August 2020
Ph.D. Dissertation

Computer-Aided Drug Modelling of the Receptor Tyrosine Kinase Inhibitors

Graduate School of Chosun University

Department of Biomedical Sciences

Swapnil Pandurang Bhujbal

Computer-Aided Drug Modelling of the Receptor Tyrosine Kinase Inhibitors

타이로신 카이네이즈 억제제의 컴퓨터를 활용한 신약 모델링 연구

August 28, 2020

Graduate School of Chosun University

Department of Biomedical Sciences

Swapnil Pandurang Bhujbal

Computer-Aided Drug Modelling of the Receptor Tyrosine Kinase Inhibitors

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

May 2020

Graduate School of Chosun University

Department of Biomedical Sciences

Swapnil Pandurang Bhujbal

CONTENTS

I. CHAPTER 1 - INTRODUCTION

II. CHAPTER 2

ABBREVIATIONS

ATP	Adenosine Triphosphate
PK	Protein Kinase
RTK	Receptor Tyrosine Kinases
DFG	Asp-Phe-Gly motif
FDA	The Food and Drug Administration
AML	Acute Myeloid Leukemia
FLT3	FMS-like tyrosine kinase 3
ITD	Internal Tandem Duplication
TKD	Tyrosine Kinase Domain
PDGFR	Platelet-derived growth Factor Receptor
RMSD	Root Mean Square Deviation
GA	Genetic Algorithm
CoMFA	Comparative Molecular Field Analysis
RF-CoMFA	Region-Focused CoMFA
CoMSIA	Comparative Molecular Similarity Indices Analysis
2D	Two Dimensional
3D	Three Dimensional
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
CCC	Concordance Correlation Coefficient
DOPE	Discrete Optimized Protein Energy
GAFF	General AMBER Force Field

HL	Hologram Length
LFO	Leave Five out
LOO	Leave Out One
NOC	Number Of Components
PLS	Partial Least Square
MerTK	Mer Receptor Tyrosine Kinase
MD	Molecular Dynamics
LGA	Lamarckian Genetic Algorithm
RET	REarranged during Transfection
MTC	Medullary Thyroid Carcinoma
NSCLC	Non-Small Cell Lung Cancer
PTC	Papillary Thyroid Carcinoma

LIST OF TABLES

CHAPTER 4

LIST OF FIGURES

Abstract

Computer-aided Drug Modelling of the Receptor Tyrosine Kinase Inhibitors

Swapnil Pandurang Bhujbal

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

Department of Biomedical Sciences

Graduate School of Chosun University

Receptor tyrosine kinases (RTKs) are a subclass of tyrosine kinases that mediate cell-to-cell communication and control a wide range of complex biological functions. A well control of cell division and morphogenesis along with intended cell death is crucial to ensure normal tissue growth and patterning. Aberrant variations that cause an imbalance of the cellular signals which are responsible for these events lead to the uncontrolled cell proliferation that is known as cancer. Since most of these signals are regulated by RTKs, it makes them an important drug target for the treatment of various types of cancer. This has created significant concern in designing and development of kinase inhibitors to treat these diseases. In our current study, we have used various tyrosine kinases such as RET, MerTK and FLT3 kinase. We utilized a combined approach of molecular docking, molecular dynamics simulation and MM/PBSA binding free energy calculations to identify and study the crucial residues that participate in inhibition of tyrosine kinases. The results of CoMFA and CoMSIA revealed structural insights in enhancing the activity of the kinase inhibitors. Based on this structural information, we have designed some new RTK inhibitors. In summary, our study could be supportive to comprehend the inhibitory mechanism of these kinases thus assists in designing more efficient and potent inhibitors.

초록

타이로신 카이네이즈 억제제의 컴퓨터를 활용한 신약 모델링 연구

스왚닐 판두랑 부즈발

지도교수: 조 승 주

의과학과

조선대학교 대학원

타이로신 카이네이즈 수용체는 넓은 범위의 복잡한 생물학적 기능을 제어하는 것과 세포사이의 통신을 매개하는데 관여한다. 의도된 세포 사멸과 함께 세포 분열 및 형태 형성의 원활한 제어는 정상적인 조직 성장 및 패턴화에 중요하다. 이러한 기능을 담당하는 세포 신호의 불균형을 유발하는 비정상적인 변화는 제어되지 않은 세포 증식, 즉, 암을 일으킨다. 이러한 신호의 대부분은 RTK에 의해 조절되므로 다양한 유형의 암 치료에 중한 약물 표적이됩니다. 따라서, 이러한 질병을 치료하기위한 키나제 억제제의 설계 및 개발에 많은 연구가 있다. 현재 연구에서 우리는 RET, MerTK 및 FLT3 키나제와 같은 다양한 타이로신 카이네이즈를 사용했다. 우리는 분자 도킹, 분자 역학 시뮬레이션 및 MM/PBSA 결합 자유 에너지 계산을 통하여 티로신 카이네이즈의 억제에 관여하는 중요한 아미노산 잔기들을 확인하고 연구했습니다. CoMFA 및 CoMSIA의 결과는 카이네이즈 억제제의 활성을 향상시키고 새로운 강력한 억제제를 설계하는 데 있어서 구조에 관계한 정보를 제공하였다. 이 구조 정보를 기반으로, 우리는 몇 가지 새로운 RTK 억제제를 설계했습니다. 결론적으로, 본 연구는 이들 카이네이즈의 억제 매커니즘을 이해하여 보다 효율적이고 강력한 억제제의 설계에 도움이 된다.

Chapter 1 - Introduction

General Introduction on Kinases

1.1. Overview on Protein Kinases

Protein kinases are the largest enzyme superfamily engaged in cell signal transduction [1]. They play an important role in controlling various cellular processes such as transcription, cell cycle progression, metabolism, cell movement, cytoskeletal rearrangement, differentiation and apoptosis [2, 3]. They are encoded by almost 2% of eukaryotic genes and 518 protein kinases are predicted using the human genome sequencing, among which 90 are tyrosine kinases, and these comprise 58 receptor tyrosine kinases [1, 3, 4]. Kinases are classified into a hierarchy of groups, families, and subfamilies primarily based on sequence comparison of their catalytic domain, structure outside of the catalytic domains and biological function [3, 5]. There are 50 distinctive kinase families among yeast, invertebrate and mammalian kinomes [2]. The kinase dendrogram is shown in figure 1 depicts the sequence similarity among catalytic domains.

Protein kinases are also classified as serine/threonine or tyrosine or dual substrate kinases based on their ability to catalyze the transfer of the terminal phosphate of ATP (adenosine triphosphate) to substrates that usually contain a serine, threonine or tyrosine residue. All known tyrosine kinases and serine/threonine kinases share an associated catalytic domain of ~270 amino acids [2]. The serine/threonine family is more diverse than the tyrosine kinase family [6].

1.2 Structure of Protein Kinase

A number of protein kinase family members possess a sequence of eleven conserved subdomains that is the catalytic core [1, 6]. The kinases can occur in an active or an inactive state. They are either present as a single domain or may be covalently linked to other regulatory domains [4]. The kinases are bilobal with the active site situated in a deep cleft between the lobes. The fold of the catalytic domain is comprised of approximately 300 residues. It consists of a smaller N-terminal lobe of nearly 80

Figure 1. A dendogram of the human kinome

residues that exhibits five-stranded β-sheets with one α-helix, the C-helix, which is crucial in regulation, and approximately 200 residues C-terminal lobe (Fig. 2). A hinge region connects the two lobes. A groove at the interface between these two lobes forms the ATP-binding pocket of the catalytic site. The contacts also contain three significant regions namely the flexible glycine-rich loop between strands β1 and β2, the C-helix, and the hinge region between the lobes. The protein substrate-binding site covers the region which is called as the activation loop usually composed of 20-30 residues between the conserved DFG motif and the conserved APE motif in the C-terminal region [3, 4, 6, 7].
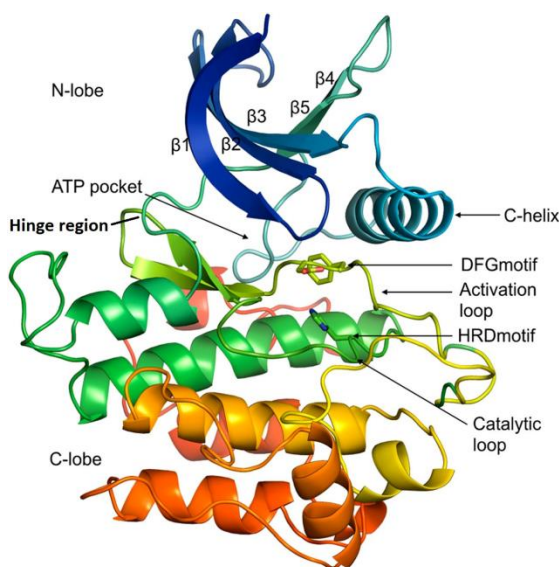
Figure 2. 3D-Structure of a protein kinase

1.3 Protein Kinases as Drug Target

Mutation of different genes (including kinases) is one of the most intense ways to the cancer development and progression. The mutated kinases can turn into unmanageably active and hence cause various cellular irregularities, leading to

cancer initiation [8]. The deregulation of kinase activity can lead to dramatic changes in cellular processes [8].

A comparison of chromosomal mapping of kinases with known disease loci showed that 164 kinases are often observed in tumors and approximately 80 kinases are implicated in other diseases [1, 4]. Over expressed receptor tyrosine kinases (RTK) result in autophosphorylation that stimulates signal transduction pathways leading to tumor metastasis [3, 5]. The human cancers are caused by an usual RTK activation which is mediated by four mechanisms such as gain-of-function mutations, chromosomal rearrangements, genomic amplification and/or autocrine activation [3, 6]. The reason that kinases play an important role in singalling pathways makes them an attractive target for anti-cancer drug design. To date, 58 genes encoding RTKs have been identified in the human genome, 30 of which have been found to be dysregulated in human cancers [1, 7]. This relevance to various diseases has encouraged us to select few tyrosine kinases which are reported to cause different types of human cancers, study their inhibitory mechanisms and design new inhibitors.

1.4 Types of kinase Inhibitors

There are mainly three types of kinase inhibitors. Type I inhibitors recognize active kinase conformation which is also known as DFG-in conformation (DFG: Asp-Phe-Gly) and blocks active kinase by competing with the ATP. On the contrary, type II inhibitors recognize the inactive conformation of the kinase which is called as DFG-out. This type of inhibitors inhibits kinases by contacting both ATP binding pocket and an adjacent allosteric site which is seen only in DFG-out conformation. They are more specific in nature due to the allosteric site they contact is less conserved among kinases than ATP binding site. Lastly, Allosteric inhibitors bind to the allosteric site which is away from the ATP-binding site and regulates kinase activity in an allosteric manner. They tend to possess high kinase selectivity.

1.5. Importance of designing new RTK drugs

Tyrosine kinases are a group of approximately 90 enzymes responsible for the activation of signal transduction cascades. These kinases act as an "on" and "off" switch for many cellular functions [6, 8]. Many of these kinases are being investigated for the treatment of cancer and various other diseases. Though many inhibitors are approved by FDA (The Food and Drug Administration) and some are under clinical trials, until now there is no drug specific to particular kinase. Most of them are multi-kinase inhibitors. It has been reported for few anticancer drugs, patients acquire resistance after prolonged treatment. The resistance usually arises from secondary mutations within the target RTK. Combination-drug therapies that target many pathways regulated by RTKs will enhance efficacy and may decrease resistance [5, 8]. Hence, it is important to design and develop new anti-cancer drugs.

Chapter 2

Design of New Therapeutic Agents Targeting FLT3 Receptor Tyrosine Kinase

1. Introduction

Acute myeloid leukemia (AML) which is characterized as a heterogeneous disease is caused by mutational or cytogenetic characteristics [9]. AML shows uncontrolled proliferation of hematopoietic cells in bone marrow. This leads to the overproduction of abnormal white blood cells, infection, organ dysfunction which in turn reduces the formation of red blood cells [10, 11]. Progression of AML is linked to the mutation or overexpression of FMS-like tyrosine kinase 3 (FLT3) [12, 13]. Two types of mutations usually occur in FLT3; an internal tandem duplication (ITD) positioned in the juxtamembrane domain and point mutation within the kinase activation loop [10, 14]. The majority of FLT3 mutations in AML are contributed by the ITD [9]. Although AML is well understood, the improvement in patient prognosis and the treatment is less in the last two decades [10, 11, 15, 16]. Even with existing treatment practices, 70% of patients died of AML within 1 year after diagnosis [17, 18]. Hence, FLT3 has been recognized as a promising therapeutic target for the treatment of AML.

FLT3 is a class III receptor tyrosine kinase (RTK) that is expressed in both normal as well as malignant lympho-hematopoietic cells [10]. It plays an important role in the immune response and development and proliferation of hematopoietic stem cells [9, 10]. Class III RTKs consist of an extracellular domain, a transmembrane (TM) helix, and an intracellular domain that comprised of a juxtamembrane (JM) segment, a tyrosine kinase domain (TKD), and a C-terminal tail [9, 19]. The binding of the FLT3 ligand to the transmembrane helix causes dimerization and subsequent FLT3 autophosphorylation of tyrosine residues within the TKD, leading to their activation and signal transduction through several signaling cascades, including RAS, SRC, and STAT5 pathways [10, 20-23].

Several FLT3 inhibitors have been reported against AML. Some of these inhibitors such as lestaurtinib [24], midostaurin [25], tandutinib [26], crenolanib [27], sorafenib [28], and sunitinib [29] have been evaluated in clinical trials. But

these inhibitors were relatively nonspecific for FLT3 and commonly inhibit class III RTKs such as KIT proto-oncogene receptor tyrosine kinase (KIT) and platelet-derived growth factor receptor (PDGFR) [30]. Subsequently, quizartinib, crenolanib, pacritinib, ponatinib and gilteritinib were developed as more potent and specific FLT3 inhibitors. Among these, quizartinib is currently in clinical trial however, drug resistance mutations are reported in response to quizartinib treatment [31]. In addition, there are several AML inhibitors in Phase I or Phase I/II clinical trials such as E6201 (dual MEK/FLT3 inhibitor), TAK-659 (dual SYK/FLT3 inhibitor,) and SKLB1028 (multikinase inhibitor) [19]. Nevertheless, none of these inhibitors have been approved as FLT3 kinase inhibitor due to their non-specificity and off-target effect. Therefore it is crucial to design potent as well as specific FLT3 kinase inhibitors.
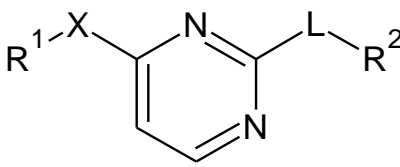
Various computational studies on FLT3 kinase have been reported. Zorn *et al*. have reported the molecular dynamics studies of apo and holo FLT3 (PDB ID 1RJB) with quizartinib to study the binding mechanism of the inhibitor and the conformational changes of the activation loop [9]. In many other studies, docking of the quizartinib and various inhibitors with FLT3 was reported to understand the quizartinib resistant mutations and the ligand-protein interactions respectively [9, 14, 31-33]. Important residues involved in FLT3 inhibition has been reported in the above studies. Residues Cys694 and Phe830 were found to be important for the binding of quizartinib to the FLT3. CEP701, PKC412 and sunitinib were reported to interact with the active site of FLT3, forming H-bonds with residues Cys694 and Glu692. Other residues such as Leu616, Val624, Ala642, Tyr693 and Gly697 were also reported to be involved in ligand binding [14, 32, 33]. In the current study, we have performed the molecular docking followed by 3D-Quantitative structure-activity relationship (3D-QSAR) on the series of 63 diaminopyrimidine derivatives to understand the structural requirements for designing the potent FLT3 antagonists.
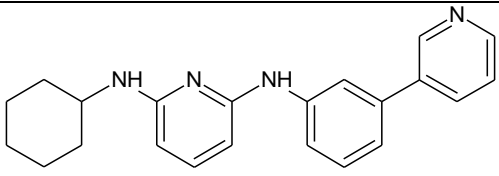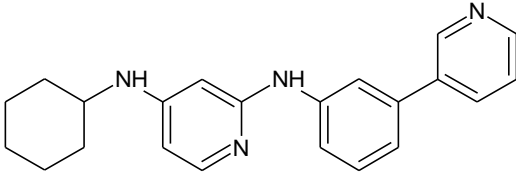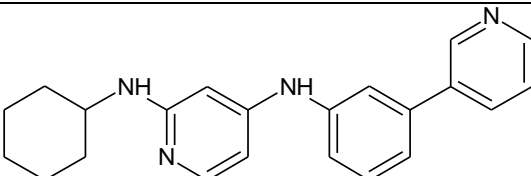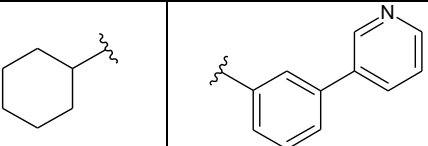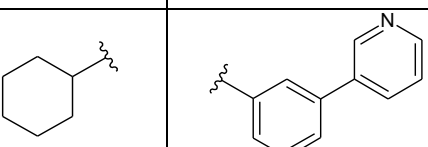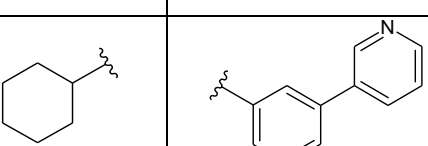
2. Materials and Methods
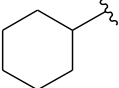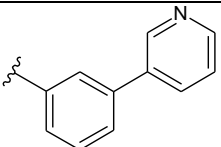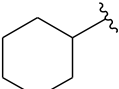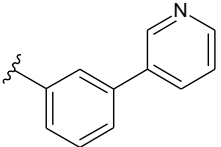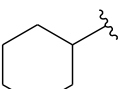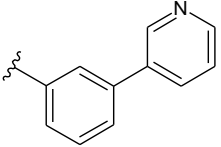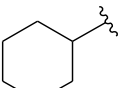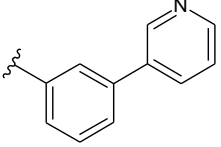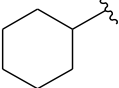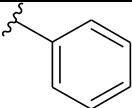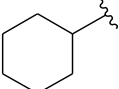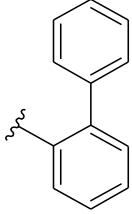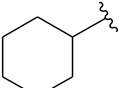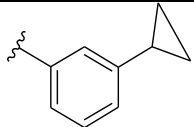
2.1 Dataset
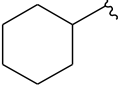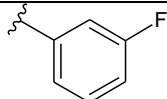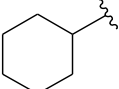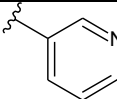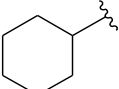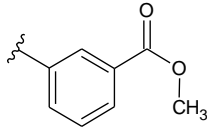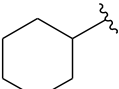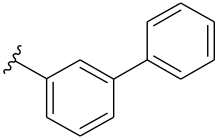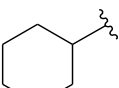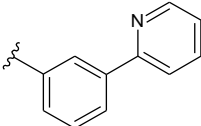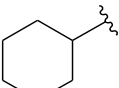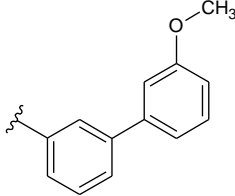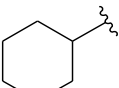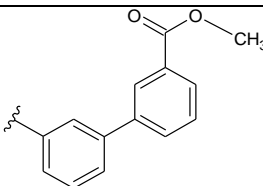
A series of 63 diaminopyrimidine derivatives as FLT3 kinase inhibitors reported by Jarusiewicz *et al* were taken for this study [10]. The experimental IC_{50} values were converted into pIC_{50} values ($-logIC_{50}$) and used as dependent variables to derive 3D-QSAR models. The dataset spans 3.6 logarithmic units which could be reasonable for QSAR studies [34]. The compounds and their corresponding activities are depicted in Table **1**. Sybyl-X 2.1 was used to sketch the structures of all the compounds and energy minimized with the tripos force field [35]. Subsequently, MMFF94 were applied as a partial charge [36] scheme. Considering the activity of compounds, the dataset was divided into 45 compounds as a training set to derive a model and 18 compounds as a test set to validate the model. The test set compounds were randomly selected while possessing various activities.
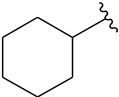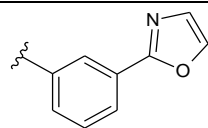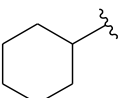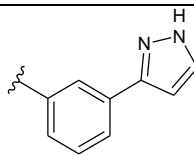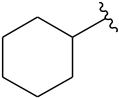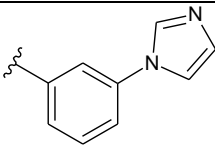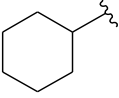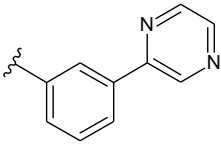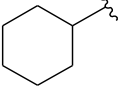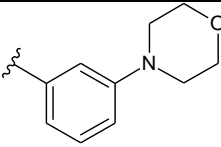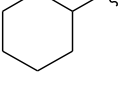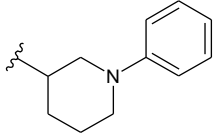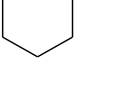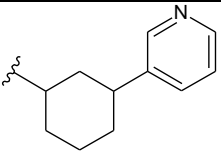
Table 1. Structure and biological activity values of diaminopyrimidine derivatives as FLT3 kinase inhibitors.
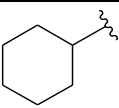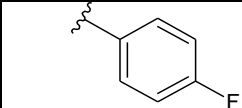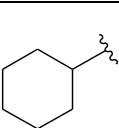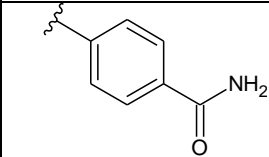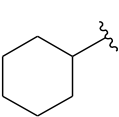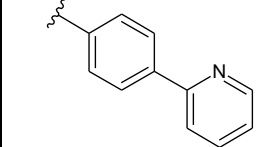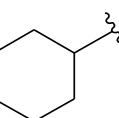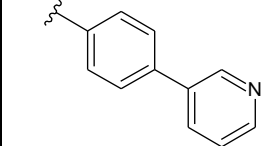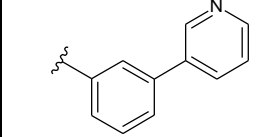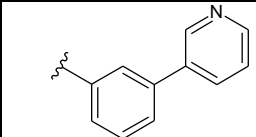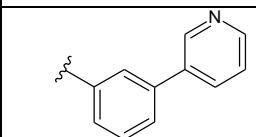
Compounds 4-63

Compound	X	L	R^1	R^2	pIC_{50}
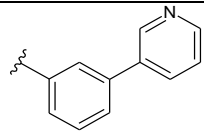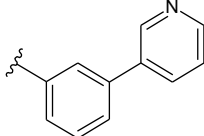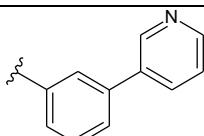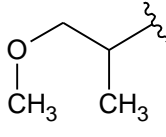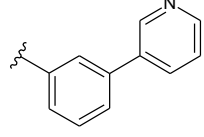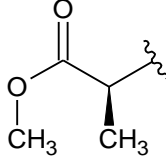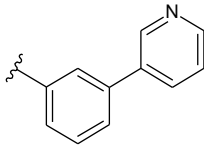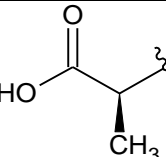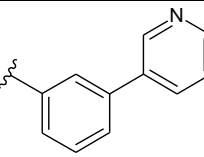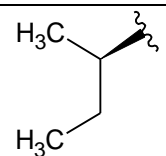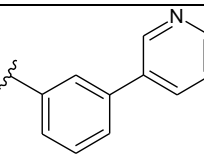
01		7.097
02		5.824
03*		6.851

04*	N-Me	NH			7.086
05	NH	N-Me			4.959
06	N-Me	N-Me			4.824

07	O	NH			6.728
08*	NH	O			5.900
09	O	O			7.310
10	NH	NH			7.495
11	NH	NH			5.724
12	NH	NH			4.886
13*	NH	NH			5.587

12
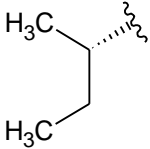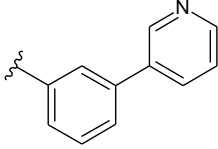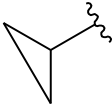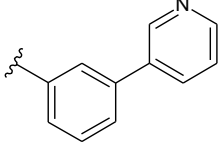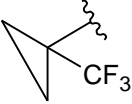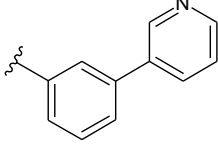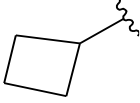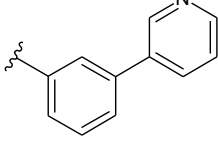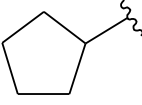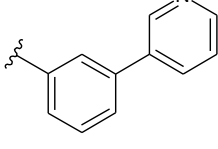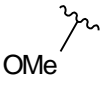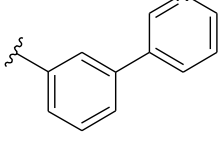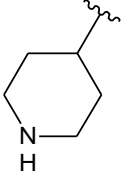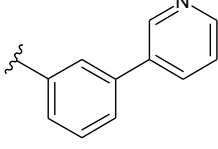
14	NH	NH			6.421
15*	NH	NH			4.721
16	NH	NH			6.010
17	NH	NH			4.921
18	NH	NH			5.767
19	NH	NH			4.611
20	NH	NH			4.876
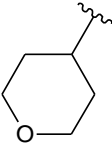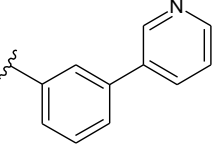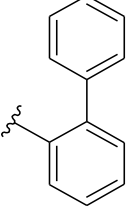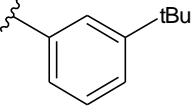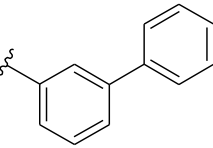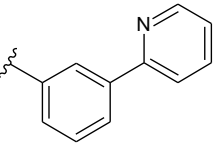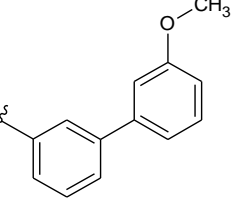
21*	NH	NH			6.474
22	NH	NH			7.018
23	NH	NH			7.796
24	NH	NH			7.699
25	NH	NH			6.142
26	NH	NH			5.487
27*	NH	NH			4.638
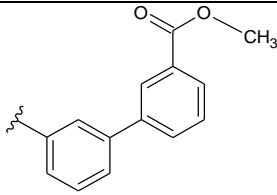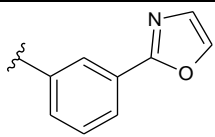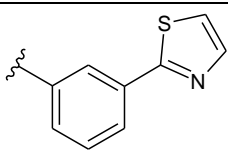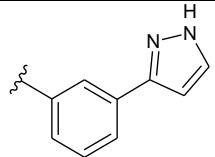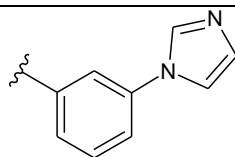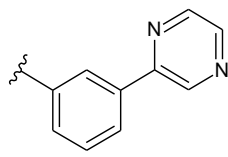
28	NH	NH			5.482
29	NH	NH			7.854
30*	NH	NH			5.726
31	NH	NH			4.509
32	NH	NH			6.620
33	NH	NH			7.323
34	NH	NH			8.222
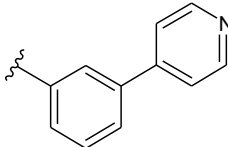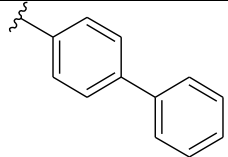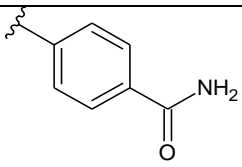
35*	NH	NH	Bu		6.910
36	NH	NH	iPr		8.222
37	NH	NH	tBu		7.886
38	NH	NH			8.222
39	NH	NH			4.658
40	NH	NH			4.638
41	NH	NH			8.097

42	NH	NH			8.222
43	NH	NH			8.222
44*	NH	NH			8.222
45	NH	NH			8.222
46*	NH	NH			7.886
47	NH	NH			4.469
48	NH	NH			8.097

17

49*	NH	NH			6.728
50	NH	NH	iPr		4.620
51	NH	NH	iPr		4.765
52*	NH	NH	iPr		4.936
53	NH	NH	iPr		5.541
54*	NH	NH	iPr		5.157

55*	NH	NH	iPr		5.353
56	NH	NH	iPr		6.037
57*	NH	NH	iPr		6.323
58*	NH	NH	iPr		7.377
59	NH	NH	iPr		8.222
60	NH	NH	iPr		7.959

61	NH	NH	iPr		6.547
62	NH	NH	iPr		4.824
63*	NH	NH	iPr		6.672

*Test set compounds

2.2 Molecular docking

The structure of FLT3 kinase co-crystallized with the inhibitor FF-10101 Succinate (FF-10101) has been reported (PDB ID 5X02). To validate the docking protocol, FF-10101 was taken out from the crystal structure and re-docked into the binding site of the protein. The root mean square deviation (RMSD) among the docked pose and crystal pose of FF-10101 was found to be 1.16 Å. The active site of FLT3 was referred from the previously published studies [10, 11, 20]. The 4 Å residues around the co-crystallized inhibitor were used as the binding site for docking studies. The dataset contains 8 most active compounds. All these compounds (Compounds **34**, **36**, **38**, **42**, **43**, **44**, **45** and **59**) were docked into the active site of FLT3 kinase to investigate the protein-ligand interactions using AutoDock 4.0 [37].

Protein preparation was carried out by the addition of Polar hydrogen atoms and Kollman charges. To avoid conformational explosion of the ligand, the number of rotatable bonds was limited to six. A 3D grid box was set utilizing the x, y, and z

coordinates of binding site and constructed with $70 \times 70 \times 70$ points. A grid spacing of 0.375 Å was used. Lamarckian algorithm was used for docking and the total number of Genetic Algorithm (GA) runs was set to 100. The docked conformation for one of the most active compounds (compound **36)** was selected depending on the lowest binding energy and interactions such as hydrogen bonding and hydrophobic interactions.

2.3 CoMFA and CoMSIA

3D-QSAR models were developed using Comparative Molecular Field Analysis (CoMFA), Region-Focused CoMFA (RF-CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) to correlate the 3D structure of the compounds with their biological activities using Sybyl-X 2.1. CoMFA uses the steric and electrostatic potential energies that are computed by Lennard-Jones and Coulombic potentials respectively [38]. The dataset compounds were aligned using a template molecule (most active compound **36**) [39].

The selection of an appropriate partial charge scheme is crucial to obtain satisfactory 3D-QSAR models [40, 41]. We have selected MMFF94 as partial charges scheme to generate 3D-QSAR models. Default parameters were utilized to develop CoMFA and RF-CoMFA models. A probe atom of sp^3 hybridized carbon with +1 charge was used along with a grid spacing of 2.0 Å. Statistically reasonable CoMFA and RF-CoMFA models were developed using the partial least squares (PLS) regression. In the PLS analysis, CoMFA descriptors and biological activity values (pIC_{50}) were used as independent variables and dependent variables respectively. PLS analysis with Leave-one-out (LOO) cross-validation was performed to assess the reliability of the generated models. PLS analysis was used to calculate an optimal number of components (ONC), the squared cross-validated correlation coefficient (q^2) and the standard deviation of prediction (SEP). A value of column filtering was kept as 2.0. Based on the obtained ONC, non-cross-validation

analysis was then performed to compute the F-test value (F), squared correlation coefficient (r^2) and standard error of estimate (SEE).

Similarly, RF-CoMFA model was developed using the PLS analysis obtained during CoMFA model development. RF-CoMFA is an iterative process that refines a built model by improving the weight for those lattice points that are most correlated to the model. This enhances the predictive capability of the PLS analysis used in RF-CoMFA [42].

Comparative Molecular Similarity Indices Analysis (CoMSIA) [43] utilizes descriptors such as electrostatic, steric, hydrophobic, hydrogen bond acceptor and hydrogen bond donor. A probe atom of radius 1.0 Å was used to calculate all the five CoMSIA similarity indices. An attenuation factor of 0.30 was used. A Gaussian function was used to calculate the CoMSIA model between the grid point and every atom of the molecule [43]. The similar lattice box employed as in CoMFA calculations was employed to derive the several CoMSIA models based on different descriptor combinations. Among the various models, a model with logical statistical standards in terms of q^2 and r^2 was chosen as the final CoMSIA model.

Among the developed 3D-QSAR models, RF-CoMFA revealed relatively better statistical results than CoMFA and CoMSIA. Hence, RF-CoMFA was selected as final model for further analyses. This model was interpreted graphically by contour map analyses. Based on the structural information obtained in the contour map analyses, we have derived a design strategy to design novel FLT3 inhibitors.

2.3.1. Model Validation

The selected RF-CoMFA model was checked for predictive ability using various validation techniques such as bootstrapping, leave-five-out (LOF), an external test set validation, concordance correlation coefficient (CCC) and rm^2 metric calculations [44]. Bootstrapping of 100 runs was carried out to validate the model's predictability [40]. The final model was validated by the predictive correlation coefficient (r^2_{pred}) as well.

3. Results and Discussion

3.1 Molecular Docking

The dataset taken for this study has 8 compounds (Compounds **34**, **36**, **38**, **42**, **43**, **44**, **45** and **59**) with the same pIC_{50} activity of 8.222 nM as shown in Table **1**. Therefore, all these compounds were considered as the most active and each compound was docked into the active site of FLT3 kinase. Initially, we selected 100 conformations for each docked compound. Among them, the reasonable binding pose was selected based on the docking score and the protein-ligand interactions for each compound. Binding energies of the selected docked poses for each compound are depicted in Table **2**. Final poses selected for each compound were compared. Compound **36** showed the lowest binding energy (-7.48 kcal/mol) and formed three H-bond interactions with the FLT3 kinase which were reported to be important for the FLT3 inhibition [13, 19, 20]. On the other hand, the rest of the docked compounds showed only two H-bond interactions and higher binding energies as compared to compound **36**.

Table 2. Binding energies of all the docked most active compounds.

Compound	Binding Energy (kcal mol^{-1})
34	-6.33
36	**-7.48**
38	-6.80
42	-7.18
43	-7.32
44	-7.03

| 45 | -7.08 |
| 49 | -6.71 |

Therefore, compound **36** was selected as the representative compound and discussed below in detail. The binding pose of the compound **36** in FLT3 kinase is shown in Fig. **(1a)**. Binding pocket of FLT3 kinase is contributed by residues Leu616, Gly617, Val624, Ala642, Tyr693, Cys694, Cys695, Tyr696, Gly697, Phe830 and Ala833. The compound **36** bound favorably in the FLT3 binding pocket by forming three hydrogen bond interactions (Fig. **(1a)**). Two hydrogen bond interactions were observed between the two amine groups around the pyrimidine ring of compound **36** with the hinge region residues Cys694 and Cys695. The H-bond with the hinge residue Cys694 is consistent with the previously reported docking studies [13, 20]. Additionally, a hydrogen bond interaction was observed between the nitrogen atom of the pyrimidine ring located between two amine groups with the residue Gly697. The H-bond interaction with the hinge residue Gly697 was also reported earlier [19].
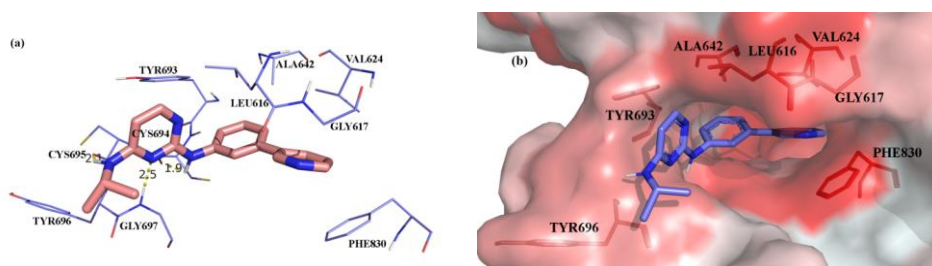
Fig. (1a). The docked conformation of the compound **36** (depicted in stick model) inside the active site of FLT3; Yellow dotted lines show hydrogen bonds formed between active site residues and compound **36**; Hydrogen bond distances are labelled in angstrom; (**b**) The most active compound **36** (depicted in stick model) inside the hydrophobic pocket of FLT3; The red colored region shows the most hydrophobic surface and white color stands for the least hydrophobic surface of the protein.

PyMOL was used to identify the hydrophobic interactions utilizing the python script 'color h' and to color the protein receptor according to the Eisenberg hydrophobicity scale [45] (Fig. **(1b)**). This colored the receptor from red which denotes the most hydrophobic to white for the least hydrophobic region. The compound **36** seems to form hydrophobic interactions with the active site residues which are located in the hydrophobic region (shown as lines in the Fig. **1b**). The phenylpyridine substitution at the R^2 position was docked into the deep hydrophobic pocket of FLT3 which was surrounded by residues Leu616, Gly617, Val624, Ala642 and Phe830. All these hydrophobic residues specifically, Phe830 which is from the conserved DFG motif, were reported to be important for the FLT3 inhibition [10, 19, 20]. Moreover, pyrimidine ring formed hydrophobic interactions with residues Tyr693 and Tyr696 which were also reported to be crucial [10, 19]. The selected pose of the compound **36** was used for further 3D-QSAR studies.

3.2 CoMFA, RF-CoMFA and CoMSIA Studies

The conformation of the compound **36** was selected as template for alignment of the rest of the dataset compounds using Sybyl-X 2.1. The alignment of the compounds was depicted in Fig. **(2)**. Full models for the CoMFA (q^2=0.506, NOC=6, r^2=0.822) and RF-CoMFA (q^2=0.566, NOC=6, r^2=0.865) were developed. The detailed statistical values of these models were shown in Table **3**. Several CoMSIA models were produced based on different combinations of the five descriptors as explained in the methodology section were shown in Table 4. The combination of steric, electrostatic and H-bond acceptor (SEA) parameters generated a statistically reasonable CoMSIA model (q^2=0.503, NOC=6, r^2=0.865). The detailed statistical values for this model were shown in Table **3**. RF-CoMFA model exhibited more reasonable statistical values as compared to the CoMFA and CoMSIA models. Hence, further analyses were carried out only for the RF-CoMFA model.
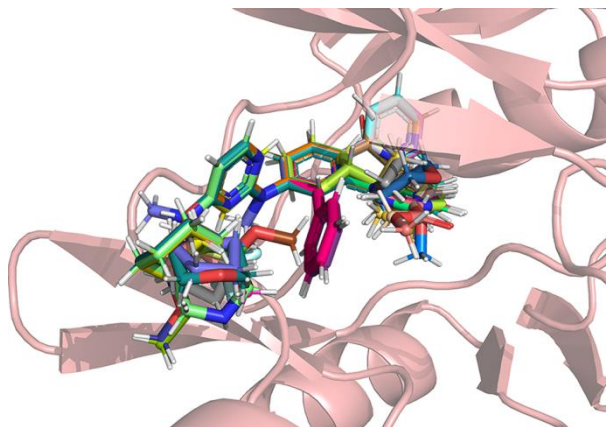
Fig. (2). Alignment of all the dataset compounds inside the active site of FLT3 kinase.

Then the RF-CoMFA model (q^2=0.629, NOC=6, r^2=0.956) for the external test set was produced. The generated RF-CoMFA model was validated through a number of validation techniques. The bootstrapping values of r^2 mean (BS-r^2) and BS-standard deviation (BS-SD) were 0.938 and 0.018 respectively which suggested the robustness of the model. The leave-five-out (LOF) of 0.644 and r^2_{pred} of 0.707 suggested that the model is predictive. The rm^2 metric calculation showed an average rm^2 of 0.655 and Delta rm^2 value of 0.197. These values exhibit the acceptable range suggested by Roy *et al*. [46]. The CCC was found to be 0.874 which indicates that the model is

Table 3. The developed 3D-QSAR models for complete dataset compounds.

Parameters	CoMFA	RF-CoMFA	CoMSIA (SEA)
q^2	0.506	0.566	0.503
ONC	6	6	6
SEP	0.955	0.895	0.957
r^2	0.882	0.865	0.865
SEE	0.466	0.499	0.499
F-value	69.811	59.763	59.792
Influence of different fields (%)			

S	52.3	57.3	22.0
E	47.7	42.7	38.8
A	-	-	39.2

q^2: cross-validated correlation coefficient; ONC: Optimal number of components; SEP: Standard Error of Prediction; r^2: non-validated correlation coefficient; SEE: Standard Error of Estimation; F value: F-test value; S: Steric; E: Electrostatic; A: H-bond acceptor.

statistically reliable [47]. The detailed statistical values of RF-CoMFA model and its validation test results are tabulated in **Table 5**. Scatter plot and contour maps of RF-CoMFA model are shown in Fig. **(3)** and Fig. **(4)** respectively. Comparison of experimental Vs predicted activity values of the model are shown in **Table 6**.

Table 4. Detailed statistical values obtained for different combination of CoMSIA descriptors.

CoMSIA	q^2	ONC	SEP	r^2	SEE	F value	Percentage contribution				
							S	E	H	A	D
S	0.243	3	1.151	0.447	0.984	15.866	100	-	-	-	-
E	0.346	5	1.088	0.680	0.761	24.246	-	100	-	-	-
H	0.191	1	1.170	0.372	1.031	36.062	-	-	100	-	-
A	0.324	6	1.116	0.604	0.854	14.255	-	-	-	100	-
D	0.018	2	1.323	0.182	1.197	4.362	-	-	-	-	100
SE	0.453	5	0.995	0.765	0.652	37.125	37.1	62.9	-	-	-
EH	0.344	2	1.063	0.579	0.851	41.199	-	47.2	52.8	-	-
EA	0.484	4	0.985	0.771	0.638	48.823	-	51	-	49	-
ED	0.388	6	1.062	0.778	0.640	32.612	-	78	-	-	22
SH	0.283	2	1.111	0.503	0.924	30.419	31.1	-	68.9	-	-

SA	0.357	3	1.060	0.669	0.761	39.694	30.7	-	-	69.3	-
SD	0.279	4	1.133	0.473	0.960	17.676	60.4	-	-	-	39.6
HA	0.256	2	1.310	0.621	0.815	32.208	-	-	48.2	51.8	-
HD	0.170	2	1.222	0.610	0.826	30.791	-	-	77	-	23
SEH	0.424	3	1.004	0.759	0.650	41.856	19.6	38.7	41.6	-	-
SEA	**0.503**	**6**	**0.957**	**0.865**	**0.499**	**59.792**	**22.0**	**38.8**	**-**	**39.2**	**-**
SED	0.470	6	0.989	0.848	0.529	52.156	28.8	56.4	-	-	14.8
EHA	0.362	3	1.070	0.725	0.694	51.761	-	35.0	34.7	30.3	-
EHD	0.367	5	1.071	0.840	0.538	60.021	-	43.4	40.3	-	16.4
SHA	0.301	3	1.106	0.675	0.754	40.817	21.8	-	37.6	40.7	-
SHD	0.278	3	1.124	0.660	0.771	38.202	31.7	-	49.5	-	18.7
EAD	0.414	6	1.039	0.830	0.560	45.509	-	46.3	-	39.3	14.5
HAD	0.246	3	1.166	0.669	0.762	39.666	-	-	38.0	46.0	16.0
SEHD	0.438	5	1.009	0.852	0.518	65.512	16.8	36.7	33.7	-	12.8
SEHA	0.409	3	1.017	0.752	0.658	59.714	13.6	30.3	29.6	26.5	-
SEAD	0.466	6	0.993	0.871	0.488	62.838	18.8	36.4	-	33.2	11.6
EHAD	0.361	2	1.049	0.755	0.655	60.578	-	33.2	27.7	26.4	12.7
SHAD	0.301	3	1.106	0.708	0.714	47.752	18.6	-	31.4	36.3	13.6
SEHAD	0.408	6	1.044	0.892	0.466	77.242	13.3	29.3	25.5	21.2	10.7

Final chosen model for CoMSIA analysis is indicated in bold font.

ONC: optimal number of components; q^2: cross-validated correlation coefficient; SEP: standard error of prediction; r^2: non-cross-validated correlation coefficient; SEE: standard error of estimate; F-value: F-test value; S: steric; E: electrostatic; H: hydrophobic; A: acceptor; D: donor.

Table 5. Statistical summary of the selected RF-CoMFA model.

Parameters	RF-CoMFA (TS-18)
q^2	0.629
ONC	6
SEP	0.903
r^2	0.956
SEE	0.429
F value	69.382
r^2_{pred}	0.707
LOF	0.644
BS-r^2	0.938
BS-SD	0.018
rm^2	0.655
Delta rm^2	0.197
CCC	0.874
S	57.8
E	42.2

q^2: cross-validated correlation coefficient; ONC: Optimal number of components; SEP: Standard Error of Prediction; r^2: non-validated correlation coefficient; SEE: Standard Error of Estimation; F value: F-test value; r^2_{pred}: Predictive r^2; LFO: leave-five-out; BS-r^2: Bootstrapping r^2 mean; BS-SD: Bootstrapping Standard deviation; Q^2: Progressive scrambling; Average rm^2 for the dataset; Delta rm^2 for the dataset; CCC: Concordance correlation coefficient; S:Steric; E:Electrostatic.
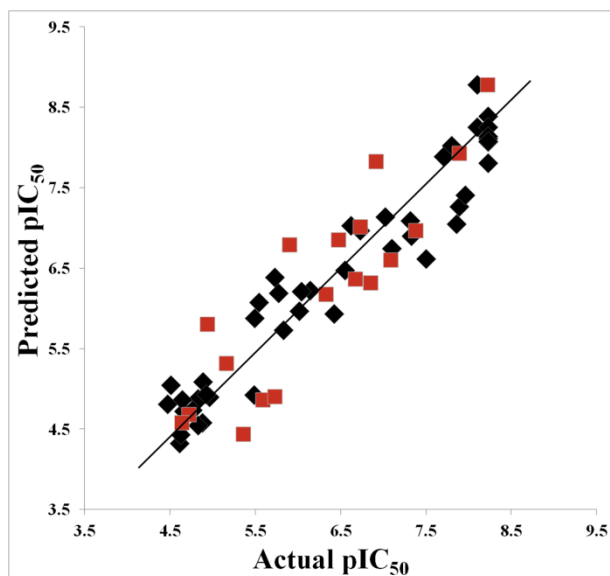
Fig. (3). Scatter plot for the selected RF-COMFA model; the plot depicts the actual pIC$_{50}$ versus predicted pIC$_{50}$ activity of the dataset for training and test sets; the training set compounds are shown as diamonds in black color; the test set compounds are shown as squares in red color.

Table 6. Experimental and predicted pIC$_{50}$ values with their residuals of selected RF-CoMFA model.

Compound	Actual pIC$_{50}$	RF-CoMFA (TS-18)	
		Predicted pIC$_{50}$	Residual
1	7.097	6.751	0.346
2	5.824	5.735	0.089
3	6.851	6.318	0.533
4	7.086	6.601	0.485
5	4.959	4.905	0.054
6	4.824	4.884	-0.060
7	6.728	6.973	-0.245
8	5.900	6.794	-0.894
9	7.310	7.095	0.215
10	7.495	6.622	0.873
11	5.724	6.388	-0.665
12	4.886	5.090	-0.204
13	5.587	4.863	0.724

14	6.421	5.937	0.484
15	4.721	4.684	0.037
16	6.010	5.969	0.041
17	4.921	4.937	-0.016
18	5.767	6.194	-0.427
19	4.611	4.325	0.286
20	4.876	4.586	0.290
21	6.474	6.854	-0.380
22	7.018	7.141	-0.123
23	7.796	8.030	-0.234
24	7.699	7.891	-0.192
25	6.142	6.227	-0.086
26	5.487	5.881	-0.394
27	4.638	4.579	0.059
28	5.482	4.931	0.551
29	7.854	7.057	0.797
30	5.726	4.909	0.817
31	4.509	5.049	-0.540
32	6.620	7.035	-0.415
33	7.323	6.902	0.421
34	8.222	7.815	0.407
35	6.910	7.828	-0.918
36	8.222	8.393	-0.171
37	7.886	7.272	0.614
38	8.222	8.076	0.146
39	4.658	4.728	-0.070
40	4.638	4.870	-0.232
41	8.097	8.261	-0.164
42	8.222	8.261	-0.039
43	8.222	8.118	0.104
44	8.222	8.779	-0.557
45	8.222	8.146	0.076
46	7.886	7.933	-0.047
47	4.469	4.812	-0.344
48	8.097	8.786	-0.689
49	6.728	7.018	-0.290
50	4.620	4.435	0.185
51	4.765	4.738	0.026
52	4.936	5.805	-0.869
53	5.541	6.078	-0.537
54	5.157	5.321	-0.164

55	5.353	4.438	0.915
56	6.037	6.215	-0.178
57	6.323	6.178	0.145
58	7.377	6.970	0.407
59	8.222	8.082	0.140
60	7.959	7.416	0.543
61	6.547	6.481	0.066
62	4.824	4.548	0.276
63	6.672	6.364	0.308

3.3 Contour Map Analysis

RF-CoMFA contour maps

The contour maps of the chosen RF-CoMFA model were represented in Fig. (**4**). The compound **36** is superimposed with RF-CoMFA contour maps within the active site of FLT3 kinase. The green and yellow colored contours in the steric contour map represented the regions for favorable and unfavorable bulky substitutions, respectively (Fig. (**4a**)). The big green colored contour observed at the R^1 position of pyrimidine ring shows that the bulky substitutions are favorable to enhance the activity. This is exemplified by the high activity values of the compounds **29, 37, 41, 46, 48, 60** and all the most active compounds including compound **36** that possess bulky group at this position. Similarly, a green colored contour observed close to the pyridine ring of R^2 substitution suggests that a bulky substitution is favorable at this position. On the contrary, two small yellow colored contours noticed near the phenylpyridine substitution at R^2 position proposed that the steric substitution at this position is not favored and could cause decrease in the inhibitory activity. This is probably explained by the low activity values of compounds **12, 19, 20, 50, 54 and 55** which exhibits bulky groups at this place.

32
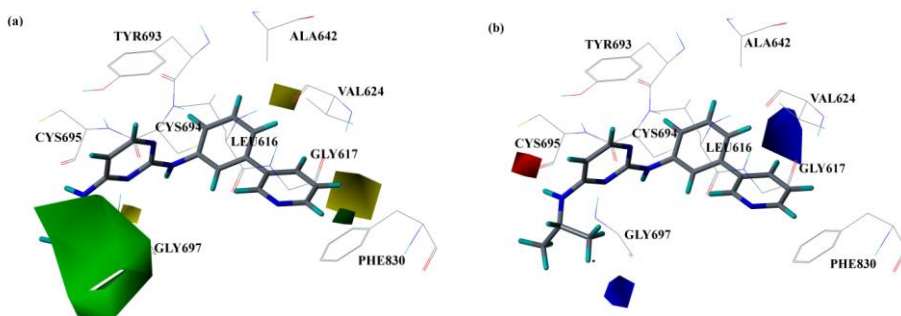
Fig. (4). Contour maps for the selected RF-CoMFA model. (**a**) RF-CoMFA steric contour map; (**b**) RF-CoMFA electrostatic contour map; Green contour shows the regions favorable for bulky substitutions and yellow contours shows the regions unfavorable for bulky substitutions; Blue contour favors electropositive substitutions while red contour favors electronegative substitutions.

RF-CoMFA electrostatic contour map is shown in Fig. (**4b**). Here, blue contours signify regions where electropositive substitutions are favorable to enhance the inhibitory activity, while red electronegative. A blue contour region at the R^1 position near isopropyl moiety and at the R^2 position near pyridine ring suggests that the positive substitutions at these areas might help in increasing the activity. On the other hand, red contour close to the amine group at R^1 site designates that negative substitutions could increase the potency. This might be justified by our docking results in which nitrogen atom in the pyrimidine ring formed hydrogen bond interaction with Gly697. This can be observed by the medium to high activities of compounds **10**, **24**, **29, 37, 41, 46, 48, 60** and all the most active compounds including compound **36** that possess electronegative group at this place.

3.4 Designing of new potent compounds
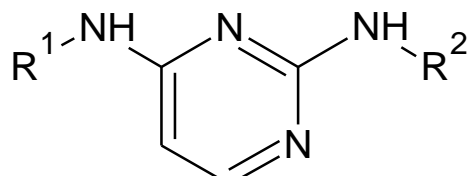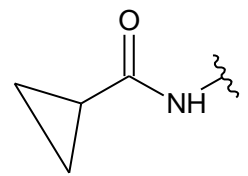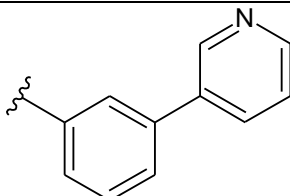
A design strategy was developed (Fig. **5**) based on the structural information acquired from the contour maps of the RF-CoMFA model. We have designed 14 novel compounds with enhanced FLT3 inhibitory activity using the design strategy. Sybyl X 2.1 was used to sketch and optimize all the designed compounds using
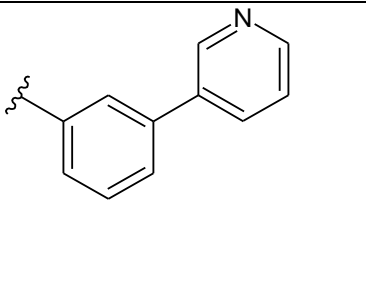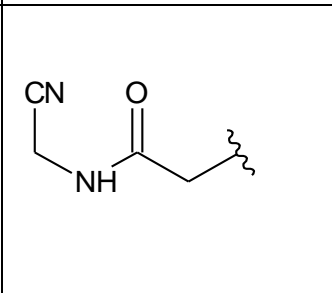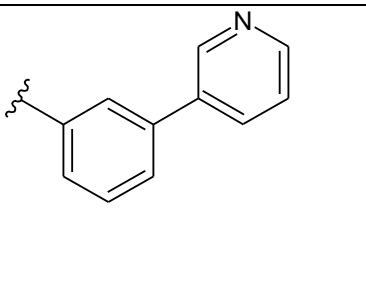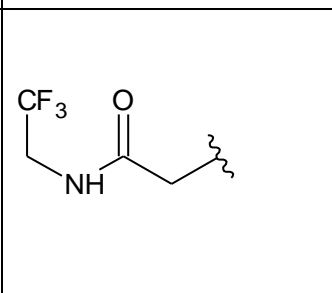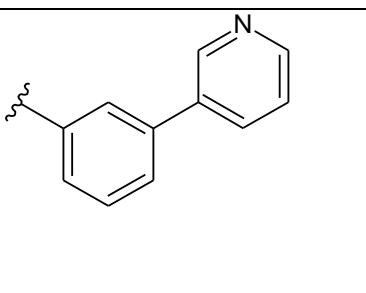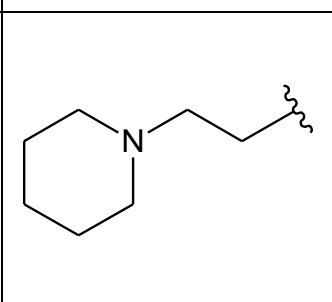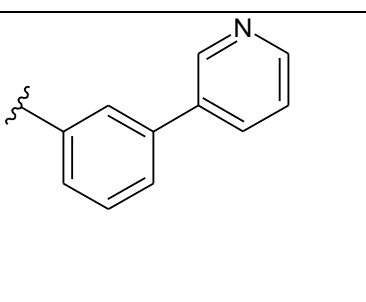
tripos force field. The final RF-CoMFA model was used to predict the activity of all the newly designed compounds. All the 14 compounds exhibited better activity than the most active compounds of the dataset taken for this study. The structure and the predicted pIC_{50} activities of all the designed compounds are depicted in Table 7.
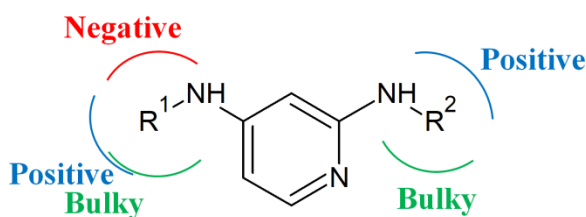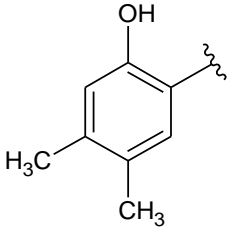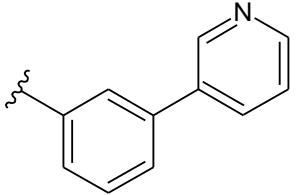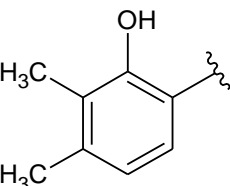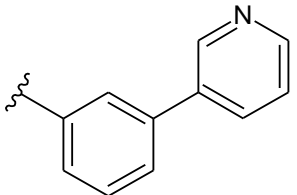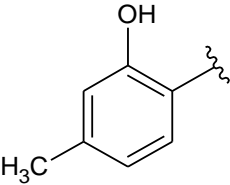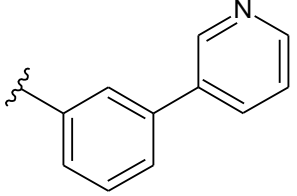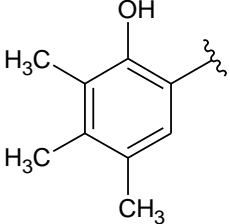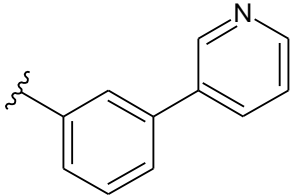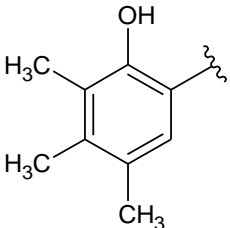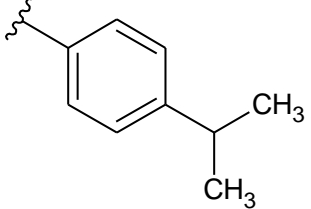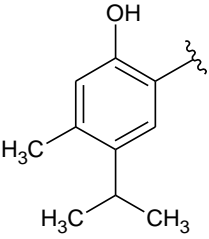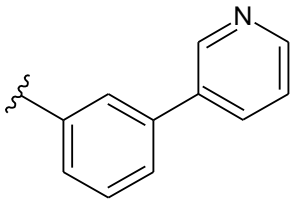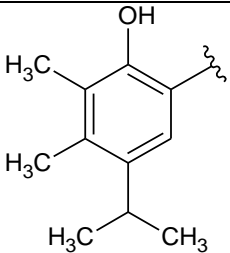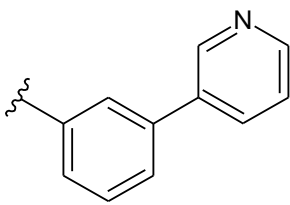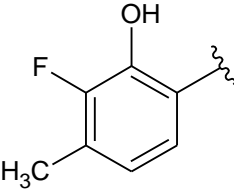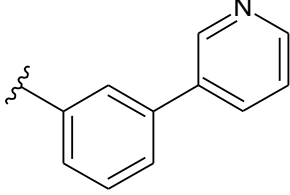
Fig. (5). Design strategy to design a series of diaminopyrimidine derivatives as FLT3 kinase inhibitors.
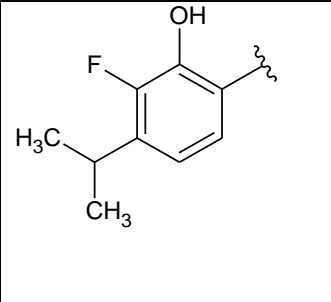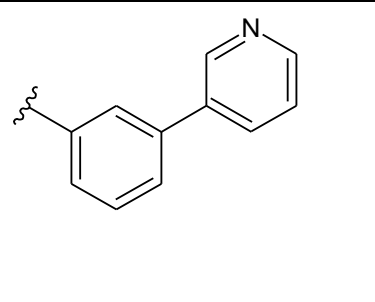
Table 7. Chemical structures and predicted pIC_{50} values of newly designed FLT3 kinase inhibitors.

Compound	R^1	R^2	Predicted pIC_{50}
D01			9.369

D02	CONO$_2$		8.862
D03			8.330
D04			8.356
D05			8.418

D06			8.360
D07			8.353
D08			8.299
D09			8.294

D10			13.654
D11			8.324
D12			8.884
D13			8.443

D14			8.758

4. Conclusion

In this work, molecular docking and 3D-QSAR studies were performed on diaminopyrimidine derivatives as FLT3 kinase antagonists. The docking analyses recognized important active site residues of FLT3 kinase which interacts with the most active compound (compound **36**). Active site residues Leu616, Gly617, Val624, Ala642, Phe830, Tyr693, Cys694, Cys695, Tyr696, and Gly697 seemed to be important in the inhibition of FLT3 kinase. All the dataset compounds were aligned inside the FLT3 active site using compound **36** as a template molecule. Subsequent alignment was used to derive CoMFA, RF-CoMFA and CoMSIA models. RF-CoMFA model yielded relatively better statistical results than CoMFA and CoMSIA so it was selected for further analyses. Validation of this model was performed by means of bootstrapping, external test set, leave-five-out, rm^2 metric calculations, concordance correlation coefficient, and r^2_{pred}. Statistically acceptable RF-CoMFA (q^2=0.629, NOC=6, r^2=0.956) model was developed. Analyses of the RF-CoMFA contour maps suggested that the bulky and electronegative substitutions are favored at R^1 position whereas bulky and electropositive substitutions are favored at R^2 position to increase the inhibitory potency of FLT3 kinase. The contour map analyses were consistent with our docking results. Using the structural information obtained from the contour maps and the protein-ligand interactions observed in

38

docking analyses, we have derived a design strategy to design more potent FLT3 inhibitors. A series of 14 new diaminopyrimidine derivatives were designed and their activities were predicted. All the designed compounds showed more inhibitory activity than the most active compounds using RF-CoMFA. Our designed compounds could be helpful to the medicinal chemists to develop more potent and selective FLT3 kinase antagonists.

Chapter 3

Macrocyclic effect on inhibitory activity of MerTK inhibitors

1. Introduction

Naturally existing macrocyclic compounds possess the ring sizes which span from 11 to 16 membered rings, most often 14 membered [48]. Structures bearing macrocyclic ring have been reported to have a favorable impact on important properties requisite for drugs for example membrane permeability [49], metabolic stability, increased binding affinity and selectivity [50], and overall pharmacokinetics [51, 52]. They are reported to display remarkably more biological activities and have balance between flexibility and conformational preorganization to attain the best possible binding properties with respect to their biological targets [53-55]. Some of the naturally occurring macrocyclic compounds are erythromycin (antibiotic), epothilone B (anticancer), tacrolimus (immunosuppressant) and bryostatins (protein kinase C inhibitor) [53-56].

Artificial compounds with large ring structures were synthesized and studied for inhibitory activity against a variety of targets. There have been different studies about the effect of ring size on the potency of inhibitors. Marsault et al. [53] reported that macrocyclic inhibitors of renin exhibited rise in inhibitory activity with the rise in ring size (from 10 to 14-membered rings). On the contrary, histone deacetylase inhibitors displayed elevated potency with a 14-membered ring over a 16-membered ring. In the similar study, the activity was lower when the ring size was larger [53]. De Clercq and co-workers [56] performed structure-activity relationship for phenylenebis(methylene) linked bistetraazamacrocycles against HIV (HIV-1 and HIV-2 replication). They noticed that increasing the ring size from 10 to 14 resulted in higher anti-HIV activity. But, additional rise in the ring size resulted in a significant reduction in the potency [56-58]. Thus, the ring size effect is very intriguing, but the nature of its interaction has not been well examined.

Receptor tyrosine kinases (RTKs) are important transmembrane receptors which regulate crucial signal transduction such as migration, apoptosis, cell proliferation and invasion of numerous cancers [59, 60]. Mer receptor tyrosine

kinase (MerTK) is a member of the TAM-family of receptor tyrosine kinases which is comprised of Tyro3, Axl and MerTK [61, 62]. TAM receptors are type I transmembrane glycoproteins, composed of an extracellular domain enclosing tandem immunoglobulin related domains, tandem fibronectin type III repeats and a protein kinase c-like intracellular kinase domain [62-64]. Members of the TAM family comprise extremely conserved kinase domain (>70% identity), and encompasses a distinctive KW(I/L)A(I/L)ES active site which differs from other RTKs [63, 65, 66].

MerTK inhibition presents exclusive clinical challenge since drugs that target MerTK would also inhibit the usual proteins in noncancerous cells. There are few compounds which have been reported to possess potent antagonism against MerTK such as UNC569 derivatives [67], ONO-9330547 [68, 69], CVO-102, R428 and BMS-777607 [51]. UNC569 derivatives and ONO-9330547 increased the sensitivity to chemotherapy and reduced colony formation of acute lymphoblastic leukemia (ALL) cells and a pediatric tumor cell line (BT12) [70]. CVO-102 has been revealed to have significant activity in preventing blood clotting and also shows capable anticancer agent in late preclinical development [71]. R428 and BMS-777607 exhibit activity against Mer but are approximately 4-fold to 15-fold more selective for Axl [71, 72].

To study MerTK and inhibitor interactions, docking, molecular dynamics and receptor based 3D-QSAR (3D-Quantitative Structure-Activity Relationship) studies were reported with a range of pyrimidine derivatives such as pyrazolo-[3,4-d]-pyrimidines [73], pyridine-substituted pyrimidines [74] pyridinepyrimidines [75] and pyrimidine-based derivatives [76]. Yu et al. and Liang et al. reported that hinge region residues from the active site, Pro672 and Met674 are extremely crucial for MerTK inhibition [74, 77]. Experimental studies showed that these two amino acids are important [78, 79]. X-ray structures (PDB ID: 5K0X & 5K0K) show that MerTK hinge region residues (Pro672 and Met674) bind to the co-crystallized inhibitors [78, 79]. In another study, aforementioned amino acids were also reported crucial by

docking analyses. In addition, 107 MerTK inhibitors were utilized to develop a ligand-based pharmacophore model. 6 compounds with new scaffolds were recognized as MerTK inhibitors by virtual screening using this model [80].

In our study, we have chosen a series of macrocyclic MerTK inhibitors containing different ring sizes (from 14 to 20-membered rings) to study ring size effect and factors affecting MerTK inhibitory activity. To this end, we employed a variety of simulation techniques such as docking, molecular dynamics and 3D-QSAR methods.

2. Materials and Methods

2.1 Overall procedure

Simple alignment schemes such as atom by atom matching could not be applied due to the complex nature of the macrocyclic ring structure. In recent studies, Hawkins [57] and Watts et al. [58] reported that sampling of low energy conformers of macrocycles is difficult due to the large ring size and flexibility of these molecules. The existing techniques such as stochastic methods based on Monte Carlo-simulated annealing, distance geometry are not convenient for conformational sampling of macrocyclic molecules [57, 58].

Molecular dynamics (MD) should be one of the reasonable simulation methods to deal with macrocyclic ring structure [51]. So, we have employed molecular dynamics for pose selection of macrocyclic compounds inside the receptor to provide reliable molecular alignments. Fortunately, there exists information about seemingly important amino acid residues such as Pro672 and Met674 [75, 78]. Hydrogen bonding interactions with these two hinge region residues mimic the interaction of ATP with MerTK, which have been considered crucial for the inhibitor recognition.

Using these hydrogen bonding interactions at the hinge region as the anchoring points, the inhibitors could be placed inside the receptor in a reasonable

way. Then the structures were optimized by molecular dynamics followed by energy minimization and structure sampling. The resultant structures of ligands were used as input structures for subsequent receptor-based 3D-QSAR studies. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were used to understand the structural requirements of MerTK inhibitors.

2.2 Molecular Docking

Prior to the docking process, the protein structure was prepared as follows: polar hydrogen was added, Kollman and Gasteiger charges and AD4 atom-types were assigned. Autodock 4.0 [37] was used to perform docking. Docking protocol was validated by re-docking of extracted co-crystallized ligand from the crystal structure (PDB code: 5K0K) into the active site of the protein. Active site was defined by the same residues as used in the previous studies [75, 76]. The most potent compound from the dataset was docked into the active site of MerTK. Active site grid was created using the x, y and z coordinates of the active site. Dimensions of this grid were extended to $80 \times 80 \times 80$ points with a grid spacing of 0.375 Å. The number of Genetic algorithm (GA) run was set to 100 and docking was executed using the Lamarckian genetic algorithm (LGA). The resultant docked conformations were clustered into groups with default RMS deviations of 0.5 Å.

A docked pose based on the docking score and the two key hydrogen bonding interactions at the hinge region (Pro672, Met674) was selected as the representative binding pose. Since the dataset comprises macrocyclic compounds with various ring sizes (from 14 to 20-membered rings), we docked all the compounds. Since Autodock does not allow flexibility of macrocyclic ring structures, we have performed molecular dynamics simulations for structural optimization and relaxation using SYBYL-X 2.1 program.

2.3 Structure optimization and sampling
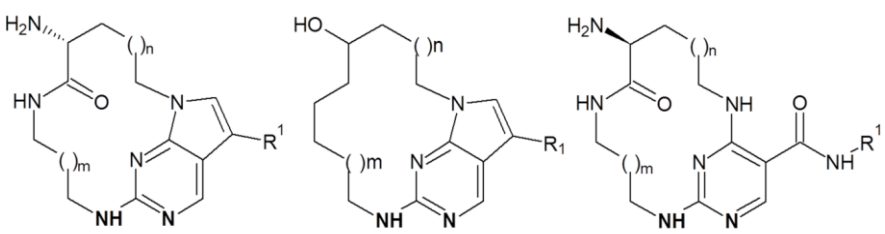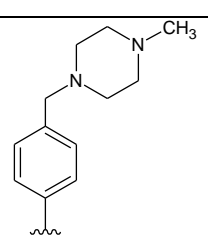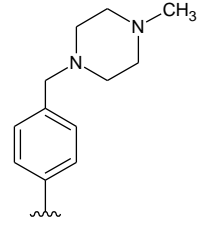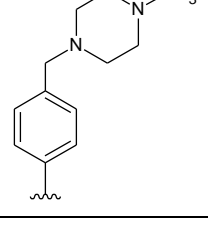
SYBYL-X 2.1 was used for MD to generate conformers of these macrocyclic ring structures. As a first step, a dynamics run for 100 ps was performed using NVT ensemble at 300 K and with a time step of 1 fs. Trajectories were recorded every 5 ps for the MD runs. Boltzmann random velocity was selected as the initial velocity. The rest of the parameters were chosen as default. Tripos force field was used for the minimization of all the compounds. Gasteiger-Marsili charges were applied to the protein-ligand complex. The important residues (Pro672, Met674) forming key hydrogen bond interactions were constrained. From the several conformers generated, one conformation was selected considering two criteria. i.e., energy and the presence of two key hydrogen bond interactions as mentioned above. Final conformers selected for the compounds were used for the 3D-QSAR study.
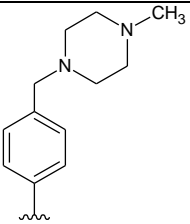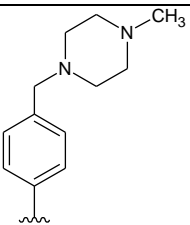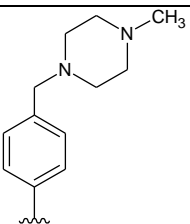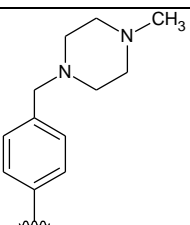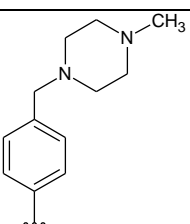
2.4 Test set / training set selection for 3D-QSAR analyses
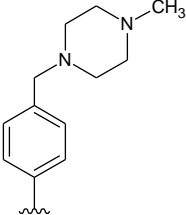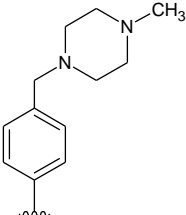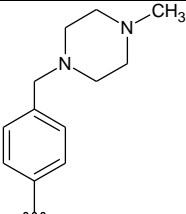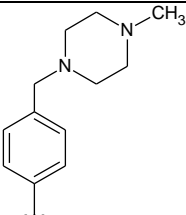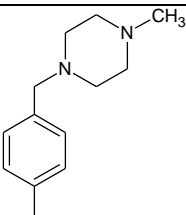
A series of 68 macrocyclic pyrimidines reported as potent MerTK inhibitors with their biological activities (IC_{50} values) were collected for this study [78, 79]. These inhibitors exhibit various ring sizes (from 14 to 20-membered rings). SYBYL-X 2.1 was used to sketch all the 68 structures and optimized by energy minimization with Tripos force field [35]. Biological data expressed as IC_{50} values were converted into pIC_{50} (-log IC_{50}) values. The pIC_{50} values cover 4.24 log units which signify the dataset is acceptable for a QSAR study [34]. The dataset was divided into a training set of **46** compounds for model derivation and a test set of **22** compounds for external predictability. The test sets were selected by considering different macrocyclic ring sizes and biological activity. The compounds were classified into high, medium, and low active ones depending on the biological activities. Subsequently, **22** compounds were randomly taken from the three activity classes and chosen as the test set. Docked conformation of the most active compound was taken as a template to sketch the rest of the compounds in the dataset. All the ligands taken for the docking studies were energy minimized to find the most reasonable

45

poses within the receptor. The chemical structures and biological activity values of all the compounds are depicted in Table **1**.
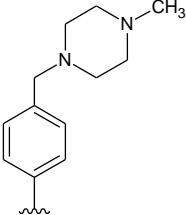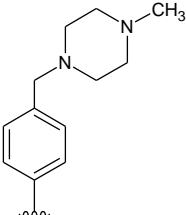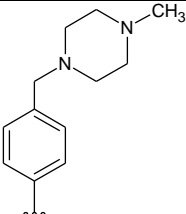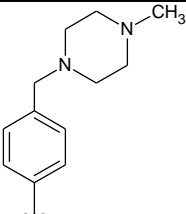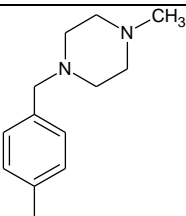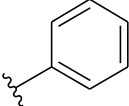
Table 1. Structure and biological activity values of macrocyclic pyrimidines as MerTK inhibitors

Group-I
Compounds 1-6

Group-II
Compounds 7-32

Group-III
Compounds 33-56
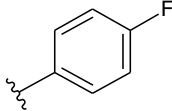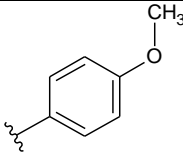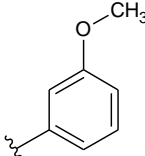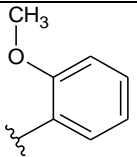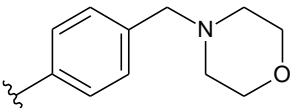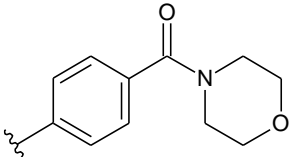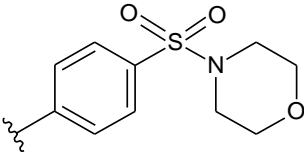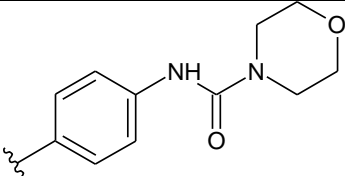
Compound	R^1	m	n	IC$_{50}$ (nm)	pIC$_{50}$ (nm)
1*		2	1	3.9	8.409
2		4	1	17	7.770
3*		6	1	30	7.523
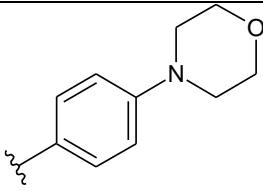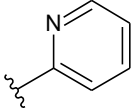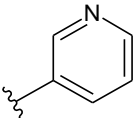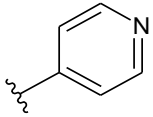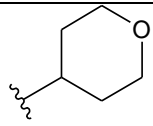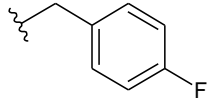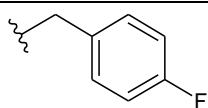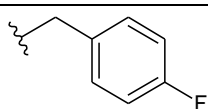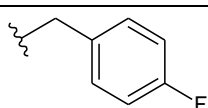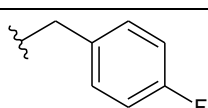
4*		2	2	8.9	8.051
5		4	2	2.2	8.658
6		6	2	0.8	9.097
7*		1	1	4.1	8.387
8		2	1	5.6	8.252
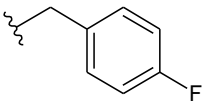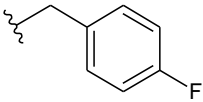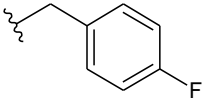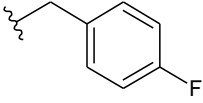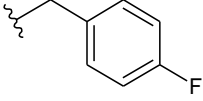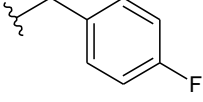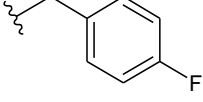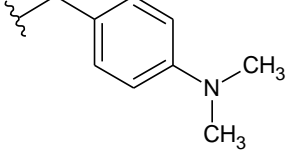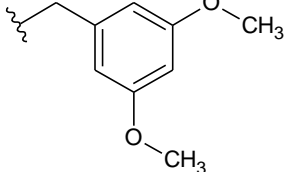
9		3	1	1.8	8.745
10*		4	1	9.6	8.018
11		1	2	8.0	8.097
12		2	2	4.0	8.398
13*		3	2	3.0	8.523
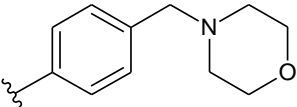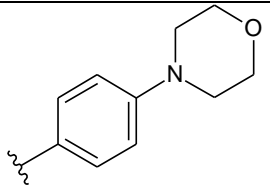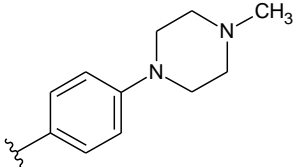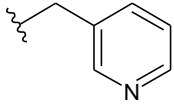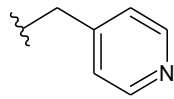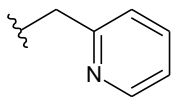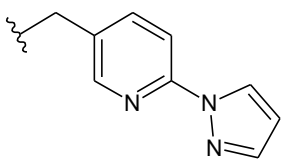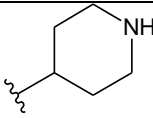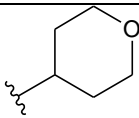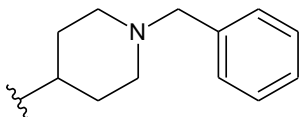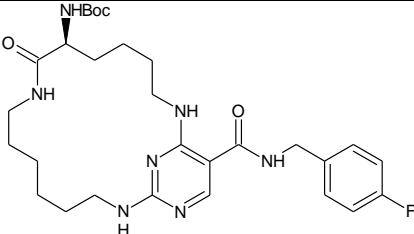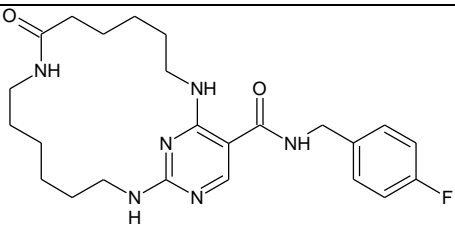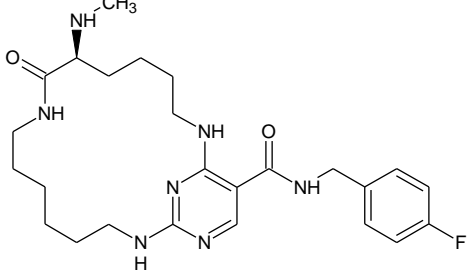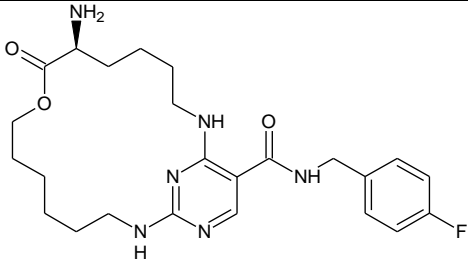
14*		4	2	2.6	8.585
15		1	3	4.1	8.387
16		2	3	5.0	8.301
17		3	3	3.1	8.509
18		4	3	4.9	8.310
19		3	2	140	6.854
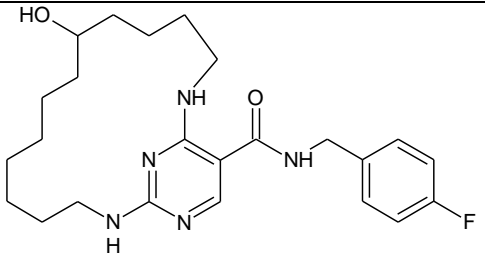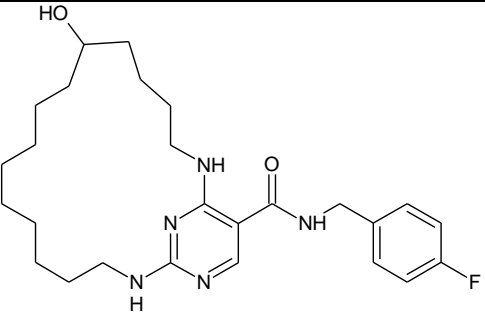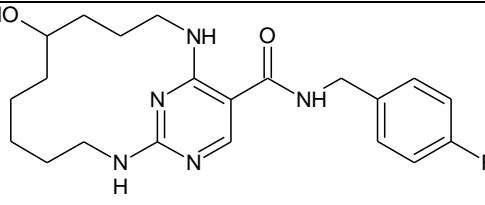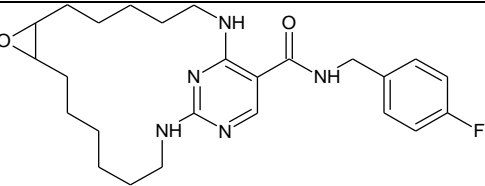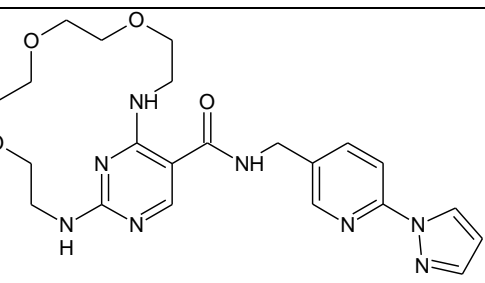
20		3	2	150	6.824
21		3	2	86	7.066
22		3	2	160	6.796
23*		3	2	170	6.770
24		3	2	16	7.796
25		3	2	8.1	8.092
26		3	2	11	7.959
27		3	2	30	7.523
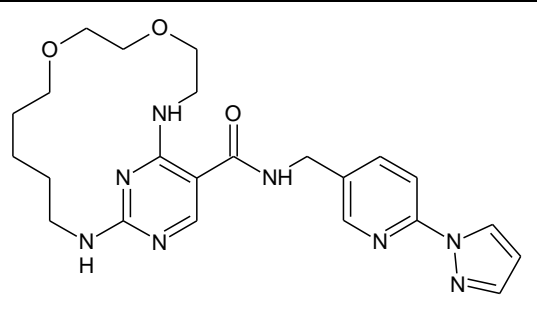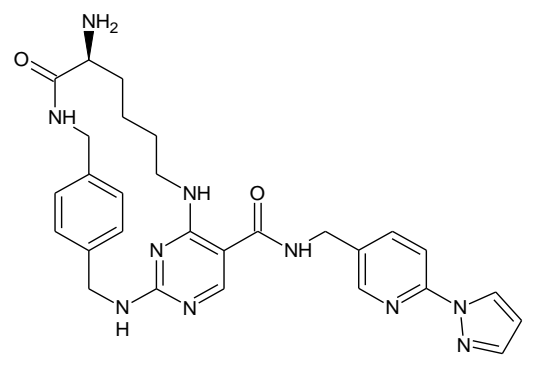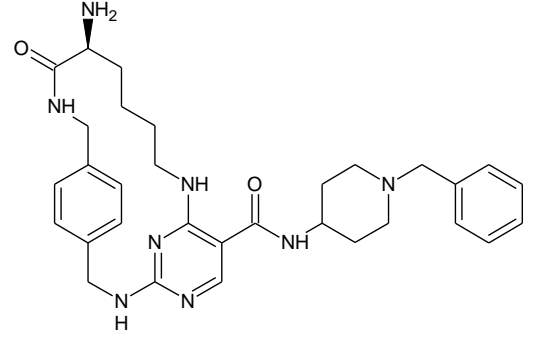
28		3	2	32	7.495
29*		3	2	54	7.268
30*		3	2	13	7.886
31		3	2	80	7.097
32		3	2	94	7.027
33		2	1	61	7.215
34		1	1	200	6.699
35*		3	1	760	6.119
36		4	1	230	6.638
37		5	1	91	7.041

	Structure				
38*	4-F-phenyl (CH_2CH_2)	6	1	62	7.208
39	4-F-phenyl (CH_2CH_2)	1	2	3200	5.495
40	4-F-phenyl (CH_2CH_2)	2	2	140	6.854
41*	4-F-phenyl (CH_2CH_2)	3	2	36	7.444
42	4-F-phenyl (CH_2CH_2)	4	2	4.4	8.357
43*	4-F-phenyl (CH_2CH_2)	5	2	4.1	8.387
44*	4-F-phenyl (CH_2CH_2)	6	2	6.2	8.208
45	4-N(CH_3)$_2$-phenyl (CH_2CH_2)	4	2	6.4	8.194
46*	3,5-di-(O$-CH_3$)-phenyl (CH_2CH_2)	4	2	25	7.602

47		4	2	6.0	8.222
48		4	2	9.5	8.022
49*		4	2	4.3	8.367
50		4	2	6.3	8.201
51*		4	2	4.4	8.357
52		4	2	22	7.658
53		4	2	1.5	8.824
54*		4	2	6.5	8.187
55		4	2	14	7.854

56		4	2	2.9	8.538
57*		-	-	54	7.268
58		-	-	59	7.229
59		-	-	11	7.959
60		-	-	3.9	8.409

61*		-	-	68	7.168
62		-	-	140	6.854
63*		-	-	21	7.678
64		-	-	480	6.319
65		-	-	770	6.114

66		-	-	14000	4.854
67		-	-	40	7.398
68		-	-	21	7.678

* Test set compounds

2.5 3D-QSAR (CoMFA and CoMSIA)

In CoMFA (Comparative Molecular Field Analysis), correlation of biological properties of compounds with steric and electrostatic potential energies are calculated using Lennard-Jones and Coulombic potentials, respectively [38]. The

dataset compounds should be aligned in their bioactive conformations to develop a reasonable model [39]. The dataset compounds were aligned using a common substructure-based alignment method with template molecule (most active compound **6**). The choice of appropriate partial charge scheme is important to obtain robust 3D-QSAR models. Previous studies reported the necessity of considering more appropriate atomic charges rather than default CoMFA charges [40, 41]. We derived 3D-QSAR models using different charge calculation schemes to study the influence of these charge schemes on the 3D-QSAR models. The partial atomic charges were applied for each ligand by using different charge schemes i.e. Gasteiger-Hückel, Hückel, Gasteiger-Marsili [81], Del-Re [82], Pullman [83], MMFF94 [36] and AM1-BCC [84].

Default parameters were used to generate the CoMFA model. A sp^3 hybridized carbon as probe atom with +1 charge and a grid spacing of 2.0 Å was used. The partial least squares (PLS) regression was used to develop statistically reasonable 3D-QSAR models. In the PLS analysis, CoMFA descriptors were used as independent variables and biological activity values (pIC_{50}) were used as dependent variables. PLS analysis with Leave-one-out (LOO) cross-validation was executed to evaluate the reliability of the models developed [85, 86]. The squared cross-validated correlation coefficient (q^2) value, an optimal number of components (ONC) and standard deviation of prediction (SEP) were calculated using PLS analysis. A column filtering value of 2.0 was used. A non-cross-validation analysis was then carried out using the obtained ONC by cross-validation to calculate the squared correlation coefficient (r^2), F-test value (F) and standard error of estimate (SEE).

In Comparative Molecular Similarity Indices Analysis (CoMSIA) [43], steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor descriptors were considered. All five CoMSIA similarity indices were calculated using a probe atom of radius 1.0 Å. An attenuation factor of 0.30 was used. The CoMSIA models were calculated between the grid point and each atom of the molecule by a Gaussian function [43]. The CoMSIA models with different

combinations were generated with the same lattice box used as in CoMFA calculations. Among the possible combinations, a model with reasonable statistical values in terms of q^2 and r^2 was selected as the final CoMSIA model. 3D-QSAR results were interpreted graphically by the field contribution maps. The developed models were checked for predictive ability and robustness using various validation techniques which includes bootstrapping, progressive scrambling, an external test set validation, rm^2 metric calculations and concordance correlation coefficient (CCC) [44]. Bootstrapping for 100 runs and the progressive scrambling of 100 runs with 2 to 10 bins were performed to validate the model's predictability [40]. The models were also validated by the predictive correlation coefficient (r^2_{pred}).

3. Results and Discussion

3.1 Activity vs ring size

The dataset comprises macrocyclic compounds with various ring sizes ranging from 14 to 20-members. Since previous reports mentioned about ring size effects over biological activity, we tried to see the ring size effects. As shown in Fig. 1, there seemed no correlation between ring size (x-axis) and biological activity (y-axis). To further investigate the factors affecting biological activity, we did docking study.

Figure 1. The scatter plot of macrocyclic ring size vs. pIC$_{50}$ activity values. Circle indicates compounds having 4 hydrogen bonds; square indicates compounds having 3 hydrogen bonds and triangle indicates compounds having 2 hydrogen bonds.

3.2 Molecular docking and pose optimization using molecular dynamics

In many cases of cyclic inhibitors, docking approaches have been particularly difficult. Most of the docking programs do not support flexibility of the cyclic inhibitors during docking. Hence, pose selection from docked structures was performed to provide flexibility to macrocyclic rings and to get more reliable binding poses for the 3D-QSAR study. Molecular dynamics (MD) was used for pose selection of macrocyclic compounds inside the receptor. The molecular dynamics for each docked compound was carried out for 100 ps with a time step of 1 fs with the above mentioned parameters using dynamics module in SybylX 2.1. The conformers were saved every 5 ps which generated a total of 20 conformers for each compound. The obtained conformers were plotted as scatter plot for each compound based on their energy. From the conformers generated for each compound, one conformation

was selected on the basis of the lowest energy and the presence of two signature hydrogen bond interactions with the important hinge region residues Pro672 and Met674. Final conformers selected were used for the alignment of dataset compounds in the 3D-QSAR study. The selected docked structures from MD simulations were used for subsequent 3D-QSAR studies.

Since diverse rings would bind differently to the same receptor, all the compounds were docked. Table 2 shows the docking analyses of all dataset compounds. First, we chose 100 poses for each docked compound. Among them, the best binding pose was selected on the basis of docking score and the key hydrogen bonding for each compound. In Fig. 2a, the binding pose of the most active compound in MerTK is shown. For this cyclic inhibitor, there are 4 H-bonds with residues at the active site.
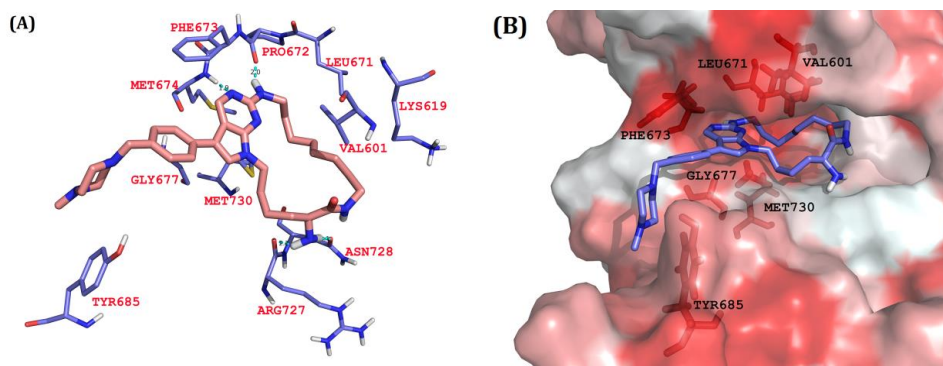
Figure. 2 a The docked conformation of the compound 06 (shown in stick model) inside the binding pocket of MerTK Cyan dotted lines represent hydrogen bonds formed between residues and compound 06. Hydrogen bond distances are labelled in angstrom **b** The most active compound (shown in stick model) inside the hydrophobic pocket of MerTK. The red colored region represents the most hydrophobic surface and white color represents the less hydrophobic surface of the protein.

The H of **NH** (bold **NH** in Fig. 3) which is adjacent to pyrimidine ring forms H-bond with the hinge region residue Pro672. Similarly, the N (bold **N** in Fig. 3) of pyrimidine ring forms H-bond with the hinge region residue Met674. X-ray crystal

structures conducted by McIver et al., [78] also have two hydrogen bonds at this hinge region. This is consistent with the previous reports that these hydrogen bonds are essential [75, 76]. These 2 key interactions were found in all the docked compounds having different ring sizes. In addition, hydrogen atoms of the amino group of the macrocyclic ring forms hydrogen bond interactions with the active site residues Arg727 and Asn728. Hydrogen bonding was assigned with PyMOL program with default parameters. From the docking results, the chiral compounds

Table 2. Hydrogen bond interactions between macrocyclic pyrimidine derivatives and MerTK (Docking analyses of the dataset compounds)

Compound	Ring Size	pIC$_{50}$	No. of H-bonds	Hydrogen bond residues			
				HB1	HB2	HB3	HB4
Cpd01	15	8.409	4	PRO672	MET674	ARG727	ASN728
Cpd02	17	7.770	4	PRO672	MET674	ARG727	ASN728
Cpd03	19	7.523	4	PRO672	MET674	ARG727	ASN728
Cpd04	16	8.051	4	PRO672	MET674	ARG727	ASN728
Cpd05	18	8.658	4	PRO672	MET674	ARG727	ASN728
Cpd06	20	9.097	4	PRO672	MET674	ARG727	ASN728
Cpd43	19	8.387	4	PRO672	MET674	ARG727	ASN728
Cpd13	17	8.523	4	PRO672	MET674	ASP741	ASN728
Cpd53	18	8.824	4	PRO672	MET674	ARG727	ASN728
Cpd42	18	8.357	4	PRO672	MET674	ARG727	ASN728
Cpd45	18	8.194	4	PRO672	MET674	ARG727	ASN728
Cpd07	14	8.387	3	PRO672	MET674	ASP741	-
Cpd09	16	8.745	3	PRO672	MET674	ASP741	-
Cpd41	17	7.444	3	PRO672	MET674	ASN728	-
Cpd37	18	7.041	3	PRO672	MET674	ARG727	-
Cpd18	19	8.310	3	PRO672	MET674	ASN728	-
Cpd40	16	6.854	3	PRO672	MET674	ASN728	-

Cpd60	18	8.409	3	PRO672	MET674	ASN728	-
Cpd44	20	8.208	3	PRO672	MET674	ASN728	-
Cpd16	17	8.301	3	PRO672	MET674	ASN728	-
Cpd34	14	6.699	3	PRO672	MET674	ASP741	-
Cpd15	16	8.387	3	PRO672	MET674	ASP741	-
Cpd12	16	8.398	2	PRO672	MET674	-	-
Cpd38	19	7.208	2	PRO672	MET674	-	-
Cpd08	15	8.252	2	PRO672	MET674	-	-
Cpd10	17	8.018	2	PRO672	MET674	-	-
Cpd11	15	8.097	2	PRO672	MET674	-	-
Cpd14	18	8.585	2	PRO672	MET674	-	-
Cpd17	18	8.509	2	PRO672	MET674	-	-
Cpd19	17	6.854	2	PRO672	MET674	-	-
Cpd20	17	6.824	2	PRO672	MET674	-	-
Cpd21	17	7.066	2	PRO672	MET674	-	-
Cpd22	17	6.796	2	PRO672	MET674	-	-
Cpd23	17	6.770	2	PRO672	MET674	-	-
Cpd24	17	7.796	2	PRO672	MET674	-	-
Cpd25	17	8.092	2	PRO672	MET674	-	-
Cpd26	17	7.959	2	PRO672	MET674	-	-
Cpd27	17	7.523	2	PRO672	MET674	-	-
Cpd28	17	7.495	2	PRO672	MET674	-	-
Cpd29	17	7.268	2	PRO672	MET674	-	-
Cpd30	17	7.886	2	PRO672	MET674	-	-
Cpd31	17	7.097	2	PRO672	MET674	-	-
Cpd32	17	7.027	2	PRO672	MET674	-	-
Cpd33	15	7.215	2	PRO672	MET674	-	-
Cpd35	16	6.119	2	PRO672	MET674	-	-
Cpd36	17	6.638	2	PRO672	MET674	-	-
Cpd39	15	5.495	2	PRO672	MET674	-	-
Cpd46	18	7.602	2	PRO672	MET674	-	-
Cpd47	18	8.222	2	PRO672	MET674	-	-

Cpd48	18	8.022	2	PRO672	MET674	-	-
Cpd49	18	8.367	2	PRO672	MET674	-	-
Cpd50	18	8.201	2	PRO672	MET674	-	-
Cpd51	18	8.357	2	PRO672	MET674	-	-
Cpd52	18	7.658	2	PRO672	MET674	-	-
Cpd54	18	8.187	2	PRO672	MET674	-	-
Cpd55	18	7.854	2	PRO672	MET674	-	-
Cpd56	18	8.538	2	PRO672	MET674	-	-
Cpd57	18	7.268	2	PRO672	MET674	-	-
Cpd58	18	7.229	2	PRO672	MET674	-	-
Cpd59	18	7.959	2	PRO672	MET674	-	-
Cpd61	18	7.168	2	PRO672	MET674	-	-
Cpd62	19	6.854	2	PRO672	MET674	-	-
Cpd63	14	7.678	2	PRO672	MET674	-	-
Cpd64	18	6.319	2	PRO672	MET674	-	-
Cpd65	16	6.114	2	PRO672	MET674	-	-
Cpd66	16	4.854	2	PRO672	MET674	-	-
Cpd67	18	7.398	2	PRO672	MET674	-	-
Cpd68	18	7.678	2	PRO672	MET674	-	-

having S configuration in the pyrimidine ring seem to be crucial to form hydrogen bond interactions with the active site residues Arg727 and Asn728. These docked poses explain why group I and III compounds have chiral carbon of S configuration. If the chiral compounds have R configuration, these hydrogen bonding interactions would not be possible.
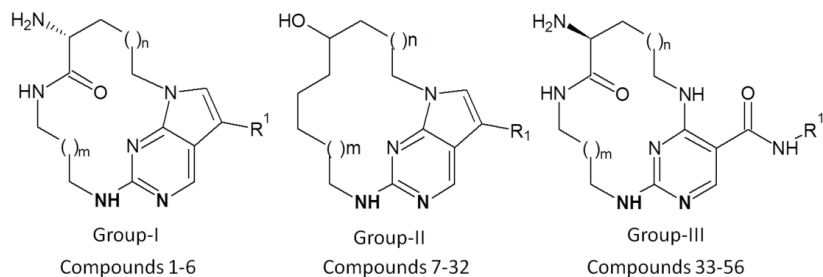
Figure. 3 Dataset compounds (Group-I, Group-II, Group-III).

Dataset compounds were categorized into Group-I, Group-II and Group-III depending on the amino group having S configuration in the pyrimidine ring and compounds having hydroxyl group in the pyrimidine ring.

Hydrophobic interactions of the most active compound with MerTK were also identified. We have used python script to color receptor according to the Eisenberg hydrophobicity scale [45] in PyMOL (Fig. 2b). This gives colouring from red for the most hydrophobic to white for the least hydrophobic. Active site residues which reside in the hydrophobic region (highlighted as sticks in the Fig. 2b) seem to form hydrophobic interactions with the most active compound. Hydrophobic interactions of phenyl ring at the R^1 position with residues Gly677 and Tyr685; and pyrimidine ring with residues Phe673 and Met730 were observed. In addition, the macrocyclic ring was docked into the deep hydrophobic pocket of MerTK which was lined with residues Val601 and Leu671.

From the overall docking results we could speculate that active site residues Pro672, Met674, Arg727, Asn728, Val601 and Leu671 are important for the binding of the macrocyclic ring inside the binding pocket of MerTK. The docked conformations of all these compounds were selected based on the presence of the two key hydrogen bond interactions with hinge region residues Pro672 and Met674 similar to that of the most active compound.

Fig. 1 shows the scatter plot of macrocyclic ring size vs. pIC_{50} activity values. This shows that the various ring-sized compounds possess various inhibitory

activities so no significant effect of the ring size over the potency was observed. However, from our docking analysis it was observed that the compounds with higher activity have more hydrogen bonding interactions whereas, compounds with lower activity values showed less hydrogen bonding interactions with the protein. From which it is quite clear that the hydrogen bond formation may have influence on the activity of the compounds rather than the ring-size effect.

3.3 CoMFA and CoMSIA

Receptor-based 3D-QSAR models were developed for the series of macrocyclic pyrimidines as potent MerTK inhibitors. Conformers obtained from the pose selection for various ring size compounds were used for the 3D-QSAR study. All the compounds were aligned with the template molecule (most active compound) using the common substructure. The common substructure is shown in Fig. 4a and the alignment of the dataset is shown in Fig. 4b. Choice of appropriate partial charges is crucial to obtain reasonable 3D-QSAR models. Thus, different charge schemes such as Gasteiger-Hückel, Hückel, Gasteiger-Marsili, Del-Re, Pullman, MMFF94 and AM1-BCC were applied.
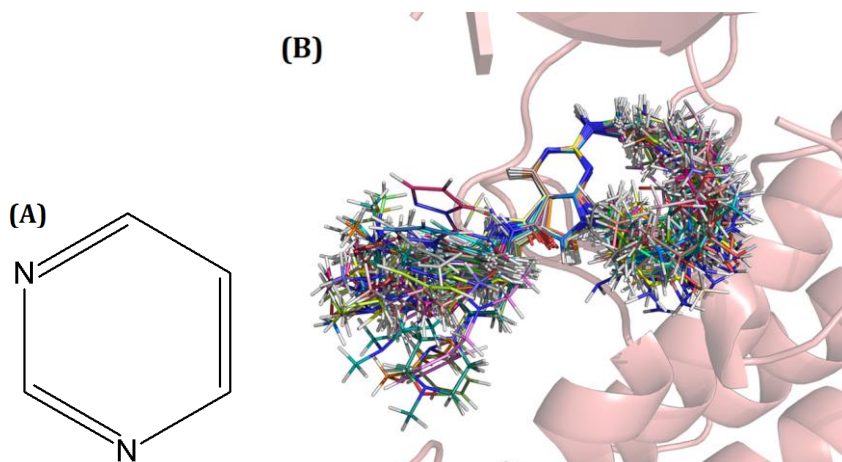
Figure. 4 a Common substructure of the dataset **b** Alignment of the dataset compounds inside the active site of MerTK.

The dataset was divided into a test set of 22 compounds and a training set of 46 compounds to develop 3D-QSAR models for the external test set. CoMFA models with steric and electrostatic field (S+E) contributions were obtained. Among the different charge schemes used, MMFF94 produced statistically reasonable CoMFA model (S+E) in terms of several statistical parameters (q^2=0.702, ONC=5, and r^2=0.979). Then we developed CoMFA models for electrostatic parameter alone (E only) to minimize the interaction between these parameters. The CoMFA models (E only) is shown in Table **3**. MMFF94 produced statistically reasonable model (q^2=0.600, ONC=6, and r^2=0.943). We compared both CoMFA models and found that model with S+E is statistically better than the model with E only. The contour maps obtained from two models (S+E and E only) were quite similar. CoMFA model obtained only with electrostatic parameters could explain clearly about the hydrogen bonding. Therefore we have selected this model to validate the notion that hydrogen bonding interactions could be crucial for the activity variation. The r^2_{pred} value of 0.724 suggests that the model is statistically stable [87]. The scatter plot and contour maps is depicted in Fig. 5 and 6 respectively. The CoMFA (E only) contour map results are consistent with our docking analysis which is discussed in detail in the contour map analysis section.

Table 3 Statistical summary of the developed CoMFA models for electrostatic descriptor with different charge schemes

Parameter		Gasteiger-Huckel	Gasteiger-Marsili	**MMFF94**	Del-re	Huckel	Pullman	AM1-BCC
E	q^2	0.297	0.315	**0.600**	0.282	0.274	0.379	0.328
	ONC	1	1	**6**	1	2	6	3
	SEP	0.758	0.748	**0.607**	0.766	0.779	0.756	0.671
	r^2	0.792	0.849	**0.943**	0.858	0.651	0.958	0.890
	SEE	0.422	0.359	**0.134**	0.349	0.553	0.197	0.341

F value	53.409	78.748	**328.045**	84.531	19.149	147.995	86.305

E: Electrostatic descriptor; 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 (These models were developed after dividing dataset into training set of 46 compounds and test set of 22 compounds)

Different combinations of steric (S), electrostatic (E), hydrophobic (H), hydrogen bond acceptor (A) and hydrogen bond donor (D) fields were used to generate the CoMSIA models. Among the charge schemes studied, MMFF94 was used as a partial charge to generate the CoMSIA models by using the similar training set and test set as used in CoMFA. The detailed statistical values for the generated CoMSIA models are depicted in Table 4. Among the probable combinations, the optimal CoMSIA model generated using electrostatic and hydrogen bond donor parameters (ED) gave relatively better statistical results (q^2=0.563, ONC=5, and r^2=0.927), therefore selected as the final model. The r^2_{pred} value of 0.672 indicates that the model is statistically reasonable [87]. The selected CoMFA (E only) and CoMSIA (ED) models helped us to confirm the hydrogen bonding interactions observed in docking analysis. It is observed that H-bond interactions with the active site residues Arg727 and Asn728 are important along with the H-bond interactions with 2 key residues Pro672 and Met674 for the inhibition of MerTK.

Table 4. Detailed statistical values for the CoMSIA models

CoMSIA	q^2	ONC	SEP	r^2	SEE	F value	Percentage contribution				
							S	E	H	A	D
S	0.473	2	0.664	0.773	0.441	47.693	100	-	-	-	-
E	0.552	5	0.618	0.919	0.196	203.003	-	100	-	-	-
H	0.538	2	0.550	0.842	0.368	74.536	-	-	100	-	-

A	0.539	2	0.620	0.833	0.378	69.947	-	-	-	100	-
D	0.429	3	0.699	0.733	0.478	38.512	-	-	-	-	100
SE	0.544	5	0.572	0.910	0.212	129.458	29.8	70.2	-	-	-
EH	0.560	3	0.534	0.907	0.283	135.865	-	44.5	55.5	-	-
EA	0.515	4	0.581	0.916	0.218	178.624	-	48.9	-	51.1	-
ED	**0.563**	**5**	**0.627**	**0.927**	**0.200**	**171.086**	**-**	**56.3**	**-**	**-**	**43.7**
SD	0.541	2	0.619	0.805	0.403	89.038	35.1	-	-	-	64.9
HD	0.533	3	0.561	0.857	0.350	83.928	-	-	38.9	-	41.1
SEH	0.548	3	0.549	0.902	0.290	128.714	17.2	38.5	44.4	-	-
SEA	0.501	3	0.584	0.903	0.288	130.178	16.9	37.0	-	46.0	-
SED	0.505	5	0.596	0.891	0.305	114.725	23.5	40.5	-	-	36.0
EHA	0.435	3	0.632	0.939	0.213	155.543	-	38.0	35.2	26.9	-
EHD	0.499	3	0.595	0.937	0.217	150.529	-	37.4	38.5	-	24.1
SHA	0.384	3	0.659	0.908	0.261	100.586	17.9	-	45.3	36.8	-
SHD	0.507	2	0.572	0.767	0.441	70.862	18.3	-	41.7	-	40.0
EAD	0.548	5	0.562	0.832	0.241	192.221	-	30.7	-	41.3	28.0
HAD	0.460	3	0.524	0.804	0.226	97.050	-	-	38.6	31.8	29.6
SEHD	0.558	3	0.617	0.841	0.310	130.913	11.7	30.4	28.4	-	28.5
SEHA	0.456	3	0.637	0.939	0.213	155.294	9.4	36.3	28.6	25.7	-

68

SEAD	0.529	3	0.540	0.912	0.214	130.262	11.6	34.3	-	29.0	25.1
EHAD	0.503	3	0.643	0.930	0.211	118.939	-	30.0	26.7	23.4	19.8
SHAD	0.428	3	0.636	0.908	0.361	110.095	13.2	-	33.0	27.7	26.1
SEHAD	0.555	3	0.640	0.915	0.259	137.423	9.0	26.4	24.1	20.9	19.5

Final selected CoMSIA model is highlighted in **bold font**.

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; S: steric, E: electrostatic, H: hydrophobic, A: acceptor, D: donor.

Table 5. Statistical values of the selected CoMFA and CoMSIA models

Parameters	CoMFA (E only)	CoMSIA (ED)
q^2	0.600	0.563
ONC	6	5
SEP	0.607	0.627
r^2	0.943	0.927
SEE	0.134	0.200
F value	328.045	171.086
BS-r^2	0.987	0.981
BS-SD	0.005	0.010
Q^2	0.413	0.554
r^2_{pred}	0.724	0.672
rm^2	0.593	0.549
$Delta\ rm^2$	0.130	0.064
CCC	0.836	0.815

E: Electrostatic; D: Hydrogen bond donor; q^2: cross-validated correlation coefficient, ONC: optimal number of components, SEP: standard error of prediction, r^2: non-validated correlation coefficient, SEE: standard error of estimation, F value: F-test value, BS-r^2: bootstrapping r^2 mean, BS-SD: bootstrapping standard deviation; Q^2: Progressive scrambling; r^2_{pred}: predictive r^2, rm^2:

average rm^2 for the dataset; *Delta rm^2*: Delta rm^2 for the dataset, CCC: concordance correlation coefficient.

Both the CoMFA and CoMSIA models were validated using bootstrapping, progressive scrambling [88], rm^2 metric calculation and CCC. The CCC value was found to be significant as per the criteria given by Gramatica et al. [44]. The detailed statistical values of the selected CoMFA and CoMSIA models are represented in Table **5**. These values suggested that the developed models were statistically reasonable. The scatter plot and contour maps were depicted in Fig. 5 & 6 respectively.
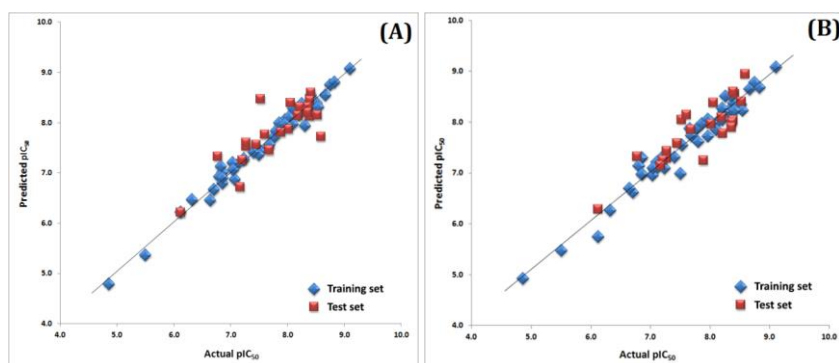
Figure. 5 a Scatter plot for the selected CoMFA model **b** Scatter plot for the selected CoMSIA model.

The plot shows the actual pIC_{50} versus predicted pIC_{50} activity of the dataset for training and test sets. The training set compounds are represented as diamonds in blue colour; the test set compounds are represented as squares in red colour.

3.4 Contour map analysis

The standard parameters for 3D-QSAR models were used to generate the 3D contour maps that illustrate the change in inhibitory activity according to the changes in physicochemical properties of the substituents. The contour maps for the final CoMFA and CoMSIA models are shown in Fig. 6. The most potent compound of the

dataset is shown superimposed with CoMFA and CoMSIA contour maps inside the receptor. Here, we discuss CoMFA model generated only with the electrostatic parameter.

The CoMFA electrostatic contour map is represented in Fig. 6a. The blue contours illustrate the regions where electropositive substitutions increase the activity while red contours illustrate the regions where electronegative substitutions increase the activity. Blue contours are observed at different positions in the macrocyclic ring which suggests that electropositive substitution at these positions increase the activity. The blue contour which is present at **NH** group near pyrimidine ring suggests that electropositive substituent at this position could interact with the Pro672. This can be justified by our docking results in which hydrogen atom from **NH** group forms a hydrogen bond with the oxygen atom of Pro672. This could be the reason for highly active compounds **17, 13, 56, 14, 5, 9**, **53 and 6** (the highest active compound) have electropositive substitutions at this region. Moreover, the presence of blue contours near phenyl ring of R^1 position signifies the region that is favorable for electropositive substitutions. In contrast, red contours are observed near the piperazine ring at R^1 position and near the phenyl ring. This could be the reason for the moderate activity of the compounds **58** and **59** which possess electronegative substituent at these positions.

Since electrostatic contour map in CoMFA (Fig. 6a) is almost similar to the CoMSIA electrostatic contour (Fig. 6b), only the hydrogen bond (H-bond) donor contour map of CoMSIA is discussed below. The H-bond donor contour map from CoMSIA is shown in Fig. 6c. The cyan contours indicate the favorable regions for H-bond donor groups, while the purple contours indicate the unfavorable regions for H-bond donor groups. Cyan contours on both sides of the amino group of macrocyclic ring suggest that the presence of H-bond donor group at these positions could increase the activity. This could be due to the fact that the H-bond donor substitutions at these regions could interact with the oxygen atoms of residues Arg727 and Asn728. These interactions are observed in our docking analyses. This

71

can also be explained by the moderate to the high activity of compounds **17, 13, 56, 14, 5, 9**, **53 and 6** (the highest active compound) which possess H-bond donor group in this region. On the contrary, the purple contours around the pyrimidine ring are unfavorable for the H-bond donor groups and could cause decrease in the efficacy. Overall contour map analysis supports the notion that the hydrogen bond formation has influence on the activity of the compounds rather than the ring-size effect.
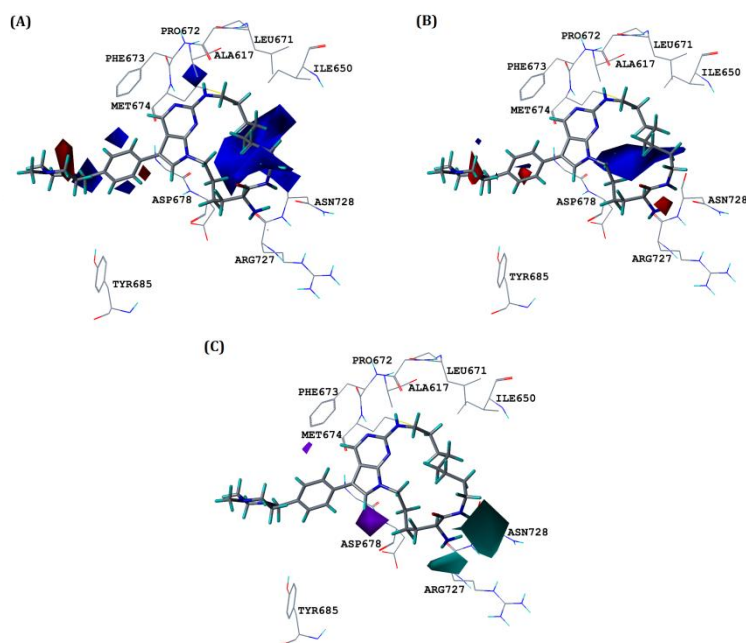
Fig. 6 Contour maps for the selected CoMFA and CoMSIA models **a** CoMFA electrostatic contour map **b** CoMSIA electrostatic contour map **c** CoMSIA hydrogen bond donor contour map. Blue contours favour electropositive substitutions while red contours favour electronegative substitutions. The Cyan contours indicate the favourable region for hydrogen bond donor substitution whereas the purple contours indicate the unfavourable region for hydrogen bond donor substitutions.

4. Conclusion

In this study, we have performed molecular docking of the macrocyclic pyrimidines followed by their pose selection and 3D-QSAR to understand the factors involving a series of macrocyclic inhibitors against MerTK. Molecular docking of these different sized compounds (14 to 20 membered rings) was carried out to check their binding affinity to MerTK. Docking studies revealed the interactions with crucial active site residues; i.e., in addition to two signature hydrogen bond interactions (Pro672 and Met674), the two hydrogen bond interactions with Arg727 and Asn728 were found essential. Furthermore, active site residues Val601, Leu671, Phe673, Gly677, Tyr685 and Met730 were observed which form hydrophobic pocket. Though the ring size effect was not observed, our results support the importance of hydrogen bonding. Moreover, the importance of number of hydrogen bonding was confirmed statistically by receptor-based CoMFA and CoMSIA models and further visualized by their contour maps. These models were validated by various statistical parameters such as bootstrapping, progressive scrambling, an external test set validation, rm^2 metric calculations and concordance correlation coefficient. The analysis of overall contour maps suggested that the electropositive substitution near the pyrimidine ring and at R^1 position are favored to increase the activity. Likewise, electropositive and H-bonding donor groups are favorable in the macrocyclic ring.

Overall, our results provide the importance of hydrogen bonding interactions for this receptor along with suggested critical residues, which could be useful for more potent MerTK inhibitor design. In addition, this work suggests the importance of hydrogen bonding for macrocyclic compounds as inhibitors, at least for the series of the compounds used in this study. However, if the dataset is different, this conclusion may not be applicable. To understand macrocyclic effects more clearly in drug discovery, further detailed studies should be desirable over this issue.

Chapter 4

Receptor-guided 3D-QSAR study to develop a design strategy for RET kinase antagonists

1.0 Introduction

Protein kinases (PKs) are the most explored targets for cancer therapy.[89] The family of receptor tyrosine kinases (TKs) is often related to cancer onset and development due to their deregulation, via mutation or overexpression.[90] RET (rearranged during transfection), one of the receptor tyrosine kinases, is essential for the normal growth, survival, maturation, and maintenance of cells and tissues.[91, 92] It is also necessary for a typical growth of the parasympathetic and peripheral sympathetic nervous systems, brain, lung, thyroid, hematopoietic progenitors, and the neuroendocrine thyroid calcitonin-making C-cells.[93] It consists of the distinctive extracellular domain, the transmembrane domain subsequent to the intracellular domain, a tyrosine kinase domain, and a carboxy-terminal end. An extracellular domain consists of 4 cadherin-like domain (CD-1/2/3/4) replicates, calcium-binding site and a cysteine-rich segment just proximal to the transmembrane domain.[94, 95]

The ligand of the glial cell line-originated neurotrophic factor (GDNF) family activates the RET kinase.[95, 96] The activation of RET requires supplementary GDNF family co-receptors (GFR1/2/3/4).[93] The glycosylphosphatidylinositol linkage helps GFR1/2/3/4 to bind to the external plasma membrane by forming the binary complexes with extracellular RET-stimulating ligands so as to consecutively bind to RET.[97, 98] Moreover, ligand binding needs calcium ions chelated onto the RET kinase extracellular domain.[97, 99] Subsequently, the RET receptor dimer is subjected to transphosphorylation which gives rise to the activity.

Aberrant kinase signaling takes place which results in a variety of cancers[100] and inflammatory diseases.[101] Oncogenic activation of RET arises through one of the four mechanisms: point mutations, chromosomal rearrangements, retroviral transduction and overexpression.[102] Point mutations in RET are usually observed in medullary thyroid carcinoma (MTC)[93, 103] and in multiple endocrine

neoplasia, whereas chromosomal rearrangements are prevalent in non-small cell lung cancer (NSCLC) and in papillary thyroid carcinoma (PTC).[104] Recently, oncogenic activation of RET has been caught up in numerous types of cancers just as prostate,[105] breast,[106] and pancreatic cancer,[99] neuroblastoma,[107] brain tumor and glioblastoma.[108] Hence, the RET proto-oncogene has been recognized as an assuring therapeutic target for the development of therapeutic agents for cancer.

Several RET antagonists have progressed as plausible agents against cancer. Recently FDA and EMA have approved vandetanib,[109] cabozantinib,[109] sorafenib,[22] and lenvatinib[110] for the treatment of MTC. But, these inhibitors are non-selective to RET and possess elevated efficacy against some other protein kinases.[7, 111] These drugs are also ineffective inhibitors against the gatekeeper mutants V804M and V804L.[112] PP1 and PP2, two pyrazolopyrimidine derivatives suppress Src and RET kinase family[107, 113] but more selective for PTK6 than for Src kinase family, and may be useful for the treatment of PTK6-positive malignant diseases such as breast cancer. Likewise, CEP-701 and CEP-751 target various kinases, for example, FLT3, TRK and JAK2 and showed inefficacy against RET.[114] Therefore it is crucial to propose potent as well as specific RET kinase inhibitors.

Few computational studies on RET kinase inhibitors have been reported. In a recent study, Pietra et al., reported the discovery of novel RET inhibitors using receptor-based virtual screening.[21] He reported a new antagonist, 5-(4-chlorophenyl)-3-((4-chlorophenylthio)methyl)-1H-1,2,4-triazole, that showed to inhibit efficiently wild-type and V804L mutant RET kinase as well.[109] In another study, a quantitative structure-activity relationship (QSAR) analysis has been reported to improve the activities of certain carbolin-1-one or indolin-2-one class derivatives as RET inhibitors.[98] We have reported the 3D-QSAR study (3D-Quantitative structure-activity relationship) on 2-substituted phenol quinazolines as RET inhibitors.[115] In the current study, we have performed molecular docking followed

76
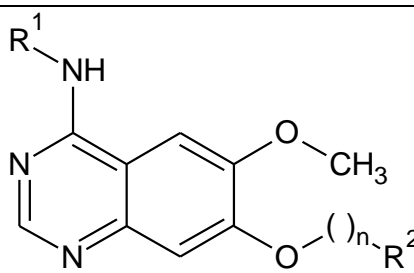
by 3D-QSAR studies on anilinoquinozoline derivatives as RET antagonists to find out the key residues and structural requirements for strong inhibition.
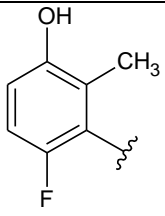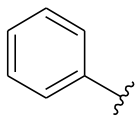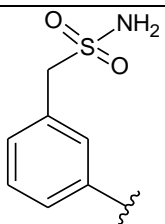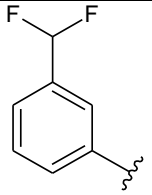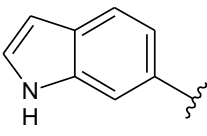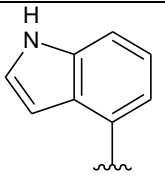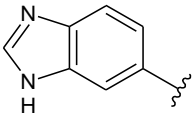
2.0 Materials and Methods
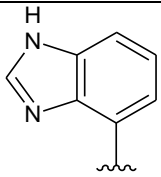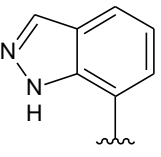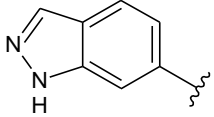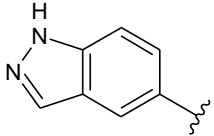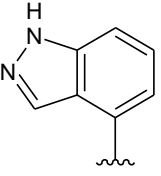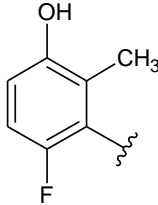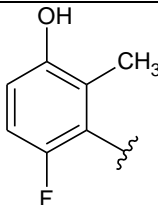
2.1 Dataset
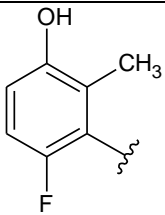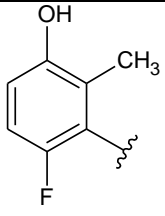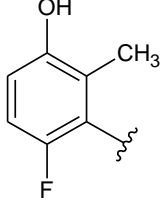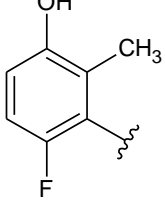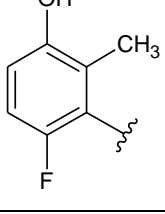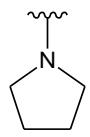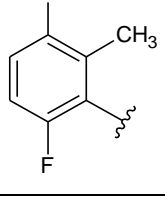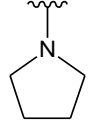
A collection of 42 anilinoquinazoline derivatives as RET kinase antagonists were taken for the study.[111] Converted pIC_{50} values ($-\log IC_{50}$) from the reported IC_{50} values (μM) of the compounds shows the log span of 3.7 which is satisfactory to perform a 3D-QSAR study.[86] Table 1 illustrates the compounds and their corresponding activities of the dataset taken for the study. The three dimensional structures of the compounds were drawn using Tripos SYBYL-X 2.1[35] followed by energy minimization with tripos force field. Gasteiger-Hückel charges were applied as partial charges. To check the external predictivity, the dataset was randomly split into 10 test set compounds and 32 training set compounds.

Table 1. Structure and biological activity values of anilinoquinazoline derivatives as RET kinase inhibitors.
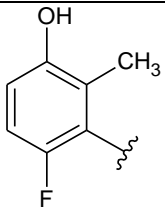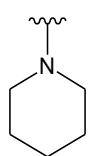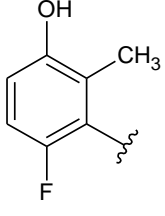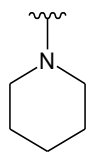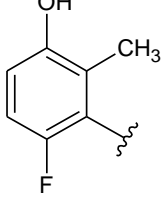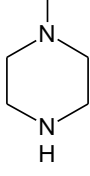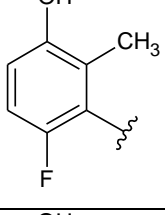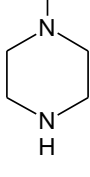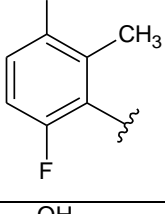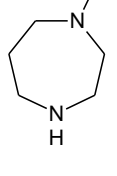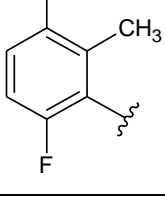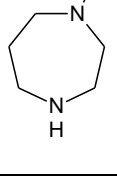
Compounds 1-42

Compound	R^1	R^2	n	IC_{50} (μM)	pIC_{50}
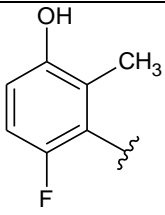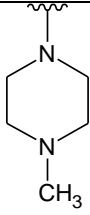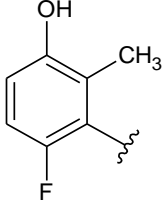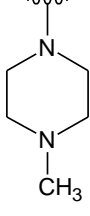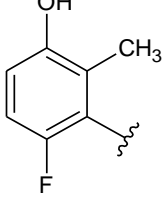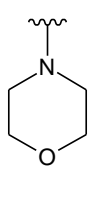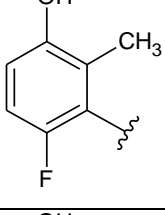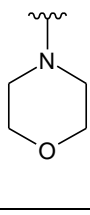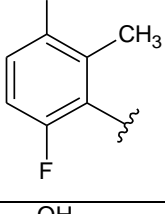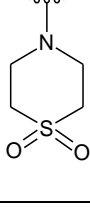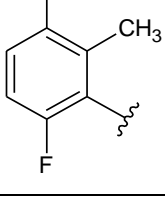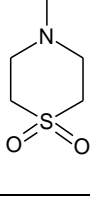
1		-	1	0.044	7.357
2		-	1	1.7	5.770
3		-	1	7.7	5.114
4		-	1	2.9	5.538
5		-	1	1.7	5.770
6*		-	1	9.1	5.041
7		-	1	>30	4.523
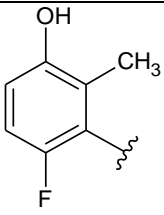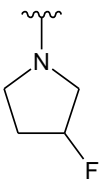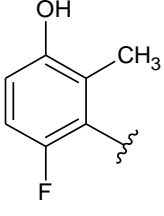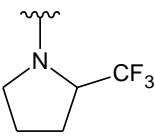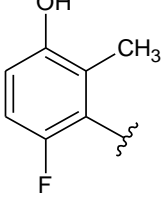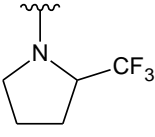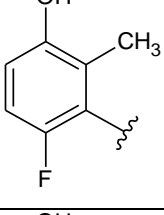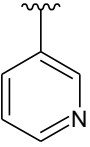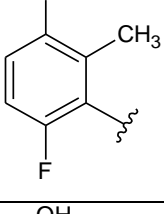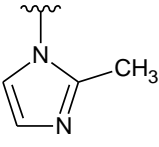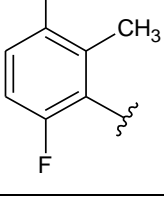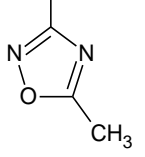
8*		-	1	4.3	5.367
9		-	1	3.3	5.482
10		-	1	0.28	6.553
11		-	1	0.60	6.222
12		-	1	0.14	6.854
13		NH_2	2	0.13	6.886
14		NH_2	3	0.048	7.319
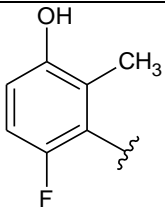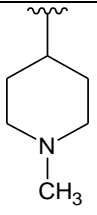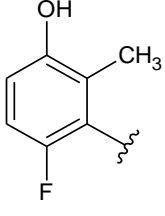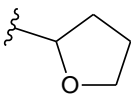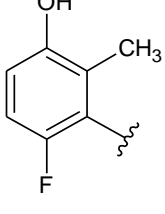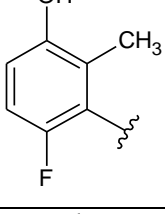
15		NHMe	2	0.10	7.000
16		NHMe	3	0.023	7.638
17		NHMe$_2$	2	0.35	6.456
18*		NHMe$_2$	3	0.007	8.155
19			2	0.10	7.000
20			3	0.015	7.824

80

21			2	0.12	6.921
22			3	0.015	7.824
23			2	0.017	7.770
24			3	0.005	8.301
25*			2	0.065	7.187
26*			3	0.011	7.959

27			2	0.044	7.357
28*			3	0.029	7.538
29			2	0.014	7.854
30			3	0.011	7.959
31			2	0.014	7.854
32			3	0.009	8.046

33*	OH, CH$_3$, F (2,6-substituted phenol)	N-pyrrolidine with F	2	0.030	7.523
34	OH, CH$_3$, F (2,6-substituted phenol)	N-pyrrolidine with CF$_3$	2	0.15	6.824
35	OH, CH$_3$, F (2,6-substituted phenol)	N-pyrrolidine with CF$_3$	3	0.062	7.208
36	OH, CH$_3$, F (2,6-substituted phenol)	pyridine	1	0.012	7.921
37*	OH, CH$_3$, F (2,6-substituted phenol)	2-methylimidazole	1	0.11	6.959
38*	OH, CH$_3$, F (2,6-substituted phenol)	5-methyl-1,2,4-oxadiazole	1	0.024	7.620

39			1	0.078	7.108
40			1	0.018	7.745
41*		OMe	2	0.014	7.854
42		OMe	3	0.038	7.420

* Test set compounds

2.2 Molecular docking

Autodock 4.0.[37] was employed for docking studies. The crystal structure of RET kinase (PDB ID: 4CKJ) was used. As a part of the docking preparations, Kollman charges and polar hydrogens were added to the protein structure. The no. of rotatable bonds of compound 24 was set to 6. The grid was placed around the active site of the protein which was reported in previous studies.[109, 111] Compound 24 (Most active molecule of the dataset) was docked into the set grid of RET kinase with 100 runs of genetic algorithm.

84

2.3 CoMFA and CoMSIA

Alignment of 3D-structures of the molecules is necessary to derive a reasonable 3D-QSAR model.[116] The dataset compounds were aligned on a common scaffold (quinazoline ring) using distill-rigid alignment method in SYBYL-X 2.1 package. 3D QSAR models with CoMFA[38] (Comparative Molecular Field Analysis) and the CoMSIA[116] (Comparative Molecular Similarity Indices Analysis) descriptors were produced using Tripos SYBYL-X 2.1. The Standard protocol was employed for calculating steric and electrostatic parameters required for CoMFA analyses. CoMSIA descriptor uses five fields (Steric, Electrostatic, Hydrophobic, H-bond Donors, H-bond acceptors). PLS regression analysis was performed taking CoMFA and CoMSIA descriptors as independent variables and biological activity (pIC_{50}) as dependent variables. Models were produced to obtain squared cross-validated correlation coefficient (q^2),[34, 85] conventional squared correlation coefficient (r^2) using non-cross validation, and standard error of estimate (SEE). To produce CoMSIA models, different field descriptor combinations were utilized and the descriptors that provided a reasonable statistical model were selected.

2.3.1 Validation of CoMFA and CoMSIA models

The internal and external predictivity of the developed models were validated using bootstrapping, leave five out (LOF), progressive scrambling and external test set validation. Furthermore, predictive correlation coefficient (r^2_{pred}), rm^2 metric calculations and concordance correlation coefficient (CCC) were also carried out.

3.0 Results and Discussion

3.1 Molecular Docking

Binding pocket of RET kinase is contributed by residues Leu730, Gly731, Gly733, Glu734, Ala807, Gly810, Ser811 and Asp874. One best-docked pose was selected from the clusters achieved from all the 100 conformations depending on binding energy, polar and non-polar interactions. The highly active compound **24** was bound favorably in the RET binding pocket (Figure 1a). The selected docked pose showed the binding energy of –7.45 kcal mol^{-1} with five hydrogen bond formations. Two hydrogen bond interactions are observed between the hydroxyl group of a phenyl ring of compound **24** with the hinge region residue Ala807. This interaction mimics the ATP's interaction with kinase which is also present in the x-ray co-crystallized ATP-RET kinase complex (PDB ID: 4CKJ), i.e., it was considered very crucial.[117] Additionally, two more hydrogen bond interactions have been monitored between the two nitrogen atoms of quinazoline ring with residue Ser811. Similarly, a hydrogen bond between the piperazine ring and Asp874 was found.
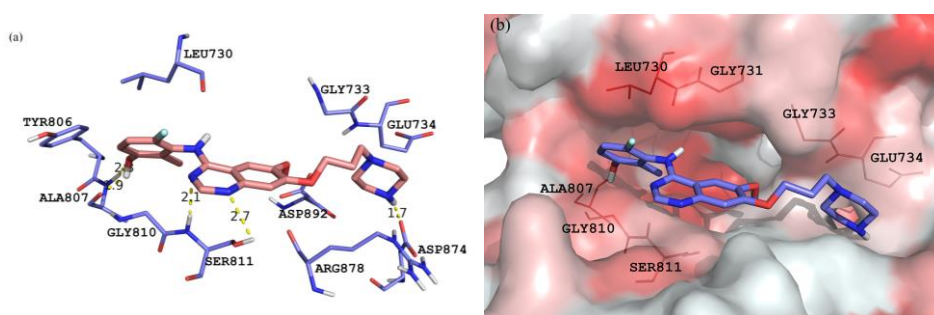
Figure 1 a. The docked conformation of the compound **24** (shown in stick model) inside the binding pocket of RET. Yellow dotted lines represent hydrogen bonds formed between residues and compound **24**. Hydrogen bond distances are labeled in angstrom. **b.** The most active compound **24** (shown in stick model) inside the hydrophobic pocket of RET. The red colored region

represents the most hydrophobic surface and white color represents the least hydrophobic surface of the protein.

Hydrophobic interactions were identified using python script 'color h' and to color the receptor according to the Eisenberg hydrophobicity scale[45] in PyMOL (Figure 1b). This gives coloring from red for the most hydrophobic to white for the least hydrophobic region. Active site residues which reside in the hydrophobic region (highlighted as lines in the figure) seem to form hydrophobic interactions with the highly potent compound **24**. The phenyl ring at the R^1 position was docked into the hydrophobic pocket of RET which was surrounded by residues Leu730, Ala807 and Gly810. Furthermore, hydrophobic interaction of quinazoline ring with residues Gly731 and Ser811 and piperazine ring with residues Gly733 and Glu734 were found. Among these, hydrophobic interaction with Gly731 and Gly733 which is from glycine-rich loop is reported to be important.[117]

We have shown the correlation between pIC_{50} values and the docking score of the remaining compounds of the dataset in Figure 2. The chosen conformation of the ligand was employed in our 3D-QSAR technique.
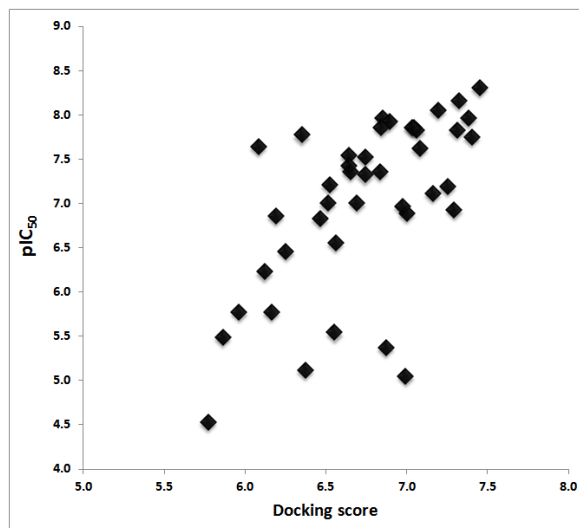
Figure 2. Scatter plot of activity (pIC$_{50}$) values vs docking score.

Here, pIC$_{50}$ = - log$_{10}$ IC$_{50}$ (µM) and docking score = binding energy (kcal mol^{-1}). All the binding energy values are negative. This scatter plot shows that the there is a correlation between pIC$_{50}$ values and the docking score. The trend is not much clear but it indicates the limitations of modeling studies. High active compounds have the lowest binding energy whereas low actives ones have high binding energy. We can also see that some of the low active compounds possess low binding energy and few medium to high active compounds possess high binding energy.

3.2 CoMFA and CoMSIA Studies

The conformation of the compound **24** was taken as template to align the rest of the dataset compounds using the distill-rigid alignment in Sybyl-X 2.1. The common core of the dataset and the alignment of the compounds are depicted in Figure 3a and Figure 3b respectively. Full model for the CoMFA (q^2=0.695, NOC=6, r^2=0.979) was developed. Then the CoMFA model for the external test set compounds of 32 training and 10 test set (q^2=0.723, NOC=5, r^2=0.980) was

produced. This model displayed more reasonable predictive power. Table 2 shows the several CoMSIA models were produced with different combination of the five field descriptors as explained in the methodology section. The field combination of electrostatic and H-bond acceptor (EA) generated a statistically sound CoMSIA model (q^2=0.679, NOC=3, r^2=0.914). Furthermore, external test set validation for combination EA yielded the most predictive CoMSIA model (q^2=0.767, NOC=6, r^2=0.967).

Table 2. Detailed statistical values for the CoMSIA models.

CoMSIA	q^2	ONC	SEP	r^2	SEE	F value	Percentage contribution				
							S	E	H	A	D
S	0.580	5	0.669	0.815	0.444	31.653	100	-	-	-	-
E	0.660	3	0.571	0.882	0.345	94.681	-	100	-	-	-
H	0.555	3	0.662	0.761	0.491	40.350	-	-	100	-	-
A	0.617	4	0.630	0.868	0.369	60.932	-	-	-	100	-
D	0.556	3	0.652	0.773	0.478	43.221	-	-	-	-	100
SE	0.646	3	0.590	0.868	0.365	83.072	26.8	73.2	-	-	-
EH	0.648	3	0.588	0.894	0.327	106.709	-	67.5	32.5	-	-
EA	**0.679**	**3**	**0.569**	**0.914**	**0.295**	**134.424**	**-**	**50.3**	**-**	**49.7**	**-**
ED	0.650	3	0.579	0.882	0.346	94.230	-	56.8	-	-	43.2
SH	0.592	3	0.641	0.774	0.477	43.367	46.5	-	53.5	-	-

SA	0.625	3	0.607	0.845	0.395	69.162	35.6	-	-	64.4	-
SD	0.623	3	0.601	0.801	0.447	51.111	33.1	-	-	-	66.9
HA	0.642	3	0.602	0.888	0.336	100.640	-	-	39.8	60.2	-
HD	0.573	3	0.648	0.834	0.409	63.520	-	-	40.9	-	59.1
SEH	0.644	2	0.591	0.868	0.365	83.020	20.5	54.2	25.3	-	-
SEA	0.675	3	0.564	0.900	0.317	114.581	18.3	39.8	-	41.8	-
SED	0.660	3	0.570	0.876	0.354	89.445	17.3	47.8	-	-	34.9
EHA	0.668	4	0.583	0.914	0.295	134.533	-	38.5	20.6	40.9	-
EHD	0.638	3	0.589	0.872	0.359	86.644	-	43.9	20.1	-	36.0
SHA	0.638	3	0.604	0.873	0.358	87.137	24.3	-	26.5	49.2	-
SHD	0.611	3	0.610	0.820	0.426	57.824	23.3	-	28.9	-	47.8
EAD	0.639	3	0.588	0.918	0.288	141.776	-	33.0	-	35.5	31.6
HAD	0.602	3	0.526	0.897	0.323	110.021	-	-	21.6	39.9	38.6
SEHD	0.647	3	0.582	0.873	0.358	87.138	14.2	38.3	17.9	-	29.6
SEHA	0.672	3	0.567	0.904	0.311	119.497	14.9	32.6	16.5	36.0	-
SEAD	0.657	3	0.573	0.915	0.293	136.367	11.3	28.4	-	34.7	25.6

EHAD	0.635	3	0.591	0.909	0.303	126.433	-	27.8	13.1	32.3	26.8
SHAD	0.621	3	0.603	0.897	0.322	110.713	13.7	-	16.1	39.9	30.3
SEHAD	0.648	3	0.581	0.909	0.304	125.771	10.2	24.6	11.6	31.2	22.4

Final selected CoMSIA model is highlighted in **bold** font.

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; S: steric, E: electrostatic, H: hydrophobic, A: acceptor, D: donor.
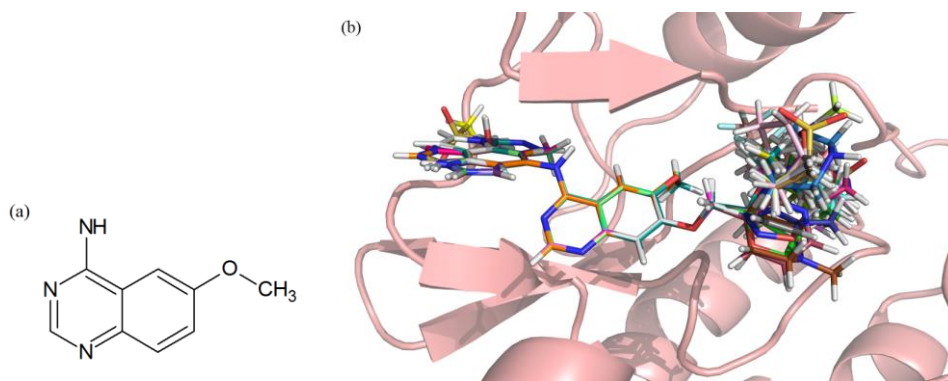
Figure 3 a. Common substructure of the dataset. **b.** Alignment of the dataset compounds inside the active site of RET.

The generated CoMFA and CoMSIA models were carefully validated through quite a few validation techniques such as Leave-Five-out, $r^2_{pred,}$ Bootstrapping, progressive scrambling (Q^2), rm^2 metric calculation and concordance correlation coefficient (CCC). All the validation results confirmed the robustness and predictive ability of the developed models. The values of the concordance correlation coefficient (CCC) for both the models were in the range suggested by Gramatica et al.[47] Similarly the average rm^2 and Delta rm^2 values of the models shows to be in the acceptable range as mentioned by Roy et al. [46] [38][40] The detailed statistical

values of both CoMFA and CoMSIA models along with their validation test results are tabulated in Table 3. Scatter plots and contour maps of CoMFA and CoMSIA models are shown in Figure 4 and Figure 5 respectively. Comparison of experimental Vs the predicted activity values of the models are shown in Table 4.

Table 3. Statistical values of the selected CoMFA and CoMSIA models.

Parameters	CoMFA	CoMSIA (EA)
q^2	0.723	0.767
ONC	5	6
SEP	0.544	0.509
r^2	0.980	0.967
SEE	0.147	0.147
F-value	253.680	212.001
BS-r^2	0.990	0.987
BS-SD	0.005	0.007
Q^2	0.560	0.529
r^2_{pred}	0.859	0.822
LOF	0.745	0.751
rm^2	0.816	0.762
Delta rm^2	0.090	0.050
CCC	0.931	0.911

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 r^2 mean; BS-SD: bootstrapping standard deviation; Q^2: progressive sampling; r^2_{pred}: predictive r^2, *LOF:* Leave-Five-out, *rm2:* average rm^2 for the dataset; *Delta rm*^{2}: Delta rm^2 for the dataset, *CCC:* concordance correlation coefficient.
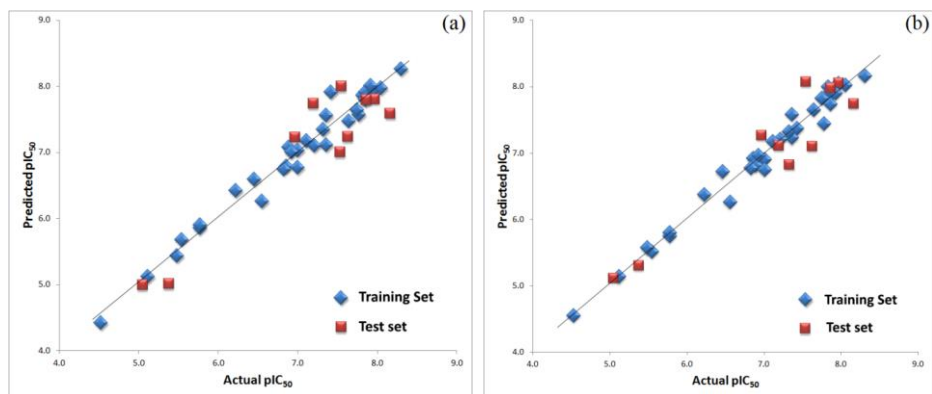
Figure 4 a. Scatter plot for the selected CoMFA model. **b.** Scatter plot for the selected CoMSIA model.

The plot shows the actual pIC_{50} versus predicted pIC_{50} activity of the dataset for training and test sets. The training set compounds are represented as diamonds in blue color; the test set compounds are represented as squares in red color.

Table 4. Actual pIC_{50} and predicted pIC_{50} with their residual values of selected CoMFA and CoMSIA models.

Compound	Actual pIC_{50}	CoMFA		CoMSIA (EA)	
		Predicted	**Residual**	**Predicted**	**Residual**
1	7.357	7.115	0.241	7.236	0.121
2	5.770	5.861	-0.091	5.753	0.017
3	5.114	5.121	-0.008	5.142	-0.029
4	5.538	5.685	-0.147	5.520	0.018
5	5.770	5.901	-0.131	5.809	-0.039
6	5.041	5.001	0.040	5.125	-0.084

7	4.523	4.427	0.096	4.557	-0.034
8	5.367	5.019	0.348	5.309	0.058
9	5.482	5.438	0.043	5.576	-0.095
10	6.553	6.266	0.287	6.268	0.285
11	6.222	6.423	-0.201	6.376	-0.154
12	6.854	6.795	0.059	6.923	-0.069
13	6.886	7.083	-0.197	6.817	0.069
14	7.319	7.350	-0.031	7.327	-0.008
15	7.000	6.777	0.223	6.912	0.088
16	7.638	7.478	0.160	7.659	-0.021
17	6.456	6.595	-0.139	6.723	-0.267
18	8.155	7.595	0.560	7.749	0.406
19	7.000	7.033	-0.033	6.755	0.245
20	7.824	7.870	-0.046	7.847	-0.023
21	6.921	7.009	-0.088	6.970	-0.049
22	7.824	7.862	-0.038	8.005	-0.181
23	7.770	7.575	0.195	7.445	0.325
24	8.301	8.264	0.037	8.167	0.134

25	7.187	7.747	-0.560	7.116	0.071
26	7.959	7.811	0.148	8.063	-0.104
27	7.357	7.562	-0.206	7.584	-0.228
28	7.538	8.005	-0.467	8.282	-0.744
29	7.854	7.912	-0.058	7.931	-0.077
30	7.959	7.938	0.021	8.055	-0.096
31	7.854	7.794	0.060	7.742	0.112
32	8.046	7.979	0.067	8.029	0.017
33	7.523	7.012	0.511	6.729	0.794
34	6.824	6.751	0.073	6.780	0.044
35	7.208	7.103	0.105	7.216	-0.008
36	7.921	8.012	-0.091	7.906	0.015
37	6.959	7.238	-0.279	7.275	-0.316
38	7.620	7.242	0.378	7.006	0.614
39	7.108	7.175	-0.067	7.176	-0.068
40	7.745	7.649	0.096	7.830	-0.085
41	7.854	7.798	0.056	7.984	-0.130
42	7.420	7.912	-0.492	7.374	0.046

3.3 Contour Map Analysis

3.3.1 CoMFA contour maps

Figures 5a and 5b illustrates the contour map of the chosen CoMFA model. The highly active compound **24** is displayed superimposed with CoMFA contour map within the active site of RET kinase. In case of steric contour map, the green and yellow colored contours designate the favorable and unfavorable steric interactions, respectively (Figure 5a). The big green colored contour observed close to R^2 position of piperazine ring expresses that the bulky substituents are ideal to raise the activity. The steric substitution at R^2 position can clarify why the compounds **26, 25, 20, 22, 23, 27** and **32** together with the highly potent compound **24** have increased activities. Similarly, a green colored contour observed close to the R^1 site of a phenyl ring suggests that a bulky substitution is favorable to increase the potency. Quite the opposite, two small yellow colored contours observed next to the methoxy group of quinazoline ring proposing that the steric substitution at this position is not favored and could cause decrease in the efficacy. Big substituent must be avoided at this place because it may sterically interact with the residue Asp892.

CoMFA Electrostatic contour maps are shown in Figure 5b. For higher inhibitory potency, blue contours symbolize regions where positive electrostatic potential are favorable, while red negative. A blue contour region at the R^1 site near a phenyl ring suggests that the positive substitutions in this area might aid in the improved potency. Presence of positive group at this site could make hydrogen bond interaction with oxygen atom of Ala807. This interaction could be explained by our docking results. Compounds bearing hydroxyl substitution at R^1 site were detected to possess higher activity that justifies the occurrence of blue colored contour. This is possibly explained by higher activities of compounds **32, 18, 26, 30, 38, 28,** and many other series of compounds including the highly potent compound **24** which contains hydroxyl group at this place. Red contours close to the nitrogen at R^1 site

near the quinazoline ring indicates that negative substituents could amplify the potency. This might be justified by our docking results where both the nitrogen atoms in the quinazoline ring formed hydrogen bond interaction with Ser811.

3.3.2. CoMSIA contour maps

An electrostatic and the H-bond acceptor descriptors were employed to generate CoMSIA contour maps. Only the H-bond acceptor field contour maps are discussed below due to the similarity between CoMSIA and CoMFA electrostatic contours. Figure 5d represents the H-bond acceptor contour map. Cyan and magenta colored contours specify the areas where H-bond acceptor and H-bond donor groups are favored respectively. Two cyan contours near the phenyl ring and piperazine ring propose that H-bond acceptor substitutions are favored. In contrast, magenta contour observed close to the hydroxyl group at R^1 site of the phenyl ring designates that presence of H-bond donor groups possibly may form hydrogen bond interaction with Ala807 which may lead to the elevated potency. This could be reason why compounds **32, 18, 26, 30, 38, 28,** and together with the highly potent compound **24** which exhibits H-bond donor group (hydroxyl group) at this site specifies increased potency. Likewise, a small magenta colored contour next to piperazine ring (R^2 site) represents the existence of H-bond donor substituents at this site might enhance the activity. This is possibly elucidated by higher activities of compounds **21, 22, 23** as well as the highly potent compound **24** that possesses hydrogen atom at this site. This can be supported with our docking study where hydrogen atom at this site establishes a hydrogen bond with oxygen atom of Asp874. Exploiting the results obtained from the
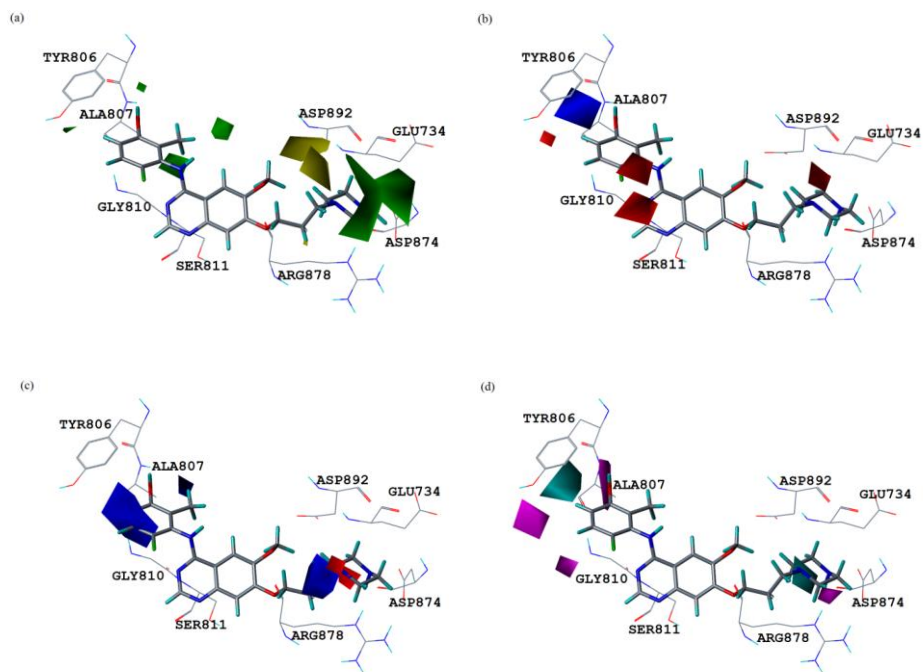
Figure 5. Contour maps for the selected CoMFA and CoMSIA models. **a.** CoMFA steric contour map. **b.** CoMFA electrostatic contour map. **c.** CoMSIA electrostatic contour map. **d.** CoMSIA hydrogen bond acceptor contour map. Green contour shows the regions favorable for bulky substitutions and yellow contours shows the regions unfavorable for bulky substitutions. Blue contours favor electropositive substitutions while red contours favor electronegative substitutions. The Cyan contours indicate the favorable region for hydrogen bond acceptor substitution whereas the magenta contours indicate the favorable region for hydrogen bond donor substitutions.

contour maps analyses, a design strategy is developed to design a series of more potent RET kinase inhibitors of anilinoquinazoline derivatives (Figure 6).
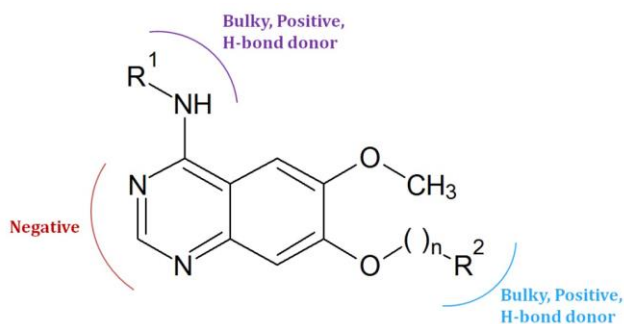
Figure 6. Design strategy to design a series of potent and selective RET kinase inhibitors of anilinoquinazoline derivatives.

4.0 Conclusion

In the present work, molecular docking followed by the 3D-QSAR studies like CoMFA and CoMSIA were carried out on anilinoquinazoline derivatives as RET kinase antagonists. Our docking analyses identified crucial residues within the active site of RET kinase, which interacts with the most active compound (compound **24** in Table 1). Residues Leu730, Gly731, Gly733, Glu734, Ala807, Gly810, Ser811 and Asp874 seemed to be significant in the inhibitory mechanism of RET kinase. The rest of the inhibitors in the dataset were also positioned inside the RET active site and the resultant poses were used to derive 3D-QSAR models. Validation of the chosen models was carried out via bootstrapping, external test set, progressive scrambling, Leave-Five-out, rm^2 metric calculations and concordance correlation coefficient. Statistically validated receptor-guided CoMFA ($q^2=0.723$, NOC=5, $r^2=0.980$) and CoMSIA ($q^2=0.767$, NOC=6, $r^2=0.967$) models were produced. Favorable regions responsible for the elevated inhibitory activity were divulged in the investigation of contour maps of the chosen models. Steric and hydrogen bond donor substitutions at R^1 and R^2 positions are favored. Positive and negative substitutions at R^1 position are favored to enhance the potency.

Furthermore, H-bond acceptor substitutions are favored in the regions where both phenyl and piperazine rings are close.

The interactions of the most potent compound with important active site residues of RET kinase such as Ala807, Ser811 and Asp874 were observed in the favorable regions ideal for the increased potency which were divulged in contour map analyses. Thus, a strong correlation among the docking analysis and outcome of contour map was identified. Our design strategy could be useful to the medicinal chemists and pharmaceutical companies to design more potent RET kinase inhibitors.

Chapter 5

Conclusion of the Study

Conclusion

Receptor Tyrosine Kinases (RTKs) are interesting drug targets due to their involvement in cancer and other diseases. They are widely considered as one of the most promising drug targets in an academic research as well as in pharmaceutical industries. Kinase inhibitor drug discovery and molecular modeling has progressed considerably in the past decade. Various computational approaches have been used in the current study to understand the inhibitory mechanism of a number of tyrosine kinases such as RET, MerTK and FLT3.

We performed a molecular modeling study combining molecular docking, three-dimensional structure-activity relationship (3D-QSAR) and molecular dynamics simulation to find the binding mode of RTK inhibitors and identify the important key active site residues that participate in the inhibition. CoMFA and CoMSIA contour maps were used to understand the structural variations essential to enhance the activity of the inhibitors and to design new molecules. These results assisted in the design of 14 novel FLT3 kinase inhibitors with the increased activity. Further *in vivo* and *in vitro* studies of the designed compounds are required to check their efficacy and safety. The overall results and knowledge acquired from our study is expected to be helpful for the scientific community and medicinal chemists to design and synthesize more potent and specific RTK inhibitors.

References

1. Shen, K., et al., *Protein kinase structure and function analysis with chemical tools.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2005. **1754**(1-2): p. 65-78.

2. Hunter, T., *[1] Protein kinase classification*, in *Methods in enzymology.* 1991, Elsevier. p. 3-37.

3. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-1934.

4. Johnson, L.N., *Protein kinase inhibitors: contributions from structure to clinical compounds.* Quarterly reviews of biophysics, 2009. **42**(1): p. 1.

5. McCance, K.L. and S.E. Huether, *Pathophysiology: The biologic basis for disease in adults and children.* 2014: Elsevier Health Sciences.

6. Bossemeyer, D., *Protein kinases—structure and function.* FEBS letters, 1995. **369**(1): p. 57-61.

7. Gaumann, A.K., et al., *Receptor tyrosine kinase inhibitors: Are they real tumor killers?* International Journal of Cancer, 2016. **138**(3): p. 540-554.

8. Cicenas, J., et al., *Kinases and cancer.* 2018, Multidisciplinary Digital Publishing Institute.

9. Zorn, J.A., et al., *Crystal structure of the FLT3 kinase domain bound to the inhibitor quizartinib (AC220).* PloS one, 2015. **10**(4): p. e0121177.

10. Jarusiewicz, J.A., et al., *Discovery of a Diaminopyrimidine FLT3 Inhibitor Active against Acute Myeloid Leukemia.* ACS omega, 2017. **2**(5): p. 1985-2009.

11. Saif, A., et al., *Acute Myeloid Leukemia: Is That All There Is?* Cureus, 2018. **10**(8).

12. Levis, M. and D. Small, *FLT3 tyrosine kinase inhibitors.* International journal of hematology, 2005. **82**(2): p. 100.

13. Sudhindra, A. and C.C. Smith, *FLT3 inhibitors in AML: are we there yet?* Current hematologic malignancy reports, 2014. **9**(2): p. 174-185.

14. Smith, C.C., et al., *Characterizing and overriding the structural mechanism of the quizartinib-resistant FLT3" gatekeeper" F691L mutation with PLX3397.* Cancer discovery, 2015: p. CD-15-0060.

15. Showel, M.M. and M. Levis, *Advances in treating acute myeloid leukemia.* F1000prime reports, 2014. **6**.

16. Takahashi, S., *Downstream molecular pathways of FLT3 in the pathogenesis of acute myeloid leukemia: biology and therapeutic implications.* Journal of hematology & oncology, 2011. **4**(1): p. 13.

17. Meyers, J., et al., *Medicare fee-for-service enrollees with primary acute myeloid leukemia: an analysis of treatment patterns, survival, and healthcare resource utilization and costs.* Applied health economics and health policy, 2013. **11**(3): p. 275-286.

18. Zhi, Y., et al., *Discovery of the selective and efficacious inhibitors of FLT3 mutations.* European journal of medicinal chemistry, 2018.

19. Sutamtewagul, G. and C.E. Vigil, *Clinical use of FLT3 inhibitors in acute myeloid leukemia.* OncoTargets and therapy, 2018. **11**: p. 7041.

20. Berenstein, R., *Class III Receptor Tyrosine Kinases in Acute Leukemia– Biological Functions and Modern Laboratory Analysis.* Biomarker insights, 2015. **10**: p. BMI. S22433.

21. Dosil, M., S. Wang, and I.R. Lemischka, *Mitogenic signalling and substrate specificity of the Flk2/Flt3 receptor tyrosine kinase in fibroblasts and interleukin 3-dependent hematopoietic cells.* Molecular and cellular biology, 1993. **13**(10): p. 6572-6585.

22. Mizuki, M., et al., *Flt3 mutations from patients with acute myeloid leukemia induce transformation of 32D cells mediated by the Ras and STAT5 pathways.* Blood, 2000. **96**(12): p. 3907-3914.

23. Schlessinger, J., *Receptor tyrosine kinases: legacy of the first two decades.* Cold Spring Harbor perspectives in biology, 2014. **6**(3): p. a008912.

24. Smith, B.D., et al., *Single-agent CEP-701, a novel FLT3 inhibitor, shows biologic and clinical activity in patients with relapsed or refractory acute myeloid leukemia.* Blood, 2004. **103**(10): p. 3669-3676.

25. Stone, R.M., et al., *Patients with acute myeloid leukemia and an activating mutation in FLT3 respond to a small-molecule FLT3 tyrosine kinase inhibitor, PKC412.* Blood, 2005. **105**(1): p. 54-60.

26. De Angelo, D.J., et al., *Phase II evaluation of the tyrosine kinase inhibitor MLN518 in patients with acute myeloid leukemia (AML) bearing a FLT3 internal tandem duplication (ITD) mutation.* 2004, Am Soc Hematology.

27. Michael, M., et al., *Phase Ib study of CP-868,596, a PDGFR inhibitor, combined with docetaxel with or without axitinib, a VEGFR inhibitor.* British journal of cancer, 2010. **103**(10): p. 1554.

28. Zhang, W., et al., *Mutant FLT3: a direct target of sorafenib in acute myelogenous leukemia.* JNCI: Journal of the National Cancer Institute, 2008. **100**(3): p. 184-198.

29. O'Farrell, A.-M., et al., *An innovative phase I clinical study demonstrates inhibition of FLT3 phosphorylation by SU11248 in acute myeloid leukemia patients.* Clinical Cancer Research, 2003. **9**(15): p. 5465-5476.

30. Lancet, J.E., *FLT3 inhibitors for acute myeloid leukemia.* Clin Adv Hematol Oncol, 2015. **13**(9): p. 573-575.

31. Smith, C.C., et al., *Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia.* Nature, 2012. **485**(7397): p. 260.

32. Ke, Y.-Y., et al., *Homology modeling of DFG-in FMS-like tyrosine kinase 3 (FLT3) and structure-based virtual screening for inhibitor identification.* Scientific reports, 2015. **5**: p. 11702.

33.   Mashkani, B., et al., *FMS-like tyrosine kinase 3 (FLT3) inhibitors: Molecular docking and experimental studies.* European journal of pharmacology, 2016. **776**: p. 156-166.

34.   Wold, S., et al., *The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses.* SIAM Journal on Scientific and Statistical Computing, 1984. **5**(3): p. 735-743.

35.   Clark, M., R.D. Cramer, and N. Van Opdenbosch, *Validation of the general purpose Tripos 5.2 force field.* Journal of computational chemistry, 1989. **10**(8): p. 982-1012.

36.   Halgren, T.A., *MMFF VI. MMFF94s option for energy minimization studies.* Journal of computational chemistry, 1999. **20**(7): p. 720-729.

37.   Huey, R., et al., *A semiempirical free energy force field with charge-based desolvation.* Journal of computational chemistry, 2007. **28**(6): p. 1145-1152.

38.   Cramer, R.D., D.E. Patterson, and J.D. Bunce, *Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins.* Journal of the American Chemical Society, 1988. **110**(18): p. 5959-5967.

39.   Kamath, S. and J.K. Buolamwini, *Receptor-guided alignment-based comparative 3D-QSAR studies of benzylidene malonitrile tyrphostins as EGFR and HER-2 kinase inhibitors.* Journal of medicinal chemistry, 2003. **46**(22): p. 4657-4668.

40.   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.

41.   Madhavan, T., et al., *Various atomic charge calculation schemes of CoMFA on HIF-1 inhibitors of moracin analogs.* International Journal of Quantum Chemistry, 2012. **112**(4): p. 995-1005.

42. Dowlati Beirami, A., Z. Hajimahdi, and A. Zarghi, *Docking-based 3D-QSAR (CoMFA, CoMFA-RG, CoMSIA) study on hydroquinoline and thiazinan-4-one derivatives as selective COX-2 inhibitors.* Journal of Biomolecular Structure and Dynamics, 2018: p. 1-8.

43. Klebe, G., U. Abraham, and T. Mietzner, *Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity.* Journal of medicinal chemistry, 1994. **37**(24): p. 4130-4146.

44. Chirico, N. and P. Gramatica, *Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection.* Journal of chemical information and modeling, 2012. **52**(8): p. 2044-2058.

45. Eisenberg, D., et al., *Analysis of membrane and surface protein sequences with the hydrophobic moment plot.* Journal of molecular biology, 1984. **179**(1): p. 125-142.

46. Roy, K., et al., *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): p. 1071-1082.

47. Chirico, N. and P. Gramatica, *Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient.* Journal of chemical information and modeling, 2011. **51**(9): p. 2320-2335.

48. Madsen, C.M. and M.H. Clausen, *Biologically Active Macrocyclic Compounds–from Natural Products to Diversity-Oriented Synthesis.* European Journal of Organic Chemistry, 2011. **2011**(17): p. 3107-3115.

49. Giordanetto, F. and J. Kihlberg, *Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties?* Journal of medicinal chemistry, 2013. **57**(2): p. 278-295.

50. Mallinson, J. and I. Collins, *Macrocycles in new drug discovery.* Future medicinal chemistry, 2012. **4**(11): p. 1409-1438.

51. Driggers, E.M., et al., *The exploration of macrocycles for drug discovery—an underexploited structural class.* Nature Reviews Drug Discovery, 2008. **7**(7): p. 608.

52. Gradillas, A. and J. Perez-Castells, *Synthesis of Natural Products Containing Macrocycles by Alkene Ring-Closing Metathesis.* Metathesis in Natural Product Synthesis: Strategies, Substrates and Catalysts, 2010: p. 149-182.

53. Marsault, E. and M.L. Peterson, *Macrocycles are great cycles: applications, opportunities, and challenges of synthetic macrocycles in drug discovery.* Journal of medicinal chemistry, 2011. **54**(7): p. 1961-2004.

54. Marsault, E. and M.L. Peterson, *Practical Medicinal Chemistry with Macrocycles: Design, Synthesis, and Case Studies*. 2017: John Wiley & Sons.

55. Wessjohann, L.A., et al., *What can a chemist learn from nature's macrocycles?–A brief, conceptual view.* Molecular diversity, 2005. **9**(1-3): p. 171-186.

56. Bridger, G.J., et al., *Synthesis and structure-activity relationships of phenylenebis (methylene)-linked bis-tetraazamacrocycles that inhibit HIV replication. Effects of macrocyclic ring size and substituents on the aromatic linker.* Journal of medicinal chemistry, 1995. **38**(2): p. 366-378.

57. Hawkins, P.C., *Conformation Generation: The State of the Art.* Journal of chemical information and modeling, 2017. **57**(8): p. 1747-1756.

58. Watts, K.S., et al., *Macrocycle conformational sampling with MacroModel.* Journal of chemical information and modeling, 2014. **54**(10): p. 2680-2696.

59. Hojjat-Farsangi, M., *Small-molecule inhibitors of the receptor tyrosine kinases: promising tools for targeted cancer therapies.* International journal of molecular sciences, 2014. **15**(8): p. 13768-13801.

60. Yu, J., et al., *Targeting Receptor Tyrosine Kinases and Their Downstream Signaling with Cell-Penetrating Peptides in Human Pulmonary Artery Smooth Muscle and Endothelial Cells.* Chemical biology & drug design, 2015. **85**(5): p. 586-597.

61. Cummings, C.T., et al., *Molecular pathways: MERTK signaling in cancer.* Clinical Cancer Research, 2013. **19**(19): p. 5275-5280.

62. Knubel, K.H., et al., *MerTK inhibition is a novel therapeutic approach for glioblastoma multiforme.* Oncotarget, 2014. **5**(5): p. 1338.

63. Evans, A.L., et al., *Antagonistic Coevolution of MER Tyrosine Kinase Expression and Function.* Molecular biology and evolution, 2017. **34**(7): p. 1613-1628.

64. Graham, D.K., et al., *Cloning and mRNA expression analysis of a novel human protooncogene, c-mer.* Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research, 1994. **5**(6): p. 647-657.

65. Lemke, G. and C.V. Rothlin, *Immunobiology of the TAM receptors.* Nature Reviews Immunology, 2008. **8**(5): p. 327.

66. Linger, R.M., et al., *TAM receptor tyrosine kinases: biologic functions, signaling, and potential therapeutic targeting in human cancer.* Advances in cancer research, 2008. **100**: p. 35-83.

67. Schlegel, J., et al., *MERTK receptor tyrosine kinase is a therapeutic target in melanoma.* The Journal of clinical investigation, 2013. **123**(5): p. 2257-2267.

68. Craven, R.J., et al., *Receptor tyrosine kinases expressed in metastatic colon cancer.* International journal of cancer, 1995. **60**(6): p. 791-797.

69. Neubauer, A., et al., *Expression of axl, a transforming receptor tyrosine kinase, in normal and malignant hematopoiesis.* Blood, 1994. **84**(6): p. 1931-1941.

70.   Nguyen, K.-Q.N., et al., *Overexpression of MERTK receptor tyrosine kinase in epithelial cancer cells drives efferocytosis in a gain-of-function capacity.* Journal of Biological Chemistry, 2014. **289**(37): p. 25737-25749.

71.   Holland, S.J., et al., *R428, a selective small molecule inhibitor of Axl kinase, blocks tumor spread and prolongs survival in models of metastatic breast cancer.* Cancer research, 2010. **70**(4): p. 1544-1554.

72.   Holland, S.J., et al., *Multiple roles for the receptor tyrosine kinase axl in tumor formation.* Cancer research, 2005. **65**(20): p. 9294-9303.

73.   Hamzah, N. and D. Tjahjono, *A Quantitative Structure-Activity Relationship Study, Compound Development, Pharmacophore Feature, and Molecular Docking of Pyrazolo-[3, 4-d]-Pyrimidine Derivatives as Mer Tyrosine Kinase Inhibitor.*

74.   Yu, Z., et al., *3D-QSAR modeling and molecular docking study on Mer kinase inhibitors of pyridine-substituted pyrimidines.* Molecular diversity, 2015. **19**(1): p. 135-147.

75.   Balupuri, A., P.K. Balasubramanian, and S.J. Cho, *Molecular modeling study on Mer kinase inhibitors using 3D-QSAR and docking approaches.* Medicinal Chemistry Research, 2015. **24**(10): p. 3730-3742.

76.   Balupuri, A., P.K. Balasubramanian, and S.J. Cho, *Determination of structural requirements of Mer kinase inhibitors and binding interaction analysis using in silico approaches.* Medicinal Chemistry Research, 2016. **25**(12): p. 3021-3029.

77.   Liang, F., et al., *Medical applications of macrocyclic polyamines.* Current medicinal chemistry, 2006. **13**(6): p. 711-727.

78.   McIver, A.L., et al., *Discovery of Macrocyclic Pyrimidines as MerTK-Specific Inhibitors.* ChemMedChem, 2017. **12**(3): p. 207-213.

79.   Wang, X., et al., *Design and synthesis of novel macrocyclic Mer tyrosine kinase inhibitors.* ACS medicinal chemistry letters, 2016. **7**(12): p. 1044-1049.

80.   Zhou, S., et al., *Pharmacophore-based 3D-QSAR modeling, virtual screening and molecular docking analysis for the detection of MERTK inhibitors with novel scaffold.* Combinatorial chemistry & high throughput screening, 2016. **19**(1): p. 73-96.

81.   Gasteiger, J. and M. Marsili, *Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges.* Tetrahedron, 1980. **36**(22): p. 3219-3228.

82.   Del Re, G., *812. A simple MO–LCAO method for the calculation of charge distributions in saturated organic molecules.* Journal of the Chemical Society (Resumed), 1958: p. 4031-4040.

83.   Berthod, H., C. Giessner-Prettre, and A. Pullman, *Sur les rôles respectifs des électrons σ et π dans les propriétés des dérivés halogénés des molécules conjuguées. Application à l'étude de l'uracile et du fluorouracile.* Theoretica chimica acta, 1967. **8**(3): p. 212-222.

84.   Jakalian, A., D.B. Jack, and C.I. Bayly, *Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation.* Journal of computational chemistry, 2002. **23**(16): p. 1623-1641.

85.   Cramer, R.D., *Partial least squares (PLS): its strengths and limitations.* Perspectives in Drug Discovery and Design, 1993. **1**(2): p. 269-278.

86.   Li, Y.-P., et al., *3D-QSAR studies of azaoxoisoaporphine, oxoaporphine, and oxoisoaporphine derivatives as anti-AChE and anti-AD agents by the CoMFA method.* Journal of Molecular Graphics and Modelling, 2013. **41**: p. 61-67.

87.   Pratim Roy, P., et al., *On two novel parameters for validation of predictive QSAR models.* Molecules, 2009. **14**(5): p. 1660-1701.

88.   Cramer, R.D., et al., *Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies.* Molecular Informatics, 1988. **7**(1): p. 18-25.

89.   Zhang, J., P.L. Yang, and N.S. Gray, *Targeting cancer with small molecule kinase inhibitors.* Nature Reviews Cancer, 2009. **9**(1): p. 28.

90.   Arora, A. and E.M. Scholar, *Role of tyrosine kinase inhibitors in cancer therapy.* Journal of Pharmacology and Experimental Therapeutics, 2005. **315**(3): p. 971-979.

91.   Jhiang, S.M., *The RET proto-oncogene in human cancers.* Oncogene, 2000. **19**(49): p. 5590.

92.   Schuchardt, A., et al., *Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret.* Nature, 1994. **367**(6461): p. 380.

93.   Mulligan, L.M., *RET revisited: expanding the oncogenic portfolio.* Nature Reviews Cancer, 2014. **14**(3): p. 173.

94.   Anders, J., S. Kjær, and C.F. Ibáñez, *Molecular modeling of the extracellular domain of the RET receptor tyrosine kinase reveals multiple cadherin-like domains and a calcium-binding site.* Journal of Biological Chemistry, 2001. **276**(38): p. 35808-35817.

95.   Sariola, H. and M. Saarma, *Novel functions and signalling pathways for GDNF.* Journal of Cell Science, 2003. **116**(19): p. 3855-3862.

96.   Airaksinen, M.S. and M. Saarma, *The GDNF family: signalling, biological functions and therapeutic value.* Nature Reviews Neuroscience, 2002. **3**(5): p. 383.

97.   Ibáñez, C.F., *Structure and physiology of the RET receptor tyrosine kinase.* Cold Spring Harbor perspectives in biology, 2013. **5**(2): p. a009134.

98.   Mologni, L., et al., *Synthesis, structure–activity relationship and crystallographic studies of 3-substituted indolin-2-one RET inhibitors.* Bioorganic and Medicinal Chemistry, 2010. **18**(4): p. 1482-1496.

99.   Knowles, P.P., et al., *Structure and chemical inhibition of the RET tyrosine kinase domain.* Journal of Biological Chemistry, 2006. **281**(44): p. 33577-33587.

100.   Sun, C. and R. Bernards, *Feedback and redundancy in receptor tyrosine kinase signaling: relevance to cancer therapies.* Trends in Biochemical Sciences, 2014. **39**(10): p. 465-474.

101.   Barnes, P.J., *New anti-inflammatory targets for chronic obstructive pulmonary disease.* Nature reviews Drug discovery, 2013. **12**(7): p. 543.

102.   Cranston, A.N., et al., *RET is constitutively activated by novel tandem mutations that alter the active site resulting in multiple endocrine neoplasia type 2B.* Cancer Research, 2006. **66**(20): p. 10179-10187.

103.   Carlomagno, F. and M. Santoro, *Thyroid cancer in 2010: a roadmap for targeted therapies.* Nature Reviews Endocrinology, 2011. **7**(2): p. 65.

104.   Plaza-Menacho, I., L. Mologni, and N. McDonald, *Mechanisms of RET signaling in cancer: current and future implications for targeted therapy.* Cellular Signalling, 2014. **26**(8): p. 1743-1752.

105.   Dawson, D.M., et al., *Altered expression of RET proto-oncogene product in prostatic intraepithelial neoplasia and prostate cancer.* JNCI: Journal of the National Cancer Institute, 1998. **90**(7): p. 519-523.

106.   Boulay, A., et al., *The Ret receptor tyrosine kinase pathway functionally interacts with the ERα pathway in breast cancer.* Cancer Research, 2008. **68**(10): p. 3743-3751.

107.   Peterson, S. and E. Bogenmann, *The RET and TRKA pathways collaborate to regulate neuroblastoma differentiation.* Oncogene, 2004. **23**(1): p. 213.

108.   Gimm, O., et al., *Mutation and deletion analysis of GFRα-1, encoding the co-receptor for the GDNF/RET complex, in human brain tumours.* British Journal of Cancer, 1999. **80**(3-4): p. 383.

109.   La Pietra, V., et al., *Challenging clinically unresponsive medullary thyroid cancer: Discovery and pharmacological activity of novel RET inhibitors.* European journal of medicinal chemistry, 2018.

110.   Marzaro, G., *Discovery of novel wtRET and V804MRET inhibitors: from hit to lead.* ChemMedChem, 2017.

111. Jordan, A.M., et al., *Anilinoquinazoline inhibitors of the RET kinase domain—Elaboration of the 7-position.* Bioorganic and Medicinal Chemistry Letters, 2016. **26**(11): p. 2724-2729.

112. Yoon, H., et al., *Identification of a novel 5-amino-3-(5-cyclopropylisoxazol-3-yl)-1-isopropyl-1H-pyrazole-4-carboxamide as a specific RET kinase inhibitor.* European journal of medicinal chemistry, 2017. **125**: p. 1145-1155.

113. Orlicek, S.L., J.H. Hanke, and B.K. English, *The src family-selective tyrosine kinase inhibitor PP1 blocks LPS and IFN-gamma-mediated TNF and iNOS production in murine macrophages.* Shock (Augusta, Ga.), 1999. **12**(5): p. 350-354.

114. Shih, K.-C., et al., *Pharmacophore modeling and virtual screening to identify potential RET kinase inhibitors.* Bioorganic and Medicinal Chemistry Letters, 2011. **21**(15): p. 4490-4497.

115. Bhujbal, S.P., P.K. Balasubramanian, and S.J. Cho, *In silico studies on 2-substituted phenol quinazoline derivatives as RET receptor tyrosine kinase antagonists.* Medicinal Chemistry Research, 2017. **26**(12): p. 3228-3239.

116. Hu, R., et al., *Receptor-and ligand-based 3D-QSAR study for a series of non-nucleoside HIV-1 reverse transcriptase inhibitors.* Bioorganic and Medicinal Chemistry, 2009. **17**(6): p. 2400-2409.

117. Plaza-Menacho, I., et al., *Oncogenic RET kinase domain mutations perturb the autophosphorylation trajectory by enhancing substrate presentation in trans.* Molecular Cell, 2014. **53**(5): p. 738-751.

Appendix A

Acknowledgement

First and foremost I would like to express my sincere gratitude to my advisor **Prof. Seung Joo Cho** for his guidance and motivation. I can't thank him enough for providing me the great opportunity to pursue my Ph.D. under his supervision. He has been a marvelous mentor to me. I'm grateful for his continuous support, patience and sharing of his immense research knowledge and ideas with me. I always enjoyed the lab meet and the discussion time with him and have learned a lot from him. His teaching and expertise have made my Ph.D. experience productive and interesting. As my advisor, his understanding and care for me made my tough times in the Ph.D. manageable. His advice on both research as well as on my career have been priceless. Throughout my Ph.D., he gave me the moral support and the freedom to pursue the research I wanted to do. Without his guidance and constant feedback, this Ph.D. would not have been possible.

My sincere thanks to the thesis committee members: **Prof. Song Yub Shin**, **Prof. Eun Ae Kim**, **Prof. Ho Joong Kim** and **Prof. Jeong Hyun Dam** for evaluating my thesis. Their insightful comments and questions have made my defense enjoyable and memorable.

I owe my gratitude to Chosun University International office team for their academic help in various stages of my Ph.D. They took us on field trips to various parts of Korea which helped me to understand and embrace the culture of Korea and warm gesture of Korean people. I thank the former coordinator **Youshin Jung** for all her help and kindness.

Words cannot express my thankfulness to **my parents**. They raised me independently and allowed to decide my career and pursue towards it. I also thank

115

---

my sisters **Supriya**, **Rupali** and my cousin brother **Sachin** for all their encouragement and support.

My deep appreciation goes to my co-guide **Dr. Pavithra K B** and my senior **Dr. Anand Balupuri** who have been supportive in many ways. There are innumerable situations we discussed our research and worked together all night to keep up with deadline. I am also grateful to my labmate **Seketoulie Keretsu** for his help and wonderful fun lab time. The time we spent together will always be remembered and cherished.

Besides my lab members, I extend my sincere gratitude to my all **Chosun University** and **Chonnam National University** friends (Dr. He Weijie, Adnan, Dinesh, Ajish, Dr. Immanuel Anna, Dr. Tamil Anna, Dr. Hari Anna, Ijaz, Chris, Yubraj, Deva, Amal, Aarthi, Ayesh, Aditya, Jaya) and Korean friends (Sujin, Vin, Jeje, Seoeun, Monica, Elena) for their help and moral support.

Last but not least, I would like to thank all my friends from India for their constant encouragement. My special thanks to Dr. Vikram, Sushant, Santosh, Aditya, Prashant, Prasad, Ganesh, Digambar, Rashmya and Preeti.

I thank everyone who helped directly and indirectly to complete my doctoral study successfully.

Appendix B

List of publications

1. Seketoulie Keretsu, **Swapnil Pandurang Bhujbal**, Seung Joo Cho. 2020. Computational Study of Pyrimidin-2-Aminopyrazol-Hydroxamate-based JAK2 Inhibitors for the Treatment of Myeloproliferative Neoplasms. Bulletin of the Korean Chemical Society. DOI: https://doi.org/10.1002/bkcs.12008.

2. **Swapnil P. Bhujbal**, Seketoulie Keretsu, Seung Joo Cho. 2020. Design of New Therapeutic Agents Targeting FLT3 Receptor Tyrosine Kinase Using Molecular Docking and 3D-QSAR Approach. Letters in Drug Design & Discovery. 17 (5): 585-596.

3. Seketoulie Keretsu, **Swapnil P. Bhujbal**, Seung Joo Cho. 2020. Docking and 3D-QSAR Studies of Hydrazone and Triazole Derivatives for Selective Inhibition of GRK2 over ROCK2. Letters in Drug Design & Discovery. 17 (5): 618-632.

4. Pavithra K. Balasubramanian, Anand Balupuri, **Swapnil P. Bhujbal**, Seung Joo Cho. 2020. 3D-QSAR Assisted Design of Novel 7-Deazapurine Derivatives as TNNI3K Kinase Inhibitors using Molecular Docking and Molecular Dynamics Simulation. Letters in Drug Design & Discovery. 17 (2): 155 – 168.

5. **Swapnil Pandurang Bhujbal**, Pavithra Kuruchi Balasubramanian, Seketoulie Keretsu, Seung Joo Cho. 2019. Receptor-guided 3D-QSAR study of anilinoquinazolines as RET receptor tyrosine kinase antagonists. Bulletin of the Korean Chemical Society. 40(3): 207-213.

6. **Swapnil Pandurang Bhujbal**, Seketoulie Keretsu, Seung Joo Cho. A Combined Molecular Docking and 3D-QSAR Studies on Tetrahydropteridin Derivatives as PLK2 Antagonists. 2019. Bulletin of the Korean Chemical Society. Vol. 40: 796–802

7. **Swapnil P. Bhujbal**, Pavithra K. Balasubramanian, Seketoulie Keretsu, Seung Joo Cho. 2019. Macrocyclic effect on inhibitory activity: a modeling study on MerTK inhibitors. Medicinal Chemistry Research, 28: 923-1938

8. Pavithra K. Balasubramanian, Anand Balupuri, **Swapnil P. Bhujbal**, Seung Joo Cho. 2019. 3D-QSAR-aided design of potent c-Met inhibitors using molecular dynamics simulation and binding free energy calculation. Journal of Biomolecular Structure and Dynamics. 37(8): 2165-2178.

9. Seketoulie Keretsu, Pavithra Kuruchi Balasubramanian, **Swapnil Pandurang Bhujbal**, Seung Joo Cho. 2017. Receptor-guided 3D-Quantitative Structure-Activity Relationship and Docking Studies of 6-Substituted 2-Arylaminopurines as CDK2 Kinase Inhibitors. Bulletin of the Korean Chemical Society. 38(11): 1275-1284.

10. **Swapnil P. Bhujbal**, Pavithra K. Balasubramanian, Seung Joo Cho. 2017. In silico studies on 2-substituted phenol quinazoline derivatives as RET receptor tyrosine kinase antagonists. Medicinal Chemistry Research, 26(12): 3228-3239.

11. Seketoulie Keretsu, Pavithra Kuruchi Balasubramanian, **Swapnil Pandurang Bhujbal**, Seung Joo Cho. 2019. Computational study of paroxetine-like

inhibitors reveals new molecular insight to inhibit GRK2 with selectivity over ROCK1. Scientific Reports, 9 (13053).

MANUSCRIPT UNDER PREPARATION

12. Seketoulie Keretsu, **Swapnil P. Bhujbal**, Seung Joo Cho. 2020. Rational Approach toward COVID-19 Main Protease Inhibitors via Various Molecular Modelling Studies. Nature Scientific Reports. (under revision)

13. **Swapnil P. Bhujbal**, Seketoulie Keretsu, Seung Joo Cho. 2020. Molecular modelling studies on pyrazole derivatives for the design of potent RET kinase inhibitors. Bulletin of the Korean Chemical Society. (Under preparation)

14. **Swapnil P. Bhujbal**, Seketoulie Keretsu, Seung Joo Cho. 2020. Molecular modeling studies of the diaminonicotinamide derivatives as IRAK4 antagonists using combined approach of molecular docking, molecular dynamics simulation and 3D-QSAR. Journal of Biomolecular Structure and Dynamics. (Under preparation)

I dedicate this thesis to my family and friends who believed in me to achieve my dream. Their unconditional love, constant support and grace has always kept me going.