ranges while maintaining their structural variations. The training set consists of 26 compounds used as a dependent variable (Training set) to construct the 3D-QSAR model, while the 9 compounds were used as the independent variable (Test set) to access the model's predictive power.

Table 1: Dataset compounds and their inhibitory activities. The test set compounds are shown by the (*) sign

#Cpds	R ₁	R ₂	R ₃	pIC_{50}	#Cpds	R ₁	R ₂	R ₃	pIC_{50}
01*			H	6.50	19	OMe	4'-NH ₂	H	7.53

02			H	5.81	20			2-Ome	5.88
03*			H	5.97	21			2-Ome	6.81
04			H	5.75	22		3'-NH ₂	H	8.21
05			H	5.54	23*		4'-NH ₂	H	8.05
06			H	5.26	24	OMe		H	6.89
07*	OMe		H	5.72	25	OMe		H	6.59
08	H		H	5.53	26		4'-NH ₂	H	7.65
09	Cl		H	5.44	27*			H	7.49
10			H	5.94	28		4'-CH ₂ NH ₂	H	7.42
11*			H	7.32	29		4'-NH ₂	H	8.08
12	OMe		H	6.96	30	OMe	4'-NH ₂	2-Me	6.45
13*	H		H	6.78	31		4'-NH ₂	3-Me	8.80
14	Cl		H	6.41	31 _(D835 y)	”	”	”	8.69
15			H	6.76	32*		4'-NH ₂	3-OMe	8.49
16*	OMe		H	5.35	33		4'-NH ₂	3-Cl	8.28
17	OMe		H	6.73	34		4'-NH ₂	3-Cl	7.55

18	OMe	3'-NH ₂	H	7.46	35		4'-NH ₂	3-F	7.44
----	-----	--------------------	---	------	----	---	--------------------	-----	------

2.8. CoMFA and CoMSIA studies

CoMFA and CoMSIA are two widely popular 3D-QSAR methods. Lennard Jones and Coulombic potential functions were used to compute the steric and electrostatic fields in CoMFA analysis[29, 30]. Each compound was placed in a spatial grid one after another during the calculation process with a grid spacing of 2.0 Å. To calculate the structural characteristics of the compounds, each grid space was assigned to the sp³ carbon atom with the vdW probe radius of 1.52 Å and a net charge of +1. The energy cut-off of 30 kcal/mol was applied, and the rest of the parameters were set to default. In the CoMSIA model, additional descriptors such as hydrophobic, H-bond acceptor, and H-bond donor were also calculated using steric and electrostatic fields. To determine the distance between compound atoms and probe atoms, a Gaussian-type function was applied in CoMSIA with the default attenuation factor(σ) to 0.3.

To produce statistically significant CoMFA and CoMSIA models, the partial least squares (PLS) method was employed to correlate the biological activity and descriptors of the compounds. The leave-one-out (LOO) method was applied in a cross-validation method to obtain the cross-validation coefficient (q^2), the optimal number of components (ONC), and the standard error of prediction (SEP) by assigning different partial charges. Then the no validation method was applied to obtain the non-cross-validation coefficient (r^2), Fisher's statistics (F value), and standard error of estimation (SEE). In the CoMSIA model, the descriptor fields, such as S, H, E, A, and D, were used in different combinations to produce the best possible statistical model[31]. To examine the internal and external validation of the 3D-QSAR model, we determined the χ^2 , RMSE, MAE, k, k', $|r_0^2 - r'_0{}^2|$, $(r^2 - r_0{}^2) / r^2$, r_m^2 , $\overline{r_m^2}$, Δr_m^2 , r_{pred}^2 , Q_{F1}^2 , Q_{F2}^2 , Q_{F3}^2 , and Q_{CC}^2 matrices as described in the previous literature[32-34].

2.9. Contour map analysis and design of new compounds

The field effects from CoMFA and CoMSIA models were presented by 3D StDev*Coeff contour maps with different color schemes. Each contour described the structural characteristic of the compound that could increase or decrease the inhibitory activity. The favorable and unfavorable regions for the steric, electrostatic, hydrophobic, H-bond acceptor, and H-bond donor were colored green, yellow, blue, white, magenta, orange, and cyan. The detailed CoMFA and CoMSIA contour maps have been summarized in a simplified SAR scheme. We designed novel compounds based on the MD and SAR and anticipated their inhibitory potency. Compounds with predictive pIC₅₀ higher than C31 were subjected to binding affinity evaluation.

3. Results and discussion

The optimal ligand-binding orientation at the active site was predicted using the molecular docking study. Docking pose verification is a critical step since it is used in molecular simulation, MM-PB/GBSA binding energy evaluation, and lastly, the generation of 3D-QSAR models. The 2D structure of pteridin-7(8*H*)-one-based compounds that have been chosen for molecular docking studies is shown in Figure 2a. None of the compounds in the dataset had FLT3 co-crystal forms; the most active compound C31 was cross-docked into the FLT3 binding pocket, which the gilteritinib molecule had previously occupied. The top-scoring ligand poses in a cross-docking experiment are not always accurate for evaluating the final docking result. Therefore, in addition to the top-scoring solution, we employed RMSD evaluation between the docked pose and gilteritinib crystal pose by the LigRMSD web server[35]. According to this study[36], RMSDs between the docked and crystal poses of 2.0-3.0 Å are an acceptable docking solution. The final docked complex should also comply with ECIDALs norms, in which the essential chemical interaction and binding configuration of the analogous ligands were thoroughly inspected across the same protein family from the PDB database. The docking interaction of the top-ranked FLT3-C31 within the binding pocket is shown in Figure 2b. Compound C31 and gilteritinib are designed to target active FLT3 in a type I inhibitor-like manner. The

RMSD between the docked ligand and crystal ligand gilteritinib was found to be 2.2 Å. In addition, compounds C31 and gilteritinib have 2-phenylaminopyrimidine and 2-phenylaminopyrazine moieties in their structure which share an identical chemical scaffold and form a bidentate H-bond interaction with the residue C694 at the hinge loop. C31 formed two additional H-bond interactions with residue L616 and the keto (-C=O) group of the DFG residue F830. Other notable interactions, such as π - π stacking and π -sigma interactions, were formed between the pteridine ring and residues Y693 and L818, respectively. Residues V624, A642, V675, and C828 formed hydrophobic interactions with the ligand. The RMSDs with gilteritinib for compounds C03, C06, C17, C22, and C28 were found to be less than 3.0 Å, indicating reasonable docking accuracy. Overall, the docking analysis suggested a satisfactory docking solution that could be utilized in the binding study of new compounds.

Since protein-ligand interaction is a highly thermodynamic process, a single docking experiment has several limitations. Besides, in the cross-docking experiment, the binding site was treated as a rigid body, and neither side-chain nor backbone movements were taken into account. Additionally, in the empirical scoring functions, water-mediated hydrogen bonding, de-solvation, and estimation of ligand binding energy by water swapping remain challenging. In many cases, the docked pose is not stable under physiological conditions. Thus, MD simulation studies were employed to validate the docking solutions and the overall stability of the protein-ligand complexes. We have manually mutated the residue D835Y in the FLT3 (Figure 1c) structure to make the FLT3_{D835Y}-C31 complex. Therefore, eight total protein-ligand systems, i.e., FLT3-C31, FLT3_{D835Y}-C31, FLT3-C01, FLT3-C03, FLT3-C06, FLT3-C17, FLT3-C17, and FLT3-C28 were subjected to 100 ns production simulations. The oscillation of the backbone α C of proteins and the heavy atoms of the ligands were plotted with respect to the simulation time in Figure 2(d-k).

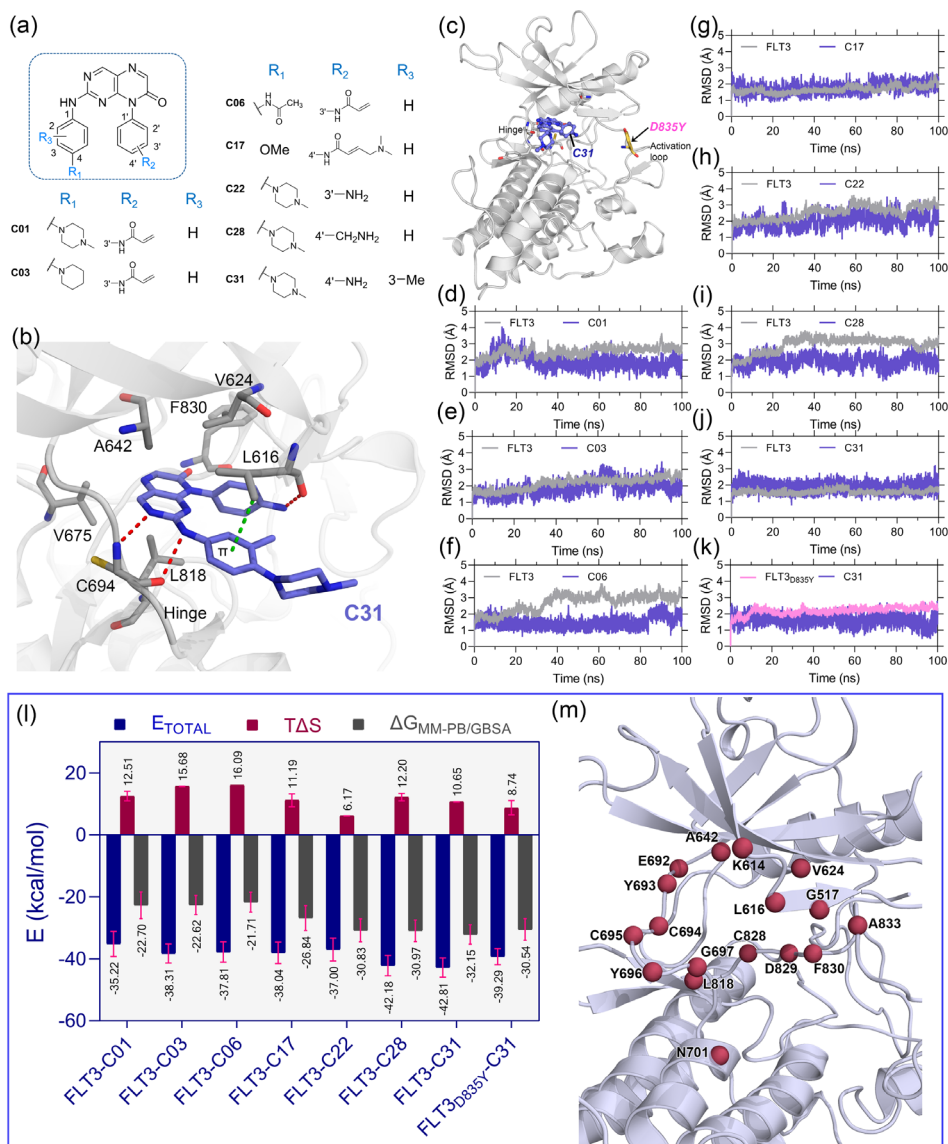


Figure 2. Molecular docking, MD simulation, and MM-PB/GBSA binding energy calculation studies. (a) 2D structure of pteridin-7(8H)-one-based compounds: C01, C03, C06, C17, C22, C28, and C31 that have been selected for molecular docking studies. (b) Binding pose orientation of C31 within the ATP pocket. The ligand is anchored to the hinge by two H-bond interactions, as shown in red dashed lines, with residue C694. The π - σ interaction with L616 is shown in green dashed lines. (c) The global structure of the C31 bound FLT3 kinase domain, depicting the D835Y mutation site in the activation loop. (d) – (k) RMSD plots of the selected complexes, i.e., FLT3-C01, FLT3-C03,

FLT3-C06, FLT3-C17, FLT3-C22, FLT3-C28, FLT3-C31, and FLT3_{D835Y}-C31, respectively. (l) A graphical comparison of the E_{TOTAL} , $-T\Delta S$, and $\Delta G_{\text{MM-PB/GBSA}}$. The binding energy terms between the complexes are colored blue, firebrick, and gray, with the standard deviation (red). (m) Residues that contributed the critical binding energies to the ligands during the per-residue MM-PB/GBSA decomposition analysis are highlighted in the firebrick sphere representation.

The RMSD plots stated that the systems were converged within the first 20 ns of the simulations. The RMSDs of the protein and ligand were in the range of 1.0-3.5 Å. The FLT3 complexes with compounds C31, C01, and C17 were stable after initial convergence. Although, the RMSDs of FLT3-C06 and FLT3-C28 suggested multiple state conversions during the MD run. The MM-PB/GBSA binding energy is a frequently used method for calculating the end-state binding free energy between protein-ligand complexes. We collected the last 2 ns trajectory or 200 snapshots from each system to compute the MM-PB/GBSA binding free energy. The entropy term ($-T\Delta S$) was calculated from the last 80 snapshots of the 2 ns trajectory. Finally, the $-T\Delta S$ was subtracted from the ΔT_{TOTAL} term to obtain the final binding energy of the MM-PB/GBSA calculation. The $\Delta G_{\text{MM-PB/GBSA}}$ was found to be -32.15 kcal/mol and -30.54 kcal/mol for the C31-bound wild-type and mutant FLT3 complexes, respectively.

Compounds C01, C03, C06, C17, C22, and C28 were estimated to have the binding energies of -22.70 kcal/mol, -22.62 kcal/mol, -21.71 kcal/mol, -26.84 kcal/mol, -30.83 kcal/mol, and -30.97 kcal/mol, respectively. In addition, we computed the residue-specific binding energy contribution within the 4.0 Å distance from the ligand atoms. Residues K614, L616, G617, V624, A642, E692, Y693, C694, L818, and F830 were found to be the major binding energy contributors in MM-PB/GBSA terms. The residue-specific binding energy decomposition analysis is graphical illustrated in Figure 1m.

FPL simulation along with the unbinding pathway was conducted, which is based on the SMD principle. It is also a relatively straightforward approach for estimating the binding affinity between protein-ligand complexes. In this method, the ligands were

forced to dissociate from the center of mass (COM) distance of the DFG residues through the caver-predicted unbinding tunnels at a distance of about 5 nm toward the X-axis (Figure 2a). Initially ($T = 0$ ps), the pulling force was minimal, and the ligand was bound to the active site cavity, referred to as the *bound* state. Over the simulation, the pulling force was gradually increased until the ligands began to dissociate from the binding pocket. At that time ($T = T_{\max}$), the pulling force reached its peak, the ligand was separated from the cavity, mobilized into the solvent, and termed rupture force (F_{\max}). The external force abruptly decreased and maintained a consistent plateau, referred to as the *unbound* state. Theoretically, the ligand with higher inhibitory activity poses a higher relative binding affinity. Thus, F_{\max} could be applied to rank the inhibitor compounds. The external pulling forces and separation distances over time are shown in Figure 2b. The average F_{\max} values for compounds C01, C03, C06, C17, C22, C28, and C31 were estimated to be 221.40, pN, 441.13 pN, 391.61 pN, 428.75 pN, 475.17 pN, 441.13 pN and 537.07 pN, respectively. In contrast, a lower F_{\max} value ($F_{\max} = 453.51$ pN) was obtained for the FLT3_{D835Y}-C31 complex compared to the C22 and C31 systems. Next, we calculated the LIE approximation over the two quasi-equilibrium states (*bound* and *unbound*) by computing the van der Waals and electrostatic interactions. For the compounds C01, C03, C06, C17, C22, C28, and C31, the absolute binding energy calculated using the LIE approximation was -28.76 kcal/mol, -27.18 kcal/mol, -28.71 kcal/mol, -30.65 kcal/mol, -30.97 kcal/mol, -30.35 kcal/mol, and -28.92 kcal/mol, respectively. In comparison to the wild-type FLT3, compound C31 had rather lower absolute binding free energy ($\Delta G_{\text{LIE}} = -28.17$ kcal/mol) to the mutant receptor. Following that, the US simulation was applied to the C31-bound FLT3 and FLT3_{D835Y} complexes to evaluate the effective binding free energy profile along with their dissociation pathway. A total of 25 evenly distributed overlapping windows were extracted from the FPL trajectories and used biased sampling simulation as described in the methodology section. The binding free energy and sufficient sampling could be traced by analyzing the PMF curve and the umbrella histogram with respect to the reaction coordinates (ξ), as shown in Figure 2(c-f).

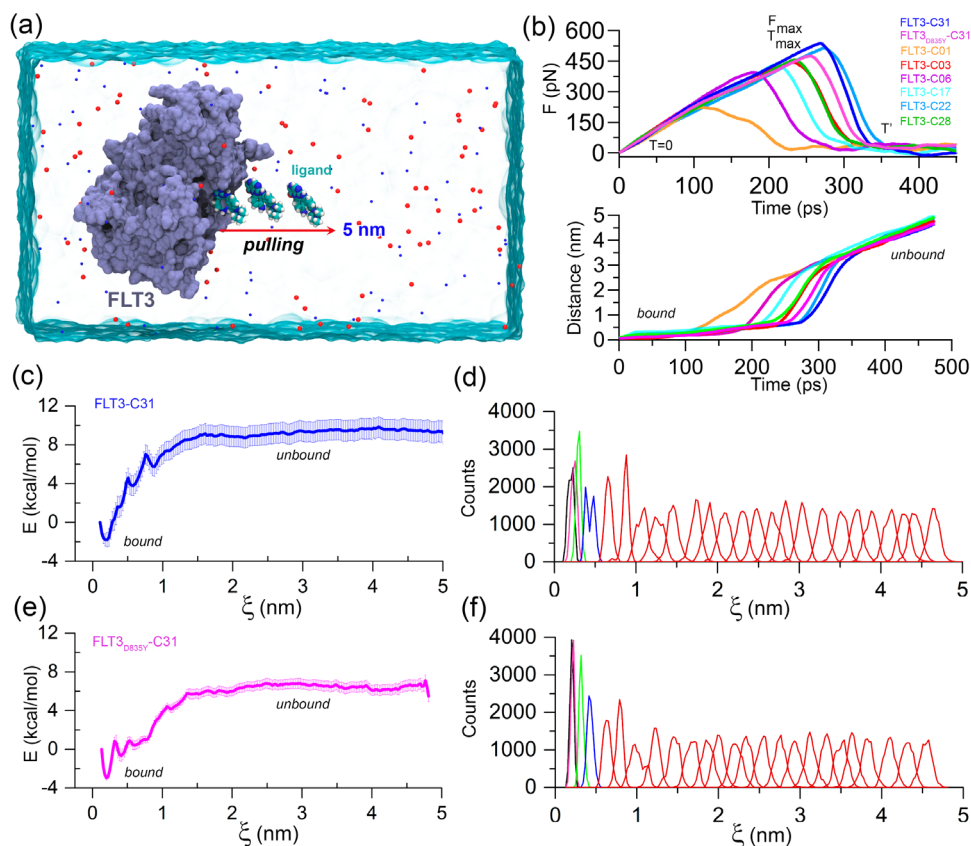


Figure 2. FPL and US simulation analysis. (a) Illustration of the FPL simulation setup. An external harmonic force is used to dissociate the ligand to a distance of 5 nm on the X-axis. (b) Graphical representation of the mean force evaluation (rupture force) and displacement of the ligands with respect to the time during the FPL runs. (c) PMF curve and (d) histogram profile of the FLT3-C31 system. (e) PMF curve and (f) histogram profile of the FLT3_{D835Y}-C31 system. The errors are calculated by using 1000 bootstrapping runs.

In PMF, the free energy began from zero and then dropped to a minimum value. After that, the energies were gradually increased to attain a stable value, where non-covalent interactions between protein and ligands were broken. The binding free energy (ΔG_{US}) of the most active compound C31 from the US was estimated to be -10.73 ± 1.27 kcal/mol and -9.49 ± 0.57 kcal/mol for wild-type and mutant FLT3, respectively. The

convergence of the calculation could be validated by the histogram profiles of overlapping neighboring windows.

The FEP simulation was carried out with the last 1 ns average MD structure of the protein-ligand complexes. The vdW and coulombic interactions of the ligands were sequentially turned on in the solute in a complex and isolated form by alter- λ simulations. The first 40% of the trajectory data were discarded to eliminate any convergence error. The final absolute binding free energy (ΔG_{FEP}) from the FEP simulation was determined to be -14.83 kcal/mol, -14.64 kcal/mol, -13.68 kcal/mol, -16.77 kcal/mol, -15.04 kcal/mol, -17.61 kcal/mol and -17.87 kcal/mol for compounds C01, C03, C06, C17, C22, C28 and C31, respectively. Compared to the wild-type complex, the FLT3_{D835Y}-C31 complex had lower binding free energy ($\Delta G_{\text{FEP}} = -16.82$ kcal/mol) in the FEP calculation.

Table 2 emphasizes the final binding free energies of the protein-ligand complexes, which are derived using MM-PB/GBSA, FPL, LIE, and FEP methods. The experimental binding energies (ΔG_{EXP}) of the compounds were deduced from their inhibitory activity (IC_{50}) and attempted to correlate with the computed binding free energies. During the correlation analysis, the binding energies of the FLT3_{D835Y}-C31 complexes were ignored. The correlation plots between the experimental binding energies and computed binding energies of the seven compounds are shown in Figure 3. A good correlation coefficient ($R_{\text{MM-PB/GBSA}} = 0.92$) was obtained between the ΔG_{EXP} and $\Delta G_{\text{MM-PB/GBSA}}$. However, the binding energies were overestimated by the MM-PB/GBSA method. In the FPL model, the F_{max} values are poorly correlated with the ΔG_{EXP} values ($R_{F_{\text{MAX}}} = -0.55$). The correlation coefficient (R_{LIE}) between ΔG_{EXP} and ΔG_{LIE} was calculated to be 0.60. Thus, the observations suggested a significant limitation and required special attention when utilizing the FPL and LIE models for ligand ranking. Although, both models were able to distinguish differences in binding affinities between C31 and receptor variants. In the FEP model, the correlation coefficient (R_{FEP}) between ΔG_{EXP} and ΔG_{FEP} was estimated to be 0.71, which is

statistically reasonable and could be used to assess the binding affinities of unknown compounds.

Table 2. Experimental binding energies and calculated binding energies from MM-PB/GBSA, FPL, LIE, and FEP calculations. Except for the rupture force (F_{MAX}), all binding energy terms are expressed in kcal/mol

Complexes	ΔG_{EXP}	$\Delta G_{MM-PB/GBSA}$ ($\pm SD$)	F_{MAX} (pN)	ΔG_{LIE} ($\pm SD$)	ΔG_{FEP} ($\pm SD$)
FLT3-C01	-8.98	-22.70 ± 4.32	221.40	-28.76 ± 4.80	-14.83 ± 0.65
FLT3-C03	-8.24	-22.62 ± 3.06	441.13	-27.18 ± 4.11	-14.64 ± 0.54
FLT3-C06	-7.26	-21.71 ± 3.25	391.61	-28.71 ± 4.61	-13.68 ± 0.43
FLT3-C17	-9.29	-26.84 ± 4.01	428.75	-30.65 ± 4.09	-16.77 ± 0.40
FLT3-C22	-11.38	-30.83 ± 3.68	475.17	-30.97 ± 3.23	-15.04 ± 0.46
FLT3-C28	-10.25	-30.97 ± 3.46	441.13	-30.35 ± 4.51	-17.61 ± 0.53
FLT3-C31	-12.15	-32.15 ± 3.13	537.07	-28.92 ± 4.51	-17.87 ± 0.43
FLT3 _{D835Y} - C31	-12.00	-30.54 ± 3.47	453.51	-28.17 ± 4.06	-16.82 ± 0.38

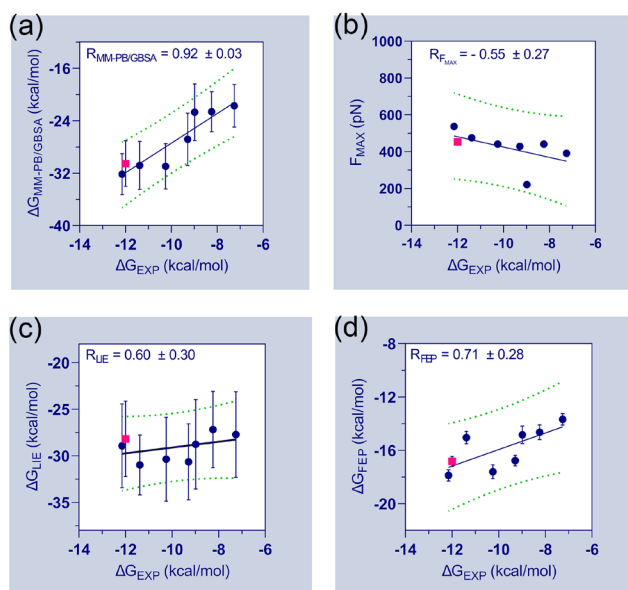


Figure 3. Relationship between experimental and estimated binding energies. In addition to the correlation coefficients of each matrix, correlation plots of (a) ΔG_{EXP} vs. $\Delta G_{\text{MM-PB/GBSA}}$, (b) ΔG_{EXP} vs. F_{MAX} (c) ΔG_{EXP} vs. ΔG_{LIE} and (d) ΔG_{EXP} vs. ΔG_{FEP} are determined. The experimental binding energies and the computed binding energies of FLT3_{D835Y}-C31 are shown in pink. Standard errors of correlation coefficients are estimated using 1000 bootstrapping runs.

The statistical analysis of CoMFA and CoMSIA is summarized in Table 3. The acceptable parameters of each statistical term are listed in the ‘Threshold values’ column based on the previously published literature. In CoMFA analysis, q^2 and r^2 were obtained as 0.768 and 0.982, greater than 0.5 and 0.6, respectively, at the ONC of 3. The steric and electrostatic contributions of the CoMFA scheme were found to be 54.2% and 45.8%, respectively. To generate the best statistically significant CoMSIA model, we used five descriptor fields, such as steric (S), electrostatic (E), hydrophobic (H), H-bond donor (D), and H-bond acceptor (A) in the permutation-combination process. The best q^2 and r^2 values of 0.844 and 0.972 were obtained in SH combination at ONC of 4, and, therefore, SH was selected as the final CoMSIA model. The final contributions of the steric and hydrophobic fields were found to be 46.8% and 53.2%, respectively. However, any QSAR models are insufficient without

being externally validated by test set compounds that were not used during model development. Thus, the external validation was carried out by estimating the predictive correlation coefficient or r_{pred}^2 . In CoMFA and CoMSIA, the values of r_{pred}^2 were determined to be 0.919 and 0.918, respectively, greater than the constrained value of 0.6, signifying that both models were statistically reliable and had good predictability. Figures 4a and 4e represent the PLS regression plots between the observed and predicted activity of compounds from the CoMFA and CoMSIA models. Besides, we also calculated other statistical parameters such as r_m^2 or $r'_m{}^2$, Q_{Fn}^2 ($n = 1, 2, 3$), and Q_{cc}^2 matrices, all of which were calculated to be within the well-accepted parameters.

Table 3. Statistical results and validation of CoMFA and CoMSIA models

Statistical parameters	CoMFA	CoMSIA SH	Threshold values	Statistical parameters	CoMFA	CoMSIA SH	Threshold values
q^2	0.768	0.844	> 0.5	k_{Test}	0.989	0.997	$0.85 \leq k$
ONC	3	4	< 6	k'_{Test}	1.009	1.000	or $k' \leq 1.15$
SEP	0.523	0.439		r^2_{Test}	0.965	0.931	
r^2	0.982	0.972	> 0.6	$r^2_{0\text{Test}}$	0.923	0.918	$\approx r^2$
SEE	0.270	0.194	$\ll 1$	$r'^2_{0\text{Test}}$	0.938	0.929	
F-value	111.678	111.663		$ r^2_{0\text{Test}} - r'^2_{0\text{Test}} $	0.009	0.011	< 0.3
BS- r^2	0.959	0.968		$\frac{r^2 - r^2_{0\text{Test}}}{r^2_{\text{Test}}}$	0.008	0.008	< 0.1
BS-SD	0.018	0.016		$\frac{r^2 - r'^2_{0\text{Test}}}{r^2_{\text{Test}}}$	0.034	0.034	
χ^2	0.056	0.066	< 1.0	$r^2_{m\text{Test}}$	0.768	0.758	r^2_m or r'^2_m
RMSE	0.322	0.352	< 0.5	$r'^2_{m\text{Test}}$	0.790	0.785	> 0.5
MAE	0.033	0.038	≈ 0	Δr^2_m	0.022	0.027	
RSS	2.69	3.23		$\overline{r^2_m}$	0.779	0.772	> 0.5

k_{Train}	0.994	0.994	$0.85 \leq k$	r_{pred}^2	0.919	0.918	
k'_{Train}	1.002	1.002	or $k' \leq 1.15$	Q_{F1}^2	0.919	0.918	> 0.6
$r_{0}^2_{Train}$	0.889	0.867	$\approx r^2$	Q_{F2}^2	0.919	0.918	
$r'_{0}{}^2_{Train}$	0.877	0.842	$\approx r^2$	Q_{F3}^2	0.919	0.918	
$ r_{0}^2 - r'_{0}{}^2 _{Train}$	0.011	0.024	< 0.1	Q_{ccc}^2	0.964	0.962	≈ 1
$\frac{r^2 - r_{0}^2}{r^2}_{Train}$	0.011	0.077	< 0.3	S (%)	54.2	46.8	
$\frac{r^2 - r'_{0}{}^2}{r^2}_{Train}$	0.018	0.104	< 0.3	E (%)	45.8	NA	
$r_m^2_{Train}$	0.683	0.686	r_m^2 or $r'_m{}^2$ > 0.5	H (%)	NA	53.2	
$r'_m{}^2_{Train}$	0.664	0.646	r_m^2 or $r'_m{}^2$ > 0.5				

q²: squared cross-validated correlation coefficient; ONC: optimal number of components; SEP: standard error of prediction; r²: squared correlation coefficient; SEE: standard error of estimation; F-value: F-test value; BS-r²: Bootstrapping squared correlation coefficient; BS-SD: standard deviation from 100 bootstrapping runs; χ^2 : chi-square value; RMSE: root mean square error; MAE: mean absolute error; k: slope of the predicted vs observed activity at zero intercepts; k': slope of the observed vs. predicted activity at zero intercepts; r₀²: squared correlation coefficient between predicted and observed activity; r'₀²: squared correlation coefficient between predicted and observed activity; r_m², r'_m²: r_m², and r'_m² matrices; $\Delta r_m^2 = |r_m^2 - r'_m{}^2|$; $\overline{r_m^2} = (r_m^2 + r'_m{}^2)/2$; r_{pred}²: predictive correlation coefficient; Q_{F1}², Q_{F2}², Q_{F3}², and Q_{ccc}²: Q_{F1}², Q_{F2}², Q_{F3}², and Q_{ccc}² matrices; S: steric; E: electrostatic; H: hydrophobic; A: H-bond acceptor; D: H-bond donor.

Next, we conducted the applicability domain analysis to visually detect the outliers[37]. It is a theoretical chemical space in which the QSAR model could reliably predict the descriptor properties of compounds. The AD analysis of CoMFA and CoMSIA is illustrated in Figs. 4b and 4f by the distance-based Williams plot. The standardized residual values of the training set and test set compounds were plotted against their leverage values within a square area of $\sigma = \pm 3$ and warning leverage (h^*). Compounds with a leverage value greater than h^* were considered outliers and significantly affected the regression slope of the QSAR models. In our study, none of

the compounds was outside the warning leverage ($h^* = 0.29$), suggesting the robustness of the 3D-QSAR models.

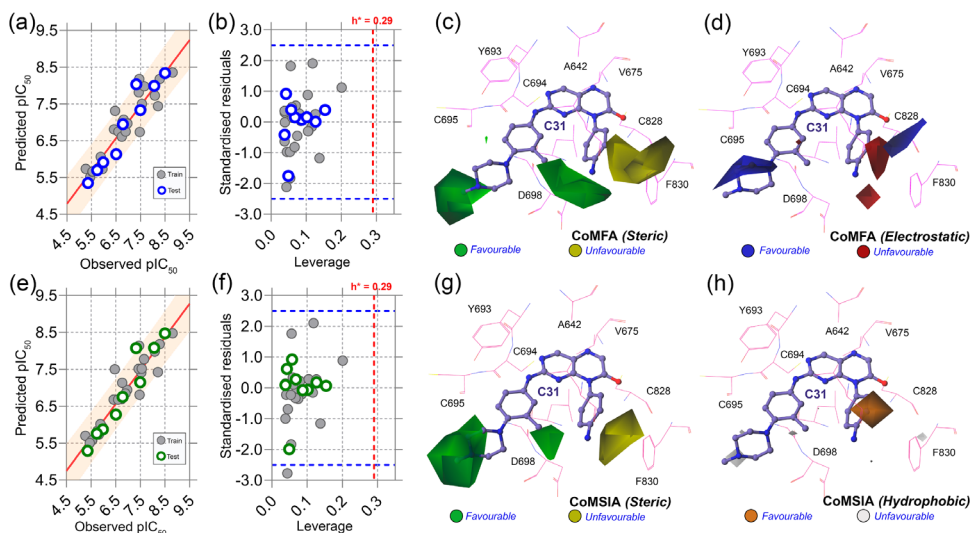


Figure 4. Correlation plots, applicability domain, and contour maps analysis. (a) PLS regression plot, (b) Williams plot, (c) steric contours, and (d) electrostatic contours from CoMFA analysis. (e) PLS regression plot, (b) Williams plot, (c) steric contours, and (d) hydrophobic contours from CoMSIA analysis. Warning leverage (h^*) is shown as red dashed lines in each AD plot. In steric maps, the green and yellow contours indicate that bulky groups are favored and unfavored. On the electrostatic map, blue and red indicate a favorable space for the electropositive and electronegative groups. The orange and light-grey color in the hydrophobic map represents favorable and unfavorable substitutions for hydrophobic chemical groups.

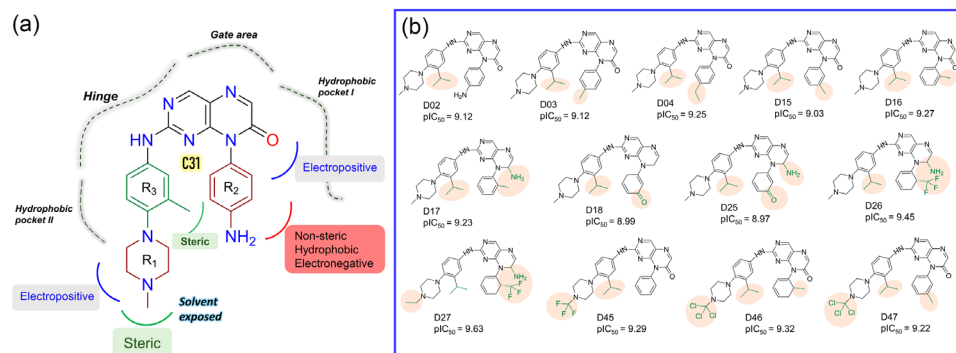


Figure 5. SAR study and development of new compounds. (a) The summary of the structure-activity relationship from the CoMFA and CoMSIA analysis around the reference compound C31. (b)

Designed compounds with higher predicted pIC_{50} in the CoMSIA model. The substitution sites are highlighted in the pale salmon color.

The contour maps analysis from the 3D-QSAR study was carried out to explore the favorable and unfavorable sites for chemical substitution. As shown in Figure 4(c-h), the field effects of the chemical descriptors from CoMFA and CoMSIA were graphically represented by contour polyhedrons around the C31-bound active site. In both CoMFA and CoMSIA, two green contours appeared at the R_1 and R_3 positions near the solvent-exposed area of the active sites, indicating that the presence of bulky steric chemical groups in that region could increase the inhibitor potency. In contrast, a large yellow contour appears near the DFG-residues, suggesting a disfavored substitution for bulky steric groups at that position. Compounds C04, C07, C08, C09, and C12 non-steric groups in their R_1 and R_3 positions exhibited lower inhibitory activity ($pIC_{50} < 0.7$) than the other dataset compounds.

On the other hand, compounds C31 and C32 consist of steric groups such as methyl (-CH₃) or methoxy (-OCH₃) groups in *meta*-position instead of *para*-position of the piperazine moiety, which allocated them in proximity to the green contours. It might favor the critical inhibitory potency of these two highest active compounds. The blue and red contours (Figure 4d) suggested a favorable substitution for the electropositive and electronegative chemical groups. In that chemical space, compounds with positively charged nitrogen (N atoms) or amine (-NH₂) groups might enhance the inhibitory activity against FLT3. An orange contour near the aniline moiety at the R_2 position towards the residue F830 suggested that a small hydrophobic substitution could be favorable (Figure 4h). Taken together, the overall observation was emphasized as a SAR scheme in Figure 5a.

In the context of SAR, we initiated the inhibitor design using substitution growth methods. The contour maps suggested a large steric substitution in the R_1 and R_3 positions, although this is not infinite. The addition of substantially bulky chemical components may result in steric clash and failure of ligand insertion into the binding pocket. Moreover, the designed compounds should satisfy Lipinski's criterion and

have a low complexity in the chemical synthesis route. Similarly, the compounds should be designed with scaffolds similar to the dataset compounds. A rather heterogeneous molecule may not be adequately evaluated by a 3D-QSAR model, causing it to be assigned outside the applicability domain or the chemical space. Earlier modeling studies have reported that D835 mutations alter the conformational changes of the phenylalanine residue (F830) of the conserved DFG motif, affecting the vdW and electrostatic interactions, which influence the binding affinity regardless of the type I or type II inhibitors[38, 39]. Our multiple binding energy computation schemes estimated a lower binding affinity of the most active compound to the mutant receptor. Therefore, growing molecular probes from the R₂ position towards DFG residues may contribute to additional steric or electrostatic interactions and ultimately improve the binding affinity of the designed compounds. This could be reinforced from the SAR study, as we obtained that the non-steric, hydrophobic, and electronegative groups could be favorable for improving the inhibitory potency of C31. By considering the above factors, we designed up to 50 new compounds (given in Table 4 as SMILES format), and their activity was predicted by the CoMSIA model. We introduced the steric substitution to the R₃ position as a first step while leaving the remaining positions unaltered. At the R₂ site, we added hydrophobic and electronegative chemical entities (-C=O, -CF₃), while the R₃ position remained unchanged. Following that, chemical probes were grown in the R₁ position, with varying degrees of substitution in the R₂ and R₃ positions. Beyond the SAR scheme, we also incorporated investigational electronegative groups like chlorine and fluorine as probe moieties.

Table 4: SAR and MD assisted designed compounds in SMILES format

SMILES	
<chem>NC1=CC=C(C=C1)N2C(C=NC3=C2N=C(NC4=CC(CC)=C(N5CCN(C)CC5)C=C4)N=C3)=O</chem>	D01
<chem>NC1=CC=C(C=C1)N2C(C=NC3=C2N=C(NC4=CC(C(C)C)=C(N5CCN(C)CC5)C=C4)N=C3)=O</chem>	D02
<chem>CC1=CC=C(C=C1)N2C(C=NC3=C2N=C(NC4=CC(C(C)C)=C(N5CCN(C)CC5)C=C4)N=C3)=O</chem>	D03
<chem>O=C1C=NC2=C(N=C(NC3=CC(C(C)C)=C(N4CCN(C)CC4)C=C3)N=C2)N1C5=CC=CC=C5</chem>	D04
<chem>CC(C)C1=C(N2CCN(CC3CC3)CC2)C=CC(NC(N=C4)=NC5=C4N=CC(N5C6=CC=CC(C)=C6)=O)=C1</chem>	D05
<chem>CC(C)C1=C(N2CCN(CC3CC3)CC2)C=CC(NC(N=C4)=NC5=C4N=CC(N5C6=CC=CC(C)=C6)=O)=C1</chem>	D06
<chem>CC(C)C1=C(N2CCN(CC3CC3)CC2)C=CC(NC(N=C4)=NC5=C4N=CC(N5C6=CC=CC(C)=O)=C1</chem>	D07

<chem>CC(C)C1=C(N2CCN(C3CC3)CC2)C=CC(NC(N=C4)=NC5=C4N=CC(N5C6=CC=CC=C6)=O)=C1</chem>	D08
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(C)=C6</chem>	D09
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(F)=C6</chem>	D10
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(Cl)=C6</chem>	D11
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(Br)=C6</chem>	D12
<chem>O=C1C=NC2=C(N=C(NC3=CC(F)=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D13
<chem>O=C1C=NC2=C(N=C(NC3=CC(Cl)=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D14
<chem>O=C1C=NC2=C(N=C(NC3=CC(C(C)C)=C(N4CCN(C)CC4)C=C3)N=C2)N1C5=CC=CC(C)=C5</chem>	D15
<chem>O=C1C=NC2=C(N=C(NC3=CC(C(C)C)=C(N4CCN(C)CC4)C=C3)N=C2)N1C5=CC=CC=C5C</chem>	D16
<chem>NC1C=NC2=C(N=C(NC3=CC(C(C)C)=C(N4CCN(C)CC4)C=C3)N=C2)N1C5=CC=CC=C5C</chem>	D17
<chem>CN(CC1)CCN1C(C=C2)=C(C(C)C)C=C2NC(N=C3)=NC4=C3N=CCN4C5=CC(C=CC5)=O</chem>	D18
<chem>O=C1C=NC2=C(N=C(NC3=CC(Br)=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D19
<chem>O=C1C=NC2=C(N=C(NC3=CC(CCl)=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D20
<chem>O=C1C=NC2=C(N=C(NC3=CC(CCl)=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=C(N)C=C6</chem>	D21
<chem>O=C1C=NC2=C(N=C(NC3=CC(CF)=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=C(N)C=C6</chem>	D22
<chem>O=C1C=NC2=C(N=C(NC3=CC(CF)=C(N4CCN(C5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(N)=C6</chem>	D23
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(F)=C6</chem>	D24
<chem>CN(CC1)CCN1C(C=C2)=C(C(C)C)C=C2NC(N=C3)=NC4=C3N=CC(N)N4C5=CC(C=CC5)=O</chem>	D25
<chem>CN(CC1)CCN1C(C=C2)=C(C(C)C)C=C2NC(N=C3)=NC4=C3N=CC(N)N4C5=CC=CC[C@H]5C(F)(F)F</chem>	D26
<chem>CC(C)C1=C(N2CCN(CC)CC2)C=CC(NC(N=C3)=NC4=C3N=CC(N)N4C5=CC=CC[C@H]5C(F)(F)F)=C1</chem>	D27
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(Cl)=C6</chem>	D28
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(Br)=C6</chem>	D29
<chem>O=C1C=NC2=C(N=C(NC3=CC(F)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D30
<chem>O=C1C=NC2=C(N=C(NC3=CC(Cl)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D31
<chem>O=C1C=NC2=C(N=C(NC3=CC(Br)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D32
<chem>O=C1C=NC2=C(N=C(NC3=CC(CCl)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC=C6</chem>	D33
<chem>O=C1C=NC2=C(N=C(NC3=CC(CCl)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=C(N)C=C6</chem>	D34
<chem>O=C1C=NC2=C(N=C(NC3=CC(CF)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=C(N)C=C6</chem>	D35
<chem>O=C1C=NC2=C(N=C(NC3=CC(CF)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC=CC(N)=C6</chem>	D36
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC(F)C=C6</chem>	D37
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC(Cl)C=C6</chem>	D38
<chem>O=C1C=NC2=C(N=C(NC3=CC=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC(Br)C=C6</chem>	D39
<chem>O=C1C=NC2=C(N=C(NC3=CC(Cl)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CCC=C6</chem>	D40
<chem>O=C1C=NC2=C(N=C(NC3=CC(F)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CCC=C6</chem>	D41
<chem>O=C1C=NC2=C(N=C(NC3=CC(Br)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CCC=C6</chem>	D42
<chem>O=C1C=NC2=C(N=C(NC3=CC(CCl)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CCC=C6</chem>	D43
<chem>O=C1C=NC2=C(N=C(NC3=CC(CF)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC(N)C=C6</chem>	D44
<chem>CC(C)C1=C(N2CCN(C(F)(F)F)CC2)C=CC(NC(N=C3)=NC4=C3N=CC(N4C5=CC=CC=C5)=O)=C1</chem>	D45

<chem>CC(C)C1=C(N2CCN(C(CI)(CI)CI)CC2)C=CC(NC(N=C3)=NC4=C3N=CC(N4C5=CC=CCC5C)=O)=C1</chem>	D46
<chem>CC(C)C1=C(N2CCN(C(CI)(CI)CI)CC2)C=CC(NC(N=C3)=NC4=C3N=CC(N4C5=CC=CC(C)=C5)=O)=C1</chem>	D47
<chem>O=C1C=NC2=C(N=C(NC3=CC(CBr)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC(N)C=C6D48</chem>	
<chem>O=C1C=NC2=C(N=C(NC3=CC(C(C)(C)C)=C(N4CCN(CC5CC5)CC4)C=C3)N=C2)N1C6=CC(N)C=C6</chem>	D49
<chem>CC(C)C1=C(N2CCN(CC(F)(F)F)CC2)C=CC(NC(N=C3)=NC4=C3N=CC(N)N4C5=CC=CCC5C)=C1</chem>	D50

Thirteen designed compounds, namely D02, D03, D04, D15, D16, D17, D18, D25, D26, D27, D45, D46, and D47, were predicted to have higher pIC₅₀ values than the most active compound (Figure 5b). The binding affinities of these 13 compounds were carried out using the MM-PB/GBSA, LIE, and FEP methods targeting wild-type and mutant FLT3 variants. The last 2 ns snapshots were extracted from the MD trajectories to calculate the MM-PB/GBSA binding free energy. Compounds D03, D15, D17-18, D25-26, and D46-47 had higher binding free energies than C31 in complex with wild-type receptors. In contrast, D02, D04, D15, D18, D25-26, and D46-47 exhibited higher binding free energies in complex with the mutant receptor. The last 1 ns average MD complexes of the designed compounds were carried out for the FPL and FEP simulation studies. Compounds D03, D04, D15-16, D27, and D46-47 had higher binding free energies in complex with FLT3 receptors than C31. The first 40% of the data is eliminated during the final FEP energy calculations to avoid the convergence error. The compounds D02, D04, D15, D18, D27, and D46-47 have stronger affinities for FLT3 receptors than C31. In Figure 6, the computed binding free energies from the MM-PB/GBSA, LIE, and FEP models are compared. The designed compounds with higher binding free energies than the most active compounds are designated by asterisks.

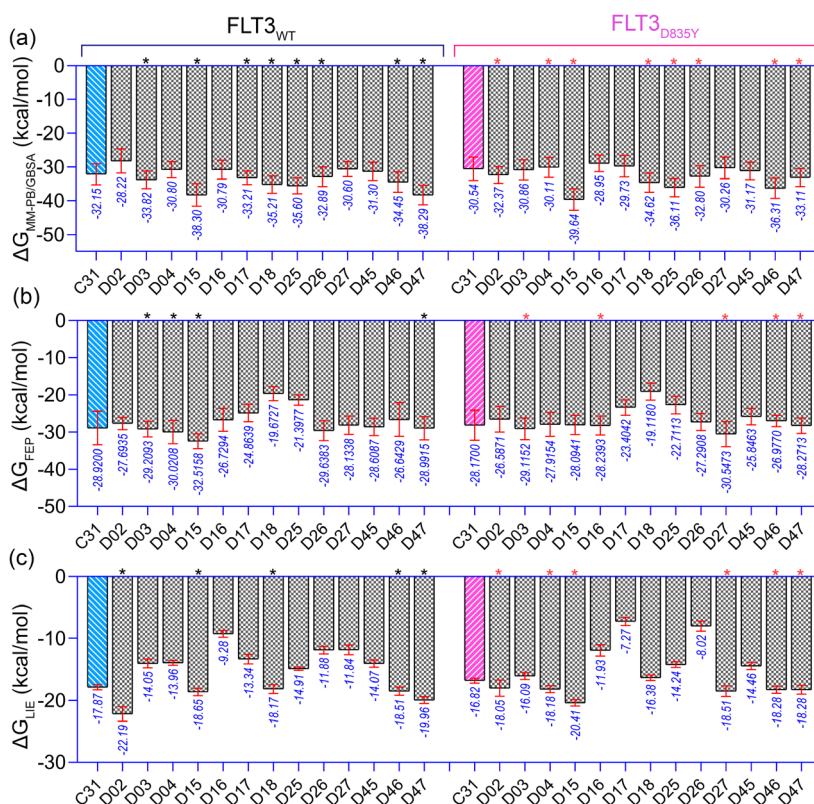


Figure 6. Comparison of the binding free energy of the designed compounds. (a) $\Delta G_{MM-PB/GBSA}$, (b) ΔG_{LIE} , and (c) ΔG_{FEP} binding free energy to wild-type (FLT3) and mutant (FLT3_{D835Y}) receptors. The higher estimated binding free energies of the designed compounds than the most active compound C31 are marked by the black (for wild-type FLT3) and pink asterisk (for FLT3_{D835Y}). The energy values shown below the standard deviation bar in blue are in kcal/mol.

3. Conclusion

In summary, overexpression and frequent mutations of FLT3 kinase remain an intriguing challenge in the treatment of AML. We have herein employed the binding studies of pteridin-7(8*H*)-one-based FLT3 inhibitors using docking, MD simulation, and multiple binding energy terms calculation. We applied MM-PB/GBSA, FPL, LIE, and FEP scoring functions to correlate the experimental binding energies and computed binding energies of the selected compounds. In MM-PB/GBSA and FEP methods, the acceptable correlation coefficients ($R_{MM-PB/GBSA} = 0.92$, and $R_{FEP} =$

0.71) were obtained, whereas the rupture force (F_{MAX}) and LIE are weakly correlated with the experimental binding energies. Although, each binding model distinguished the free energy differences between wild-type and mutant complexes of the most active compounds. Besides, we used a more rigorous biased sampling (US) simulation to evaluate the effective binding free energies between the FLT3-C31 and FLT3_{D835Y}-C31 complexes from the PMF curve. Following that, the statistically significant CoMFA ($q^2 = 0.768$, $r^2 = 0.982$) and CoMSIA ($q^2 = 0.844$, $r^2 = 0.972$) models were generated which showed the strong correlations between the observed and predicted inhibitory activity of the dataset compounds. The developed 3D-QSAR models had a satisfactory predictive power ($r^2_{pred} > 0.6$) and could be used to assess the inhibitory potential of the unknown compounds with analogous scaffolds. We designed several new compounds based on the SAR scheme by growing the chemical entities from the reference molecule C31. Thirteen compounds were predicted to have higher pIC_{50} than the most active compounds and were subjected to binding affinity evaluation by MM-PB/GBSA, LIE, and FEP calculations. Multiple compounds were determined to have greater binding free energy to wild-type and mutant receptors than most active compounds in different scoring models. Overall, these designed compounds have the potential to be lead molecules in future biochemical assays.

References

1. Wei, J.; Hui, A.-M., Role of Biomarkers in FLT3 AML. *Cancers* 2022, 14, (5), 1164.
2. Friedman, R., The molecular mechanisms behind activation of FLT3 in acute myeloid leukemia and resistance to therapy by selective inhibitors. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 2022, 1877, (1), 188666.
3. Almatani, M. F.; Ali, A.; Onyemaechi, S.; Zhao, Y.; Gutierrez, L.; Vaikari, V. P.; Alachkar, H., Strategies targeting FLT3 beyond the kinase inhibitors. *Pharmacology & Therapeutics* 2021, 225, 107844.
4. Yamatani, K.; Ai, T.; Saito, K.; Suzuki, K.; Hori, A.; Kinjo, S.; Ikeo, K.; Ruvolo, V.; Zhang, W.; Mak, P. Y., Inhibition of BCL2A1 by STAT5 inactivation overcomes resistance to targeted therapies of FLT3-ITD/D835 mutant AML. *Translational oncology* 2022, 18, 101354.

5. Nitika, J. W.; Hui, A.-M., Role of Biomarkers in FLT3 AML. *Cancers* 2022, 14, (5).
6. Kawase, T.; Nakazawa, T.; Eguchi, T.; Tsuzuki, H.; Ueno, Y.; Amano, Y.; Suzuki, T.; Mori, M.; Yoshida, T., Effect of Fms-like tyrosine kinase 3 (FLT3) ligand (FL) on antitumor activity of gilteritinib, a FLT3 inhibitor, in mice xenografted with FL-overexpressing cells. *Oncotarget* 2019, 10, (58), 6111.
7. Cilibrasi, V.; Spanò, V.; Bortolozzi, R.; Barreca, M.; Raimondi, M. V.; Rocca, R.; Maruca, A.; Montalbano, A.; Alcaro, S.; Ronca, R., Synthesis of 2H-Imidazo [2', 1': 2, 3][1, 3] thiazolo [4, 5-e] isoindol-8-yl-phenylureas with promising therapeutic features for the treatment of acute myeloid leukemia (AML) with FLT3/ITD mutations. *European Journal of Medicinal Chemistry* 2022, 235, 114292.
8. Smith, C. C.; Levis, M. J.; Perl, A. E.; Hill, J. E.; Rosales, M.; Bahceci, E., Molecular profile of FLT3-mutated relapsed/refractory patients with AML in the phase 3 ADMIRAL study of gilteritinib. *Blood advances* 2022, 6, (7), 2144-2155.
9. Perl, A. E.; Pratz, K. W., Incorporation of FLT3 Inhibitors Into the Treatment Regimens for FLT3 Mutated Acute Myeloid Leukemia: The Case for Total Therapy. *The Cancer Journal* 2022, 28, (1), 14-20.
10. Ahn, J.-S.; Kim, H.-J., FLT3 mutations in acute myeloid leukemia: a review focusing on clinically applicable drugs. *Blood research* 2022, 57, (Suppl 1), 32.
11. Arai, Y.; Chi, S.; Minami, Y.; Yanada, M., FLT3-targeted treatment for acute myeloid leukemia. *Int. J. Hematol.* 2022, 1-13.
12. Marensi, V.; Keeshan, K. R.; MacEwan, D. J., Pharmacological impact of FLT3 mutations on receptor activity and responsiveness to tyrosine kinase inhibitors. *Biochem. Pharmacol.* 2021, 183, 114348.
13. Ghosh, S.; Keretsu, S.; Cho, S. J., Molecular Modeling Studies of N-phenylpyrimidine-4-amine Derivatives for Inhibiting FMS-like Tyrosine Kinase-3. *International Journal of Molecular Sciences* 2021, 22, (22), 12511.
14. Im, D.; Jun, J.; Baek, J.; Kim, H.; Kang, D.; Bae, H.; Cho, H.; Hah, J.-M., Rational design and synthesis of 2-(1 H-indazol-6-yl)-1 H-benzo [d] imidazole derivatives as inhibitors targeting FMS-like tyrosine kinase 3 (FLT3) and its mutants. *J. Enzyme Inhib. Med. Chem.* 2022, 37, (1), 472-486.
15. Ghosh, S.; Keretsu, S.; Cho, S. J., 3D-QSAR, Docking and Molecular Dynamics Simulation Study of C-Glycosylflavones as GSK-3 β Inhibitors. *Journal of the Chosun Natural Science* 2020, 13, (4), 170-180.
16. Ghosh, S.; Keretsu, S.; Cho, S. J., Computational Modeling of Novel Phosphoinositol-3-kinase γ Inhibitors Using Molecular Docking, Molecular Dynamics, and 3D-QSAR. *Bulletin of the Korean Chemical Society* 2021, 42, (8).
17. Ghosh, S.; Cho, S. J., Structural Insights from Molecular Modeling of Isoindolin-1-One Derivatives as PI3K γ Inhibitors against Gastric Carcinoma. *Biomedicines* 2022, 10, (4).

-
18. Ghosh, S.; Keretsu, S.; Cho, S. J., Designing of the N-ethyl-4-(pyridin-4-yl)benzamide based potent ROCK1 inhibitors using docking, molecular dynamics, and 3D-QSAR. *PeerJ* 2021, 9.
 19. Sousa da Silva, A. W.; Vranken, W. F., ACPYPE-Antechamber python parser interface. *BMC research notes* 2012, 5, (1), 1-8.
 20. Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E., gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *Journal of Chemical Theory and Computation* 2021, 17, (10), 6281-6291.
 21. Duan, L.; Liu, X.; Zhang, J. Z. H., Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein–Ligand Binding Free Energy. *Journal of the American Chemical Society* 2016, 138, (17), 5722-5728.
 22. Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P., CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. 2012.
 23. Tam, N. M.; Nguyen, M. T.; Ngo, S. T., Evaluation of the absolute affinity of neuraminidase inhibitor using steered molecular dynamics simulations. *J. Mol. Graphics Model* 2017, 77, 137-142.
 24. Ngo, S. T.; Vu, K. B.; Bui, L. M.; Vu, V. V., Effective estimation of ligand-binding affinity using biased sampling method. *ACS omega* 2019, 4, (2), 3887-3893.
 25. Ngo, S. T.; Nguyen, H. M.; Quan, P. M.; Truong, V. K.; Tung, N. T.; Vu, V. V., Assessing potential inhibitors of SARS-CoV-2 main protease from available drugs using free energy perturbation simulations. *RSC Advances* 2020, 10, (66), 40284-40290.
 26. Ghosh, S.; Cho, S. J., Structure–activity relationship and in silico development of c-Met kinase inhibitors. *Bulletin of the Korean Chemical Society* 2022, 43, (6), 882-892.
 27. Keretsu, S.; Ghosh, S.; Cho, S. J., Molecular Modeling Study of c-KIT/PDGFR α Dual Inhibitors for the Treatment of Gastrointestinal Stromal Tumors. *International journal of molecular sciences* 2020, 21, (21), 8232.
 28. Keretsu, S.; Ghosh, S.; Cho, S. J., Computer aided designing of novel pyrrolopyridine derivatives as JAK1 inhibitors. *Scientific reports* 2021, 11, (1), 1-12.
 29. Bang, S. J.; Cho, S. J., 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.
 30. Gadhe, C. G.; Kothandan, G.; Cho, S. J., 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.
 31. Gadhe, C. G.; Madhavan, T.; Kothandan, G.; Cho, S. J., In silico quantitative structure-activity relationship studies on P-gp modulators of

- tetrahydroisoquinoline-ethyl-phenylamine series. *BMC structural biology* 2011, 11, (1), 5.
32. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N., Some case studies on application of “rm2” metrics for judging quality of quantitative structure–activity relationship predictions: emphasis on scaling of response data. *Journal of computational chemistry* 2013, 34, (12), 1071-1082.
 33. Sahigara, F.; Ballabio, D.; Todeschini, R.; Consonni, V., Assessing the validity of QSARs for ready biodegradability of chemicals: an applicability domain perspective. *Current computer-aided drug design* 2014, 10, (2), 137-147.
 34. Todeschini, R.; Ballabio, D.; Grisoni, F., Beware of unreliable Q 2! A comparative study of regression metrics for predictivity assessment of QSAR models. *Journal of Chemical Information and Modeling* 2016, 56, (10), 1905-1913.
 35. Velázquez-Libera, J. L.; Durán-Verdugo, F.; Valdés-Jiménez, A.; Núñez-Vivanco, G.; Caballero, J., LigRMSD: a web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. *Bioinformatics* 2020, 36, (9), 2912-2914.
 36. Ramírez, D.; Caballero, J., Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules (Basel, Switzerland)* 2018, 23, (5).
 37. Hadni, H.; Bakhouch, M.; Elhallaoui, M., 3D-QSAR, molecular docking, DFT and ADMET studies on quinazoline derivatives to explore novel DHFR inhibitors. *J. Biomol. Struct. Dyn.* 2021, 1-15.
 38. Smith, C. C.; Lin, K.; Stecula, A.; Sali, A.; Shah, N. P., FLT3 D835 mutations confer differential resistance to type II FLT3 inhibitors. *Leukemia* 2015, 29, (12), 2390-2392.
 39. Wang, Z.; Hu, B.; An, Y.; Wang, J., Exploring the Resistance Mechanisms of Distal D835V Mutation in FLT3 to Inhibitors. *Oxidative medicine and cellular longevity* 2022, 2022.

Suparna Ghosh Ph.D. Thesis

Chosun University, Department of Biomedical Sciences

PART III

Comparative binding affinity analysis of dual CDK2/FLT3 inhibitors

1. Introduction

AML can be characterized by the abnormal proliferation of immature hematopoietic stem cells and bone marrow cells[1]. Overexpression and frequent mutations of FLT3 have been reported in approximately 30% of AML cases in patients with a poor prognosis. Although pharmacological inhibition of FLT3 appears to be a promising therapy for leukemia, several issues, notably drug resistance during clinical trials, limit its therapeutic advantages[2, 3].

Recent research has shown that the knockdown of CDK2 induces AML differentiation without losing the embryo model. CDK2 activates the G1/S border checkpoint and leads the cell cycle through the S phase after binding to cyclins E and A[4-6]. This opens up the possibility of targeting CDK2 in specific AML types. However, the effective treatment of AML by suppressing myeloid differentiation rather than proliferation remains a significant challenge. Most FDA-approved drugs aim to achieve antiproliferative effects against abnormal cells. The high similarity across the active sites of CDK members and other kinases makes ATP competitive inhibitors nonspecific or very toxic[7-9]. Therefore, further development of selective CDK2 inhibitors is required.

Investigating the underlying mechanisms of ligand binding and a systematic search for lead-like compounds play a crucial role in the discovery of kinase drugs. Several in silico methods have been developed, including docking, MM-PB/GBSA, linear interaction energy, virtual screening, QSAR, and free energy perturbation to rank the binding potential of the ligands, thus lowering the cost of therapeutic development[10]. Herein, we carried out molecular modeling studies of 3*H*-Pyrazolo[4,3-*f*]quinoline-containing compounds as anti-AML inhibitors reported by Dayal et al. [11]. These compounds exhibited a wide range of inhibitory activities against CDK2 and FLT3-ITD. Therefore, the compounds might be promising therapeutics in heterogeneous leukemia where multiple kinases are overexpressed. The experimental binding energy (ΔG_{EXP}) of the compounds were deduced from their IC_{50} values. However, when the $-\log IC_{50}$ (pIC_{50}) of CDK2 and FLT3 were correlated

with each other, the squared correlation (r^2) was found to be only 0.29, suggesting that the selectivity and inhibitory potency of the compounds are partially independent from each other. In addition, compounds C24, C25, C26 and C30 have nonspecific inhibitory activities (or $IC_{50} > 20 \mu M$), although they exhibited the inhibitory potency against CDK2. These phenomena have led us to postulate that more structural investigations utilizing computational modeling may enhance the feasibility of developing novel chemical entities that are selective to the CDK2 receptor.

The structure-activity relationship (SAR) study was conducted using CoMSIA, a well-known 3D-QSAR method to identify the key structural features of inhibitor compounds, which could increase the inhibitory potency and selectivity of CDK2 over FLT3. Binding free energy estimations between protein and small molecules are roughly categorized into three broad classes: 1. Endpoint methods, 2. Alchemical methods and 3. Pathway methods[12]. We employed MM-GBSA and US simulations as endpoint and pathway methods for nine structurally diverse ligands for the binding energy calculation, which are computationally inexpensive compared to the alchemical routes. Thereafter, the computed energies using the above methods were correlated with the experimental binding free energies. The correlation coefficients between the computed and experimental binding energies were determined.

2. Methodology

2.1. Dataset preparation

The inhibitory activities (IC_{50} values) of the compounds were converted into the $-\log IC_{50}$ or pIC_{50} values to prepare the data set. Compound 18 is available in the CDK2 bound co-crystallized form (PDB ID 7B7S), which is considered a biological 3D conformer. Thus, C18 was selected as the template molecule and 3*H*-Pyrazolo[4,3-*f*]quinoline as the common skeleton. Based on the template, the remaining compounds were sketched, minimized, and added partial charges in Sybyl X 2.1 according to our

previously described study[13]. Some compounds that were reported to have nonspecified activity, i.e., $IC_{50} > 20\mu M$, were discarded during model development. To develop the CoMSIA model of CDK2 fifty-two datapoints were used, whereas forty-eight datapoints were considered for FLT3 CoMSIA model. The compounds were aligned on the common skeleton one after another, using C18 as a reference. The pIC_{50} values spanned over three log units and were classified into low, medium, and high activity groups. From each group, the compounds were selected randomly by maintaining structural diversity to build the test set, while the remaining compounds were treated as a training set to achieve a final ratio between training and test set compound 3:1.

2.2. CoMSIA model development and Contour map analysis

In CoMSIA, descriptor fields were calculated in a 3D cube with a grid spacing of 1 Å. A gaussian function was applied to calculate the similarity indices for all grid points. Five descriptor fields, steric (S), electrostatic (E), hydrophobic (H), H-bond donor (D), and H-bond acceptor (A) were calculated in the permutation-combination process. Partial least square (PLS) method was used to correlate the physicochemical descriptors and pIC_{50} of the compounds. Cross-validation and no-cross-validation methods were used consecutively to build a statistically reliable CoMSIA model. Additional methodological details can be found in our previous studies[14-17]. The predictive correlation coefficient (r^2_{pred}) was determined to evaluate the predictive power of the developed model. The r^2_{pred} can be calculated as follows:

$$r^2_{pred} = (SD-PRESS)/SD \quad (1)$$

SD is the sum of the squared deviations between the actual activities of the test set and the average activity values of the training set compounds. The PRESS is the sum of the squared deviations between the actual and predicted activities of the test set compounds.

The field effects of the chemical descriptors of the CoMSIA model were visually represented as 3D StDev*Coeff contour polyhedrons with different color codes. Each

color signifies the favorable or unfavorable chemical substitution that might increase or decrease the inhibitory potency of the compounds in the data set.

2.3. Structure preparation and molecular docking

Compound C18 bound CDK2 crystal structure (PDB ID 7B7S) was retrieved from the RCSB PDB database. Water molecules, ions, crystal ligands, and the Cyclin-A2 polypeptide chain were removed. The missing residues and loops of the CDK2 chain were modeled by MODELLER and Chimera 1.14[18]. Compound C18 was self-docked into the predefined binding pocket with grid dimensions of 30×40×40 points in X, Y, and Z directions and grid spacing of 0.375 Å using Autodock4.2 in the graphical version of AutoDockTools[19]. More methodological details about receptors, ligands, and grid-box preparation can be found in our previous papers[20, 21]. One hundred docking runs were performed, and the final docking solution was collected from the lowest energy and RMSD cluster. During docking pose validation, polar, nonpolar, and RMSD from the crystal pose were considered. A similar protocol was applied during the cross-docking simulations of compounds C01, C11, C16, C19, C24, C36, C38, and C48. The final solution of each individual docked complex was undertaken for the MD simulation study.

2.4. MD simulation and binding energy calculations

GROMACS 2019.5[22] was used for the MD simulation study using Amber ff99SB force field according to our previous studies[23, 24]. ACEPYPE[25] was used to parameterize the ligand molecules with AM1-BCC charges. The protein-ligand complexes were immersed in a cubic box with a minimum periodic boundary of 10 Å in the X, Y, and Z directions of the protein atoms. TIP3P water models were used to dissolve the system, and an adequate amount of Na⁺ and Cl⁻ counterions were used to neutralize the increased ion concentration to 0.15M. The system was carried out for 10000 steps minimization, 200 ps NVT simulation to attain the temperature of 300 K, and 400 ps NPT simulation runs to attain the pressure of 1 bar, followed by 60 ns of

production simulation. Positional restraint was applied during the NVT and NPT ensemble and was omitted during the production simulation. Finally, the RMSDs of the protein-ligand complexes were calculated using the built-in '*gmx rms*' function in gromacs.

According to our previous study[26], the binding free energy ($\Delta G_{MM-GBSA}$) and entropy term ($-T\Delta S$) as Interaction Entropy (IE) were calculated using the *gmx_MMPBSA* package[27]. The binding free energy from the MM-GBSA calculation can be written as:

$$\Delta G_{MM-GBSA} = E_{RL} - E_R - E_L \quad (2)$$

where, the free energy of the complex (RL), receptor (R), and ligand (L) are represented by E_{RL} , E_R , and E_L respectively. The $\Delta G_{MM-GBSA}$ is further defined as:

$$\Delta G_{MM-GBSA} = \Delta H - T\Delta S = \Delta E_{MM} + \Delta E_{SOL} - T\Delta S \quad (3)$$

The ΔH stands for the enthalpy, whereas ΔE_{MM} and ΔE_{SOL} represents molecular mechanics energy in the gas phase and solvation-free energy, respectively.

The entropy term ($-T\Delta S$) is expressed as,

$$-T\Delta S = KT \ln \langle e^{\beta(\Delta E_{ele} + \Delta E_{vdw})} \rangle \quad (4)$$

where K, T, and β represent the Boltzmann constant, the simulation temperature (300 K), and the inverse of temperature ($1/kT$), respectively. Electrostatic and van der Waals interactions between receptor and ligand are expressed by ΔE_{ele} and ΔE_{vdw} , respectively. The ensemble average of receptor-ligand interactions is denoted by $\langle \rangle$.

2.5. US simulation study

The US simulation was carried out in a two-step process according to this study[28]. In the first step, the unbinding pathway was determined by caver analysis. The complexes were placed in the center of a periodic water box of $10 \times 8 \times 10 \text{ nm}^3$ and ionized with 0.15 M NaCl. Following that, Minimization, NPT, and SMD simulations were performed. In SMD, the ligands were dissociated at a distance of 3 nm using an external harmonic force at a velocity (v) of 0.006 nm/ps and a cantilever spring constant (k) of 400 kJ/mol/nm².

In the second step, the unbinding snapshots were extracted from the SMD trajectory with an increasing spacing distance of 0.2 nm between the center-of-mass (COM) distance of DFG residues and ligands. However, an additional frame was allocated, resulting in a 0.1 nm spacing between the first three windows. A total of 13 snapshots were taken for the umbrella sampling simulation. Each window was subjected to Minimization, 500 ps NPT, and 5 ns US simulation runs. The weighted histogram analysis method (WHAM) was used to determine the potential mean force (PMF) using the built-in 'gmx wham' function by the following equation:

$$w(\xi) = -\beta \ln \left[\frac{P(\xi)}{P(\xi_0)} \right] \quad (5)$$

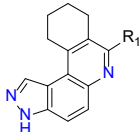
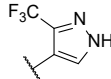
In this equation, $\beta = k_B T$, where k_B represents Boltzmann constant and T is the simulation temperature. The terms $P(\xi)$ and $P(\xi_0)$ signifies the probability density of ξ and reference point ξ_0 (i.e. $w(\xi_0) = 0$), respectively. The binding free energy (ΔG_{US}) from the US simulation was estimated as the difference between the highest and lowest values from the PMF curve.


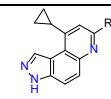

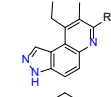

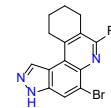

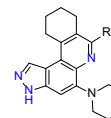

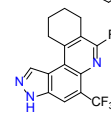

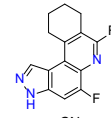

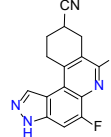

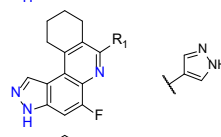

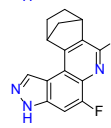
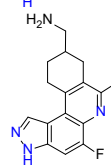
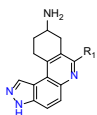
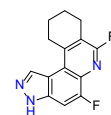
3. Results and Discussion

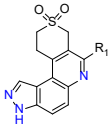
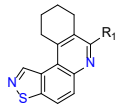
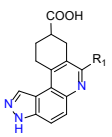
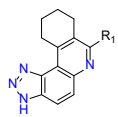
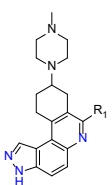
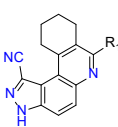
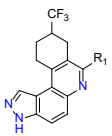
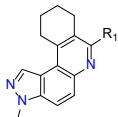
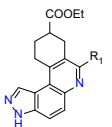
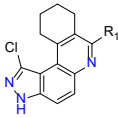
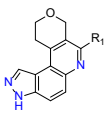
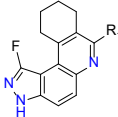


3.1. Statistical analysis of CoMSIA models

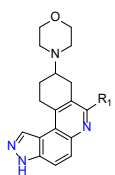
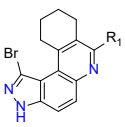
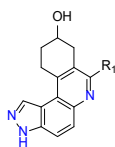
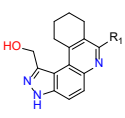
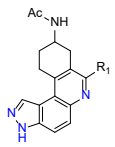
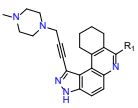
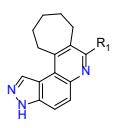
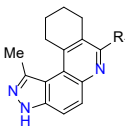

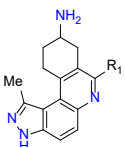
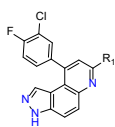
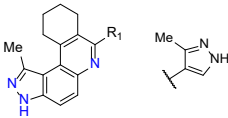
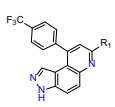
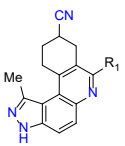
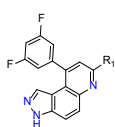
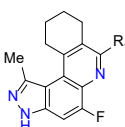
The 2D structure of all compounds and their respective inhibitory activity (pIC_{50}) for CDK2 and FLT3 are shown in Table 1. The compounds aligned with the common core taking C18 as a reference.

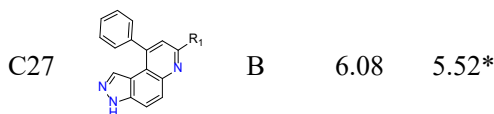
Table 1. Dataset compounds and their inhibitory activities against CDK2 and FLT3

#Cpds	Structure	R ₁	Structure (A)		#Cpds	Structure	R ₁	Structure (B)	
			CDK2 (pIC ₅₀)	FLT3-ITD (pIC ₅₀)				CDK2 (pIC ₅₀)	FLT3-ITD (pIC ₅₀)
									

C01	A		6.25	5.69	C28		B	7.06	6.43
C02	A		6.62	5.50	C29		B	7.33	8.39
C03	A		6.74*	6.87	C30		B	4.96	NA
C04	A		7.14	7.04	C31*		B	5.24*	5.90
C05	A		5.83	5.41	C32*		B	6.97*	7.56
C06	A		5.94	5.43	C33		B	7.40	7.92
C07	A		6.25	6.26	C34		B	6.95	7.06
C08	A		7.05	6.67*	C35		B	7.40	7.36*
C09	A		6.97	6.17	C36		B	7.30	7.79
C10	A	B	7.40	8.30	C37		B	7.23*	7.72*
C11	B		5.22	5.50	C38		B	7.76	8.15

C12		B	6.63	6.09*	C39*		B	5.28	5.07
C13		B	5.84	6.24	C40		B	5.89*	6.38
C14		B	6.57*	6.85	C41		B	NA	NA
C15		B	6.85	6.89	C42		B	5.48	5.77
C16		B	6.95	6.92*	C43		B	5.63	7.38
C17		B	6.72	6.96	C44		B	6.56	7.50*
C18		B	7.39	7.76*	C45		B	7.20	7.67

C19		B	6.04	7.79	C46		B	6.48	8.52
C20		B	7.69	7.79	C47		B	6.85*	8.52*
C21		B	6.49	7.69	C48		B	5*	7.74
C22		B	7.50*	8.52	C49		B	6.36	8.22
C23		B	5.76	7.46*	C50		B	6.19	6.67
C24		B	5.89	NA	C51		B	6.14*	7.37
C25		B	5.08	NA	C52		B	6.47	7.82*
C26		B	6.32*	NA	C53		B	6.80	9.00



(*) Test set

The S, E, H, A, and D descriptor fields were used in different combinations to obtain the statistically robust CoMSIA models. While developing the CoMSIA model using the CDK2 training set, the PLS analysis yielded the highest q^2 and r^2 values of 0.672 and 0.933 at ONC of 5 in the SEAD combination. The values of q^2 and r^2 were also higher than the minimum accepted criterion ($q^2 > 0.5$, $r^2 > 0.6$)[29], indicating a good correlation between the physicochemical descriptors and the inhibitory activities of the compounds. Therefore, the SEAD combination was selected as the final CoMSIA model for CDK2 inhibitors. We determined the χ^2 value for internal model validation, which was obtained as 0.097, satisfying the accepted parameters of <1.0 . The external validation and predictive power of the QSAR model are then determined by calculating r^2_{pred} from the predicted pIC_{50} values of the chemicals in the test set. The r^2_{pred} was found to be 0.888 in the CoMSIA/SEAD model, which also satisfied the acceptable norms, i.e., $r^2_{\text{pred}} > 0.6$, indicating an excellent predictive power of the model.

Continuing that, we developed the CoMSIA model for FLT3 by taking the training set compounds and their inhibitory activity (pIC_{50}) to FLT3. The SEHD combination yielded the best q^2 and r^2 values of 0.631 and 0.859 at ONC of 3, the highest among the other combinations. The internal validation parameter χ^2 was obtained as 0.326, whereas the external validation or predictive power of the model r^2_{pred} was anticipated to be 0.887. Overall, the statistics of the SEHD combination followed the acceptable norms, and thus selected as the final CoMSIA model for FLT3. The statistical results of the CoMSIA models and PLS regression plots are shown in Table 2 and Figure 1a,b.

Table 2. Statistical result and validation of CDK2 and FLT3 CoMSIA models

Statistical parameters	CDK2 SEAD	FLT3 SEHD	Threshold values
q^2	0.672	0.631	> 0.5
ONC	5	3	
SEP	0.482	0.661	
r^2	0.933	0.859	> 0.6
SEE	0.218	0.409	<< 1
F-value	91.912	64.984	
BS- r^2	0.958	0.904	
BS-SD	0.016	0.034	
χ^2	0.097	0.326	< 1.0
r_{pred}^2	0.888	0.887	> 0.6
S (%)	35.8	10.7	
E (%)	19.2	27.8	
H (%)	-	34.5	
A (%)	30.6	-	
D (%)	14.4	17.0	

q^2 : squared cross-validated correlation coefficient; ONC: optimal number of components; SEP: standard error of prediction; r^2 : squared correlation coefficient; SEE: standard error of estimation; F-value: F-test value; BS- r^2 : Bootstrapping squared correlation coefficient; χ^2 : Chi-square value; r_{pred}^2 : predictive correlation coefficient; S(%): steric contribution; E(%): electrostatic contribution; H(%): Hydrophobic contribution; A(%): H-bond acceptor contribution; D(%): H-bond donor contribution.

3.2. Contour map analysis and SAR study

In addition to predicting the activity of the novel compounds, CoMSIA in 3D-QSAR vividly illustrated favorable and unfavorable chemical descriptor fields as distinctive colored contour maps. In the CoMSIA/SEAD model of CDK2, the steric, electrostatic, hydrophobic, and H-bond donors contributed 25.8%, 19.2%, 30.6%, and 24.4% in the field distribution, respectively. Together, the steric and hydrophobic descriptors account for more than 60% of the overall field contribution. Figure 1c illustrates the favorable and unfavorable space for steric and hydrophobic chemical group substitutions, using C18 as a reference. The green contour at the *meta*-position of the thiopyran ring indicated that bulky steric groups at that position might enhance the

inhibitory potency. On the contrary, yellow contours were present near the thiopyran and indazole rings, suggesting that bulky steric groups could reduce the inhibitory activity. A large magenta contour appeared at the *para*-position of the thiopyran ring, indicating that an H-bond accepting group would be advantageous in that region. Two brownish-grey contours near the indazole and 3-(trifluoromethyl)-1*H*-pyrazole moieties indicated the unfavorable substitutions of H-bond acceptors at these positions.

In the FLT3 CoMSIA/SEHD model, the field distributions of steric, electrostatic, hydrophobic, and H-bond donor descriptors were found to be 10.7%, 29.8%, 24.5%, and 35.0%, respectively. Electrostatic and H-bond donor fields account for more than 60% of the total field contribution and are shown in Figure 1d. The blue and red contours indicated favorable positions for electropositive and electronegative chemical groups, respectively, which could increase the bioactivity of inhibitor compounds. Similarly, the orange and dark-gray contours represented the favorable and unfavorable positions for hydrophobic chemical groups.

Considering the observations above, chemical group modification of compound C18 according to the CoMSIA/SEAD contours of CDK2 would likely promote enhanced bioactivity. While alteration of the chemical groups in the area designated according to the CoMSIA/SEHD of FLT3 contours might improve CDK2 selectivity over FLT3.

3.3. Docking, MD simulation, and binding energy calculation

To study molecular interactions and protein-ligand binding stability, we carried out the docking and MD simulation studies of the compounds C01, C11, C16, C18, C19, C24, C36, C38, and C48. These nine compounds are selected considering their diverse bioactivities while maintaining their structural diversities. Compound C18 was self-docked into the ATP pocket of CDK2, and the docked pose was compared with the crystal-bound pose. The RMSD between the crystal pose and docked pose was obtained to 1.07 Å, reflecting the satisfactory validation of the overall docking protocol. The ligand was anchored to the hinge loop by forming an H-bond interaction

with residue L83. Residue F80 and catalytic residue K33 formed the π - π and π - σ interaction with the indazole and pyrazole rings, respectively. Additional hydrophobic interactions were observed with residues I10, A33, and L134.

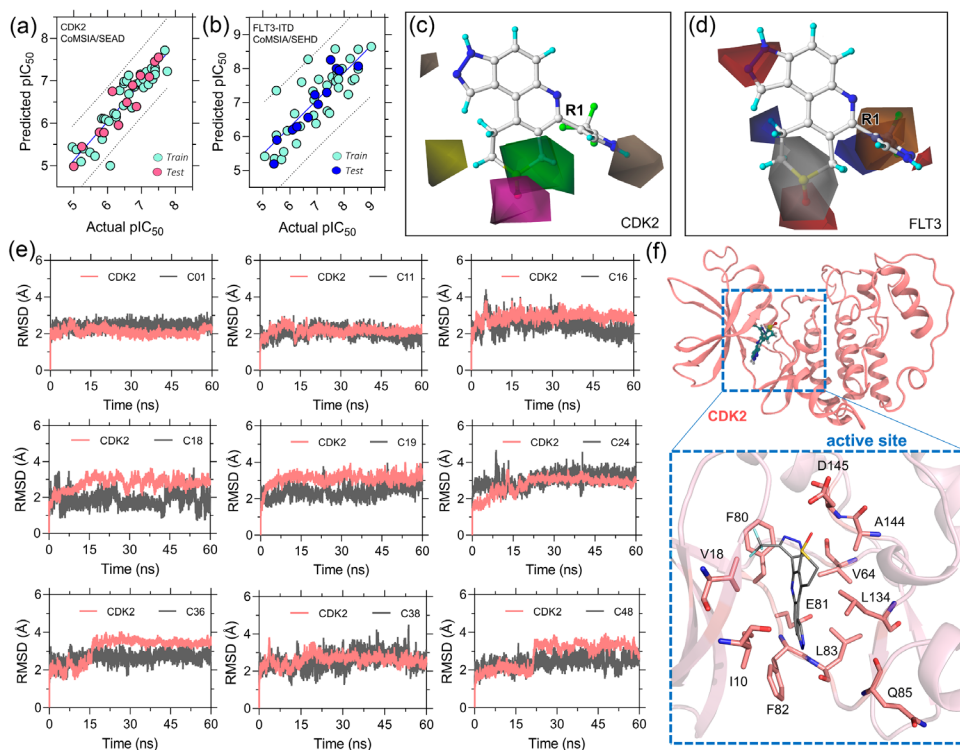


Figure 1. CoMSIA analysis, RMSD plots, and residue-specific binding energy contribution to the ligand. PLS regression plots of (a) CoMSIA/SEAD model of CDK2 system and (b) CoMSIA/SEAD model of FLT3 system. (c) Contour map analysis of the CoMSIA/SEAD model of CDK2. The steric (S) and H-bond acceptor (A) are shared 35.8%, and 30.6% (>60%) of the field contribution are shown in colored contours. Favorable and unfavorable positions for steric substitutions are shown in green and yellow contours. Similarly, favorable and unfavorable H-bond acceptor positions are depicted in magenta and brownish-grey contours. (d) Contour map analysis of CoMSIA/SEAD model of FLT3 in which electrostatic (E) and hydrophobic (H) have shared 27.4% and 34.5% (>0.60%) field contributions. The favorable substitutions for electropositive and electronegative groups are shown by blue and red contours, whereas orange-grey contours depict the favorable-unfavorable position for hydrophobic groups. (e) RMSD analysis of protein backbones and ligand heavy atoms for the 60 ns production run in salmon and dark gray. (f) Common residues that decompose the critical binding energies to the ligand within the active site are shown in the stick representation.

The rest of the compounds were cross-docked into the binding pocket of CDK2. Figure 1e depicts the RMSD changes of the docked complexes employed in all-atom production simulations for 60 ns. Throughout the MD run, the compounds remained stable in the complex with FLT3, and no significant variations in RMSD curves were observed. The MM-GBSA binding free energies were calculated using the last 2.5 ns frames of each MD trajectory. Compounds C01, C11, C16, C18, C19, C24, C36, C38, and C48 exhibited the binding energy of -26.32 kcal/mol, -25.81 kcal/mol, -29.55 kcal/mol, -27.27 kcal/mol, -21.27 kcal/mol, -23.89 kcal/mol, -32.68 kcal/mol, -32.59 kcal/mol, and -23.81 kcal/mol, respectively. The residue-specific binding energy decomposition analysis to the ligand is important information for the drug development process. The residues within 4\AA distance from the ligand atoms that contributed to the critical binding free energy are reported and shown in the stick representation around compound C18 in Figure 1f.

3.4. US simulation analysis

The binding affinity of two biomolecules is often estimated using atomistic simulation approaches rather than quantum chemical calculations, which requires less computational resources. However, in conventional MD, it is still impossible to sample the whole phase space of large conformational changes of biomolecules, which often requires microseconds to milliseconds of simulation. To precisely estimate the binding affinity, enhanced sampling techniques such as steered MD (SMD), metadynamics (MetaD), scaled MD, and replica-exchanged MD (REMD) seem to be more effective. Many kinases feature solvent-exposed active sites, which allow ligands to move in and out of the binding pocket without inducing conformational changes to the receptors. Thus, US simulation is a popular choice for determining the binding affinity of ligands with reasonable accuracy. The unbinding direction of the ligands from the ATP pocket was determined by caver analysis as void channels in which the molecules could be forced to mobilize in the solvent. The two predicted channels with the highest ranking (P1 and P2) are shown in Figure 2. The ligands were then forced to dissociate

from the active site at a distance of 3 nm with a constant velocity to the X-axis direction by SMD simulation, as illustrated in Figure 3a. Initially, the pulling force started from zero and gradually increased over the simulation time. The ligands were dissociated entirely from the active site at the maximum force potential (F_{\max}), also known as the rupture force. Then, the force was steeply decreased when the ligand was mobilized into the solvent. Theoretically, F_{\max} could be correlated with the binding affinity of the ligands because a higher binding affinity required a higher rupture force (F_{\max}) for ligand dissociation. However, in the current study, we only used the SMD simulation to generate input for the US simulations. The free energy in the US simulation was derived as a PMF curve along with the reaction coordinates (ξ) at X-axis by WHAM analysis in Figure 3b.

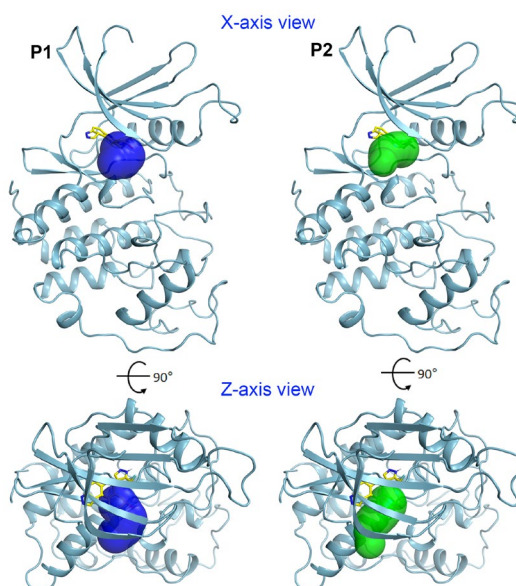


Figure 2. Cover predicted unbinding direction of the ligand. The top two predicted results (P1 and P2) are shown in blue and green vdW channels.

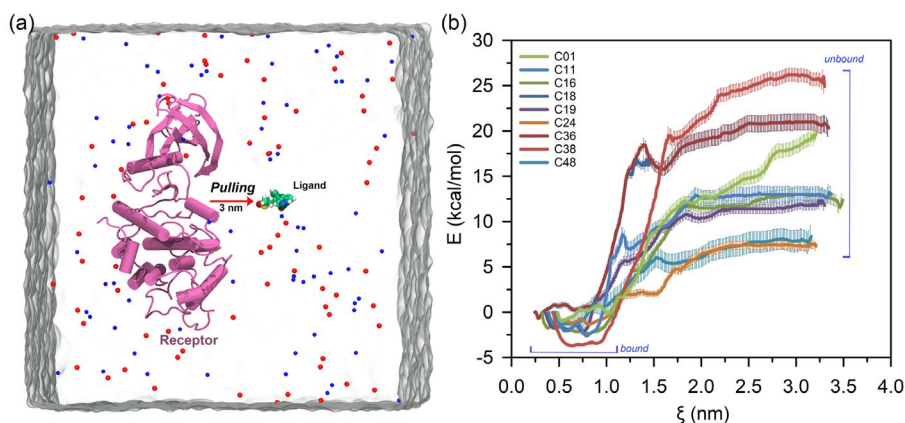


Figure 3. Overview of the unbinding direction and estimation of the PMF curve in the US simulation process. (a) The ligand molecule is forced to mobilize into the solvent at a distance of 3 nm from the X-axis direction from the active site. The constant velocity of 0.006 nm/ps and a cantilever spring constant of 400 kJ/mol/nm². (b) The PMF energy curves of the compounds C01, C11, C16, C18, C19, C24, C36, C38, and C48 in the complex with CDK2 were obtained by WHAM analysis. The error was estimated using 1000 rounds of bootstrapping runs.

The free energy in the PMF curve starts at zero and drops to a minimum value, termed the *bound* state. Subsequently, the curve increased to a maximum energy value (E) at the ξ of 1.5-2.0 nm and remained steady. From that ξ onwards, the nonbonded interactions between the protein and ligands were diminished, termed the *unbound* state. The binding energy change from the US simulation (ΔG_{US}) was roughly estimated as the difference between the maximum and minimum values of the PMF curves. Compounds C01, C11, C16, C18, C19, C24, C36, C38, and C48 exhibited binding energy (ΔG_{US}) of -19.46 kcal/mol, -15.67 kcal/mol, -15.30 kcal/mol, -21.60 kcal/mol, -14.43 kcal/mol, -8.79 kcal/mol, -21.62 kcal/mol, -29.79 kcal/mol and -9.62 kcal/mol, respectively.

The experimental binding energies were deduced from the inhibitory activity (IC_{50}) of the compounds and shown along with the computed binding free energies in Table 2. In Figure 4, the computed binding energies of the MM-GBSA and US simulation are correlated with the experimental binding energies. The correlation coefficient ($R_{MM-GBSA}$) between $\Delta G_{MM-GBSA}$ and ΔG_{EXP} was obtained as 0.77. Similarly, the

correlation coefficient (R_{US}) was determined to be 0.81 between ΔG_{US} and ΔG_{EXP} . In both MM-GBSA and US methods, the binding free energies were overestimated compared to the experimental values. Therefore, in MM-GBSA methods, other entropic calculations such as normal mode (nmode) or quasi-harmonic (QH) computation might be useful instead of the IE method. On the other hand, the initial configuration of the receptor-ligand complex is crucial for SMD setup. Furthermore, many frames were discarded during input generation, and only countable frames were acquired for biased sampling simulation. As a result, there could be a higher chance of non-equilibrium phase space sampling error. In such a case, analyzing the free energy landscape of the complexes may be informative; but, as previously noted, significant computing resources would be needed for convergence sampling.

Table 3. Binding free energy calculation from MM-GBSA and US simulation in kcal/mol

Complexes	ΔG_{EXP}	$\Delta G_{MM-GBSA}$ ($\pm SD$)	ΔG_{US} ($\pm SD$)
CDK2-C01	-8.63	-26.32 \pm 2.27	-19.46 \pm 0.74
CDK2-C11	-7.21	-25.81 \pm 3.56	-15.67 \pm 0.87
CDK2-C16	-9.60	-29.55 \pm 2.36	-15.30 \pm 0.66
CDK2-C18	-10.21	-27.27 \pm 3.02	-21.60 \pm 0.91
CDK2-C19	-8.34	-21.27 \pm 4.50	-14.43 \pm 0.62
CDK2-C24	-8.13	-23.89 \pm 2.82	-8.79 \pm 0.46
CDK2-C36	-10.08	-32.68 \pm 2.55	-21.62 \pm 0.89
CDK2-C38	-10.72	-32.59 \pm 3.39	-29.79 \pm 0.70
CDK2-C48	-6.90	-23.81 \pm 2.62	-9.62 \pm 1.20

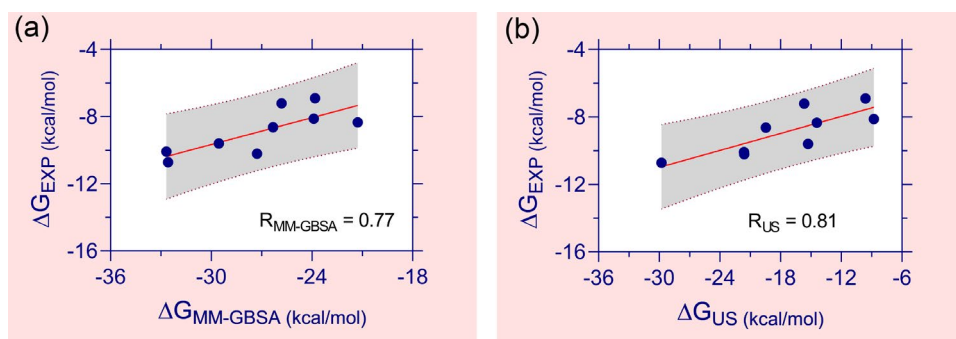


Figure 4. Correlation plots between experimental binding energies and computed binding energies. (a) The correlation coefficient ($R_{MM-GBSA}$) between the $\Delta G_{MM-GBSA}$ and ΔG_{EXP} is estimated to be 0.77. (b) The correlation coefficient (R_{US}) between the ΔG_{US} and ΔG_{EXP} is estimated to be 0.81.

4. Conclusion

To summarize, the current work established statistically reliable CoMSIA models of CDK2 and FLT3 inhibitors. The predictive power of both models was evaluated by determining the r^2_{pred} using the test set compounds. The dominant field contributors in the CDK2 CoMSIA model are steric and H-bond acceptor descriptors, as evidenced by the green and magenta contours surrounding the thiopyran ring. In contrast, electrostatic and hydrophobic descriptors dominated the field contributions CoMSIA model of FLT3 and appeared near the indazole, pyrazole, and thiopyran rings. Therefore, chemical substitution along the CDK2 CoMSIA contours might enhance bioactivity, while bypassing the chemical space according to the FLT3 contours could promote CDK2 selectivity.

Binding affinity predictions of inhibitor molecules to receptors are crucial steps in computer-aided drug development. As a result, we employed binding-free energy estimation using MM-GBSA and US simulation methods. Acceptable correlation coefficients ($R_{MM-GBSA} = 0.77$, $R_{US} = 0.81$) were obtained between the computed and experimental binding energies with reasonable accuracy. Furthermore, the experimental binding energy could be expressed by the following equations -

$$\Delta G_{\text{EXP}} = 0.2664 \times \Delta G_{\text{MM-GBSA}} - 1.670 \quad (6)$$

$$\Delta G_{\text{EXP}} = 0.1679 \times \Delta G_{\text{US}} - 5.953 \quad (7)$$

Using the MM-GBSA and US methods described here and in combination with the above equations (6) and (7), the binding affinity of the new compound could be reliably estimated. The findings might aid in developing more potent and selective CDK2 targeting inhibitors in the treatment of AML by suppressing the myeloid differentiation pathway.

Reference

1. Drula, R., et al., Exploiting the ubiquitin system in myeloid malignancies. From basic research to drug discovery in MDS and AML. *Blood Reviews*, 2022: p. 100971.
2. Bewersdorf, J.P. and O. Abdel-Wahab, Translating recent advances in the pathogenesis of acute myeloid leukemia to the clinic. *Genes & Development*, 2022. **36**(5-6): p. 259-277.
3. Singh, A., et al., The relationship between emotional well-being and understanding of prognosis in patients with acute myeloid leukemia (AML). *Supportive Care in Cancer*, 2022. **30**(1): p. 897-906.
4. Bazzar, W., et al., Pharmacological inactivation of CDK2 inhibits MYC/BCL-XL-driven leukemia in vivo through induction of cellular senescence. *Cell Cycle*, 2021. **20**(1): p. 23-38.
5. Wang, L., et al., Discovery of a first-in-class CDK2 selective degrader for AML differentiation therapy. *Nature Chemical Biology*, 2021. **17**(5): p. 567-575.
6. Chantkran, W., et al., Interrogation of novel CDK2/9 inhibitor fadraciclib (CYC065) as a potential therapeutic approach for AML. *Cell death discovery*, 2021. **7**(1): p. 1-14.
7. Osman, A.E. and M.W. Deininger, Chronic myeloid leukemia: modern therapies, current challenges and future directions. *Blood Reviews*, 2021. **49**: p. 100825.
8. Saleemizadeh Parizi, M., et al., Myeloid-derived suppressor cells (MDSCs) in brain cancer: challenges and therapeutic strategies. *Inflammopharmacology*, 2021. **29**(6): p. 1613-1624.
9. Chen, S., et al., TANK-binding kinase 1 inhibitor GSK8612 enhances daunorubicin sensitivity in acute myeloid leukemia cells via the AKT-CDK2 pathway. *American journal of translational research*, 2021. **13**(12): p. 13640.
10. Nazarova, A.L. and V. Katritch, It all clicks together: In silico drug discovery becoming mainstream. *Clinical and Translational Medicine*, 2022. **12**(4).

11. Dayal, N., et al., 3H-Pyrazolo[4,3-f]quinoline-Based Kinase Inhibitors Inhibit the Proliferation of Acute Myeloid Leukemia Cells In Vivo. *Journal of Medicinal Chemistry*, 2021. **64**(15): p. 10981-10996.
12. Reif, M.M. and M. Zacharias, Computational Tools for Accurate Binding Free-Energy Prediction. *Methods Mol Biol*, 2022. **2385**: p. 255-292.
13. Ghosh, S., S. Keretsu, and S.J. Cho, 3D-QSAR, Docking and Molecular Dynamics Simulation Study of C-Glycosylflavones as GSK-3 β Inhibitors. *Journal of the Chosun Natural Science*, 2020. **13**(4): p. 170-180.
14. Ghosh, S., S. Keretsu, and S.J. Cho, Designing of the N-ethyl-4-(pyridin-4-yl)benzamide based potent ROCK1 inhibitors using docking, molecular dynamics, and 3D-QSAR. *PeerJ*, 2021. **9**.
15. Gadhe, C.G., G. Kothandan, and S.J. Cho, 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): p. 861-871.
16. Gadhe, C.G., et al., In silico quantitative structure-activity relationship studies on P-gp modulators of tetrahydroisoquinoline-ethyl-phenylamine series. *BMC structural biology*, 2011. **11**(1): p. 5.
17. San Juan, A.A. and S.J. Cho, 3D-QSAR study of microsomal prostaglandin E 2 synthase (mPGES-1) inhibitors. *Journal of Molecular Modeling*, 2007. **13**(5): p. 601-610.
18. Yang, Z., et al., UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *Journal of structural biology*, 2012. **179**(3): p. 269-278.
19. Huey, R. and G.M. Morris, Using AutoDock 4 with AutoDocktools: a tutorial. *The Scripps Research Institute, USA*, 2008. **8**: p. 54-56.
20. Ghosh, S. and S.J. Cho, Structural Insights from Molecular Modeling of Isoindolin-1-One Derivatives as PI3K γ ; Inhibitors against Gastric Carcinoma. *Biomedicines*, 2022. **10**(4).
21. Ghosh, S., S. Keretsu, and S.J. Cho, Computational Modeling of Novel Phosphoinositol-3-kinase γ Inhibitors Using Molecular Docking, Molecular Dynamics, and 3D-QSAR. *Bulletin of the Korean Chemical Society*, 2021. **42**(8).
22. Van Der Spoel, D., et al., GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 2005. **26**(16): p. 1701-1718.
23. Ghosh, S., S. Keretsu, and S.J. Cho, Molecular Modeling Studies of N-phenylpyrimidine-4-amine Derivatives for Inhibiting FMS-like Tyrosine Kinase-3. *International Journal of Molecular Sciences*, 2021. **22**(22): p. 12511.
24. Ghosh, S. and S.J. Cho, Binding Studies and Lead Generation of Pteridin-7 (8H)-one Derivatives Targeting FLT3. *International Journal of Molecular Sciences*, 2022. **23**(14): p. 7696.
25. Da Silva, A.W.S. and W.F. Vranken, ACPYPE-Antechamber python parser interface. *BMC research notes*, 2012. **5**(1): p. 1-8.

26. Ghosh, S. and S.J. Cho, Structure–activity relationship and in silico development of c-Met kinase inhibitors. *Bulletin of the Korean Chemical Society*, 2022.
27. Valdés-Tresanco, M.S., et al., gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *Journal of Chemical Theory and Computation*, 2021. **17**(10): p. 6281-6291.
28. Ngo, S.T. and M.Q. Pham, Umbrella Sampling-Based Method to Compute Ligand-Binding Affinity, in *Computational Methods for Estimating the Kinetic Parameters of Biological Systems*. 2022, Springer. p. 313-323.
29. Méndez-Lucio, O., et al., 3D-QSAR studies on purine-carbonitriles as cruzain inhibitors: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). *MedChemComm*, 2011. **2**(11): p. 1058-1065.

APPENDIX

A. List of publications

1. **Ghosh S**, Cho SJ. *Comparative binding affinity analysis of dual CDK2 inhibitors over FLT3*. (Accepted, Bull. Korean Chem. Soc.)
2. **Ghosh S**, Cho SJ. *Binding studies and lead generation of pteridine-7(8H)-one derivatives targeting FLT3*. Int J Mol Sci. 2022, 23, 7696. doi: 10.3390/ijms23147696. (IF: 6.20, 2022).
3. **Ghosh S**, Cho SJ. *Structure-Activity Relationship and in Silico Development of c-Met Kinase Inhibitors*. Bull. Korean Chem. Soc. 2022; 43:882-892. doi: 10.1002/bkcs.12537. (IF: 1.24, 2022).
4. **Ghosh S**, Cho SJ. *Structural Insights from Molecular Modeling of Isoindolin-1-One Derivatives as PI3K γ Inhibitors against Gastric Carcinoma*. Biomedicines. 2022; 10:813. doi: 10.3390/biomedicines10040813. (IF: 6.02, 2020).
5. Keretsu S, **Ghosh S**, Cho SJ. *Computer aided designing of novel pyrrolopyridine derivatives as JAK1 inhibitors*. Sci Rep. 2021 Nov 29;11(1):23051. doi: 10.1038/s41598-021-02364-2. (IF: 4.99, 2022).
6. Park JY, Lee D, Lee JJ, Gim J, Gunasekaran TI, Choi KY, Kang S, Do AR, Jo J, Park J, Park K, Li D, Lee S, Kim H, Dhanasingh I, **Ghosh S**, Keum S, Choi JH, Song GJ, Sael L, Rhee S, Lovestone S, Kim E, Moon SH, Kim BC, Kim S, Saykin AJ, Nho K, Lee SH, Farrer LA, Jun GR, Won S, Lee KH; Alzheimer's Disease Neuroimaging Initiative. *A missense variant in SHARPIN mediates Alzheimer's disease-specific brain damages*. Transl Psychiatry. 2021 Nov 16;11(1):590. doi: 10.1038/s41398-021-01680-5. (IF: 7.98, 2022).
7. **Ghosh S**, Keretsu S, Cho SJ. *Molecular Modeling Studies of N-phenylpyrimidine-4-amine Derivatives for Inhibiting FMS-like Tyrosine Kinase-3*. Int J Mol Sci. 2021 Nov 19;22(22):12511. doi: 10.3390/ijms222212511. (IF: 6.20, 2022).
8. **Ghosh S**, Keretsu S, Cho SJ. *Designing of the N-ethyl-4-(pyridin-4-yl)benzamide based potent ROCK1 inhibitors using docking, molecular dynamics, and 3D-QSAR*. PeerJ. 2021 Aug 9;9: e11951. doi: 10.7717/peerj.11951. (IF: 3.06, 2022).

9. **Ghosh S**, Keretsu S, Cho SJ. *Computational Modeling of Novel Phosphoinositol-3-kinase γ Inhibitors Using Molecular Docking, Molecular Dynamics, and 3D-QSAR*. Bull. Korean Chem. Soc. 2021, 42, doi:10.1002/bkcs.12305. **(IF: 1.24, 2022)**.
10. **Ghosh S**, Keretsu S, Cho SJ. *3D-QSAR, Docking and Molecular Dynamics Simulation Study of C-Glycosylflavones as GSK-3 β Inhibitors*. Journal of the Chosun Natural Science. 2020 Dec 31;13(4):170-80. Doi: 10.13160/RICNS.2020.13.4.170.
11. Cao TP, Yi H, Dhanasingh I, **Ghosh S**, Choi JM, Lee KH, Ryu S, Kim HS, Lee SH. *Non-catalytic-Region Mutations Conferring Transition of Class A β -Lactamases Into ESBLs*. Front Mol Biosci. 2020 Nov 27; 7:598998. doi: 10.3389/fmolb.2020.598998. **(IF: 6.11, 2022)**.
12. Keretsu S, **Ghosh S**, Cho SJ. *Molecular Modeling Study of c-KIT/PDGFR α Dual Inhibitors for the Treatment of Gastrointestinal Stromal Tumors*. Int J Mol Sci. 2020 Nov 3;21(21):8232. doi: 10.3390/ijms21218232. **(IF: 6.20, 2022)**.
13. **Ghosh S**, Dhanasingh I, Ryu J, Kim SW, Lee SH. *Crystal Structure of Cytochrome cL from the Aquatic Methylophilic Bacterium Methylophaga aminisulfidivorans MPT*. J Microbiol Biotechnol. 2020 Aug 28;30(8):1261-1271. doi: 10.4014/jmb.2006.06029. **(IF: 3.27, 2022)**.

Conference Presentation

14. **Ghosh S**, Dhanasingh I, Ryu J, Kim SW, Lee SH. *Structural Insights of Cytochrome cL from the Aquatic Methylophilic Bacterium Methylophaga aminisulfidivorans^{MPT} Reveals Distinctive Features for Efficient Electron Transfer during Methanol Oxidation*. 한국미생물학회 학술대회논문집. 2020 Oct:76-76.

Manuscript under preparation

15. **Ghosh S**, Cho SJ. *Structural study and relative binding affinity estimation of nuclear focal adhesion kinase inhibitors*. 2022-2023.

B. Acknowledgement

Firstly, I would like to express my sincere gratitude to my advisor Prof. Seung Joo Cho for his immense support and valuable advice. He has been a wonderful mentor to me. I applaud his continuous encouragement, patience, and willingness to share his vast research expertise and thoughts with me. Throughout my Ph.D., he extended moral support as well as the freedom to pursue the research I desired. This Ph.D. would not have been feasible without his mentoring and frequent feedback.

My sincere thanks to the thesis committee members for evaluating my thesis. Their insightful comments and questions have made my defense enjoyable and memorable.

I am also grateful to Dr. Seketoulie Keretsu, my senior lab mate, for his assistance and fantastic pleasant lab time. The time we spent together will always be remembered and cherished.

I am grateful to my family and friends who have supported me during this precious journey. All these years, their encouragement has been a source of motivation for me.