T.C.
BAHCESEHIR UNIVERSITY
GRADUATE SCHOOL OF EDUCATION
THE DEPARTMENT OF BIOENGINEERING

E. SAYYAH

NOVEL RESISTANCE-FREE RET TYROSINE KINASE INHIBITOR DISCOVERY THROUGH DYNAMIC STRUCTURE-BASED PHARMACOPHORE AND QSAR MODELING AND VIRTUAL SCREENING OF ULTRA LARGE LIGAND LIBRARIES

MASTER'S THESIS

EHSAN SAYYAH

BAU 2023

ISTANBUL, 2023

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF EDUCATION

THE DEPARTMENT OF BIOENGINEERING

NOVEL RESISTANCE-FREE RET TYROSINE KINASE INHIBITOR DISCOVERY THROUGH DYNAMIC STRUCTURE-BASED PHARMACOPHORE AND QSAR MODELING AND VIRTUAL SCREENING OF ULTRA LARGE LIGAND LIBRARIES

MASTER'S THESIS

EHSAN SAYYAH

THESIS ADVISOR

PROF. DR. SERDAR DURDAĞI

ISTANBUL, 2023

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

June/14/2023

MASTER THESIS APPROVAL FORM

Program Name:	BIOENGINEERING PROGRAM
Student's Name and Surname:	Ehsan SAYYAH
Name of The Thesis:	Novel Resistance-Free RET Tyrosine Kinase Inhibitor Discovery through Dynamic Structure-Based Pharmacophore and QSAR Modeling and Virtual Screening of Ultra Large Ligand Libraries
Thesis Defense Date:	June/21/2023

This thesis has been approved by the Graduate School and has fulfilled the necessary conditions as a master's thesis.

Assist. Prof. Dr. Fatma Elif ÇETİN
Institute Director

This thesis was read by us, quality and content as a Master's thesis have been seen and accepted as sufficient.

	Title/Name	**Signature**
Thesis Advisor's	Prof. Dr. Serdar Durdağı	
2nd Member	Assoc. Prof. Dr. Kader Şahin	
3rd Member	Assist. Prof. Dr. Enes Kotil	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Ehsan Sayyah

Signature:

ABSTRACT

NOVEL RESISTANCE-FREE RET TYROSINE KINASE INHIBITOR DISCOVERY THROUGH DYNAMIC STRUCTURE-BASED PHARMACOPHORE AND QSAR MODELING AND VIRTUAL SCREENING OF ULTRA LARGE LIGAND LIBRARIES

Sayyah, Ehsan

Bioengineering Master's Program

Thesis Advisor: Prof. Dr. Serdar Durdağı

June 2023, 157 pages

Gene fusion and point mutations cause to activate RET tyrosine kinase where the gene fusion is responsible for non-small cell lung cancer and papillary thyroid cancers and the point mutation on the RET proto-oncogene, which is the receptor tyrosine kinase, causes multiple endocrine neoplasia type 2A and 2B (MEN2A, MEN2B) and Familial medullary thyroid cancer. Furthermore, small molecules as inhibitors bind to the binding site of RET kinase domain to block their enzymatic activity. On the other hand, a single amino acid change on the RET kinase position can provide resistance to the tyrosine kinase inhibitors, so it is essential to find a compound with activity against RET mutants. So, we hypothesize that with in-silico E-pharmacophore modeling of molecular dynamics trajectories and predicting pIC50 values of small molecules from several libraries using QSAR models, we can discover new small molecules that can inhibit the activation of RET proto-oncogene and their signaling pathways, so it can stop the growth of tumor sizes with the lowest resistance to that inhibitor.

Keywords: E-Pharmacophore; Medullary Thyroid Cancer; Non-Small Cell Lung Cancer; RET proto-oncogene; QSAR

ÖZET

ULTRA BÜYÜK LİGAND KÜTÜPHANELERİNİN SANAL TARAMASI VE DİNAMİK YAPI TEMELLİ FARMAKOFOR VE QSAR MODELLEMESİ İLE YENİ REZİSTANS GÖSTERMEYEN RET TİROZİN KİNAZ İNHİBİTÖRÜNÜN KEŞFİ

Ehsan Sayyah

Biyomühendislik Yüksek Lisans Programı

Tez Danışmanı: Prof. Dr. Serdar Durdağı

Haziran 2023, 157 sayfa

Gen füzyonları ve nokta mutasyonları, gen füzyonunun non-small cell lung cancer ve papiller tiroid kanserlerinden sorumlu olduğu ve RET proto-onkogenindeki nokta mutasyonunun, reseptör tirozin kinazı olan RET proto-onkogeninin aktivasyonuna neden olduğu, multiple endokrin neoplazi tip 2A ve 2B (MEN2A, MEN2B) ve familial medüller tiroid kanseriyle ilişkilidir. Ayrıca, inhibitör olarak kullanılan küçük moleküller RET kinaz etki alanına bağlanarak enzimatik aktivitelerini engellerler. Bununla birlikte, RET kinaz pozisyonunda tek bir amino asit değişikliği, tirozin kinaz inhibitörlerine direnç sağlayabilir, bu nedenle RET mutasyonlarına karşı etkili bir bileşik bulmak önemlidir. Bu nedenle, moleküler dinamik trajelerin e-farmakofor modellemesi ve QSAR modelleri kullanılarak küçük moleküllerin çeşitli kütüphanelerinden elde edilen pIC_{50} değerlerinin tahmin edilmesi ile RET proto-onkogenin aktivasyonunu ve sinyal yolaklarını inhibe edebilen yeni küçük moleküller keşfedebileceğimizi hipotez ediyoruz. Bu şekilde, tümör büyümesini ve inhibitöre karşı en düşük direnci olan tümör boyutlarını durdurabiliriz.

Anahtar Kelimeler: E-Farmakofor; Küçük Hücreli olmayan Akciğer Kanseri; Medüller Tiroid Kanseri; RET proto-onkogeni; QSAR

I dedicate this thesis to my advisor, Prof. Dr. Serdar Durdağı, whose guidance and support have been invaluable throughout my journey as a master's student in the laboratory. I also dedicate it to all DurdagiLab members, whose collaboration and expertise have significantly contributed to the success of this project. To my family and friends, thank you for your unwavering support and encouragement. Lastly, I dedicate this thesis to all the individuals who have dedicated their lives to the pursuit of scientific knowledge, may our collective efforts make a difference and inspire future generations of scientists.

ACKNOWLEDGEMENTS

I would like to extend my heartfelt appreciation to all those who have supported and contributed to the successful completion of this thesis. Foremost, I am profoundly grateful to my advisor, Prof. Dr. Serdar Durdağı, for their invaluable guidance, encouragement, and expertise throughout the entire research process. Their insightful feedback and unwavering support have played a pivotal role in shaping this work.

I would also like to express my gratitude to all the members of the Durdağı lab for their support, encouragement, and engaging discussions during this journey. Their camaraderie and shared experiences have made this research endeavor more enriching.

Lastly, I am deeply thankful to my family for their constant love, understanding, and unwavering encouragement throughout my academic pursuits. Their unwavering support and belief in my abilities have been a constant driving force.

I extend my heartfelt thanks to each and every one of you for your support, guidance, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

ABBREVIATIONS

2D	2 – Dimensional
3D	3 – Dimensional
ALA	Alanine
ARG	Arginine
ASP	Asparagine
ATC	Anaplastic Thyroid Carcinoma
CCDC6	Coiled-Coil Domain-Containing Protein 6
CDK5	Cyclin-Dependent Kinase 5
CYS	Cysteine
DTC	Differentiated Thyroid Carcinomas
FDA	Food and Drug Administration
FP	False Positive
FTC	Follicular Thyroid Carcinoma
IFD	Induced Fit Docking
GDNF	Glial Cell Line-Derived Neurotrophic Factor
GFL	GDNF family of ligands
GLIDE	Grid-based ligand docking with energetics
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HTVS	High Throughput Virtual Screening
IGF-1	Immunoglobulin Factor I
ILE	Isoleucine
KIF5B	Kinesin Family Member 5B
LEU	Leucine
MAPK	Mitogen-Activated Protein Kinase
MCC	Matthews Correlation Coefficient
MD	Molecular Dynamics

MEN2	Multiple Endocrine Neoplasia Type 2
MET	Methionine
MM	Molecular Mechanics
MTC	Medullary Thyroid Carcinoma
mTOR	Mammalian Target of Rapamycin
N-terminal	Amino Terminal
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NMR	Nuclear Magnetic Resonance
NPT	Isothermal – Isobaric Ensemble
NSCLC	Non-Small Cell Lung Cancer
OPLS	Optimized Potential for Liquid Simulations
PD	Parkinson Disease
PDB	Protein Databank
PHE	Phenylalanine
PLS	Partial Least Square
PTC	Papillary Thyroid Carcinoma
QSAR	Quantitative Structure – Activity Relationship
RET	Rearranged During Transfection
RMSD	Root-Mean-Square Deviation
RMSE	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
MM/GBSA	Molecular Mechanics the Generalized Born Surface Area
TIP3P	Transferable Intermolecular Potential 3P
TP	Tanimato Prioritization
SCLC	Small Cell Lung Cancer
SD	Standard Deviation
SER	Serine
SERM	Selective Estrogen Modulators
SP	Standard Precision
SPC	Single Point Charge
THR	Threonine
TN	True Negative

TRP	Tryptophan
TYR	Tyrosine
UTC	Undifferentiated Thyroid Carcinomas
VAL	Valine
WHO	World Health Organization
XP	Extra Precision

Chapter 1

Introduction

1.1 Purpose of Thesis

Thyroid cancer and non-small cell lung cancer (NSCLC) are among the most frequent types of cancer worldwide and can be associated with high morbidity and mortality rates. Despite recent advances in cancer treatment, effective therapies for these cancers remain limited, highlighting the need for new and innovative approaches to cancer drug discovery.

One potential target for drug discovery in these cancers is the RET tyrosine kinase protein. RET, a transmembrane receptor tyrosine kinase situated within the cell membrane, assumes a pivotal function in fostering cellular proliferation, differentiation, and viability. Activating mutations in RET have been found to be associated with the development and progression of several types of cancer, including medullary thyroid cancer and NSCLC.

Medullary thyroid cancer is a rare type of thyroid cancer that arises from the C cells of the thyroid gland. It accounts for approximately 5-10% of all thyroid cancers and can be associated with familial genetic syndromes. In contrast, NSCLC is a heterogeneous group of lung cancers that account for approximately 85% of all lung cancer cases. NSCLC is often diagnosed at advanced stages and has a poor prognosis, highlighting the need for new therapeutic options.

RET has garnered attention as a prospective focus for the discovery of cancer drugs, specifically in relation to medullary thyroid cancer and non-small cell lung cancer (NSCLC). Preclinical studies have shown that inhibition of RET signaling can lead to reduced tumor growth and improved survival outcomes in animal models. Additionally, recent clinical trials have shown promising results for RET inhibitors in patients with medullary thyroid cancer and NSCLC, highlighting the potential of this approach for cancer treatment.

Selpercatinib (LOXO-292) and Pralsetinib (BLU-667) are two highly selective RET inhibitors that have shown great potential in clinical trials. Both medications have received approval from the FDA for treating metastatic RET fusion-positive NSCLC

and medullary thyroid cancer, showcasing remarkable rates of response and long-lasting effects.

The main problem with the currently available RET inhibitors, such as selpercatinib and pralsetinib, is the emergence of resistance mutations. While these drugs have shown promising results in clinical trials, some patients have developed resistance to the therapy over time, leading to disease progression and reduced treatment efficacy (Drilon et al. 2020; V. Subbiah et al. 2018).

The emergence of resistance mutations can occur through various mechanisms, including changes in the binding affinity of the drug to its target, alterations in downstream signaling pathways, and activation of compensatory signaling pathways. For example, a common resistance mutation in RET is the gatekeeper mutation (V804M), which affects the binding of the drug to its target site and reduces its inhibitory effect (Drilon et al. 2017).

Another challenge with the use of Selpercatinib and Pralsetinib is their selectivity for RET, which may limit their efficacy in patients with tumors that have multiple driver mutations. In these cases, combination therapies with other drugs may be necessary to achieve optimal treatment outcomes (Gainor et al. 2021).

To address these challenges, there is a need for the development of new RET inhibitors that can overcome resistance mutations and have broader selectivity for tumors with multiple driver mutations.

Despite these advances, the development of RET inhibitors has been challenging due to the high degree of homology between RET and other tyrosine kinases. Developing inhibitors that are both potent and selective for RET has been a major obstacle in drug discovery efforts. In recent years, in-silico drug discovery methods have emerged as powerful tools for identifying novel RET inhibitors (Lanzi et al. 2009; Lin and Shaw 2016).

Computer aid drug discovery methods such as QSAR and pharmacophore modeling can be used to screen large libraries of compounds for potential RET inhibitors. Machine learning algorithms can be trained on large data sets of known RET inhibitors and non-inhibitors to predict novel compounds that are likely to bind to RET with high affinity. Pharmacophore modeling can be used to identify common structural features among known RET inhibitors and use these features to screen virtual compound libraries (Brogi et al. 2020; Zhang, Chung, and Oldenburg 1999).

Molecular dynamics simulations can also be used to generate MD trajectories for analysis, which can provide valuable insights into the binding mode and specificity of potential inhibitors. By combining computational methods with experimental biophysical and pharmacological studies, it may be possible to identify novel RET inhibitors with improved therapeutic properties (Drilon et al. 2017).

1.2 Scientific Aims of the Thesis

The primary aim of this thesis is to develop novel RET tyrosine kinase inhibitors that can effectively halt the growth of medullary thyroid cancer and non-small cell lung cancer (NSCLC) by identifying hit molecules using computational methods such as quantitative structure-activity relationship (QSAR) and e-pharmacophore modeling. The study also aims to gain insights into the mechanism of RET inhibition by analyzing the binding properties of the most promising compounds to the RET tyrosine kinase protein. By comparing the results of QSAR and e-pharmacophore modeling, this research aims to determine the most effective approach for identifying potential RET inhibitors. In addition to identifying hit molecules for RET inhibitors, this thesis also aims to develop inhibitors that are less prone to resistance. Previous studies have shown that some RET inhibitors, such as Selpercatinib and Pralsetinib, may develop resistance due to mutations in the binding site of the inhibitor. Therefore, a key objective of this research is to design inhibitors that can effectively bind to the RET tyrosine kinase protein and overcome resistance caused by mutations. By doing so, this research aims to contribute to the development of more effective and long-lasting treatments for medullary thyroid cancer and NSCLC (Drilon et al. 2018a; Lin and Shaw 2016; Vidal et al. 2005). The ultimate goal of this research is to contribute to the development of more efficient drug discovery pipelines for RET inhibitors, thereby improving the therapeutic options available to patients suffering from medullary thyroid cancer and NSCLC.

Chapter 2

Literature Review

2.1 Cancer

Cancer is a disease that results from the uncontrolled growth and spread of abnormal cells in the body. The first recorded case of cancer was found in an Egyptian textbook written around 1600 BCE. Throughout history, cancer has been known by many names, including "the emperor of all maladies", "the king of terrors", and "the great crippler". It is a major public health issue worldwide, with millions of new cases and deaths each year (Siddhartha Mukherjee 2011). As per the World Health Organization (WHO), cancer stands as the second most prevalent cause of mortality worldwide, resulting in approximately 10 million deaths in the year 2020 (Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M 2020). In Turkey, cancer is also a significant health problem. According to the Turkish Ministry of Health, cancer was the second leading cause of death in 2019, accounting for 21.6% of all deaths (Dr. BİRİNCİ et al. 2020). In the early 20th century, researchers discovered that cancer was caused by changes, or mutations, in the DNA of cells. This led to the development of treatments such as chemotherapy, radiation therapy, and surgery. In recent years, advances in cancer research have led to the development of targeted therapies and immunotherapies, which have exhibited encouraging outcomes in the treatment of some types of cancer. Furthermore, computer-aided drug discovery techniques have played a crucial role in the development of targeted therapies and immunotherapies for cancer treatment (The American Cancer Society 2021).

2.1.1 Thyroid Cancer.

Thyroid cancer is a type of cancer that originates in the thyroid gland, a butterfly-shaped gland located in the neck that secretes hormones that govern the body's metabolic processes. The incidence of thyroid cancer has been increasing in recent decades, and some experts believe that improved access to care and advances in diagnostic technology have contributed to this trend (Davies and Welch 2014; Morris et al. 2013).

The molecular pathogenesis of thyroid cancer is complex and not fully understood, but genetic alterations that activate oncogenic pathways are thought to play a key role in the development of the disease. (2,6) Common genetic alterations in

thyroid cancer include mutations in genes such as BRAF, RAS, and RET, which promote cell proliferation and survival (Xing 2013).

The typical diagnosis of thyroid cancer involves a blend of physical examination, imaging tests, and biopsy of thyroid tissue. To aid in treatment decisions, the Bethesda System for Reporting Thyroid Cytopathology is extensively employed as a classification system for interpreting the results of thyroid biopsies (Cibas and Ali 2009). The 2015 American Thyroid Association Management Guidelines provide recommendations for the evaluation and management of thyroid nodules and differentiated thyroid cancer (Haugen et al. 2016).

The treatment approach for thyroid cancer is contingent upon various factors, including the cancer's type, stage, and patient-specific considerations like age and overall health. Possible treatment options encompass surgical intervention to remove a portion or the entirety of the thyroid gland, radioiodine therapy to eliminate thyroid cells, and external beam radiation therapy (Haddad et al. 2022).

In conclusion, thyroid cancer is a complex disease with a rising incidence. Understanding its molecular pathogenesis and appropriate management strategies are crucial for improving outcomes for patients with thyroid cancer.

2.1.1.1 ***Classification of Thyroid Cancer.*** Thyroid cancer can be classified into four main histological types: papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), medullary thyroid carcinoma (MTC), and anaplastic thyroid carcinoma (ATC) (2). Among them, PTC and FTC are differentiated thyroid carcinomas (DTC), while MTC and ATC are undifferentiated thyroid carcinomas (UTC) (Nikiforov and Nikiforova 2011).

2.1.1.1.1 *Papillary Thyroid Carcinoma.* The BRAF V600E mutation is the most common genetic alteration found in PTC, occurring in around 40-45% of cases (Xing 2013). This mutation results in a constitutively activated BRAF protein, which in turn activates the downstream MEK-ERK signaling pathway, leading to increased cell proliferation and survival. The RET/PTC rearrangements, which are present in around 10-20% of cases, involve the fusion of the RET proto-oncogene with various partner genes, resulting in a constitutively activated RET protein. The RAS mutations, which are present in around 10-15% of cases, result in a constitutively activated RAS protein, leading to the activation of the same downstream MEK-ERK signaling pathway as the BRAF mutation (Agrawal et al. 2014). Exposure to ionizing radiation is a well-known risk factor for developing PTC. This can occur due to exposure to radiation during childhood or adolescence, as the thyroid gland is more sensitive to radiation damage during these periods of development. In addition, certain inherited genetic conditions, such as familial adenomatous polyposis and Cowden syndrome, can increase the risk of developing PTC. Exposure to certain chemicals or environmental toxins, such as asbestos and certain pesticides, may also increase the risk of PTC (Crnčić et al. 2020).

2.1.1.1.2 *Follicular Thyroid Carcinoma.* Follicular thyroid carcinoma (FTC) ranks as the second most prevalent form of thyroid cancer, accounting for approximately 10-15% of cases (Nikiforov and Nikiforova 2011). FTC arises from the cells that produce and store thyroid hormones, known as follicular cells and is characterized by the formation of abnormal, cancerous follicles in the thyroid gland. FTC is more common in women than in men and typically occurs in individuals over the age of 40 (Ashorobi and Lopez 2023). The majority of FTC cases are sporadic, meaning they occur without a known genetic cause. However, certain genetic mutations have been identified as risk factors for developing FTC. One of the most well-studied mutations is the RAS gene mutation, which is found in approximately 40-50% of FTC cases (Xing 2013). Mutations in RAS lead to the persistent activation of the RAS protein, consequently triggering the downstream MAPK signaling pathway. This activation contributes to heightened cell proliferation and survival. Other genetic mutations that have been implicated in the development of FTC include the PAX8-PPARγ rearrangement and mutations in the TERT promoter region (Nikiforov and Nikiforova 2011).

FTC development is associated with a recognized risk factor, which is exposure to ionizing radiation. Like PTC, FTC can occur due to exposure to radiation during childhood or adolescence, as the thyroid gland is more sensitive to radiation damage during these periods of development. In addition, certain inherited genetic conditions, such as Cowden syndrome and Carney complex, can increase the risk of developing FTC (Xing 2013). FTC is typically diagnosed through a combination of imaging studies, such as ultrasound and radioactive iodine scans, and biopsy of thyroid nodules. Treatment options for FTC include surgery to remove the affected thyroid tissue, and radioactive iodine therapy to destroy any remaining cancerous cells. In some cases, chemotherapy and external radiation therapy may also be used (Ashorobi and Lopez 2023).

2.1.1.1.3 *Medullary Thyroid Cancer.* Medullary thyroid carcinoma (MTC) is an uncommon form of thyroid cancer originating from the parafollicular cells, also referred to as C-cells. These cells are responsible for producing the hormone calcitonin. MTC accounts for approximately 5-10% of all thyroid cancers (Wells et al. 2015a). Unlike other types of thyroid cancer, MTC is not typically associated with exposure to ionizing radiation, and there are several known genetic mutations that can cause this cancer. MTC can occur sporadically or as part of an inherited syndrome. Approximately 75% of all MTC cases are sporadic and occur without a known genetic cause. The remaining 25% of cases are inherited and are associated with germline mutations in the RET proto-oncogene, which encodes a receptor tyrosine kinase that regulates cell growth and differentiation (Moura et al. 2011). Multiple endocrine neoplasias (MEN) type 2 is the most well-known inherited syndrome associated with MTC and is caused by mutations in the RET gene. MEN2 is further divided into two subtypes: MEN2A, which is characterized by the presence of MTC, pheochromocytoma (a tumor of the adrenal gland), and hyperparathyroidism; and MEN2B, which is characterized by the presence of MTC, pheochromocytoma, and a variety of other clinical features (Wells et al. 2015a). In sporadic MTC cases, somatic mutations in the RET gene have been identified in up to 50% of cases, as well as mutations in other genes such as RAS and TP53 (Moura et al. 2011).

These mutations lead to the activation of the RET signaling pathway, which promotes cell growth and proliferation. Inherited mutations in the RET gene typically involve the loss of function of the RET protein, which can predispose individuals to the development of MTC. MTC is typically diagnosed through the measurement of serum calcitonin levels and imaging studies such as ultrasound, CT, and MRI scans. Treatment options for MTC include surgery to remove the affected thyroid tissue, as well as any lymph nodes that may be involved. In cases where cancer has spread beyond the thyroid, chemotherapy and radiation therapy may also be used (Romei, Ciampi, and Elisei 2016).

2.1.1.1.4 *Anaplastic Thyroid Carcinoma.* Anaplastic thyroid carcinoma (ATC) is a rare, aggressive, and highly malignant cancer that accounts for less than 2% of all thyroid cancers. It arises from the follicular cells of the thyroid gland and is characterized by the rapid and uncontrolled growth of cancerous cells that invade the surrounding tissues and organs. ATC typically presents in elderly individuals and carries an unfavorable prognosis due to its aggressive characteristics and limited treatment options. Although the precise cause of ATC is not entirely comprehended, it is believed, similar to other forms of thyroid cancer, to be associated with genetic mutations. Several mutations have been identified in ATC, including mutations in the TP53, BRAF, and PIK3CA genes (Agrawal et al. 2014). TP53 is a tumor suppressor gene that helps prevent the development of cancer by regulating cell growth and division. Mutations in TP53 can result in the loss of its tumor suppressor activity, leading to uncontrolled cell growth and the development of cancer. BRAF and PIK3CA are oncogenes that promote cell growth and proliferation. Mutations in these genes can activate their signaling pathways, leading to the development and progression of cancer. ATC is usually diagnosed based on a combination of imaging studies and biopsy results. Treatment options for ATC include surgery, radiation therapy, and chemotherapy, either alone or in combination. However, due to its aggressive nature and resistance to conventional therapies, the prognosis for ATC is generally poor, with a five-year survival rate of less than 10% (Smallridge et al. 2012).

In recent years, there have been advances in targeted therapies and immunotherapies that show promise in the treatment of ATC. These therapies aim to specifically target the genetic mutations that drive the growth of cancer cells or to

stimulate the patient's own immune system to attack the cancer cells. Clinical trials are ongoing to evaluate the effectiveness of these new treatments in improving the outcomes for patients with ATC.

2.1.2 **Lung Cancer.** Lung cancer ranks among the primary contributors to cancer-related fatalities on a global scale. It is a type of cancer that starts in the cells of the lung tissue, usually in the cells lining the air passages. The disease can be traced back to ancient times, with evidence of lung cancer found in mummies from Egypt and Peru. However, it was not until the 20th century that the disease became recognized as a major public health concern (Herbst 2013). In the early 1900s, lung cancer was relatively rare, with most cases occurring in people who worked in the mining and manufacturing industries. However, with the rise of cigarette smoking in the mid-20th century, the incidence of lung cancer began to increase rapidly. Studies in the 1950s and 1960s established a strong link between smoking and lung cancer, and smoking remains the primary cause of the disease today (Doll and Peto 1976). When a person inhales cigarette smoke, the carcinogenic compounds in the smoke can damage the cells lining the lungs. Over time, this damage can lead to the development of cancerous cells that can grow and spread throughout the lung tissue and beyond. Other risk factors for lung cancer include exposure to radon gas, asbestos, and air pollution, as well as a family history of the disease (The American Cancer Society medical and editorial content team 2019). Lung cancer can be difficult to detect in its early stages, as symptoms often do not appear until the disease has progressed. Frequent indications of lung cancer encompass coughing, chest pain, difficulty breathing, and coughing up blood. The available treatment alternatives for lung cancer consist of surgery, radiation therapy, chemotherapy, and targeted therapy, which are determined based on the cancer's stage and type (PDQ Adult Treatment Editorial Board 2002). Unfortunately, lung cancer has a high mortality rate, with only about 20% of patients surviving for 5 years after diagnosis. This is largely due to the fact that the disease is often diagnosed at an advanced stage, when it has already spread to other parts of the body. In addition, the aggressive nature of lung cancer and its resistance to many conventional treatments make it a difficult disease to manage (Molina et al. 2008).

2.1.2.1 ***Classification of Lung Cancer.*** Lung cancer is typically categorized into two primary types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) (PDQ Adult Treatment Editorial Board 2002). Non-small cell lung cancer (NSCLC) prevails as the most prevalent form, comprising approximately 85% of all diagnosed lung cancer cases (Lukeman 2015). NSCLC can be further classified into three subtypes based on the type of cells found in the tumor: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Travis et al. 2015). The classification of NSCLC is important because it can help determine the appropriate treatment approach, as well as the prognosis and potential outcomes of the disease (Detterbeck et al. 2017). SCLC is less common than NSCLC, accounting for about 15% of all lung cancer cases. It is characterized by small cells that grow rapidly and form large tumors that can quickly spread to other parts of the body (PDQ Adult Treatment Editorial Board 2002). Overall, the classification of lung cancer is based on the type of cells found in the tumor and is important for guiding treatment decisions and predicting outcomes.

2.1.2.1.1 *Small Cell Lung Cancer.* Small cell lung cancer (SCLC) is a highly malignant type of lung cancer that arises from neuroendocrine cells in the bronchial epithelium. SCLC is known for its aggressive nature and tendency to rapidly spread to other parts of the body (Gazdar, Bunn, and Minna 2017). The majority of SCLC cases are caused by smoking, with up to 98% of patients having a history of smoking (PDQ Adult Treatment Editorial Board 2002). Other risk factors for SCLC include exposure to radon, asbestos, and other environmental toxins (Molina et al. 2008). Mutations in several genes have been linked to the development and progression of SCLC. The most commonly mutated genes in SCLC are TP53, RB1, and PTEN, which are known tumor suppressor genes (George et al. 2015). Other frequently mutated genes in SCLC include MYC, NOTCH, and the SWI/SNF chromatin remodeling complex genes (Rudin and Poirier 2016). SCLC is highly responsive to chemotherapy and radiation therapy, but it often becomes resistant to treatment and relapses quickly. Immunotherapy has shown promising results in the treatment of SCLC, with the PD-1 inhibitor pembrolizumab being approved as a first-line treatment for SCLC in certain cases (Horn et al. 2018).

2.1.2.1.2 *Non-Small Cell Lung Cancer.* Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancer cases and can be further classified into three subtypes: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Siegel, Miller, and Jemal 2020). Adenocarcinoma is the most common subtype of NSCLC and often arises from cells that line the small air sacs (alveoli) in the lungs. It is more likely to occur in people who have never smoked and is often associated with certain genetic mutations, such as EGFR, ALK, and ROS1 (Lichtenfels et al. 2018). Squamous cell carcinoma, also known as epidermoid carcinoma, arises from the cells that line the airways in the lungs. It is strongly associated with tobacco smoking and may be caused by mutations in genes such as TP53, CDKN2A, and PIK3CA. Large cell carcinoma, also known as undifferentiated carcinoma, is a less common subtype of NSCLC and is named for the large size of the tumor cells. It can arise from any part of the lung and is often associated with genetic mutations such as TP53 and KRAS (Siegel et al. 2020). Like SCLC, NSCLC can also be caused by mutations in various genes that are involved in cell growth and division, DNA repair, and cell death. These mutations can occur spontaneously or be caused by exposure to carcinogens such as tobacco smoke, air pollution, and radiation (Lichtenfels et al. 2018). Recent studies have identified RET proto-oncogene as a potential therapeutic target in NSCLC. RET is a receptor tyrosine kinase that plays a crucial role in cell proliferation, differentiation, and survival. RET mutations are found in approximately 1-2% of lung adenocarcinomas and have been shown to drive tumor growth and progression (Gainor and Shaw 2013). Several small molecule inhibitors targeting RET have been developed and are currently undergoing clinical trials in NSCLC. Early results have shown promising activity of these inhibitors in patients with RET-mutant NSCLC, suggesting that RET inhibition may be a viable treatment option for this subset of patients (Drilon et al. 2016, 2018b). In addition to RET mutations, other genetic alterations such as EGFR, ALK, and ROS1 are also being targeted with specific inhibitors in NSCLC (Paik et al. 2011). With the increasing availability of targeted therapies, the molecular classification of NSCLC has become increasingly important for identifying patients who may benefit from these therapies and for designing personalized treatment regimens (Travis et al. 2015).

2.2 Tyrosine Kinase

Tyrosine kinases are a class of enzymes that play a crucial role in cell signaling and are important targets for cancer therapy (Roskoski 2016). They facilitate the transfer of a phosphate group from ATP to a tyrosine residue on a protein substrate, thereby initiating the activation of subsequent signaling pathways that govern cell growth, differentiation, and survival (Lemmon and Schlessinger 2010). Dysregulated activation of tyrosine kinases has been associated with the onset and advancement of various cancer types, such as leukemia, breast cancer, lung cancer, and gastrointestinal stromal tumors (GIST) (Blume-Jensen and Hunter 2001). Mutations or overexpression of tyrosine kinase genes can lead to constitutive activation of downstream signaling pathways, promoting uncontrolled cell proliferation and survival (Weinstein and Joe 2008). Targeting tyrosine kinases has become a major focus of cancer therapy, and several tyrosine kinase inhibitors (TKIs) have been developed and approved for clinical use in various types of cancer. TKIs work by binding to the ATP-binding site of the tyrosine kinase domain, inhibiting kinase activity and downstream signaling (Wong, Siah, and Lo 2019). Examples of FDA-approved TKIs include imatinib for the treatment of chronic myeloid leukemia (CML) and GIST, erlotinib for non-small cell lung cancer (NSCLC), and trastuzumab for HER2-positive breast cancer (Joo, Visintin, and Mor 2013). Despite the clinical success of TKIs, resistance to these agents can develop through various mechanisms, including acquisition of secondary mutations in the target kinase domain or activation of alternative signaling pathways (Johnson et al. 2016). Understanding the underlying molecular mechanisms of resistance and developing strategies to overcome resistance is an active area of research in cancer therapeutics.

Tyrosine kinases are divided into two categories: receptor tyrosine kinase (RTK) and non-receptor tyrosine kinase (NRTK). Receptor tyrosine kinases are transmembrane receptors that are expressed on cell surface. Binding the ligands on their extracellular domain cause to activation of RTKs. On the other hand, non-receptor tyrosine kinases (NRTKs) are placed in cytoplasm or bound to cell membrane (Siveen et al. 2018). RTKs have 58 subfamilies, such as, fibroblast growth factor receptor (FGFR), epidermal growth factor receptor (EGFR), insulin receptor (IR), rearranged during transfection (RET), etc. (Hubbard 1999).

RTKs are activated by ligand-induced dimerization, which occurs when two receptor monomers bind to the same ligand. This binding leads to a conformational change in the receptor that brings the intracellular domains into close proximity, allowing them to auto phosphorylate each other on specific tyrosine residues. This autophosphorylation leads to the recruitment of downstream signaling molecules that bind to the phosphorylated tyrosine residues, leading to the activation of downstream signaling pathways (Lemmon and Schlessinger 2010). The ligand-receptor interaction is highly specific, and each RTK has a unique set of ligands that activate it. For example, the epidermal growth factor receptor (EGFR) is activated by binding to EGF or transforming growth factor-alpha (TGF-α), while the platelet-derived growth factor receptor (PDGFR) is activated by binding to PDGF (Yarden and Sliwkowski 2001). The activated RTKs initiate downstream signaling pathways that regulate various cellular functions. The two major signaling pathways activated by RTKs are the Ras-MAPK pathway and the PI3K-Akt pathway. The Ras-MAPK pathway is activated by RTKs and leads to the activation of the mitogen-activated protein kinase (MAPK) cascade, which ultimately results in the activation of transcription factors that regulate gene expression. The Ras-MAPK pathway is set in motion through the recruitment of the adaptor protein Grb2 to the phosphorylated tyrosine residues on the activated receptor tyrosine kinase (RTK). Grb2 subsequently recruits the guanine nucleotide exchange factor SOS, which triggers the activation of the small GTPase Ras. Activated Ras, in turn, initiates the MAPK cascade, ultimately resulting in the activation of transcription factors like c-Fos and c-Jun (Chang and Karin 2001). The PI3K-Akt pathway is also activated by RTKs and regulates various cellular functions, including cell survival and proliferation. The activation of the PI3K-Akt pathway is initiated by the recruitment of the adaptor protein Gab1 to the phosphorylated tyrosine residues on the activated RTK. Gab1 then recruits the regulatory subunit of PI3K, leading to the activation of PI3K. Activated PI3K then produces the lipid second messenger phosphatidylinositol (Chang and Karin 2001; Manning and Cantley 2007; Mendelsohn and Baselga 2000)-trisphosphate (PIP3), which recruits Akt to the plasma membrane. Akt is then phosphorylated and activated by upstream kinases, leading to the regulation of various downstream effectors that regulate cell survival and proliferation (Manning and Cantley 2007). Dysregulation of RTK signaling pathways has been implicated in various cancers. In many cases, cancer cells overexpress or mutate RTKs, leading to constitutive activation of downstream signaling pathways. For example, the EGFR is

frequently overexpressed in various cancers, including lung, colon, and breast cancer. Overexpression of EGFR leads to constitutive activation of the Ras-MAPK and PI3K-Akt pathways, resulting in increased cell proliferation, survival, and metastasis (Mendelsohn and Baselga 2000). Genes can also lead to constitutive activation of downstream signaling pathways. For example, mutations in the PDGFR gene have been identified in gastrointestinal stromal tumors (GISTs). These mutations give rise to the ligand-independent activation of the PDGFR, causing a persistent activation of the Ras-MAPK and PI3K-Akt pathways (Heinrich et al. 2003). The dysregulation of RTK signaling pathways in cancer has made them attractive targets for therapeutic intervention. Several RTK inhibitors have been developed that target various RTKs, including EGFR, PDGFR, and vascular endothelial growth factor receptor (VEGFR). These inhibitors have shown promising results in clinical trials and have been approved for the treatment of various cancers (Roskoski 2021). For example, the EGFR inhibitor gefitinib has been approved for the treatment of non-small cell lung cancer (NSCLC) with activating mutations in the EGFR gene. Gefitinib targets the intracellular tyrosine kinase domain of EGFR, inhibiting downstream signaling pathways and leading to decreased cell proliferation and survival (Maemondo et al. 2010). Similarly, imatinib, a small molecule inhibitor of PDGFR, has been approved for the treatment of GISTs. Imatinib targets the ATP-binding site of PDGFR, inhibiting its tyrosine kinase activity and downstream signaling pathways (Joensuu 2006).

2.3 Classification of Tyrosine Kinase Inhibitor

Tyrosine kinase inhibitors (TKIs) are a class of drugs that selectively inhibit the activity of tyrosine kinases, which play a key role in various signaling pathways involved in cancer development and progression. TKIs can be classified based on their chemical structure, target specificity, and mechanism of action. The initial generation of TKIs consists of imatinib, which focuses on the BCR-ABL fusion protein in chronic myeloid leukemia (CML), and sunitinib, which targets receptors such as vascular endothelial growth factor receptor (VEGFR), platelet-derived growth factor receptor (PDGFR), and c-Kit in renal cell carcinoma (RCC). These inhibitors are ATP-competitive and bind to the catalytic site of the kinase, preventing ATP from binding and inhibiting kinase activity (Druker et al. 2001; Motzer et al. 2007). The second generation of TKIs includes dasatinib, which targets BCR-ABL and Src kinases in CML, and axitinib, which targets VEGFR in RCC. These inhibitors are also ATP-

competitive but have a higher binding affinity and broader target specificity compared to first-generation inhibitors. Dasatinib has been shown to induce deeper and faster responses than imatinib in CML patients, including those with imatinib-resistant disease (Kantarjian et al. 2006). Axitinib has been shown to significantly prolong progression-free survival and overall survival in patients with advanced RCC who failed first-line therapy with sunitinib or cytokine therapy (Rini et al. 2011). The third generation of TKIs includes ponatinib, which targets BCR-ABL, and several other kinases involved in CML, and osimertinib, which targets epidermal growth factor receptor (EGFR) with the T790M resistance mutation in non-small cell lung cancer (NSCLC). These inhibitors are both irreversible and can overcome resistance to earlier generation inhibitors. Ponatinib has been shown to induce deep and durable responses in heavily pretreated CML patients, including those with BCR-ABL mutations that confer resistance to other TKIs (Cortes et al. 2013). In NSCLC patients with the T790M mutation, osimertinib has demonstrated substantial enhancements in progression-free survival and overall survival when compared to platinum-based chemotherapy (Mok et al. 2017). In addition to these classifications, TKIs can also be categorized based on their target specificity. For example, crizotinib and ceritinib target anaplastic lymphoma kinase (ALK) in NSCLC, and dabrafenib and vemurafenib target BRAF in melanoma. These inhibitors have shown significant clinical benefit in patients with their respective molecular targets (Chapman et al. 2011; Shaw et al. 2013). Despite their efficacy, TKIs can cause various adverse effects, such as diarrhea, nausea, skin rash, and cardiovascular toxicity. In addition, the development of resistance to TKIs is a major challenge in the treatment of cancer, highlighting the need for the development of new and improved inhibitors.

2.4 Rearranged During Transfection (RET) Kinase

RET (rearranged during transfection) is a proto-oncogene that encodes a receptor tyrosine kinase (RTK) is located on chromosome 10q11.2 which is a pericentromeric region and it is involved in various signaling pathways, including cell survival, proliferation, and differentiation. RET was first identified as a fusion gene with the H4 gene in a thyroid cancer sample in 1985 (Takahashi, Ritz, and Cooper 1985; Wells and Santoro 2009). Since then, RET mutations and rearrangements have been identified in various cancer types, including papillary and medullary thyroid carcinoma, lung adenocarcinoma, and colorectal cancer (V15Subbiah et al. 2018).

The RET protein is a receptor tyrosine kinase situated across the cell membrane, featuring distinct components including an extracellular domain, a single transmembrane domain, and an intracellular domain. Within the extracellular domain, there are four cadherin-like repeats and a cysteine-rich domain, responsible for ligand binding, receptor dimerization, and subsequent activation, as depicted in Figure 1. The intracellular domain contains a juxtamembrane domain, a catalytic domain, and a C-terminal tail, which are responsible for signal transduction and downstream signaling pathways (Mulligan 2014; Plaza-Menacho, Mologni, and McDonald 2014).

The crystal structure of the RET extracellular domain has been determined in complex with glial cell line-derived neurotrophic factor (GDNF), its cognate ligand, and the GDNF family receptor alpha 1 (GFRα1), a co-receptor that enhances ligand binding and specificity. The complex structure reveals a dimeric arrangement of the RET extracellular domains, with GDNF and GFRα1 binding at the interface between the two monomers. The structure also provides insights into the mechanism of ligand-induced dimerization and activation of the receptor (Lu et al. 2015; Parkash et al. 2008).

The crystal structure of the RET kinase domain has also been determined in complex with ATP and a small molecule inhibitor, revealing the active conformation of the kinase and the binding site of the inhibitor. The structure provides insights into the mechanism of kinase activation and inhibition, as well as the design of selective and potent RET inhibitors for cancer therapy (Chen et al. 2016; DeLano et al. 2000).

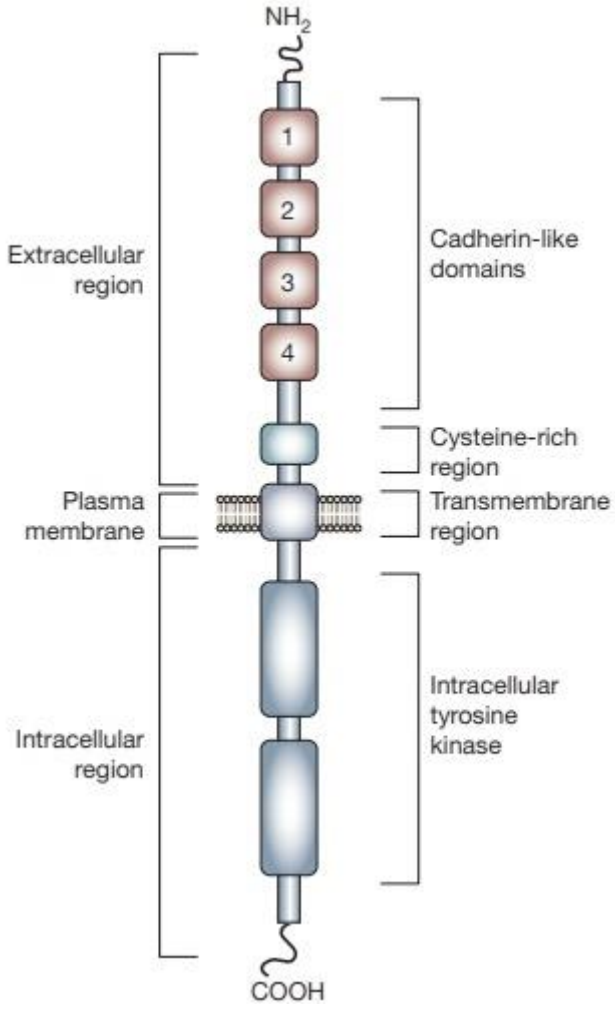


Figure 1. The RET receptor tyrosine kinase's domains

The RET receptor tyrosine kinase consists of an arrangement comprising an extracellular domain encompassing four cadherin-like domains and a cysteine-rich region, a transmembrane region spanning the plasma membrane, and an intracellular domain that houses a substantial tyrosine kinase domain

(Drosten and Pützer 2006).

The activation of RET entails the interaction between its extracellular domain and ligands, including glial cell line-derived neurotrophic factor (GDNF) and other ligands from the GDNF family. This interaction is facilitated by the presence of Ca^{2+} ions (Wells and Santoro 2009). The ligand binding induces dimerization of RET with its co-receptors, the GDNF family receptor alpha (GFRα) proteins, leading to autophosphorylation of the intracellular tyrosine residues in the RET kinase domain. GDNF, Neurturin, Artemin and Persephin are the four glial derived neurotrophic factor ligands which are bound to the RET in conjunction with GFRα1, GFRα2, GFRα3,

GFRα4 respectively and generate a ternary complex which is an important complex for RET signaling and it shown in Figure 2. and Figure 3. This GFL-GFRα-RET complex is caused to dimerization and activation of the RET tyrosine kinase (De Falco et al. 2017; Wells and Santoro 2009).

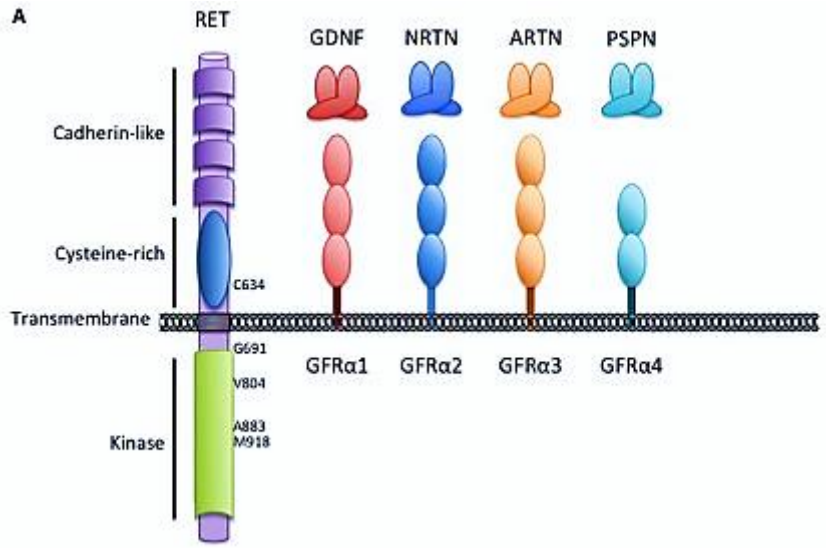


Figure 2. The RET receptor tyrosine kinase, its interaction with GDNF family ligands and coreceptors

The diagram illustrates the RET receptor tyrosine kinase, its engagement with GDNF family ligands, and the involvement of vital functional domains. This includes the transmembrane receptor function of the four GFLs (glial cell line-derived neurotrophic factor family) and the binding of GFLs to RET via interactions with cell surface coreceptors from the GFRα family

(Mulligan 2018).

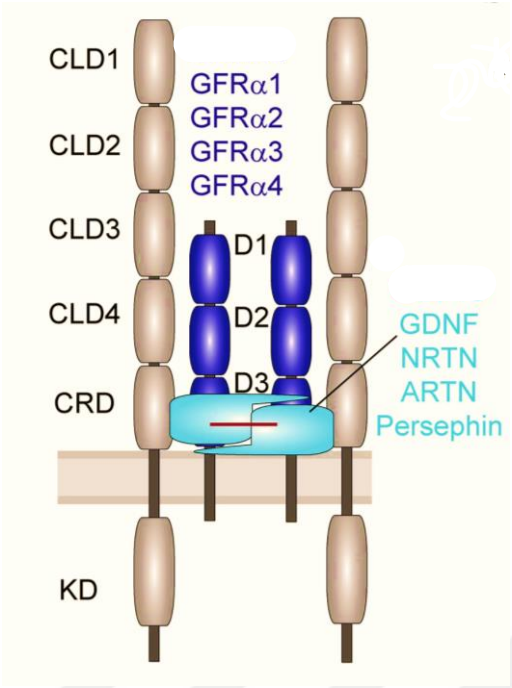


Figure 3. The domain structure of the RET receptor–coreceptor–ligand complex

The provided figure showcases the domain arrangement of the 2/2/2 complex involving the RET receptor, coreceptor, and ligand. The receptor, depicted in gray, exhibits its specific structural domains. The coreceptors, GFRAL and GFRα1–4, are displayed in blue, while the ligands GDF15, GDNF, NRTN, ARTN, and persephin are depicted in cyan

(Trenker and Jura 2020).

Autophosphorylation of the tyrosine residues in the kinase domain activates the RET receptor by inducing a conformational change that exposes the docking sites for downstream signaling molecules, such as the adaptors Shc, Grb2, and Gab1, and the effector enzymes PI3K, PLCγ, and Src. These signaling molecules are recruited to the activated RET receptor and initiate downstream signaling pathways, including the RAS/RAF/MAPK and PI3K/AKT pathways, which regulate cell growth, survival, differentiation, and migration (Mulligan 2014; Plaza-Menacho 2018).

The activation of RET is tightly regulated by multiple mechanisms, including the interaction with its ligands and co-receptors, the activity of its negative regulators, such as protein tyrosine phosphatases (PTPs) and Sprouty proteins, and the formation of intracellular protein complexes that modulate the signaling output. Dysregulation of RET signaling, due to genetic mutations, overexpression, or aberrant activation of its downstream effectors, has been implicated in various human diseases, including

cancer and neurodegenerative disorders (Plaza-Menacho et al. 2007; Trenker and Jura 2020).

Phosphorylation of RET is a key post-translational modification that regulates the activity of the receptor and downstream signaling pathways. The intracellular domain of RET contains multiple tyrosine residues that can be phosphorylated by various tyrosine kinases, including RET itself (autophosphorylation) and non-RET kinases such as c-Src and focal adhesion kinase (FAK) (Mulligan 2014; Plaza-Menacho 2018).

Phosphorylation of specific tyrosine residues in RET creates docking sites for downstream signaling molecules, such as the adaptor protein SHC, the docking protein GRB2, and the cytoplasmic protein tyrosine phosphatase SHP2. These molecules can activate downstream signaling pathways, including the RAS-MAPK and PI3K-AKT pathways, which are involved in cell proliferation, survival, and differentiation (Mulligan 2019).

Phosphorylation of other tyrosine residues in RET can also create negative feedback loops that downregulate the activity of the receptor and prevent excessive signaling. For example, phosphorylation of tyrosine 981 in RET creates a binding site for the adaptor protein NCK, which can recruit the E3 ubiquitin ligase CBL to the receptor and induce its degradation (5,6). Also, autophosphorylation of tyrosine 1062 in the c-tail of the RET protein plays a crucial role in signaling. This process facilitates the recruitment of intracellular adaptors like SHC, FRS2, and IR1/2. Consequently, RET activates the RAS-MAPK and PI3K-AKT-mTOR cascades, which serve as important pathways for signal transduction (De Falco et al. 2017).

In addition to tyrosine phosphorylation, RET can also be phosphorylated on serine and threonine residues by various serine/threonine kinases, including PKC, PKA, and GSK3. These phosphorylation events can also regulate the activity of the receptor and downstream signaling pathways (Östman and Böhmer 2001).

Overall, phosphorylation of RET is a dynamic and complex process that plays a critical role in the regulation of cellular signaling and physiological processes. Dysregulation of RET phosphorylation has been implicated in various diseases, including cancer and neurodegenerative disorders.

In the thyroid gland, RET is essential for the development and function of the parafollicular C-cells, which produce calcitonin, a hormone that regulates calcium metabolism. Activating mutations in RET are the major cause of hereditary medullary

thyroid carcinoma (MTC), a rare form of thyroid cancer that arises from the C-cells. In sporadic MTC, somatic mutations in RET are also frequently found, indicating that RET is a key driver of this cancer type (Mulligan 2014).

In the lung, RET has been implicated in the development of non-small cell lung cancer (NSCLC), the most common form of lung cancer. Rearrangements of the RET gene, which result in fusion with other genes such as KIF5B or CCDC6, are found in a small subset of NSCLC patients, and RET inhibitors have shown promising results in preclinical and clinical studies (Kumi Kawai 2020).

Furthermore, within the nervous system, RET is expressed in diverse neuronal populations and plays a crucial role in the formation and sustenance of the enteric and sympathetic nervous systems, along with the central nervous system. Loss-of-function mutations in RET are the cause of Hirschsprung disease, a congenital disorder characterized by the absence of enteric ganglia in the distal bowel, which leads to severe constipation and intestinal obstruction. Moreover, RET has been linked to the development of various neurodegenerative disorders, such as Parkinson's disease and Alzheimer's disease, although the precise mechanisms involved are not yet comprehensively understood (Jing et al. 1996). Due to (Kramer and Liss 2015) RET plays a crucial role in the survival of dopamine neurons and provides neuroprotection, particularly when challenged by neurotoxins. On the other hand, Parkinson’s disease adversely affects dopamine neurons, leading to a decline in dopamine levels. Although GFL have shown promise in promoting the survival of midbrain dopamine neurons, their effectiveness in clinical trials with PD patients has been limited. Consequently, targeting RET presents an opportunity for PD treatment. The main challenge lies in delivering GFL to the brain of patients, which can be addressed by developing small molecules that specifically target RET and possess favorable pharmacodynamic properties to penetrate the blood-brain barrier. Such an approach represents a disease-modifying strategy that could enable the inclusion of early-stage PD patients in clinical trials.

Mutations in the RET gene can lead to various diseases, including cancer and inherited disorders. In cancer, RET mutations are found in multiple types of tumors, including papillary thyroid carcinoma (PTC), lung cancer, and multiple endocrine neoplasia type 2 (MEN2) syndromes. These mutations can result in constitutive activation of the RET protein, leading to aberrant signaling and tumor growth (Mulligan 2014).

In PTC, the most common RET mutation is a point mutation in codon 634, which accounts for approximately 90% of all RET mutations in PTC. This mutation leads to a single amino acid substitution of cysteine to arginine (C634R) in the extracellular domain of the RET protein, resulting in ligand-independent dimerization and activation of the receptor (Marsh, Learoyd, and Robinson 1995; Romei et al. 2016).
In MEN2 syndromes, RET mutations are inherited in an autosomal dominant manner and can lead to the development of medullary thyroid carcinoma (MTC), a rare and aggressive form of thyroid cancer. The mutations in MEN2 are clustered in specific regions of the RET gene, including codons 609, 611, 618, 620, 634, and 918, and can result in constitutive activation of the RET protein and downstream signaling pathways (Wells et al. 2015b).

Several RET inhibitors have been developed to target RET-driven cancers, such as vandetanib and cabozantinib for MTC and selpercatinib for RET fusion-positive cancers. However, the development of resistance to these inhibitors is a major challenge, highlighting the need for the development of new and improved therapies for RET-mutant cancers (Drilon et al. 2018a; Vivek Subbiah et al. 2018).

2.5 Acquired resistance.

Acquired resistance has been a major challenge in the clinical use of RET kinase inhibitors in the treatment of medullary thyroid cancer (MTC) and non-small cell lung cancer (NSCLC). The development of acquired resistance can ultimately lead to disease progression and relapse in patients. This resistance can either be primary, where the tumor does not respond to the treatment due to intrinsic or patient-specific factors, or acquired, where the cancer obtains the ability to resist the activity of an inhibitor to which it was previously susceptible (Drilon et al. 2018a).

There are several mechanisms of acquired resistance in RET inhibition, including on-target resistance, pathway-driven resistance, and alternative pathway-driven resistance. On-target resistance can occur due to target amplification or secondary resistance mutations, which decrease the affinity of the inhibitor to the target. Pathway-driven resistance can occur when mutations downstream of the RET signaling pathway cause constitutive activation of the pathway. Alternative pathway-driven resistance can occur when alternative RTKs are activated in tumor cells to bypass the inhibition and confer acquired resistance (Zhang et al. 2022).

In MTC, acquired resistance to RET inhibition is a significant clinical issue, with studies identifying various mechanisms of resistance, such as mutations in the RET kinase domain and activation of alternative pathways, including the PI3K/Akt/mTOR pathway. In NSCLC, RET fusions are the primary target for RET inhibition, with acquired resistance occurring due to mechanisms such as secondary mutations in the RET kinase domain and activation of alternative signaling pathways, including the EGFR and MET pathways (Pan et al. 2021).

Acquired resistance typically involves modifying the RET kinase, the target molecule, or acquiring mutations that allow for bypass signaling. Within the RET kinase domain, specific single amino acid changes have been identified as contributors to target-mediated resistance against various tyrosine kinase inhibitors (TKIs). These changes occur in the hinge segment of the RET kinase, with mutations found in RET V804, Y806, and G810, as well as in the activation segment with S904. Notably, RET kinase V804 plays a crucial role in what is known as the 'gatekeeper' position, controlling the accessibility of the hydrophobic ATP-binding and drug-binding pocket. This positioning can potentially impact the binding affinity of the RET kinase for ATP and most frequent single mutations in this residue which are identified in patients with MTC and NSCLC, is mutation of valine to methionine or leucine (V804M/L). other mutations that mostly caused resistance to selective RET inhibitors are G810S located at floor of solvent-front, Y806C/N mutation located at hinge, V738A mutation located at β2 strand and L730V/I mutations located at roof of the solvent-front site of the RET proto-oncogene (Salvatore, Santoro, and Schlumberger 2021).

Several studies have identified the presence of specific mutations in medullary thyroid cancer (MTC) patients. These include the M918T mutation as well as double mutations such as V804M/G810S, V804M/Y806C, and Y806C/V738A (Liu et al. 2020; Xia and Ou 2020).

To overcome acquired resistance in RET inhibition, several strategies have been proposed, including change of doses and schedules, combination therapy, and development of new inhibitors. Combination therapy can be used with other rational inhibitors or with traditional chemotherapy drugs, targeting and stopping the bypass tracks. Developing novel potent inhibitors with higher specificity for mutant types is also a potential strategy to overcome on-target driven resistance mechanisms (Casals et al. 2017).

2.6 RET Kinase Inhibitors

Inhibitors are molecules that can bind to a protein or enzyme and decrease its activity. In the case of RET kinase inhibitors, these molecules are designed to bind specifically to the ATP-binding pocket of the RET kinase domain, which is necessary for its activity (Vodopivec and Hu 2022). By binding to this pocket, the inhibitor can prevent ATP from binding to the kinase, leading to inhibition of its activity and downstream signaling pathways. There are several classes of RET kinase inhibitors, including multikinase inhibitors that target multiple receptor tyrosine kinases, and selective inhibitors that target RET specifically. These inhibitors have shown promise in preclinical and clinical studies for the treatment of RET-driven cancers, including lung cancer and medullary thyroid cancer (Li et al. 2023; Santoro et al. 2020). The FDA has approved two selective RET inhibitors for the treatment of RET-driven cancers: selpercatinib (Retevmo) and pralsetinib (Gavreto).

2.6.1 **Selpercatinib.** Selpercatinib is a small molecule inhibitor designed to target the ATP-binding pocket of the RET kinase domain. Its chemical structure consists of a pyrrolopyridine core with multiple substitutions and functional groups, which are responsible for its binding specificity and potency (Drilon et al. 2018a).

Selpercatinib, marketed as Retevmo, is a targeted RET kinase inhibitor that obtained approval from the U.S. Food and Drug Administration (FDA) in May 2020. It is indicated for the treatment of RET fusion-positive non-small cell lung cancer (NSCLC), RET-mutant medullary thyroid cancer (MTC), and RET fusion-positive thyroid cancer. Selpercatinib was developed based on the structure of the RET kinase domain and its interaction with ATP and inhibitors. The molecule was designed to bind specifically to the ATP-binding pocket of RET kinase and inhibit its activity, leading to the inhibition of downstream signaling pathways (FDA 2020b). In preclinical and clinical studies, selpercatinib has demonstrated potent activity against RET fusion-positive and mutant tumors, including NSCLC, MTC, and thyroid cancer, with high response rates and durable responses. Selpercatinib was approved based on the results of two Phase I/II clinical trials, LIBRETTO-001 and LIBRETTO-531, which evaluated the efficacy and safety of the drug in patients with advanced RET fusion-positive NSCLC and RET-mutant MTC, respectively. The studies showed that selpercatinib had significant clinical activity and an acceptable safety profile in these patient populations (Drilon et al. 2018a). In addition to its use in RET fusion-positive

NSCLC and MTC, selpercatinib is also being evaluated in other RET-driven cancers, including papillary thyroid cancer, solid tumors with RET gene fusions, and other advanced solid tumors (Vivek Subbiah et al. 2018).

Selpercatinib has shown remarkable clinical activity against RET fusion-positive non-small cell lung cancer (NSCLC) and thyroid cancer. However, resistance to selpercatinib has been observed in some patients, particularly those with acquired mutations in the RET kinase domain, such as gatekeeper mutations V804L and G810R. These mutations are known to confer resistance to multiple kinase inhibitors by hindering drug binding to the ATP-binding pocket. Studies have shown that selpercatinib can still exhibit partial inhibitory activity against these mutations, albeit at higher concentrations. The drug is thought to achieve this by binding it to an allosteric site near the kinase domain, which stabilizes the inactive conformation of the RET kinase and reduces its activity even in the presence of activating mutations (Rosen et al. 2021).

One study demonstrated that in vitro treatment with selpercatinib could effectively suppress RET signaling and induce apoptosis in cell lines expressing gatekeeper mutations, albeit at higher concentrations than those required for wild-type RET. Another study showed that in patients with RET fusion-positive cancers who developed acquired resistance to selpercatinib, combination therapy with a second-generation RET inhibitor, pralsetinib, could overcome the resistance and achieve durable responses (Shabbir et al. 2023).

2.6.2 **Pralsetinib.** Gavreto, also known as pralsetinib, is an additional selective RET kinase inhibitor that has received FDA approval for the treatment of cancers driven by RET alterations. Developed by Blueprint Medicines, it was granted accelerated approval in September 2020 for the management of advanced or metastatic RET-mutant medullary thyroid cancer and RET fusion-positive thyroid cancer (GAVRETOTM (pralsetinib) Prescribing Information (U.S.) 2020).

Like selpercatinib, pralsetinib also targets the ATP-binding pocket of the RET kinase domain and inhibits its activity. It has demonstrated potent and selective inhibition of RET in preclinical studies and has shown promising clinical activity in patients with RET-driven cancers, including lung cancer, thyroid cancer, and other solid tumors.

The structure of pralsetinib is similar to that of selpercatinib, with a pyrazole-pyridine scaffold and a 3,5-dimethylpiperazine group for binding to the ATP-binding pocket of the RET kinase domain. However, there are differences in the chemical substitutions and modifications, which may contribute to differences in their pharmacological and therapeutic properties (V. Subbiah et al. 2018).

Pralsetinib, also known as BLU-667, is a selective RET kinase inhibitor that has shown promising results in preclinical and clinical studies for the treatment of RET-driven cancers, particularly non-small cell lung cancer and medullary thyroid cancer (Griesinger et al. 2022; Taylor et al. 2019). Like other RET inhibitors, pralsetinib binds to the ATP-binding pocket of the RET kinase domain, preventing the binding of ATP and subsequent activation of downstream signaling pathways (Wang 2013). In Figure 4. it shown that Pralsetinib bind to the ATP binding site of RET protein.

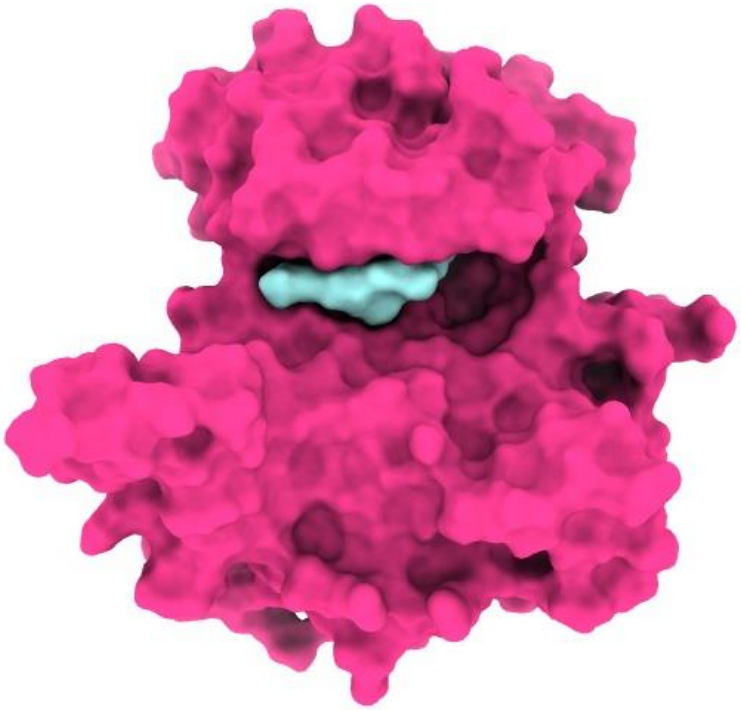


Figure 4. Surface view of the Pralsetinib in the ATP binding site of RET protein.

Pralsetinib was developed using a structure-based drug design approach, in which the crystal structure of the RET kinase domain was used to guide the design of small molecule inhibitors that can bind specifically to the ATP-binding pocket of the kinase (Knowles et al. 2006; Luo et al. 2021). Preclinical studies have shown that pralsetinib has potent inhibitory activity against several RET fusion variants and mutations, including those that are resistant to other RET inhibitors such as vandetanib and cabozantinib (Drilon et al. 2013).

In a phase I/II clinical trial, pralsetinib demonstrated high response rates and durable clinical benefits in patients with RET fusion-positive non-small cell lung cancer and medullary thyroid cancer, with manageable side effects (Drilon et al. 2019; Griesinger et al. 2022). Based on these results, pralsetinib was granted accelerated approval by the U.S. Food and Drug Administration (FDA) in 2020 for the treatment of adult patients with metastatic RET fusion-positive non-small cell lung cancer (FDA 2020a).

Pralsetinib has also shown activity against RET fusion proteins and mutations that confer resistance to other RET inhibitors, such as selpercatinib. However, some mutations, such as RET V738A, Y806C/N, G810S and L730V/I have been shown to confer resistance to pralsetinib as well (the L730V/I RET roof mutations display different activities toward pralsetinib an selpercatinib) (Drilon et al. 2018c; Vivek Subbiah et al. 2018). Pralsetinib acts as an ATP-competitive inhibitor by binding to the ATP-binding pocket of the RET kinase domain, thereby preventing ATP from binding and inhibiting downstream signaling pathways (Li et al. 2023). Resistance to pralsetinib can occur through mutations in the RET kinase domain that alter the conformation of the ATP-binding pocket or increase its affinity for ATP, thereby reducing the binding of the inhibitor (V. Subbiah et al. 2018).

Chapter 3
Theoretical Background

3.1 Quantitative Structure – Activity Relationship (QSAR)

Quantitative structure-activity relationship (QSAR) is a computational technique used in drug discovery and development to predict the biological activity of new compounds based on their structural properties. The first QSAR study was conducted in 1868 by Crum-Brown and Fraser, who established a correlation between the chemical structures of a series of barbiturates and their sedative effects (Brown and Fraser 1868). Since then, QSAR has been extensively used in drug design to predict the biological activity, toxicity, and other properties of novel compounds.

QSAR is based on the assumption that the biological activity of a compound is related to its physicochemical properties and structural features. The goal of QSAR is to develop a mathematical model that can predict the biological activity of new compounds based on their structural and physicochemical properties (Arthur, Ejeh, and Uzairu 2020). To achieve this, QSAR models are trained using a set of compounds with known biological activity, and their physicochemical and structural properties are characterized using various descriptors such as molecular weight, solubility, and lipophilicity. These descriptors are then used to generate a mathematical model that can predict the biological activity of new compounds.

QSAR has several advantages over traditional drug discovery methods. It can significantly reduce the time and cost required to develop new drugs by allowing researchers to predict the biological activity of novel compounds before they are synthesized and tested in the laboratory (Fujita 1995). Additionally, QSAR models can be used to optimize the properties of existing drugs, such as improving their selectivity and reducing their toxicity.

3.2 Structure-Based Pharmacophore Modeling

Structure-based pharmacophore is a computational technique used in drug discovery to identify the essential chemical features required for ligand binding to a target protein. The process of pharmacophore generation involves the identification of ligand-protein interactions, which are then translated into a three-dimensional (3D) pharmacophore model (Jain 1996). This model represents the spatial arrangement of

chemical features, including hydrogen bond donors, hydrogen bond acceptors, hydrophobic regions, and ionic interactions, that are critical for ligand binding to the target protein. Structure-based pharmacophore generation is based on the hypothesis that structurally diverse ligands that bind to the same protein share common chemical features that are essential for binding (Wolber and Langer 2005).

The pharmacophore hypothesis can be created by analyzing a set of structurally diverse ligands that bind to the target protein. By identifying the common chemical features present in these ligands, a 3D pharmacophore model can be generated that represents the essential chemical features required for ligand binding to the target protein. This model can then be used to screen large databases of compounds to identify potential lead compounds with similar chemical features (Jain 1996).

To evaluate the fitness of a compound to the pharmacophore model, a scoring function is used. The scoring function is a mathematical approach that quantifies the fit between the pharmacophore model and the ligand. The fitness score is calculated based on the interactions between the ligand and the pharmacophore model, such as the distance between the ligand and the pharmacophore features, the angles between the ligand and the pharmacophore features, and the orientation of the ligand within the pharmacophore model (Wang, Fu, and Lai 1997).

The fitness score is calculated using a mathematical formula that takes into account the degree of match between the ligand and the pharmacophore features. In pharmacophore modeling, the fitness score (FS) is calculated based on the following equation:

$$\boldsymbol{FS} = \boldsymbol{n} + (\boldsymbol{n_p} \times \boldsymbol{w_p}) + (\boldsymbol{n_{hb}} \times \boldsymbol{w_{hb}}) + (\boldsymbol{n_r} \times \boldsymbol{w_r}) \qquad \textit{Equation 3. 1}$$

Where n is the number of matching features, $\boldsymbol{n_p}$ is the number of matched pharmacophore points, $\boldsymbol{w_p}$ is the weight assigned to the pharmacophore points, $\boldsymbol{n_{hb}}$ is the number of matched hydrogen bond acceptors and donors, $\boldsymbol{w_{hb}}$ is the weight assigned to hydrogen bonds, $\boldsymbol{n_r}$ is the number of rotatable bonds in the ligand, and $\boldsymbol{w_r}$ is the weight assigned to rotatable bonds. This formula allows for a quantitative evaluation of the degree of fit between a ligand and a pharmacophore, which can be used to prioritize or filter large virtual compound libraries and to predict the biological activity of compounds (Jain 1996).

An e-pharmacophore, or energy-optimized structure-based pharmacophore, is a computational method used in drug discovery. It has proven to offer superior virtual screening speed and enrichments when compared to molecular docking utilizing Glide's "standard precision" (SP) scoring function (Moumbock et al. 2021).

3.3 Z-Score

The Z-score is a statistical measure that evaluates the deviation of a value from the mean in terms of standard deviation. It is calculated by subtracting the mean from the value of interest and dividing it by the standard deviation:

$$\boldsymbol{Z - score} = (\boldsymbol{value} - \boldsymbol{mean}) \,/\, \boldsymbol{standard\ deviation} \qquad \textit{Equation 3. 2}$$

In drug discovery, the Z-score is used to evaluate the potency of a compound relative to a reference dataset. It allows researchers to compare the activity of a compound against the activity distribution of a large number of compounds. The Z-score is particularly useful when working with datasets that have a non-normal distribution, as it provides a standardized measure of deviation that is not affected by the distribution's shape.

In drug discovery, the Z-score is used in a variety of applications, such as virtual screening, hit identification, and lead optimization. It is commonly used to evaluate the potency of a compound in vitro and in vivo, as well as its pharmacokinetic and pharmacodynamic properties (Andrade 2021).

3.4 METACORE™/METADRUG™

The METACORE™/METADRUG™ tool offered by Clarivate Analytics is a popular online platform in the field of drug discovery that provides an extensive profile of pharmacokinetic and pharmacodynamic properties for a selected compound or group of compounds. By utilizing 25 common diseases binary QSAR models and 26 toxicities binary QSAR models, the Tanimoto Prioritization (TP) value of a compound can be determined based on its similarity to the training and test sets of the QSAR models. The primary objective of this platform is to predict the major metabolites of an input structure and their impact on the key enzymes of the human body (Ekins et al. 2006). The METACORE™/METADRUG™ platform utilizes QSAR to forecast

the pharmacokinetic and pharmacodynamic characterization of small molecules. The Cooper statistical parameters, including sensitivity, specificity, accuracy, and Matthews Correlation Coefficient (MCC), are used to verify the effectiveness of the QSAR models in this platform. These parameters are used to evaluate the fitness, robustness, and predictive power of the QSAR models (Dogan and Durdagi 2021). The MCC is a correlation coefficient between experimental values and predicted values, while sensitivity measures correctly predicted positives, specificity measures correctly predicted negatives, and accuracy measures how close the model can get to the true value. Overall, the METACORE™/METADRUG™ platform is a reliable tool for pharmacokinetic and pharmacodynamic prediction and evaluation in drug discovery research (Myshkin et al. 2012).

3.5 Molecular Docking

Molecular docking is a computational method used to study the binding between a protein receptor and a small molecule ligand. The main aim of molecular docking is to predict the binding affinity and orientation of a ligand to its target protein receptor, which can then be used to design novel drugs or optimize existing ones. The method involves generating multiple conformations of the ligand and protein receptor and then evaluating their binding energy scores to obtain the most favorable conformation.

The scoring functions used in molecular docking are typically based on the principles of thermodynamics and include empirical, physics-based, and hybrid approaches. One commonly used scoring function is the empirical scoring function, which assigns weights to various factors such as hydrogen bonding, van der Waals interactions, and electrostatics. The most popular empirical scoring function is AutoDock, which uses a combination of Lamarckian genetic algorithm and a local search algorithm for docking simulations (Morris et al. 1639).

Another scoring function used in molecular docking is the physics-based force field approach, which uses the principles of molecular mechanics to calculate the binding energy of the protein-ligand complex. This approach is more accurate than the empirical scoring function but requires significant computational resources.

In recent years, machine learning algorithms have been applied to improve the accuracy of molecular docking. These methods use large datasets to train a predictive model that can estimate the binding affinity of new ligands to a protein receptor. One

such method is DeepDock, which uses a deep neural network to predict the binding affinity of protein-ligand complexes (Jiménez-Luna et al. 2020).

Molecular docking is a computational technique used to determine the structure of a complex formed between a drug candidate and the target molecule, which is referred to as a docking pose. In many molecular docking simulations, the ligand is allowed to have complete conformational flexibility, while the target is restricted to a few flexible residues at the binding site or is considered entirely rigid to minimize computational resources. Multiple docking poses can be generated by considering different orientations of the ligand within the binding site, which necessitates the use of a score to rank the different docking poses. This score, known as the docking score, is computed based on the binding free energy (ΔG_{bind}) but is subject to certain assumptions to accelerate computer calculations (Suhandi et al. 2021).

$$\boldsymbol{\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}} \qquad \textit{Equation 3. 3}$$

The formula used in the calculation of docking scores involves the enthalpy change ($\boldsymbol{\Delta H_{bind}}$) and entropy change ($\boldsymbol{\Delta S_{bind}}$) due to binding, along with the temperature (T). However, during this process, it is assumed that the binding of a ligand to the target does not result in a significant change in entropy. Therefore, the calculation of $\boldsymbol{\Delta G_{bind}}$ could be accomplished through the direct computation of $\boldsymbol{\Delta H_{bind}}$. To compute the enthalpy change, one can determine the change in internal energy, $\boldsymbol{\Delta U_{bind}}$, and account for the effects of pressure, $\boldsymbol{P}$, and volume changes, $\boldsymbol{\Delta V}$, using the equation:

$$\boldsymbol{\Delta H_{bind} = \Delta U_{bind} - P\Delta V} \qquad \textit{Equation 3. 4}$$

In molecular docking, it is generally assumed that when a smaller ligand binds to a protein, it does not have a significant effect on the volume of the protein. This means that the second term in equation 3.4, which accounts for changes in volume, can be disregarded, and the binding free energy $\Delta\boldsymbol{G_{bind}}$ can be assumed to be equal to the enthalpy change $\boldsymbol{\Delta H_{bind}}$. Therefore, the internal energy changes due to binding,

$\boldsymbol{\Delta U_{bind}}$, can be used to directly calculate $\Delta \boldsymbol{G_{bind}}$, and is typically employed in the calculation of docking scores.

$$\Delta \boldsymbol{G_{bind}} = \boldsymbol{\Delta G_0} + \boldsymbol{\Delta G_{h-bond}} \sum_{\boldsymbol{h-bond}} \boldsymbol{f}(\Delta \boldsymbol{r}, \Delta \boldsymbol{\alpha}) + \Delta \boldsymbol{G_{ionic}} \sum_{\boldsymbol{ionic}} \boldsymbol{f}(\Delta \boldsymbol{r}, \Delta \boldsymbol{\alpha}) + \Delta \boldsymbol{G_{arom}} \sum_{\boldsymbol{aromatic}} \boldsymbol{f}(\Delta \boldsymbol{r}, \Delta \boldsymbol{\alpha}) + \Delta \boldsymbol{G_{lipo}} \sum_{\boldsymbol{lipo}} \left| \boldsymbol{A_{lipo}} \right| + \Delta \boldsymbol{G_{rot}} \boldsymbol{N_{rot}} \quad \textit{Equation 3. 5}$$

Equation 3.5 comprises a constant term $\boldsymbol{\Delta G_0}$ that does not depend on the system and several terms that correspond to ideal hydrogen bonds, ionic interactions, aromatic interactions, and lipophilic interactions. These terms are multiplied by a penalty function f $(\Delta \boldsymbol{r}, \Delta \boldsymbol{\alpha})$. The term $\Delta \boldsymbol{G_{rot}}$ is multiplied by $\boldsymbol{N_{rot}}$, which is the count of rotatable bonds in the ligand. This multiplication accounts for the decrease in free energy when a rotatable bond in the ligand is assumed to be fixed during binding. This approach was described by (Christopher J. 2004).

In Maestro, various docking algorithms such as Glide / HTVS, Glide / SP, and Glide / XP were used for molecular docking analysis (Halgren et al. 2004). The Glide/HTVS algorithm, which does not have advanced settings, provides a practical docking method for scanning libraries containing a large number of molecules. Since scanning a large number of molecules in docking simulations using advanced settings and overly sensitive algorithms is costly and time-consuming, first scanning with Glide/HTVS provides practical and convenient results for the rough filtering of large molecule libraries. Glide/HTVS is a method that contributes to a significant reduction in the cost of drug development by enabling the rapid screening of the biological activity of large chemical libraries. It is a preferred computational strategy to accelerate structure-based drug design studies (Halgren et al. 2004).

The Glide/SP algorithm offers more advanced settings and improved parameters compared to Glide/HTVS. To reduce errors in Coulomb and vows interactions, net ionic charge interactions on charged groups such as carboxylates and guanidiniums have been reorganized. Interactions energies have been improved by reducing the interaction energies of directly contributing atoms, thereby reducing inequalities in interaction energies. The interaction energy between two charged groups has also been improved to some extent in the Glide/SP algorithm. The most significant variable introduced in the Glide/SP algorithm is the solvent model component. If charged and polar groups of protein and ligand molecules are not properly soluble, serious

constraints occur in ligand poses. In this case, the physical effects of water molecule interactions are neglected, making the complex physically unrealistic. The accessibility of charged groups to solvent must be evaluated very carefully, and the trapping of water molecules in hydrophobic pockets by the ligand can prevent interactions and should be avoided. The location and surroundings of water molecules in the protein active site when the ligand is bound are very important, but existing continuity solvent models have difficulty capturing these details. To prevent these errors, open water regions are placed for each ligand pose, and exposure of various groups to open water is measured using empirical scoring terms and solvation effects are also included in the calculations. The open water approach employed in Glide/SP reliably rejects the majority of false positives that arise in any empirical docking calculation. Although Glide/SP is a more precisely developed algorithm than Glide/HTVS, it has been developed relatively less precisely than Glide/XP, which is required to scan a large number of molecules. Therefore, some false positives are relatively ignored in Glide/XP due to the need to consider the algorithm's speed, which enables the safe virtual screening of tens or hundreds of thousands of ligands. For example, errors due to solubility in Glide/SP are evaluated with fewer penalty coefficients than in Glide/XP.

The Glide/XP algorithm differs from Glide/SP in that it uses a more comprehensive funnel to achieve binding structures with greater differences. Glide/SP is utilized to initiate processing with binding, while Glide/XP utilizes various parts of the molecule as anchor points to generate better scoring by considering each anchor point. In the Glide/XP protocol, the explicit water model used in Glide/SP is further developed to better evaluate serious physical principal violations during binding and to provide higher penalties. GlideScore XP particularly scores the occupancy rate of hydrophobic pockets that can interact with the hydrophobic groups of the ligand. The most important physical effects that are against binding are ligand, protein, or both's strain energy, loss of entropy of ligand and protein, and insolubility (Friesner et al. 2006).

3.6 Molecular Dynamics (MD) Simulations

MD simulation is a crucial tool in the field of in-silico study of biological systems. This approach involves a time-dependent treatment of an atomic system to

provide valuable insights into the stability and conformational changes of proteins and nucleic acids. The origins of molecular dynamics can be traced back to the late 1950s when Alder and Wainwright used this approach to study hard spheres (Alder and Wainwright 1959). The next significant advancement came in 1964 when Rahman introduced the use of force fields in MD simulations. The first application of MD simulations to study proteins was conducted by Mc Cammon and his colleagues in their work on bovine pancreatic trypsin inhibitors (McCammon, Gelin, and Karplus 1977).

MD simulations have emerged as a popular method to investigate the conformation, mobility, and thermodynamics of large biological systems. Additionally, MD methods are employed in crystallography and NMR tools to enhance the quality of the structure. In the pursuit of developing novel drugs, MD simulations are particularly useful in elucidating ligand-protein complexes. The simulations are based on Newton's second law, F=ma, and integrate the equation of motion to determine the positions, velocities, and accelerations of the system.

The equation of motion known as Newton's second law can be expressed as:

$$\boldsymbol{F_i = m_i a_i} \qquad \textit{Equation 3. 6}$$

where F represents the net force acting on an object, m represents its mass, and a represents the resulting acceleration.

The gradient of the potential energy can be used to describe the force,

$$\boldsymbol{Fi = -\nabla iV} \qquad \textit{Equation 3. 7}$$

The two equations are merged, resulting in:

$$\boldsymbol{\frac{-dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}} \qquad \textit{Equation 3. 8}$$

where V represents potential energy.

The force calculations in MD simulations are typically done using empirical potential energy functions or force fields, which are mathematical functions that describe the interactions between atoms and molecules. The force fields used in MD simulations are parameterized using experimental data or quantum mechanical

calculations and are essential for accurately simulating the system's behavior (Leach 1996).

MD simulations have revolutionized the study of biological systems and have been used to investigate the structure, stability, and dynamics of proteins, nucleic acids, and other biomolecules. The approach has been applied to a wide range of biological problems, including protein folding, ligand binding, enzyme catalysis, and membrane transport. However, MD simulations are computationally intensive and require substantial computational resources, and the accuracy of the results depends on the accuracy of the force field used (Karplus and McCammon 2002).

3.7 MD Trajectory Analysis

MD trajectory analysis is a critical component of the molecular dynamic simulation workflow, as it allows for the interpretation and extraction of meaningful information from the large amounts of data generated during the simulation (Dror et al. 2012).

3.7.1 **Root-Mean-Square Deviation (RMSD).** The Root Mean Square Deviation (RMSD) is a commonly used analysis method for MD trajectories. It involves calculating the RMSD of all frames in the trajectory relative to the initial conformation using the following formula:

$$\boldsymbol{RMSD} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\delta_i^2}} \qquad \textit{Equation 3. 9}$$

This equation calculates the RMSD by taking the distance between N pairs of equivalent atoms, represented by the symbol δ.
The RMSD is defined as follows for two sets of n points (r_i' and r_i):

$$\boldsymbol{RMSD_x} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{r_i'(t_x)} - \boldsymbol{r_i(t_{ref})})^{\mathbf{2}}} \qquad \textit{Equation 3. 10}$$

The equation involves selecting the N number of atoms and comparing their positions in different frames of the simulation trajectory. The reference time, t_{ref} is usually set as the time of the first frame (t=0), and the position of the selected atoms in subsequent frames is superimposed onto the reference frame. The position of the

selected atoms in each frame is denoted by r', and is recorded at a specific time t_x. This process is repeated for every frame in the simulation trajectory (Wu, Chandler, and Chandler 1988).

3.7.2 Root-Mean-Square Fluctuation (RMSF). The RMSF is a valuable tool to analyze local fluctuations in the protein chain. It provides information on the variability of each residue along the simulation trajectory. The formula to calculate the RMSF of residue i is:

$$\boldsymbol{RMSF_i} = \sqrt{\frac{1}{N}\sum_{t=1}^{T} < (\boldsymbol{r_i'(t)} - \boldsymbol{r_i(t_{ref})})^2 >} \qquad \textit{Equation 3. 11}$$

Root Mean Square Fluctuation (RMSF) can be used to assess the local changes that occur along the protein chain. To calculate the RMSF for a given residue i, the position of that residue (r_i) is compared to the position of the atoms in that residue after they have been superimposed on a reference structure (r_i'). The average of the square distance over the selection of atoms in the residue is then taken, and this value is divided by the trajectory time t_{ref}. The procedure is repeated for each residue in the protein chain (Wu et al. 1988).

3.8 Molecular Mechanics the Generalized Born Solvent Accessible Surface Area (MM/GBSA)

The calculation of binding free energy for biomolecular systems is crucial in comprehending the thermodynamic properties that are also measured in experiments. The use of MD simulations is currently one of the most advanced methods in tracking biomolecular systems and estimating binding free energies. There are different approaches available to compute free energies using computational methods. Although some methods can be highly accurate and provide results comparable to experimental values, they require sophisticated, time-consuming, and expensive computations. However, there are also faster and computationally cheaper methods available that may not provide accurate binding free energies but can be used for ranking different compounds. An example of such an approach is the end-point free energy computations, which are computationally efficient and only consider the final state of the system. The MM/GBSA approach is one such end-point free energy computation method generally used to estimate binding free energies for protein-ligand complexes,

aiding in scoring and ranking different ligands against the protein targets (Genheden and Ryde 2015; Wang et al. 2019).

The MM/GBSA score is calculated using molecular mechanics (MM) and implicit solvation approaches to determine the polar and non-polar contributions. The calculation of the MM/GBSA score involves two terms, namely the change in internal energy (ΔEMM) calculated by MM and the solvation-free energy ($\Delta GSOL$) determined by the implicit solvation approach, while also considering the entropy contribution (ΔS). The formula for MM/GBSA computations can be expressed as:

$$\Delta \boldsymbol{G_{bind}} = \Delta \boldsymbol{E_{MM}} + \Delta \boldsymbol{G_{SOL}} - \boldsymbol{T}\Delta \boldsymbol{S} \qquad \textit{Equation 3. 12}$$

where ΔE_{MM} represents the change in internal energy that includes contributions from $\Delta E_{int} + \Delta E_{ele} + \Delta E_{vdw}$. This approach is frequently used to calculate binding free energies for protein-ligand complexes and to score and rank different ligands against protein targets. It is computationally efficient, allowing for the rapid ranking of different compounds (Genheden and Ryde 2015; Wang et al. 2019).
The implicit solvation methods estimate the solvation energy and express it as follows:

$$\Delta \boldsymbol{G_{SOL}} = \Delta \boldsymbol{G_{GB}} + \Delta \boldsymbol{G_{SA}} \qquad \textit{Equation 3. 13}$$

In the MM/GBSA approach, the nonpolar contributions to the free energy are represented by ΔG_{SA}, which is estimated using the solvent-accessible surface area (SASA) method. The polar contribution, or the electrostatic solvation energy, is denoted by ΔG_{GB} and is calculated by the generalized Born surface area (GB) approach, as described by (Wang et al. 2019).

$$\Delta \boldsymbol{G_{SA}} = \boldsymbol{\gamma}.\boldsymbol{SASA} + \boldsymbol{c} \qquad \textit{Equation 3. 14}$$

In the given equation, the term SASA represents the solvent-accessible surface area. The symbol gamma (γ) is related to the surface tension, while c denotes the constant free energy contribution of the vacuum.

Chapter 4
Methodology

The aim of this research was to identify potential inhibitors for the RET proto-oncogene with the lowest resistance to mutations compared to the reference molecule, through an extensive screening of over 500 million molecules from the ZINC and Chemdiv databases. The methodology employed in this study was multifaceted, involving various computational techniques.

The first step in our workflow involved text mining of the ZINC database to extract relevant information on potential ligands for RET. Using the AutoQSAR module, a QSAR model was developed to predict the pIC_{50} values of a fragment-based ligand library. The resulting molecules were then subjected to a z-score analysis, with a threshold range of 2-3 based on the number of molecules. The hits that passed this filter were then evaluated for their cancer score using the MetaCore™/MetaDrug™ platform.

To prepare the ligands for virtual screening, the Schrödinger-Maestro's LigPrep tool was employed, and the RET structure was prepared with a grid center based on the Pralsetinib grid. Molecular dynamics (MD) simulation and MM/GBSA calculation were then conducted. As a secondary method, a long MD simulation was carried out for 7JU5, and an e-pharmacophore for all frames was generated to obtain a hypothesis. A Python script was used to analyze and visualize the hypothesis in frames with the lowest RMSD, and the coordinates with the lowest RMSD were employed to screen the fragment-based ligand libraries of the ZINC and Chemdiv databases.

To filter the molecules, 2D structures were initially filtered using the Lipinski Rules. 2D pharmacophore features were then calculated for the molecules using RDKit in Python, and the resulting hits were filtered based on the minimum features obtained from the hypothesis generated from long MD. The filtered molecules were then screened using the Phase tool to obtain fitness scores, and a z-score analysis was conducted on these scores. The molecules with a z-score of 2 or higher were retained, and if the count of molecules was less than 20, the top 20 molecules with the highest fitness scores were selected, and glide SP docking was performed on them. Molecules with the lowest docking scores were subjected to MD simulation and MM/GBSA calculation.

Next, the molecules with MM/GBSA scores above a predefined threshold were subjected to long MD simulation and MM/GBSA calculation to determine if there were significant differences. Furthermore, eight critical RET mutations identified in the literature were evaluated one by one. The corresponding protein structures were prepared, and a grid center was created for each mutation. The molecules that passed the MM/GBSA threshold were then subjected to Glide SP docking and MD simulation and MM/GBSA calculation, and the resulting graphs were analyzed and visualized.

The methodology employed in this study was comprehensive, utilizing various computational techniques to identify potential inhibitors for the RET proto-oncogene that have the lowest resistance to mutations compared to Pralsetinib, as a reference inhibitor. The workflow was optimized to filter out non-relevant hits and select the most promising candidates for further analysis. The complete method is represented in a schematic in Figure 5.

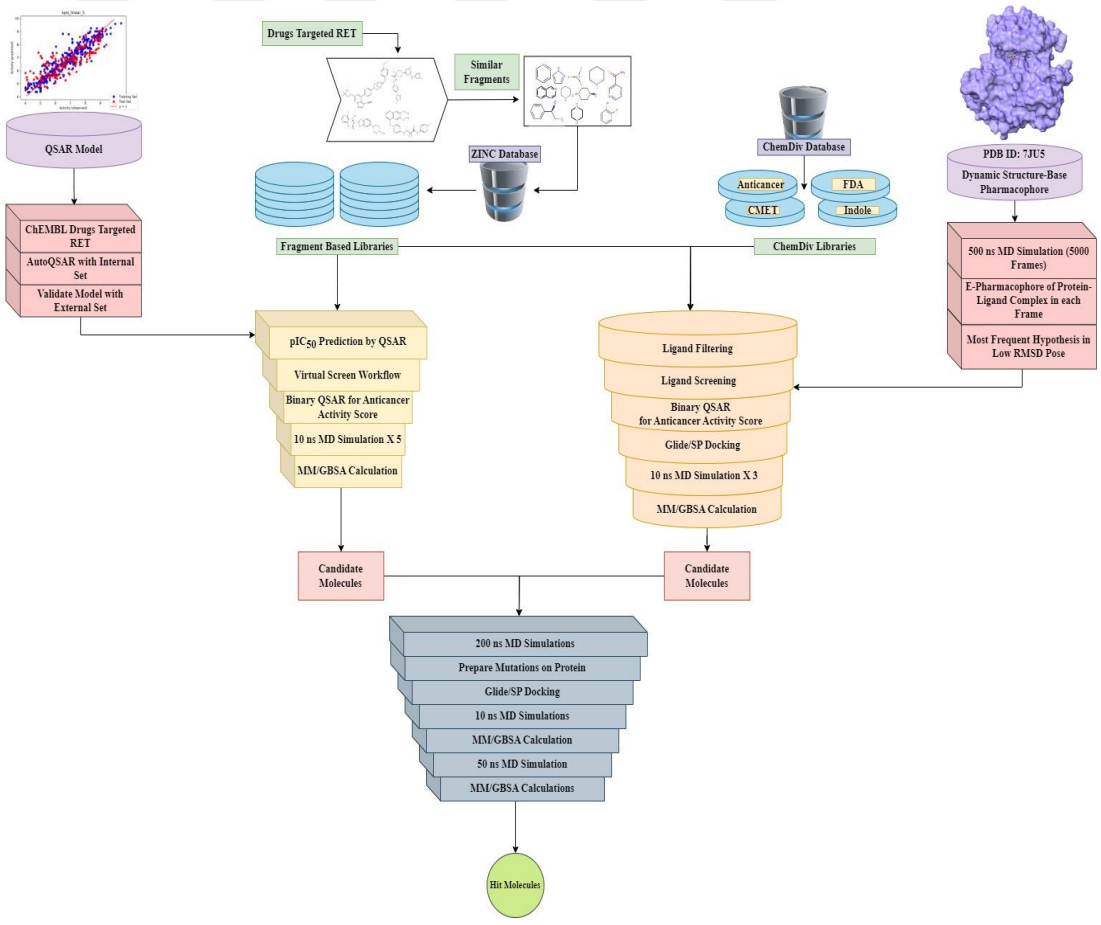


Figure 5. Schematic representation of the complete method for the discovery of a novel resistance-free RET tyrosine kinase inhibitor.

The method includes dynamic structure-based pharmacophore and QSAR modeling, as well as a virtual screening of ultra-large ligand libraries.

4.1 Method 1: A Fusion of Text Mining and QSAR for Predictive Drug Discovery

Our first method focuses on leveraging text mining techniques and QSAR modeling to predict pIC_{50} values for small molecules. This method includes several steps that allow us to extract relevant data from the literature, preprocess, curate the data, and build QSAR models using the curated data. Additionally, we utilize docking and molecular dynamics simulation with MMGBSA calculations to validate our predictions. These steps are performed using a combination of open-source and commercial software tools, and the resulting data are analyzed using statistical methods to identify the most promising small molecules for further study. By leveraging this methodology, we are able to rapidly identify compounds with the potential for therapeutic applications in drug discovery.

4.1.1 Text mining. In order to perform text mining on small molecules that inhibit the RET proto-oncogene protein, a filtering process was applied to select 120 molecules with IC50 values of 5 nM and below from the ChEMBL database as shown in Appendix (Table A.1). ChEMBL is a database maintained by the European Molecular Biology Laboratory that provides information on biological activities of small molecules, their interactions with protein targets, and ADMET properties, and is widely used by researchers in drug discovery (Mendez et al. 2019). The 2D structures and IUPAC names of these molecules were then compiled, and similarities between the structures were analyzed. The most frequently occurring fragments within these molecules were identified. Additionally, a list of RET inhibitor drugs that are approved or in clinical trials was created and displayed. The similarities between the structures of these drugs and the most common fragments found in the small molecules from the ChEMBL database were analyzed, and the fragments that were most likely to inhibit the RET protein were identified.

These fragments were then used to conduct text mining on a library of 500 million small molecules with known IUPAC names obtained from the ZINC database. ZINC is a freely available database of over 500 million commercially available compounds used for virtual screening and drug discovery (Sterling and Irwin 2015). A Python script was developed to search for each fragment name within the IUPAC names of the 500 million molecules to determine how many of the compounds

contained the fragment in their structure. All of the IUPAC names that contained the same fragment were then added to a text file. The IUPAC names were converted into their SMILES and SDF formats using the MolConverter tool from Marvin Suite, a tool from ChemAxon. Finally, all of the SDF files were converted into the Mae format, which is compatible with Schrödinger-maestro software, using the SDConvert tool from the command line.

4.1.2 Development and Application of QSAR Model for Predicting pIC_{50}.

We aimed to create a reliable QSAR model for RET protein-related molecules using the ChEMBL database. For this purpose, a total of 885 molecules with known IC_{50} values were extracted from the ChEMBL database. Among these, 793 molecules were selected as the internal set, while the remaining 92 molecules were used as the external set. The IC_{50} values of all molecules were converted to pIC_{50} values using the following equation:

$$\mathbf{pIC_{50}} = -\mathbf{log}(\mathbf{IC_{50}}) \qquad \textit{Equation 4. 1}$$

The AutoQSAR tool in the Schrödinger Maestro software was used to develop and validate the QSAR models. The compounds were divided into a 30% test set (238 compounds) and a 70% training set (555 compounds). The QSAR models were evaluated on the test set, and the model with the highest R-square value and highest R-square value after the prediction was chosen as the final model. The selected model was then used to predict the pIC_{50} values of all molecules in our created libraries. Due to the large number of molecules, the libraries were divided into several files to reduce their size and speed up the predictions. Finally, the predicted results were obtained in CSV file format and were read using a Python script with the Pandas library. Pandas is a powerful data manipulation and analysis library for Python (McKinney 2010). After merging the predicted pIC_{50} values into a Data Frame using the Pandas library, the normal distribution curve of the predicted data was visualized to analyze the distribution pattern. Z-Scores of pIC50 values were also calculated to evaluate the predicted values and select the most promising molecules for further investigation. The Z-Score is a statistical measure that indicates the deviation of a data point from the mean of the population. In this study, compounds with Z-Scores of 2 or 3 and over,

were selected as promising molecules. The threshold for the Z-Score selection was based on the number of molecules, where 2 is usually taken as the threshold. However, if a large number of molecules had Z-Scores over 2, then the threshold was raised to 3 and over to ensure that the selected molecules were truly significant.

4.1.1 **METACORE™/METADRUG ™ Analysis.** All the molecules obtained from the Z-Score calculation were separated into different files in SDF format. so that each file contains a maximum of 500 molecules. Then files are uploaded to the METACORETM platform to screen the "cancer therapeutic activity prediction QSAR". This functionality allows for the prediction of a compound's likelihood to exhibit anticancer activity. The screening process involves comparing input compounds to those with known high anticancer activity and assigning a probability value between 0 and 1, with values above 0.5 indicating potential anticancer activity. To construct the cancer therapeutic activity QSAR model, a set of descriptors and a training set of 886 compounds, achieving a sensitivity of 0.95, specificity of 0.92, accuracy of 0.93, and MCC of 0.87 were used. Also, the model was validated with a test set of 167 compounds, achieving a sensitivity of 0.89, specificity of 0.83, accuracy of 0.86, and MCC of 0.72. Based on the initial screening results, a cutoff value of 0.5 was established, and only molecules with probabilities ≥ 0.5 selecting for further consideration while filtering out the rest.

4.1.2 Protein Preparation. The RCSB Protein Data Bank was used to obtain the X-Ray diffraction structure of RET (PDB ID: 7JU5 (Subbiah et al. 2021)), which was co-crystallized with Pralsetinib, a selective RET TKs inhibitor, and has a resolution of 1.90 Å. Before using the protein for any computational experiments such as docking or MD simulations, it needs to be modified from its tense X-Ray diffracted form. The protein was processed using the Protein Preparation Wizard (Madhavi Sastry et al. 2013) to assign bond orders, add hydrogens, create zero-order bonds to metals, and recreate disulfide bonds. Missing side chains or loops were filled in with Prime (Jacobson et al. 2004), and water beyond 5Å of hetero groups was removed. The protonation states of the protein at pH assigned by PROPKA and side chain atoms were minimized with the OPLS3e force field (Harder et al. 2016).

4.1.3 Receptor Grid Generation. A receptor grid can be created to limit the movement of the ligand within a cubic region that includes the active site, so that only interactions with the active region are calculated. Partial flexibility can be achieved by allowing some rotatable bonds in the receptor active region to rotate during grid generation (Friesner et al. 2006; Halgren et al. 2004). The active part of the receptor is defined with respect to the reference ligand, and matching atoms are placed at the same coordinates as in the reference ligand. The internal coordinates of atoms outside the grid are rearranged while preserving their positions. At this stage, rotatable bonds are roughly scored. Then, a minimization process is applied to refine the torsion angles not found in the active site. Finally, post-docking minimization is performed. Binding modes that do not fully meet all constraints during grid application are not considered.

The receptor grid for the RET was generated using the Receptor Grid Generation Tool with the prepared structure. The grid was generated based on the co-crystallized native ligand, Pralsetinib, with a default measurement of 10 Å on all sides. No constraints were applied during the process, and rotation of hydroxyl or thiol groups of THR729, THR753, THR754, SER765, SER767, SER774, TYR791, TYR806, SER811, SER819, SER891, SER896, and THR946 was allowed. The (x, y,z) coordinates of the grid center are (21.76, 13.093, -23.514).

4.1.4 Ligand Preparation. To generate 3D structures for each molecule, the LigPrep module of maestro was employed (Chen and Foloppe 2010). During preparation, Epik was used to predict the ionization states of molecules at a pH of 7.0 ±2.0 (Shelley et al. 2007). The molecules with chiral centers were generated as stereoisomers. At most 4 structures with possible stereoisomers and ionization states were generated for each compound. The force field used for the preparation was OPLS3e (Roos et al. 2019).

4.1.5 **Virtual Screen Workflow.** Virtual screening workflow is a computational technique used to identify potential ligands for a target protein from a large library of compounds. One of the most widely used virtual screening methods is molecular docking, which involves the prediction of the binding mode and binding affinity of small molecule ligands to the target protein.

In molecular docking, different scoring functions are used to evaluate the binding affinity of the ligands to the protein. These scoring functions are based on various physical and chemical properties of the ligands and the protein-ligand complex. There are different types of molecular docking methods available, including High-Throughput Virtual Screening (HTVS), Standard Precision (SP) docking, and extra Precision (XP) docking. These methods differ in their computational complexity and accuracy. In the virtual screen workflow tool, all these methods are performed in order and considering the percentage of impact.

To perform the HTVS docking method, the tool imports the ligand library and a grid center file created based on Pralsetinib. Epik state penalties are added to the docking score, and a partial charge cutoff of 0.15 is set. Post-docking minimization is carried out with strain correction terms during the final scoring. Additionally, the tool generates up to one pose per compound, and 25% of the best-compounds are retained and passed through SP docking.

Then, SP docking occurs by adding Epik state penalties to the docking score, set 0.15 as a cutoff for the partial charge, and post-docking minimization was performed by applying strain correction terms during the final scoring. Also, it keeps generating up to 1 pose per compound and after docking 25% of the best compounds with good scoring states are retained and pass through XP docking.

Finally, the XP docking method is used to refine the top hits from HTVS and SP docking. Epik state penalties are added to the docking score, and the partial charge cutoff is set to 0.15. Post-docking minimization is performed with the tool generating up to one pose per compound. Only 10% of the best compounds with the best scoring states are retained after docking, but if 10% contain less than 10 compounds then the percentage increases to 20%. The top 10 molecules with the best docking scores from each library are selected for further analysis.

4.1.6 **Molecular Dynamics (MD) Simulations.** The 10 molecules with the lowest docking scores from each library were selected for all-atom MD simulations, which were conducted using the Desmond software (Bowers et al. 2006). The protein RET (PDB ID: 7JU5 (Subbiah et al. 2021)) was placed in a solvation box of the TIP3P model (Jorgensen et al. 1983), with a buffer of 10 Å around the protein in an orthorhombic shape. To ensure thermodynamic equilibrium, the isothermal-isobaric ensemble NPT was used, with a constant pressure of 1.01325 bar and temperature of 310 K. The system was initially neutralized with Na+ atoms and a salt concentration of 0.15 M (NaCl) at a physiological pH of 7.4. The simulations were conducted for each compound individually using the Nose-Hoover thermostat and Martyna-Tobias-Klein barostat (Martyna, Klein, and Tuckerman 1992). The RESPA integrator was used for bonded, near non-bonded, and far non-bonded interactions, with time steps of 2.0, 2.0, and 6.0 femtoseconds, respectively. A cut-off of 9 Å was set for short-range electrostatics and van der Waals interactions, while the particle mesh Ewald method (Essmann et al. 1995) and Periodic Boundary Conditions were applied for long-range electrostatic interactions. The OPLS3e force field was used to calculate the potential energy of the system. Before the simulations commenced, Desmond relaxed and minimized the system. In addition to the simulation described above, five additional simulations were performed on the same system, each with a different seed number to account for the effect of random initialization on the results. These simulations were run for a shorter duration of 10 ns but were repeated five times to obtain a total simulation time of 50 ns. From each of these simulations, 1000 frames were collected for analysis.

4.1.7 Molecular Mechanics the Generalized Born Solvent Accessible Surface Area (MM/GBSA) Calculations. MM/GBSA used in Prime was employed to compute the binding energies. For each 10 ns MD simulation, 100 protein-ligand complexes were extracted from the trajectories. For every 10 frames, 1 frame was used to calculate MM/GBSA.

The implicit solvation model, VSBG 2.0 (Li et al. 2011) was utilized to define the dielectric constant as a variable under the OPLS3e forcefield. Thus, the internal dielectric constant can range between 1.0 and 4.0 while the outside system is treated as a water system with a constant dielectric constant of 80 (Li et al. 2011). Following

the calculations of each complex, average values and standard deviations were calculated for each compound.

4.2 Method 2: Leveraging E-Pharmacophore Hypothesis and Long MD Trajectories for Accelerated Ligand Screening in Drug Discovery

In addition to our first method, we have developed a second approach in our research that focuses on ligand screening using the e-pharmacophore hypothesis generated from trajectories with the lowest RMSD of long MD simulation. This method involves several steps that allow us to identify potential ligand molecules for a specific target protein using molecular dynamics simulations and e-pharmacophore hypotheses. These ligands are then subjected to further analysis using statistical methods to identify the most promising candidates for further study. By using this methodology, we are able to rapidly screen large libraries of compounds and identify the most promising ones for further study, leading to the discovery of novel drug candidates.

4.2.1 Protein Preparation. To utilize the protein in any computational experiments, including docking or MD simulations, it requires modifications from its tense X-Ray diffracted form which is obtained from the RCSB Protein Data Bank using PDB ID: 7JU5 (Subbiah et al. 2021). Also, the ligand was removed from the protein-ligand complex and got it as wildtype RET. Furthermore, eight mutated RET obtain by changing eight different residues individually: G810S, V804L, V804M, V738A, Y806C, Y806N, L730V, and L730I. Also, another three mutated RET which include Y806C-V738A, V804M-Y806C, and V804M-G810S mutations are obtained by changing residues. So, we obtain eleven mutated and one wildtype RET protein and a RET protein in complex with Pralsetinib.

The Protein Preparation Wizard (Madhavi Sastry et al. 2013) was employed to process the protein-Ligand complex, wildtype RET, and eleven mutated by assigning bond orders, adding hydrogens, creating zero-order bonds to metals, and regenerating disulfide bonds. Any missing side chains or loops were filled in using Prime (Jacobson et al. 2004), and any water located beyond 5Å of hetero groups were eliminated. The protein's protonation states at the assigned pH by PROPKA, as well as its side chain atoms, were minimized using the OPLS3e force field (Harder et al. 2016).

4.2.2 Receptor Grid Generation. The Receptor Grid Generation Tool was used to generate a receptor grid for the RET-Pralsetinib complex, using the prepared structure and based on the co-crystallized native ligand, Pralsetinib, with a default measurement of 10 Å on all sides. The process did not involve any constraints, and the rotation of hydroxyl or thiol groups of THR729, THR753, THR754, SER765, SER767, SER774, TYR791, TYR806, SER811, SER819, SER891, SER896, and THR946 was allowed. Moreover, the receptor grid for eleven mutated RET and the wildtype generated based on the co-crystal pose of native Pralsetinib with the same process we used for the complex one. Also, for each protein suggested rotatable residues are allowed to rotate. The resulting grid center (x, y, z) coordinates for all the prepared proteins are (21.76, 13.093, -23.514).

4.2.3 **Long Molecular Dynamics (MD) Simulations.** The RET protein in complex with Pralsetinib (PDB ID: 7JU5) was subjected to all-atoms molecular dynamics (MD) simulations using the Desmond software (Bowers et al. 2006). The protein complex was solvated using the TIP3P model (Jorgensen et al. 1983) with an orthorhombic shape and a buffer of 10 Å around the protein. To maintain thermodynamic equilibrium, the isothermal-isobaric ensemble NPT was utilized with a constant pressure of 1.01325 bar and temperature of 310 K. The system was neutralized with Na+ ions and a salt concentration of 0.15 M (NaCl) at pH 7.4. The Nose-Hoover thermostat and Martyna-Tobias-Klein barostat (Martyna et al. 1992) were used for temperature and pressure control. Bonded, near non-bonded, and far non-bonded interactions were integrated using the RESPA integrator with time steps of 2.0, 2.0, and 6.0 femtoseconds, respectively. Short-range electrostatics and van der Waals interactions were truncated at 9 Å, while long-range electrostatic interactions were calculated using the particle mesh Ewald method (Essmann et al. 1995) with periodic boundary conditions. The OPLS3e force field was used to calculate the potential energy of the system. Prior to the simulations, the system was relaxed and minimized using Desmond. The simulations were performed for 500 ns, and 5000 frames were collected.

4.2.4 E-Pharmacophore Modeling. Our aims are to generate multiple pharmacophore hypotheses from the long MD simulation's numerous frames using the e-pharmacophore tool of PHASE. For this purpose, we extracted 5000 frames from the trajectory and used the e-pharmacophore method to create pharmacophore models based on the receptor-ligand complex from PHASE. During the process, we set the maximum number of features that can be created by this method as 7, with a minimum distance of 2 Å between features and a minimum distance of 4 Å between features of the same type. Additionally, we used a receptor-based excluded volume shell, considering the Van der Waals radii of the receptor atoms and fixing their scaling factor at 0.5.

However, considering a large number of frames, we needed to minimize the estimated time for creating hypotheses and getting results in CSV format files to simplify data visualization. We achieved this by executing the process from the Schrödinger-Maestro command line through a bash script. Firstly, the script created the grid center of Pralsetinib in each frame, split the receptor and ligand from the complex, then create a hypothesis with the e-pharmacophore method by applying in-place docking, and created a pharmacophore hypothesis for each frame, and then generated output in CSV format files.

Additionally, we developed a Python script to visualize the hypotheses and identify which hypotheses were generated the most during the long MD simulation, especially in frames with the lowest backbone RMSD while the protein aligned on frame 0. After identifying the selected hypotheses, we utilized them in ligand screening.

4.2.5 Ligand Filtering. In this approach, we employed ligand libraries which consist of fragment-based libraries obtained from a previous method and four different libraries from the Chemdiv database. These libraries contain over 300 million compounds in total. To ensure high-quality compounds, we utilized filtering methods such as Lipinski’s rule of five (Lipinski et al. 2001), in combination with 2D pharmacophore features. A Python script was developed to read the SDF file of each compound, calculate their physicochemical descriptors, and apply filtering based on these values.

Compounds that meet the criteria of Lipinski's rule of five, including a molecular weight below 500 Da, logP below 5, no more than 5 hydrogen bond donors (HBD), and no more than 10 hydrogen bond acceptors (HBA), were selected for the creation of 2D pharmacophore features using RDKit library in Python. This library can calculate the 2D pharmacophores of ligands based on atom pair descriptors and topological distances between them. Moreover, compounds that pass Lipinski's rules of five, are filtered based on features selected in the e-pharmacophore modeling step. So, those compounds that contain at least the same number of each feature in our reference feature hypothesis are passed the filtering part.

4.2.6 **Ligand Preparation.** To generate 3D structures for each compound, we used maestro's LigPrep module (Chen and Foloppe 2010). In order to predict the correct ionization states for a pH range of 7.0 ±2.0, Epik was employed during preparation, as recommended by Shelley et al. (2007). Molecules with chiral centers were created as stereoisomers based on their 3D atomic geometry, and a maximum of four structures with possible stereoisomers and ionization states were generated for each compound. The OPLS3e force field was utilized for preparation, according to Roos et al. (2019).

4.2.7 Ligand Screening. In order to search for potential ligands, the PHASE software is utilized to perform ligand and database screening. This process Screening the ligand libraries based on the hypothesis selected from the e-pharmacophore analyzing part to find the ligands highly matched on the reference hypothesis by using existing conformers. Phase screen score is selected as the scoring function of screening, no constraints are involved, and rejection criteria are set as default. Specify the number of hits to return per molecule in the Return at most 1 hit per molecule by sorting the hits by decreasing their phase screen score.

4.2.8 METACORE™/METADRUG ™ Analysis. After the screening process based on the hypothesis and data sorted by decreasing the phase screen score, the data for each library was placed into a Pandas library Data Frame. The distribution pattern of the predicted data was analyzed by visualizing the normal distribution curve. Z-Scores were calculated for the phase screen score values to evaluate the fitness score values and select the most promising molecules for further investigation. Promising molecules were selected based on a Z-Score threshold of 2 or higher. However, if the number of molecules with Z-Scores over 2 was less than 20, then the top 20 molecules were selected to predict their anticancer activity scores.

The molecules obtained from the Z-Score calculation were separated into separate SDF files, each containing a maximum of 500 molecules. These files were uploaded to the METACORE^{TM} platform to screen for potential cancer therapeutic activity. This screening process involves comparing the input compounds to those with known high anticancer activity and assigning a probability value between 0 and 1. Values above 0.5 indicate potential anticancer activity. For the construction of the cancer therapeutic activity QSAR model, a collection of 886 compounds was utilized along with a set of descriptors as the training set. The model exhibited notable performance metrics, including a sensitivity of 0.95, specificity of 0.92, accuracy of 0.93, and MCC of 0.87. To validate the model, an additional test set of 167 compounds was employed, resulting in a sensitivity of 0.89, specificity of 0.83, accuracy of 0.86, and MCC of 0.72. Based on the initial screening results, a cutoff value of 0.5 was established, and only molecules with probabilities greater than or equal to 0.5 were selected for further consideration, while the rest were filtered out.

4.2.9 Molecular Docking. The molecular docking process utilized the Glide program, which employs force field-based scoring functions (Friesner et al. 2006; Halgren et al. 2004). The docking jobs were configured with the OPLS3e force field (Roos et al. 2019) and set to Standard Precision (SP). The default settings from Glide/SP docking protocol were used, with the addition of Epik state penalties to the docking score, and post-docking minimization performed for five poses, rejecting minimization for poses within an energy window of 0.50 kcal/mol.

To expand the number of docking poses, the ring sampling energy window was established at 2.5 kcal/mol. In the initial docking phase, 50,000 poses per ligand were

retained, and subsequently, the top 4,000 poses per ligand were subjected to energy minimization. Additionally, expanded sampling was employed to bypass the elimination of poses during the rough scoring stage. After identifying the best poses with the lowest docking scores, the top 10 molecules from each library were selected for the next step in the docking protocol.

4.2.10 **Molecular Dynamics (MD) Simulations.** The top 10 molecules of each library that passed the glide/SP docking step were subjected to all-atoms molecular dynamics (MD) simulations using the Desmond software (Bowers et al., 2006). The protein complex was solvated with an orthorhombic shape using the TIP3P model (Jorgensen et al. 1983) and a buffer of 10 Å around the protein. To maintain thermodynamic equilibrium, the isothermal-isobaric ensemble NPT was used, with a constant pressure of 1.01325 bar and temperature of 310 K. The system was neutralized with Na+ ions, and a salt concentration of 0.15 M (NaCl) at pH 7.4 was applied. Temperature and pressure control were achieved using the Nose-Hoover thermostat and Martyna-Tobias-Klein barostat (Martyna et al. 1992), respectively. The interactions were integrated using the RESPA integrator with time steps of 2.0, 2.0, and 6.0 femtoseconds for bonded, near non-bonded, and far non-bonded interactions, respectively. Short-range electrostatics and van der Waals interactions were truncated at 9 Å, and long-range electrostatic interactions were calculated using the particle mesh Ewald method (Essmann et al. 1995) with periodic boundary conditions. The OPLS3e force field was used to calculate the potential energy of the system. The system was relaxed and minimized using Desmond before simulations. Three repeats of 10 ns MD simulations were performed, and 1000 frames were collected.

4.2.11 Molecular Mechanics the Generalized Born Solvent Accessible Surface Area (MM/GBSA) Calculations. The binding energies were calculated using MM/GBSA in Prime. 100 protein-ligand complexes were extracted from the trajectories of each 10 ns MD simulation, and MM/GBSA was computed for every 10 frames, resulting in a total of 1000 complexes per simulation. The VSBG 2.0 implicit solvation model was used with the OPLS3e forcefield, allowing the internal dielectric constant to vary between 1.0 and 4.0, while the outside system was treated as a water system with a constant dielectric constant of 80. Li et al. (2011) described this method in detail. The average values and standard deviations were calculated for each compound after the calculations for each individual complex were completed.

4.2.12 **Long Molecular Dynamics Simulations.** The MM/GBSA calculation results were used to set a threshold of -65 and molecules with an average MM/GBSA score lower than this threshold were chosen for further analysis. These selected hit molecules underwent all-atoms long molecular dynamics (MD) simulations using Desmond software, with the protein complex being solved using the TIP3P model and a buffer of 10 Å around the protein. To maintain thermodynamic equilibrium, the NPT ensemble was employed with a constant pressure of 1.01325 bar and temperature of 310 K, and the system was neutralized with Na+ ions and a salt concentration of 0.15 M (NaCl) at pH 7.4. Temperature and pressure control were maintained using the Nose-Hoover thermostat and Martyna-Tobias-Klein barostat, respectively. The RESPA integrator was used to integrate the interactions with time steps of 2.0, 2.0, and 6.0 femtoseconds for bonded, near non-bonded, and far non-bonded interactions, respectively. Short-range electrostatics and van der Waals interactions were truncated at 9 Å, and long-range electrostatic interactions were calculated using the particle mesh Ewald method with periodic boundary conditions. The OPLS3e force field was used to calculate the potential energy of the system. Before the simulations, the system was relaxed and minimized using Desmond. MD simulations were performed with 200 ns for each hit molecule, and 2000 frames were collected.

4.2.13 Molecular Docking on Mutated Proteins. To investigate the potential resistance of our hit molecules compared to Pralsetinib, we sought to dock them onto the RET protein, which is known to show resistance to some TK inhibitors due to mutations on its binding site. We carried out individual docking experiments for each of the eleven different mutations on the RET protein using the glide/SP protocol to determine the docking score and the docking poses of our hit molecules to observe how the interactions between our hit molecules and the RET protein change in the context of these mutations and analyze effects of them on molecules behavior by MD simulation analysis.

The docking jobs were configured using the OPLS3e force field and set to Standard Precision (SP), with default settings and the addition of Epik state penalties to the docking score. Post-docking minimization was performed for ten poses, rejecting minimization for poses within an energy window of 0.50 kcal/mol, and expanded sampling was employed to generate more poses for docking by setting the energy window for ring sampling to 2.5 kcal/mol. During the initial phase of docking, 50,000 poses per ligand were retained, and the best 4,000 poses per ligand were used for energy minimization. After identifying the best poses with the lowest docking scores for each hit molecule, that pose was selected for the MD simulation analysis. This approach allowed us to investigate how the interactions of our hit molecules change their behavior compared to our reference molecule, Pralsetinib, and whether they exhibit resistance to mutations on the binding site of RET tyrosine kinase.

4.2.14 **Molecular Dynamics (MD) Simulation for Mutated Proteins.** All the hit molecules and Pralsetinib were subjected to all-atoms molecular dynamics (MD) simulations with their docking poses in each mutation using the Desmond software (Bowers et al. 2006). Each hit molecule in a complex with mutated proteins was solvated with an orthorhombic shape using the TIP3P model (Jorgensen et al. 1983) and a buffer of 10 Å around the protein. To maintain thermodynamic equilibrium, the isothermal-isobaric ensemble NPT was used, with a constant pressure of 1.01325 bar and temperature of 310 K.

The system was neutralized with Na+ ions, and a salt concentration of 0.15 M (NaCl) at pH 7.4 was applied. Temperature and pressure control were achieved using the Nose-Hoover thermostat and Martyna-Tobias-Klein barostat (Martyna et al. 1992),

respectively. The interactions were integrated using the RESPA integrator with time steps of 2.0, 2.0, and 6.0 femtoseconds for bonded, near non-bonded, and far non-bonded interactions, respectively. Short-range electrostatics and van der Waals interactions were truncated at 9 Å, and long-range electrostatic interactions were calculated using the particle mesh Ewald method (Essmann et al. 1995) with periodic boundary conditions. The OPLS3e force field was used to calculate the potential energy of the system. The system was relaxed and minimized using Desmond before simulations. MD simulations in the length of 10 ns were performed for each complex, and 1000 frames were collected. Furthermore, 50 ns MD simulations with collecting 5000 frames was performed for molecules with low resistance against mutated RET in comparison with Pralsetinib.

4.2.15 Molecular Mechanics the Generalized Born Solvent Accessible Surface Area (MM/GBSA) Calculations for Mutated Proteins. The binding free energies were calculated using MM/GBSA in Prime. 100 protein-ligand complexes were extracted from the trajectories of each 10 ns MD simulation, and 500 complexes were extracted from trajectories of molecules with 50 ns MD simulations. So, MM/GBSA was computed for every 10 frames, resulting in a total of 1000 complexes per simulation, and also for every 10 frames in a total of 5000 complexes that were extracted from trajectories.

The VSBG 2.0 implicit solvation model was used with the OPLS3e forcefield, allowing the internal dielectric constant to vary between 1.0 and 4.0, while the outside system was treated as a water system with a constant dielectric constant of 80. Li et al. (2011) described this method in detail. The average values and standard deviations were calculated for each compound after the calculations for each complex were completed.

4.2.16 METACORE™ Toxicity Analysis. To perform toxicity analysis, the platforms rely on a variety of data sources, including public databases of known toxicants, as well as in-house data generated through experimental testing. Then it applies machine learning algorithms and other analytical techniques to these data sets in order to generate predictions of toxicity for new compounds. So, hit molecules were subjected to binary toxicity QSAR test in the MetaCore™/MetaDrug™ platform. All molecules test in 26 toxicity QSAR models and each of them has a cutoff value, mostly a cutoff value is 0.5, and molecules with values higher than 0.5 indicate potential toxicity.

Chapter 5

Results

5.1 Results for First Method

In our first method, various approaches were employed such as text mining and the QSAR model, and their outcomes are presented below, categorized by relevant subheadings.

5.1.1 Text Mining and QSAR. To do the text mining, after collecting the top 120 molecules with IC50 value of 5 nM and under it from the ChEMBL database then their 2D structures and IUPAC names are listed as shown in Table A.1. After analyzing similarities between those small molecules, fragments that are contained in most of the small molecules are listed and shown in Table 1.

Table 1.

Similar Fragments Between Top 120 ChEMBL Molecules That Target RET.

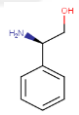
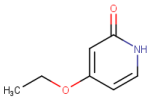
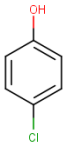
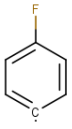
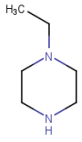
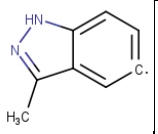
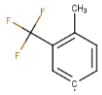
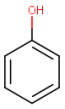
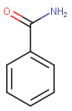
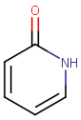
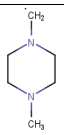
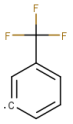
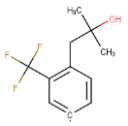
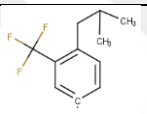
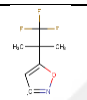
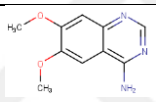
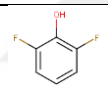
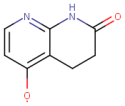
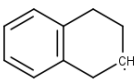
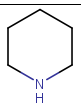
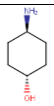
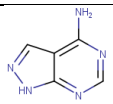
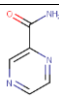
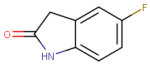
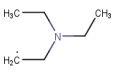
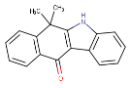
2D structure	IUPAC	Number of fragments	2D structure	IUPAC	Number of fragments
	[(1R) -2-hydroxy-1-phenylethyl] amino	4		5-ethoxy-6-oxo-1,6-dihydro pyridin-3-yl	15
	pyridin	31		trifluoromethyl	29
	4-chloro-phenol	5		fluorophenyl	19
	4-ethylpiperazin	43		3-methyl-1H-indazol	3

Table 1. (cont.d)

	4-methyl-3-(trifluoromethyl) phenyl	17		phenol	17
	phenyl	98		benzamide	10
	6-oxo-1,6-dihydropyridin	17		4-methylpiperazin	40
	(trifluoromethyl) phenyl	24		4-(2-hydroxy-2-methylpropyl) -3-(trifluoromethyl)phenyl	1
	4-(2-methyl propyl) -3-(trifluoromethyl)phenyl	6		oxazol	13
	3-(1,1,1-trifluoro-2-methylpropan-2-yl) -1,2-oxazol-5-yl	3		6,7-dimethoxyquinazolin	13
	Difluorophenol	3		fluorophenol	11
	(7-oxo-5,6,7,8-tetrahydro-1,8-naphthyridin-4-yl) oxy	2		1,2,3,4-tetrahydronaphthalen	2
	piperidin	26		[(1r,4r) -4-hydroxycyclohexyl] amino	17
	1H-pyrazolo[3,4-d] pyrimidin-4-amine	8		pyrazine-2-carboxamide	36
	5-fluoro-2-oxo-2,3-dihydro-1H-indole	2		2-(diethyl amino)ethyl	2
	6,6-dimethyl-11-oxo-5H,6H,11H-benzo[b]carbazole	4		5-cyclopropyl-1,2-oxazol-3-yl	5
	pyrazol	12			

Then, RET inhibitor drugs which are approved or in a clinical trial that are shown in Table 2. are used to detect the most frequent fragments between them by detecting their IUPAC names in text format, and seventeen fragments are mostly seen between them as shown in Table 3.

Table 2.

2D Structure of Approved Drugs and In Clinical Trail Drug Targeted RET Protein

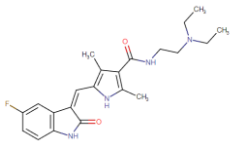
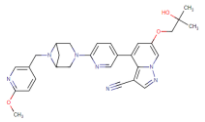
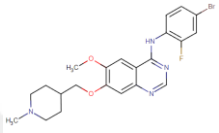
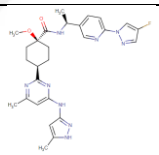
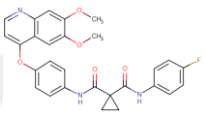
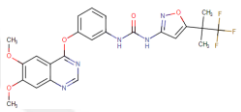
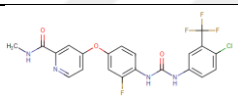
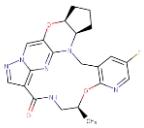
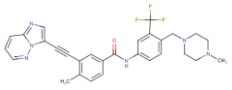
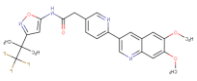
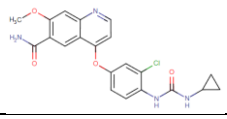
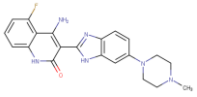
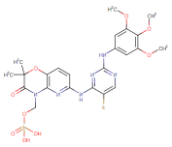
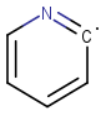
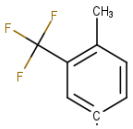
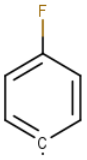
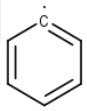
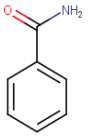
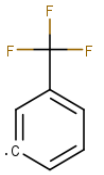
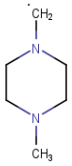
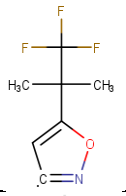
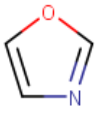
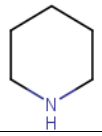
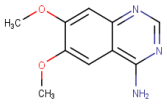
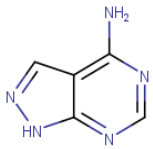
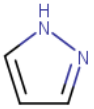
Drug Name	2D Structure	Drug Name	2D Structure
Sunitinib		**Selpercatinib**	
Vandetanib		**Pralsetinib (BLU-667)**	
Cabozantinib		**Agerafenib (RXDX-105)**	
Regorafenib		**Enbezotinib (TPX-0046)**	
Ponatinib		**Zeteletinib (BOS172738)**	
Lenvatinib		**dovitinib (TKI258)**	
Fostamatinib			

Table 3.

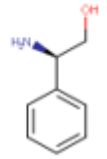
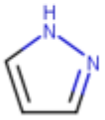
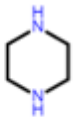
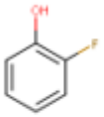
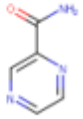
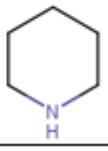
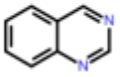
Similar Fragments Between Approved and In Clinical Trail Drugs Targeted RET Protein.

2D structure	IUPAC	Number of fragments	2D structure	IUPAC	Number of fragments
	pyridine	5		trifluoromethyl	4
	4-methyl-3-(trifluoromethyl)phenyl	1		Fluorophenyl	2
	phenyl	7		Benzamide	1
	(trifluoromethyl)phenyl	3		4-methylpiperazin	2
	3-(1,1,1-trifluoro-2-methylpropan-2-yl)-1,2-oxazol-5-yl	2		Oxazole	2
	Piperidine	1		6,7-dimethoxyquinazolin	2
	1H-pyrazolo[3,4-d]pyrimidin-4-amine	1		Pyrazole	1

So, by analyzing similar fragments between ChEMBL structures and RET inhibitor which are approved or in clinical trial, 10 of those fragments are chosen as fragments that can play an important role to inhibit the RET protein as you can see in Table 4.

Table 4.

Most Frequent Fragments Between 120 ChEMBL Compounds And 13 Approved and In Clinical Trials Drugs.

2D structure	IUPAC	2D structure	IUPAC
	[(1R) -2-hydroxy-1-phenylethyl] amino		pyrazol
	pyridin		trifluoromethyl
	piperazin		fluorophenol
	pyrazine-2-carboxamide		oxazol
	piperidin		quinazolin
	[(1r,4r) -4-hydroxycyclohexyl] amino		

So, it is decided to go through these fragments to do text mining inside the ZINC data library. This data library contains IUPAC names of 500 million small molecules got from the ZINC database. So, by using a python script it is found that each fragment name is included in how many compounds between the IUPAC names of 500 million small molecules as shown in Table 5. Then, all the IUPAC names that contain those fragment names are added to a text file to convert IUPAC names to SMILES and SDF format files using the MolConverter tool. Moreover, these SDF files are all converted to Mae format, with the SDConvert tool of Maestro from the command line.

Table 5.

The Number of Compounds Includes Fragments Name in Their IUPAC Names.

Fragments	Count of the fragment in 500Million ZINC database
Piperidin	88.836.619
Pyridin	83.144.820
Pyrazol	59.598.058
Oxazol	31.169.364
Piperazin	23.438.212
Trifluoromethyl	13.169.364
Quinazolin	1.954.454
Pyrazin-2-carboxamid	809.172
Fluorophenol	89.172
[(1r,4r)-4-hydroxycyclohexyl] amino	2.450

In contrast, the AUTOQSAR tool in maestro was utilized to generate various QSAR models based on 885 ChEMBL compounds that target RET protein. Using linear fingerprints of internal sets (793 compounds), the KPLS algorithm produced the best model with an R-square value of 0.8769 for the training set and a Q-square value of 0.6864 for the test sets, as illustrated in Figure 6.

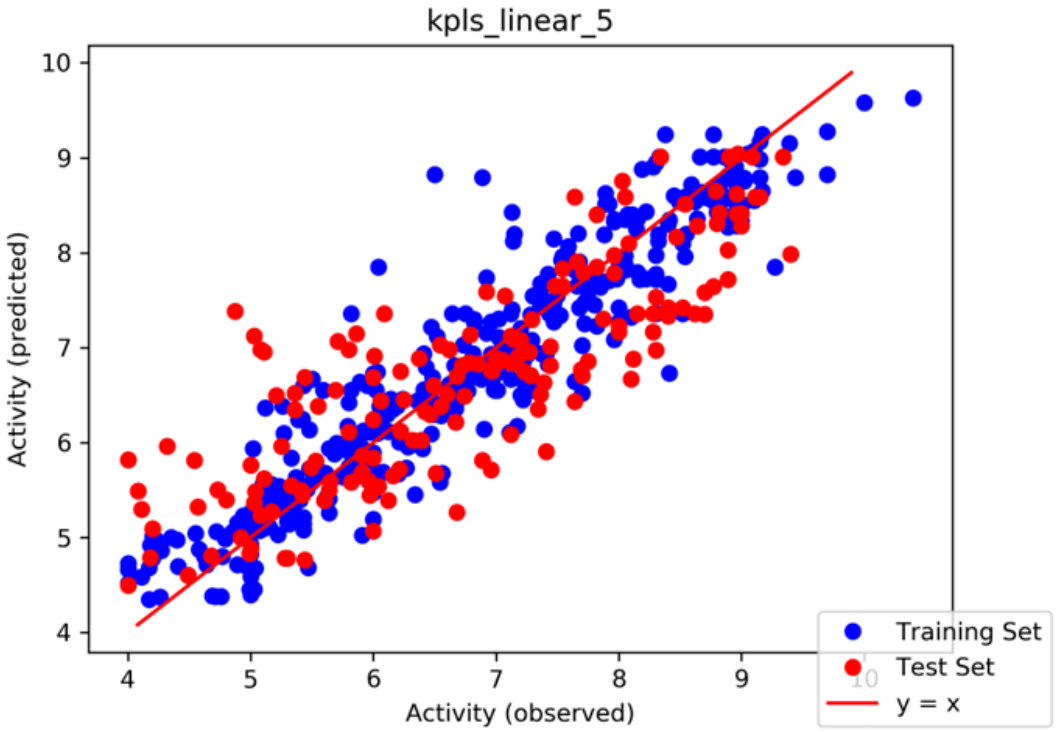


Figure 6. QSAR model created using AUTOQSAR tool of the maestro

Also, the model was validated by predicting pIC_{50} values of the external set (92 compounds), and the R-square of the model for the external set was obtained as 0.6521 as shown in Figure 7.

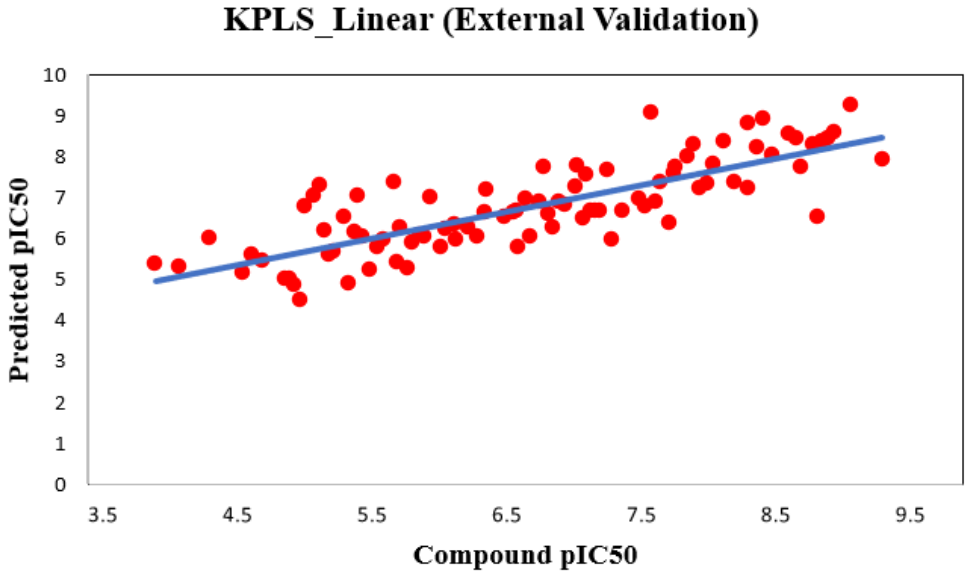


Figure 7. Scatter plot for predicted pIC_{50} of compounds in external set (R-square is equal to 0.6521).

The model was utilized to anticipate pIC_{50} values for eight out of ten libraries based on frequent fragments. For two libraries with a small number of compounds, all of their compounds were selected for analysis. Due to the high number of compounds in each library, a Python script was developed to read and calculate the mean of the predicted pIC_{50} values for each library. The normal distribution and z-score of each compound were also calculated to analyze the distribution of predicted pIC_{50} values. Compounds with a z-score value above a threshold of two or three (depending on the number of compounds) were selected for further analysis. Table 6. displays the number of compounds from each library selected and their corresponding z-score threshold.

Table 6.

Number of Compounds That Selected from Each Library Based on Their Cutoff Values.

Library name	**cutoff**	**Number of compounds**	**Library name**	**cutoff**	**Number of compounds**
Piperidin	3	5.488	Trifluoromethyl	2	1.660
Pyridin	3	14.326	Quinazolin	2	5.813
Pyrazol	3	25.289	Pyrazin-2-carboxamid	3	2.671
Oxazol	3	2.204	Fluorophenol	-	All compounds
Piperazin	2	10.908	[(1r,4r)-4-hydroxy cyclohexyl] amino	-	All compounds

5.1.2 **METACORE™/METADRUG ™ Results.** All compounds selected through the z-score analysis were subjected to QSAR models in the METACORETM platform to predict their anticancer activity score. The cutoff value for the score was set to 0.5, with compounds having scores above this value considered to have high anticancer activity. Therefore, this cutoff value was employed as a filtering criterion, and compounds with scores exceeding 0.5 from each library were selected for further analysis. In Table 7. the number of compounds over the cutoff value is shown.

Table 7.

The number of compounds in each library with anticancer activity scored over 0.5.

Library name	**Number of compounds**	**Library name**	**Number of compounds**
Piperidin	2.756	Trifluoromethyl	1.101
Pyridin	12.974	Quinazolin	5.626
Pyrazol	17.316	Pyrazin-2-carboxamid	2.413
Oxazol	1.756	Fluorophenol	63.537
Piperazin	7.576	[(1r,4r)-4-hydroxy cyclohexyl] amino	2.136

5.1.3 **Virtual Screen Workflow.** All molecules with an anticancer activity score above 0.5 in each library underwent HTVS, SP, and XP docking using the virtual screen workflow tool of maestro. The top 25% of compounds were retained for HTVS and SP docking, while the top 10% were retained for XP docking, except for cases where there were fewer compounds, in which case the retention rate was increased to 20%. From the XP docking results, the top 10 molecules with the lowest docking scores for each library were selected for further analysis through MD simulation, as presented in Tables 8-17.

Table 8.

Top 10 Molecules of Library Including Piperidin Fragment with Their Docking Scores.

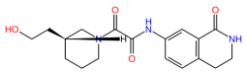
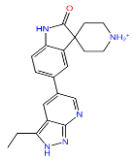
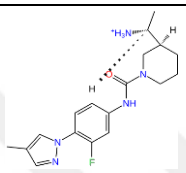
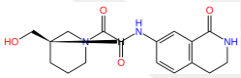
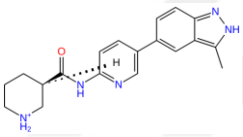
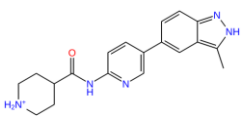
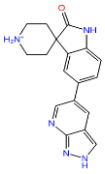
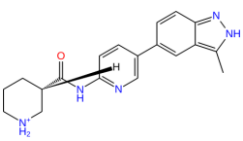
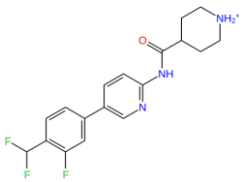
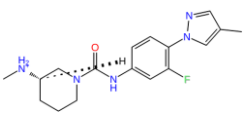
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	2-[(3S)-3-(2-hydroxyethyl) piperidin-1-yl]-2-oxo-N-(1-oxo-1,2,3,4-tetrahydroisoquinolin-7-yl) acetamide	-9.326
	5-{3-ethyl-2H-pyrazolo[3,4-b] pyridin-5-yl}-2-oxo-1,2-dihydrospiro[indole-3,4'-piperidin]-1'-ium	-8.983
	(1R)-1-[(3S)-1-{[3-fluoro-4-(4-methyl-1H-pyrazol-1-yl) phenyl] carbamoyl} piperidin-3-yl] ethan-1-aminium	-8.797
	2-[(3R)-3-(hydroxymethyl) piperidin-1-yl]-2-oxo-N-(1-oxo-1,2,3,4-tetrahydroisoquinolin-7-yl) acetamide	-8.791
	(3R)-3-{[5-(3-methyl-2H-indazol-5-yl)pyridin-2-yl]carbamoyl}piperidin-1-ium	-8.606
	4-{[5-(3-methyl-2H-indazol-5-yl)pyridin-2-yl]carbamoyl}piperidin-1-ium	-8.587
	2-oxo-5-{2H-pyrazolo[3,4-b]pyridin-5-yl}-1,2-dihydrospiro[indole-3,4'-piperidin]-1'-ium	-8.498
	(3S)-3-{[5-(3-methyl-2H-indazol-5-yl)pyridin-2-yl]carbamoyl}piperidin-1-ium	-8.410
	4-({5-[4-(difluoromethyl)-3-fluorophenyl]pyridin-2-yl}carbamoyl)piperidin-1-ium	-8.288
	(3S)-1-{[3-fluoro-4-(4-methyl-1H-pyrazol-1-yl)phenyl]carbamoyl}-N-methylpiperidin-3-aminium	-8.280

Table 9.

Top 10 Molecules of Library Including Pyridine Fragment with Their Docking Scores.

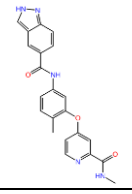
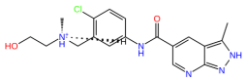
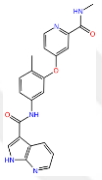
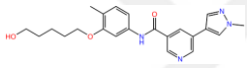
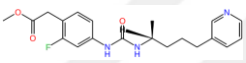
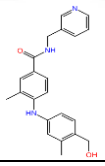
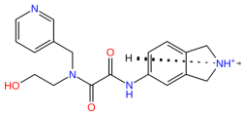
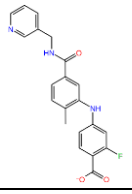
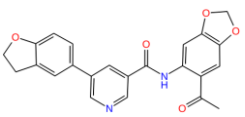
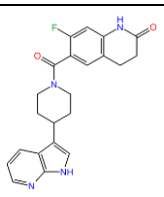
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	N-(4-methyl-3-{[2-(methylcarbamoyl) pyridin-4-yl] oxy}phenyl)-2H-indazole-5-carboxamide	-10.137
	N-(4-chloro-3-{[(2-hydroxyethyl) (methyl) amino] methyl}phenyl)-3-methyl-2H-pyrazolo[3,4-b]pyridine-5-carboxamide	-9.772
	N-methyl-4-(2-methyl-5-{1H-pyrrolo[2,3-b]pyridine-3-amido}phenoxy)pyridine-2-carboxamide	-9.724
	N-{3-[(5-hydroxypentyl)oxy]-4-methylphenyl}-5-(1-methyl-1H-pyrazol-4-yl)pyridine-3-carboxamide	-9.509
	methyl 2-[2-fluoro-4-({[(2S)-5-(pyridin-3-yl)pentan-2-yl]carbamoyl}amino)phenyl]acetate	-9.457
	4-{[4-(hydroxymethyl)-3-methylphenyl]amino}-3-methyl-N-[(pyridin-3-yl)methyl]benzamide	-9.389
	N'-(2-hydroxyethyl)-N-(2-methyl-2,3-dihydro-1H-isoindol-5-yl)-N'-[(pyridin-3-yl)methyl]ethanediamide	-9.370
	2-fluoro-4-[(2-methyl-5-{[(pyridin-3-yl)methyl]carbamoyl}phenyl)amino]benzoic acid	-9.339
	N-(6-acetyl-2H-1,3-benzodioxol-5-yl)-5-(2,3-dihydro-1-benzofuran-5-yl)pyridine-3-carboxamide	-9.287
	7-fluoro-6-(4-{1H-pyrrolo[2,3-b] pyridin-3-yl} piperidine-1-carbonyl)-1,2,3,4-tetrahydroquinolin-2-one	-9.249

Table 10.

Top 10 Molecules of Library Including Pyrazol Fragment with Their Docking Scores.

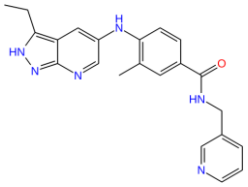
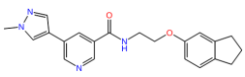
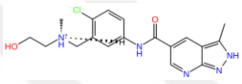
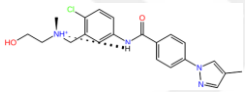
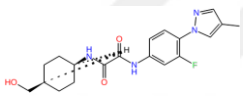
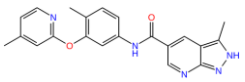
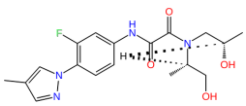
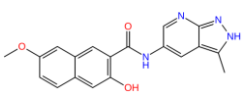
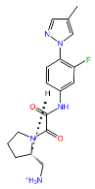
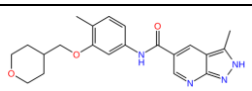
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	4-({3-ethyl-2H-pyrazolo[3,4-b]pyridin-5-yl}amino)-3-methyl-N-[(pyridin-3-yl)methyl]benzamide	-10.270
	N-[2-(2,3-dihydro-1H-inden-5-yloxy)ethyl]-5-(1-methyl-1H-pyrazol-4-yl)pyridine-3-carboxamide	-9.991
	N-(4-chloro-3-{[(2-hydroxyethyl)(methyl)amino]methyl}phenyl)-3-methyl-2H-pyrazolo[3,4-b]pyridine-5-carboxamide	-9.772
	N-(4-chloro-3-{[(2-hydroxyethyl)(methyl)amino]methyl}phenyl)-4-(4-methyl-1H-pyrazol-1-yl)benzamide	-9.627
	N-[3-fluoro-4-(4-methyl-1H-pyrazol-1-yl)phenyl]-N'-[(1s,4s)-4-(hydroxymethyl)cyclohexyl]ethanediamide	-9.610
	3-methyl-N-{4-methyl-3-[(4-methylpyridin-2-yl)oxy]phenyl}-2H-pyrazolo[3,4-b]pyridine-5-carboxamide	-9.553
	N'-[3-fluoro-4-(4-methyl-1H-pyrazol-1-yl)phenyl]-N-[(2S)-1-hydroxypropan-2-yl]-N-[(2S)-2-hydroxypropyl]ethanediamide	-9.517
	3-hydroxy-7-methoxy-N-{3-methyl-2H-pyrazolo[3,4-b]pyridin-5-yl}naphthalene-2-carboxamide	-9.451
	[(2S)-1-({[3-fluoro-4-(4-methyl-1H-pyrazol-1-yl)phenyl]carbamoyl}carbonyl)pyrrolidin-2-yl]methanaminium	-9.335
	3-methyl-N-{4-methyl-3-[(oxan-4-yl)methoxy]phenyl}-2H-pyrazolo[3,4-b]pyridine-5-carboxamide	-9.332

Table 11.

Top 10 Molecules of Library Including Oxazol Fragment with Their Docking Scores.

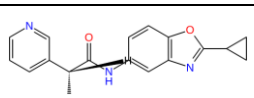
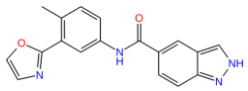
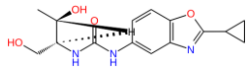
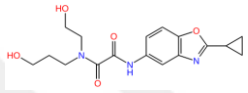
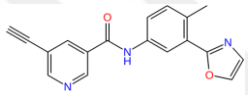
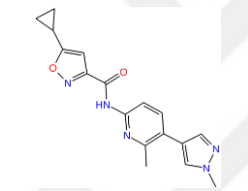
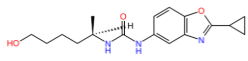
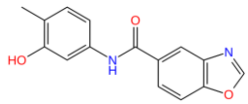
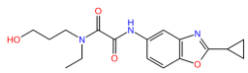
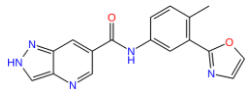
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	(2R)-N-(2-cyclopropyl-1,3-benzoxazol-5-yl)-2-(pyridin-3-yl)propanamide	-8.993
	N-[4-methyl-3-(1,3-oxazol-2-yl)phenyl]-2H-indazole-5-carboxamide	-8.953
	1-(2-cyclopropyl-1,3-benzoxazol-5-yl)-3-[(2R,3R)-1,3-dihydroxybutan-2-yl]urea	-8.721
	N'-(2-cyclopropyl-1,3-benzoxazol-5-yl)-N-(2-hydroxyethyl)-N-(3-hydroxypropyl)ethanediamide	-8.541
	5-ethynyl-N-[4-methyl-3-(1,3-oxazol-2-yl)phenyl]pyridine-3-carboxamide	-8.371
	5-cyclopropyl-N-[6-methyl-5-(1-methyl-1H-pyrazol-4-yl)pyridin-2-yl]-1,2-oxazole-3-carboxamide	-8.323
	1-(2-cyclopropyl-1,3-benzoxazol-5-yl)-3-[(2R)-6-hydroxyhexan-2-yl]urea	-8.093
	N-(3-hydroxy-4-methylphenyl)-1,3-benzoxazole-5-carboxamide	-8.021
	N'-(2-cyclopropyl-1,3-benzoxazol-5-yl)-N-ethyl-N-(3-hydroxypropyl)ethanediamide	-7.913
	N-[4-methyl-3-(1,3-oxazol-2-yl)phenyl]-2H-pyrazolo[4,3-b]pyridine-6-carboxamide	-7.896

Table 12.

Top 10 Molecules of Library Including Piperazin Fragment with Their Docking Scores.

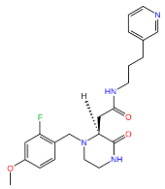
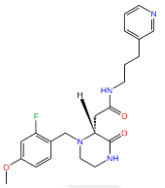
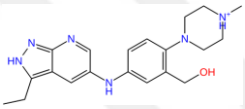
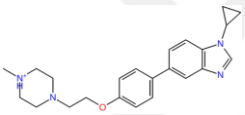
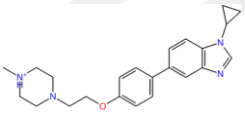
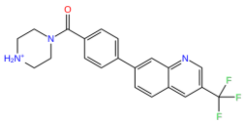
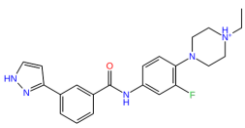
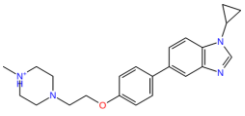
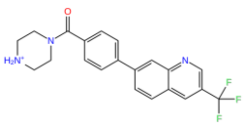
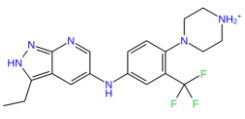
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	2-[(2S)-1-[(2-fluoro-4-methoxyphenyl) methyl]-3-oxopiperazin-2-yl]-N-[3-(pyridin-3-yl) propyl] acetamide	-9.192
	2-[(2R)-1-[(2-fluoro-4-methoxyphenyl)methyl]-3-oxopiperazin-2-yl]-N-[3-(pyridin-3-yl)propyl]acetamide	-9.181
	[5-({3-ethyl-2H-pyrazolo[3,4-b]pyridin-5-yl}amino)-2-(4-methylpiperazin-1-yl)phenyl]methanol	-8.988
	1-{2-[4-(1-cyclopropyl-1H-1,3-benzodiazol-5-yl)phenoxy]ethyl}-4-methylpiperazin-1-ium	-8.965
	1-cyclopropyl-5-{4-[2-(4-methylpiperazin-1-yl)ethoxy] phenyl}-1H-1,3-benzodiazole	-8.880
	7-[4-(piperazine-1-carbonyl)phenyl]-3-(trifluoromethyl)quinoline	-8.859
	N-[4-(4-ethylpiperazin-1-yl)-3-fluorophenyl]-3-(1H-pyrazol-3-yl)benzamide	-8.731
	4-{2-[4-(1-cyclopropyl-1H-1,3-benzodiazol-5-yl) phenoxy]ethyl}-1-methylpiperazin-1-ium	-8.725
	4-{4-[3-(trifluoromethyl)quinolin-7-yl]benzoyl} piperazin-1-ium	-8.697
	4-[4-({3-ethyl-2H-pyrazolo[3,4-b]pyridin-5-yl}amino)-2-(trifluoromethyl)phenyl]piperazin-1-ium	-8.694

Table 13.

Top 10 Molecules of Library Including Trifluorophenol Fragment with Their Docking Scores.

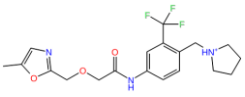
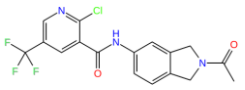
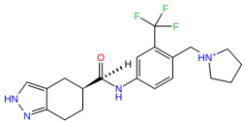
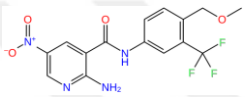
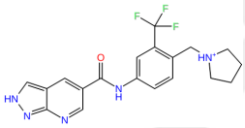
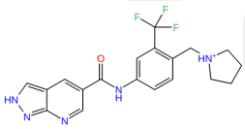
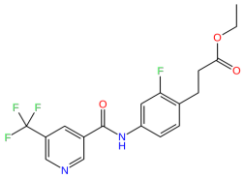
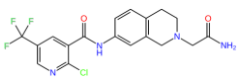
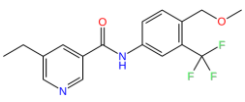
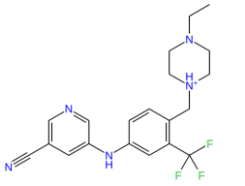
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	2-[(5-methyl-1,3-oxazol-2-yl) methoxy]-N- {4-[(pyrrolidine-1-yl)methyl]-3-(trifluoromethyl) phenyl}acetamide	-8.505
	N-(2-acetyl-2,3-dihydro-1H-isoindol-5-yl)-2-chloro-5-(trifluoromethyl)pyridine-3-carboxamide	-8.244
	(5S)-N-{4-[(pyrrolidin-1-yl)methyl]-3-(trifluoromethyl)phenyl}-4,5,6,7-tetrahydro-2H-indazole-5-carboxamide	-8.218
	2-amino-N-[4-(methoxymethyl)-3-(trifluoromethyl) phenyl]-5-nitropyridine-3-carboxamide	-8.128
	N-{4-[(pyrrolidin-1-yl)methyl]-3-(trifluoromethyl) phenyl}-2H-pyrazolo[3,4-b]pyridine-5-carboxamide	-8.047
	1-[(4-{2H-pyrazolo[3,4-b]pyridine-5-amido}-2-(trifluoromethyl)phenyl)methyl]pyrrolidin-1-ium	-8.047
	ethyl 3-{2-fluoro-4-[5-(trifluoromethyl)pyridine-3-amido]phenyl}propanoate	-7.959
	N-[2-(carbamoylmethyl)-1,2,3,4-tetrahydroisoquinolin -7-yl]-2-chloro-5-(trifluoromethyl)pyridine-3-carboxamide	-7.949
	5-ethyl-N-[4-(methoxymethyl)-3-(trifluoromethyl) phenyl]pyridine-3-carboxamide	-7.937
	4-({4-[(5-cyanopyridin-3-yl)amino]-2-(trifluoro methyl)phenyl}methyl)-1-ethylpiperazin-1-ium	-7.904

Table 14.

Top 10 Molecules of Library Including Quinazolin Fragment with Their Docking Scores.

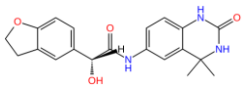
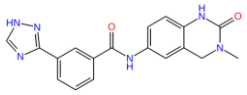
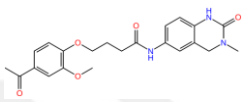
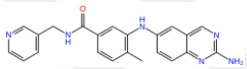
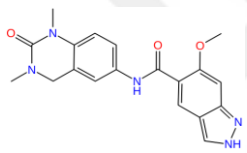
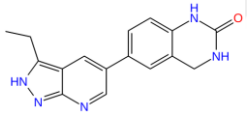
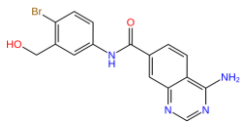
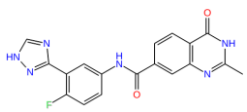
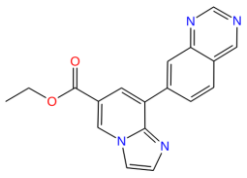
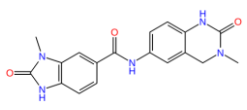
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	(2R)-2-(2,3-dihydro-1-benzofuran-5-yl)-N-(4,4-dimethyl-2-oxo-1,2,3,4-tetrahydroquinazolin-6-yl)-2-hydroxyacetamide	-9.295
	N-(3-methyl-2-oxo-1,2,3,4-tetrahydroquinazolin-6-yl)-3-(1H-1,2,4-triazol-3-yl)benzamide	-9.264
	4-(4-acetyl-2-methoxyphenoxy)-N-(3-methyl-2-oxo-1,2,3,4-tetrahydroquinazolin-6-yl)butanamide	-9.111
	3-[(2-aminoquinazolin-6-yl)amino]-4-methyl-N-[(pyridin-3-yl)methyl]benzamide	-9.088
	N-(1,3-dimethyl-2-oxo-1,2,3,4-tetrahydroquinazolin-6-yl)-6-methoxy-2H-indazole-5-carboxamide	-9.032
	6-{3-ethyl-2H-pyrazolo[3,4-b]pyridin-5-yl}-1,2,3,4-tetrahydroquinazolin-2-one	-8.929
	4-amino-N-[4-bromo-3-(hydroxymethyl)phenyl]quinazoline-7-carboxamide	-8.905
	N-[4-fluoro-3-(4H-1,2,4-triazol-3-yl)phenyl]-2-methyl-4-oxo-3,4-dihydroquinazoline-7-carboxamide	-8.769
	ethyl 8-(quinazolin-7-yl)imidazo[1,2-a]pyridine-6-carboxylate	-8.596
	3-methyl-N-(3-methyl-2-oxo-1,2,3,4-tetrahydroquinazolin-6-yl)-2-oxo-2,3-dihydro-1H-1,3-benzodiazole-5-carboxamide	-8.583

Table 15.

Top 10 Molecules of Library Including Pyrazin-2-carboxamid Fragment with Their Docking Scores.

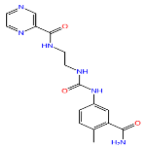
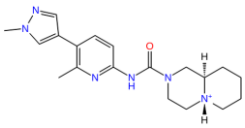
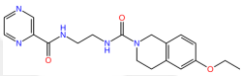
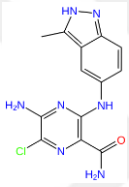
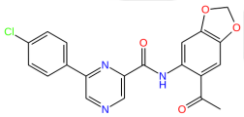
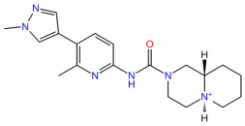
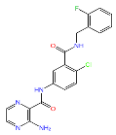
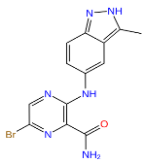
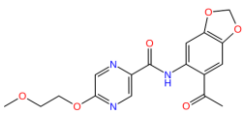
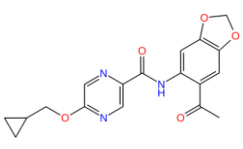
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	N-(2-{[(3-carbamoyl-4-methylphenyl)carbamoyl] amino}ethyl)pyrazine-2-carboxamide	-9.455
	(9aR)-N-[6-methyl-5-(1-methyl-1H-pyrazol-4-yl)pyridin-2-yl]-octahydro-1H-pyrido[1,2-a]pyrazine-2-carboxamide	-9.154
	N-{2-[(6-ethoxy-1,2,3,4-tetrahydroisoquinoline-2-carbonyl)amino]ethyl}pyrazine-2-carboxamide	-9.111
	5-amino-6-chloro-3-[(3-methyl-2H-indazol-5-yl) amino]pyrazine-2-carboxamide	-9.061
	N-(6-acetyl-2H-1,3-benzodioxol-5-yl)-6-(4-chlorophenyl)pyrazine-2-carboxamide	-9.054
	(9aS)-N-[6-methyl-5-(1-methyl-1H-pyrazol-4-yl)pyridin-2-yl]-octahydro-1H-pyrido[1,2-a]pyrazine-2-carboxamide	-8.986
	3-amino-N-(4-chloro-3-{[(2-fluorophenyl)methyl] carbamoyl}phenyl)pyrazine-2-carboxamide	-8.949
	6-bromo-3-[(3-methyl-2H-indazol-5-yl)amino] pyrazine-2-carboxamide	-8.676
	N-(6-acetyl-2H-1,3-benzodioxol-5-yl)-5-(2-methoxyethoxy)pyrazine-2-carboxamide	-8.464
	N-(6-acetyl-2H-1,3-benzodioxol-5-yl)-5-(cyclopropylmethoxy)pyrazine-2-carboxamide	-8.326

Table 16.

Top 10 Molecules of Library Including Fluorophenol Fragment with Their Docking Scores.

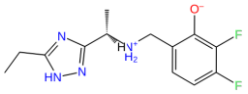
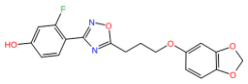
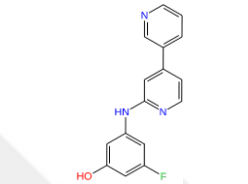
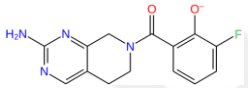
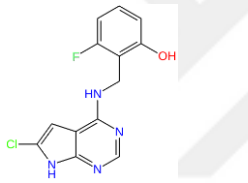
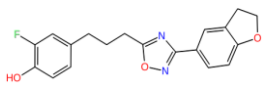
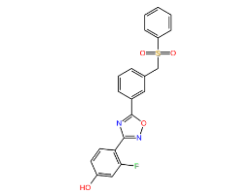
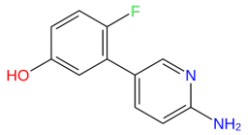
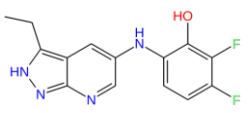
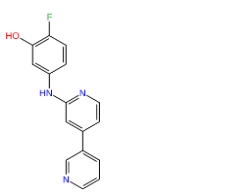
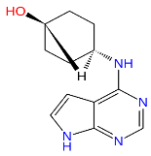
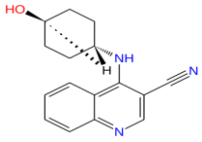
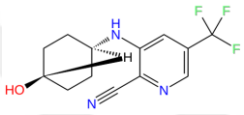
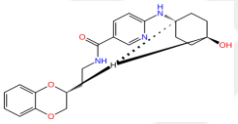
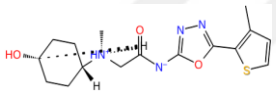
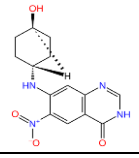
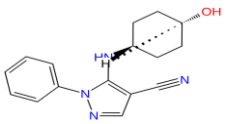
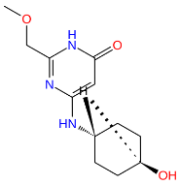
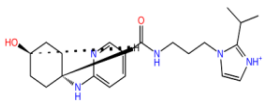
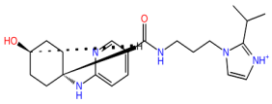
Structure	**IUPAC Name**	**docking score** (kcal/mol)
	6-({[(1S)-1-(5-ethyl-4H-1,2,4-triazol-3-yl)ethyl] amino}methyl)-2,3-difluorophenol	-9.954
	4-{5-[3-(2H-1,3-benzodioxol-5-yloxy)propyl]-1,2,4-oxadiazol-3-yl}-3-fluorophenol	-9.6247
	3-({[3,4'-bipyridin]-2'-yl}amino)-5-fluorophenol	-9.461
	2-{2-amino-5H,6H,7H,8H-pyrido[3,4-d]pyrimidine-7-carbonyl}-6-fluorophenol	-9.448
	2-[({6-chloro-7H-pyrrolo[2,3-d]pyrimidin-4-yl}amino) methyl]-3-fluorophenol	-9.185
	4-{3-[3-(2,3-dihydro-1-benzofuran-5-yl)-1,2,4-oxadiazol-5-yl]propyl}-2-fluorophenol	-9.144
	4-(5-{3-[(benzenesulfonyl)methyl]phenyl}-1,2,4-oxadiazol-3-yl)-3-fluorophenol	-9.105
	3-(6-aminopyridin-3-yl)-4-fluorophenol	-8.981
	6-({3-ethyl-2H-pyrazolo[3,4-b]pyridin-5-yl}amino)-2,3-difluorophenol	-8.957
	5-({[3,4'-bipyridin]-2'-yl}amino)-2-fluorophenol	-8.955

Table 17.

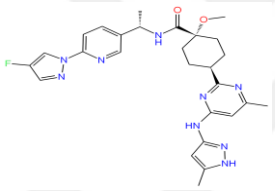
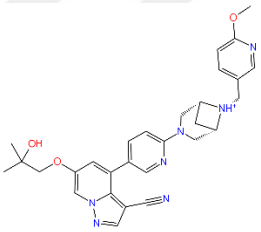
Top 10 Molecules of Library Including [(1r,4r)-4-hydroxycyclohexyl]amino Fragment with Their Docking Scores.

Structure	**IUPAC Name**	**docking score** (kcal/mol)
	4-{[(1r,4r)-4-hydroxycyclohexyl]amino}-7H-pyrrolo [2,3-d]pyrimidin-3-ium	-8.312
	4-{[(1r,4r)-4-hydroxycyclohexyl]amino}quinoline-3-carbonitrile	-8.265
	3-{[(1r,4r)-4-hydroxycyclohexyl]amino}-5-(trifluoromethyl)pyridine-2-carbonitrile	-7.780
	N-{2-[(2R)-2,3-dihydro-1,4-benzodioxin-2-yl]ethyl}-6-{[(1r,4r)-4-hydroxycyclohexyl]amino}pyridine-3-carboxamide	-7.765
	2-{methyl[(1r,4R)-4-hydroxycyclohexyl]amino}-N-[5-(3-methylthiophen-2-yl)-1,3,4-oxadiazol-2-yl]acetamide	-7.690
	6-nitro-7-{[(1r,4r)-4-hydroxycyclohexyl]amino}-3,4-dihydroquinazolin-4-one	-7.627
	1-phenyl-5-{[(1r,4r)-4-hydroxycyclohexyl]amino}-1H-pyrazole-4-carbonitrile	-7.465
	2-(methoxymethyl)-6-{[(1r,4r)-4-hydroxycyclohexyl] amino}-3,4-dihydropyrimidin-4-one	-7.227
	2-(propan-2-yl)-1-{3-[(6-{[(1r,4r)-4-hydroxy cyclohexyl]amino}pyridin-3-yl)formamido]propyl}-1H-imidazol-3-ium	-7.125
	N-{3-[2-(propan-2-yl)-1H-imidazol-1-yl]propyl}-6-{[(1r,4r)-4-hydroxycyclohexyl]amino}pyridine-3-carboxamide	-7.125

Ultimately, in order to compare the obtained results with reference molecules, redocking experiments were conducted using glide/XP to Pralsetinib and Selpercatinib. The purpose was to determine whether the redock poses of these molecules were consistent with their co-crystal poses. The redock pose with the lowest RMSD between the redocked and co-crystal poses was chosen as the first pose for MD simulation. The corresponding docking score was used as the reference docking score. Table 18. presents the redocking scores of the reference molecules and their RMSD values between the co-crystal structures. Additionally, Figure 5.3 illustrates the structures of the reference molecules and their redocked poses.

Table 18.

2D Structure of Pralsetinib and Selpercatinib, Their RMSD Value with Co-Crystal Structure, and Their Glide/XP Score.

Structure	**name**	**docking score** (kcal/mol)	**RMSD (Å)**
	**Pralsetinib**	**-9.463**	**2.254**
	**Selpercatinib**	**-9.164**	**1.202**

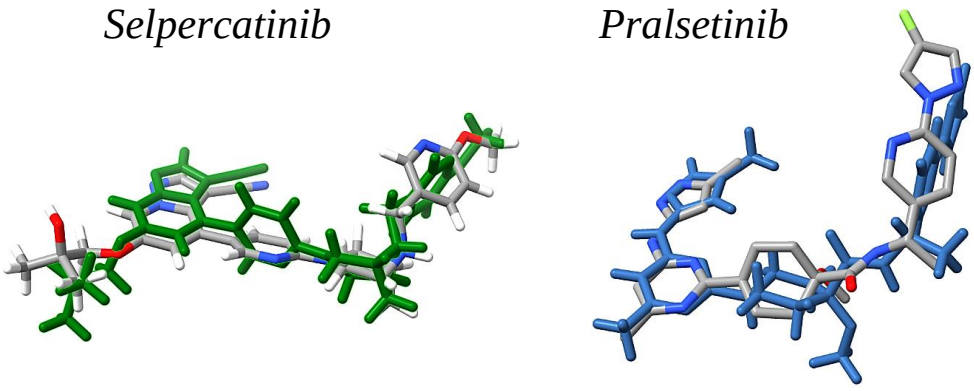


Figure 8. Structure of reference molecules with their redock poses.

Left side co-crystal structure of Selpercatinib with its redock poses (RMSD of 1.20 Å) which is shown in green. Right side co-crystal structure of Pralsetinib with the redock pose (RMSD of 2.25 Å) which is shown in blue.

5.1.4 Molecular Dynamics Simulations and MM/GBSA calculation. The molecular dynamics simulations were conducted for the top 10 molecules from each library, as well as Selpercatinib and Pralsetinib. The simulations were run for a duration of 10 ns, with 1000 frames collected from each simulation. Each simulation was repeated five times with different seed numbers. Once all the simulations were completed, 100 protein-ligand complex structures were extracted from the trajectories of each simulation, and MM/GBSA calculations were performed. The average MM/GBSA of the 100 structures for all repeats was calculated, and the average for the five repeats was obtained. These averages were obtained for all the molecules, and the total MM/GBSA average for Pralsetinib, Selpercatinib, and the top 4 hit molecules with the lowest MM/GBSA averages are presented in Table 19.

Table 19.

Total Average of MM/GBSA For Two Reference Molecules and Top 4 Hit Molecules.

name	**Library**	**Total average MM/GBSA (kcal/mol)**
Pralsetinib	Reference	-75.544 ± 3.415
Selpercatinib	Reference	-75.677 ± 5.096
N-(6-acetyl-2H-1,3-benzodioxol-5-yl)-6-(4-chlorophenyl) pyrazine-2-carboxamide	Pyrazine-2-carboxamide	-61.673 ± 4.167
N-(4-methyl-3-{[2-(methylcarbamoyl) pyridin-4-yl] oxy} phenyl)-2H-indazole-5-carboxamide	pyridine	-60.273 ± 3.976
N-{3-[(5-hydroxypentyl) oxy]-4-methylphenyl}-5-(1-methyl-1H-pyrazol-4-yl) pyridine-3-carboxamide	pyridine	-60.737 ± 4.609
N-(6-acetyl-2H-1,3-benzodioxol-5-yl)-5-(2,3-dihydro-1-benzofuran-5-yl) pyridine-3-carboxamide	pyridine	-60.121 ± 5.567

5.2 Results for Second Method

The second approach is based on the e-pharmacophore hypothesis derived from poses exhibiting the lowest RMSD from MD simulation, and the results are presented in a step-by-step manner.

5.2.1 **E-Pharmacophore Modeling Results.** For this approach, 500 ns MD simulation was performed for RET protein in complex with Pralsetinib (PDB ID: 7JU5) by collecting 5000 frames. Then, all of the 5000 frames are extracted as protein-ligand complex from the trajectory and by developing a script, the grid center of all the structures are determined as (x,y,z) coordinates. So, by using these coordinates e-pharmacophore method is used to create pharmacophore models between protein-ligand complexes with a maximum of seven features. Furthermore, all the hypotheses created by the e-pharmacophore method pass through further analysis by developing another python script. The script visualizes all the hypotheses and analyzes the frequency of the hypotheses between the frames with low RMSD values, as can be seen in Table 20. However, it is understood that the most frequent hypothesis among the hypothesis containing five features is ADDRR which is generated 2,892 times during Md simulation frames and the second most frequent hypothesis is ADRR which contains four features and is seen in 895 frames, as it is shown in Figure 5.4 and 5.5. Also, it is found that both of the hypotheses are generated in frames with low RMSD. ADRR was generated in frame 4338 with an RMSD value of 1.093 Å and ADDRR was generated in frame 4291 with an RMSD value of 1.098 Å.

Table 20.

All Hypotheses Created by the E-Pharmacophore Method with The Total Number of That Hypothesis in Total Frames.

Features	**number of hypotheses**
ARR	23
ADRR	895
DRRR	9
AHRR	4
ADDRR	2892
AADRR	5
ADHRR	266
DDRRR	19
DHRRR	4
ADRRR	5
ADDRRR	26
AADDRR	16
ADDHRR	812
DDHRRR	13
ADHRRR	1
ADDHRRR	4
ADDDHRR	2
AADDHRR	5

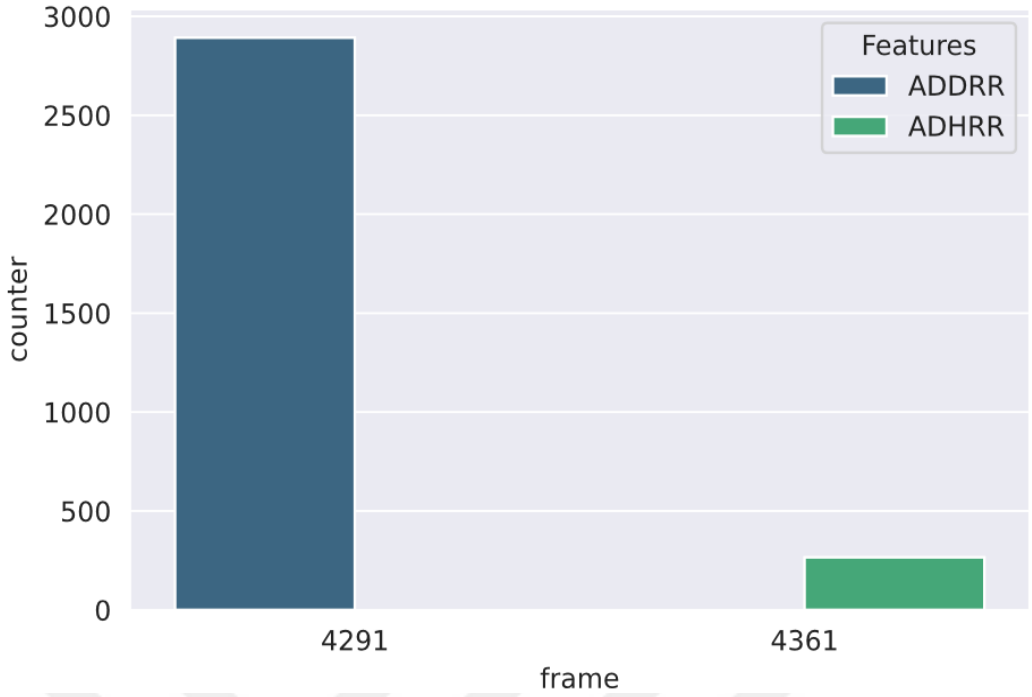


Figure 9. Bar chart comparison of top 2 five-feature hypotheses and their frequency in frames with low RMSD

(Number in the x-axis represents the frame number with the lowest RMSD value containing that hypothesis).

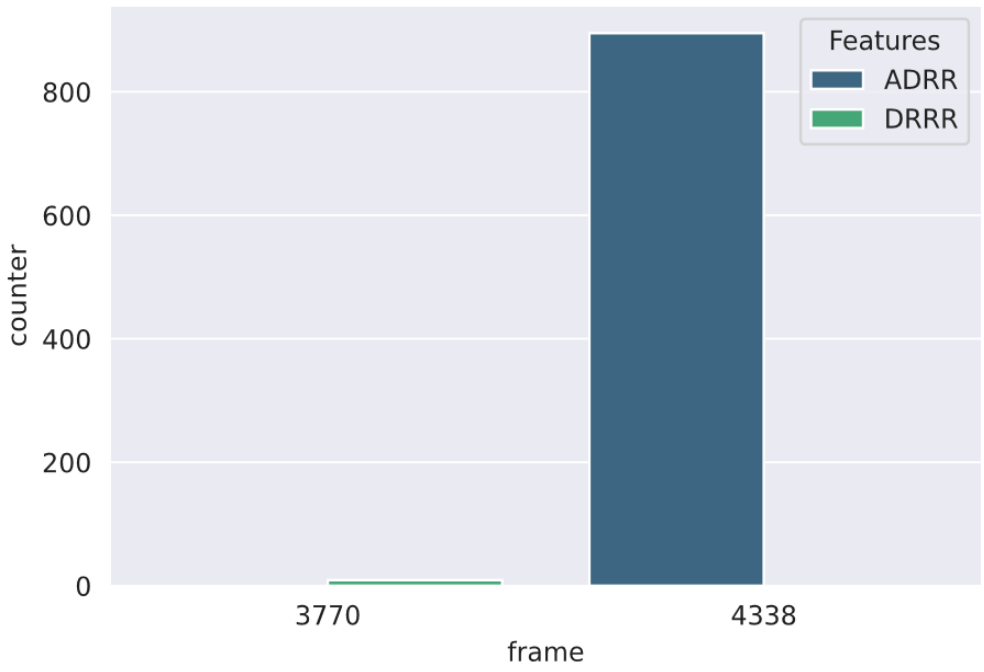


Figure 10. Bar chart comparison of top 2 four-feature hypotheses and their frequency in frames with low RMSD

(Number in the x-axis represents the frame number with the lowest RMSD value containing that hypothesis).

As a result of the aforementioned analyses, the ADDRR hypothesis generated in frame 4291 with its feature coordinates was selected for the screening of ligands that contain similar features in similar coordinates. Figure 11. displays the ligand pose of Pralsetinib bound to the RET protein with the ADDRR hypothesis overlaid on it for reference.

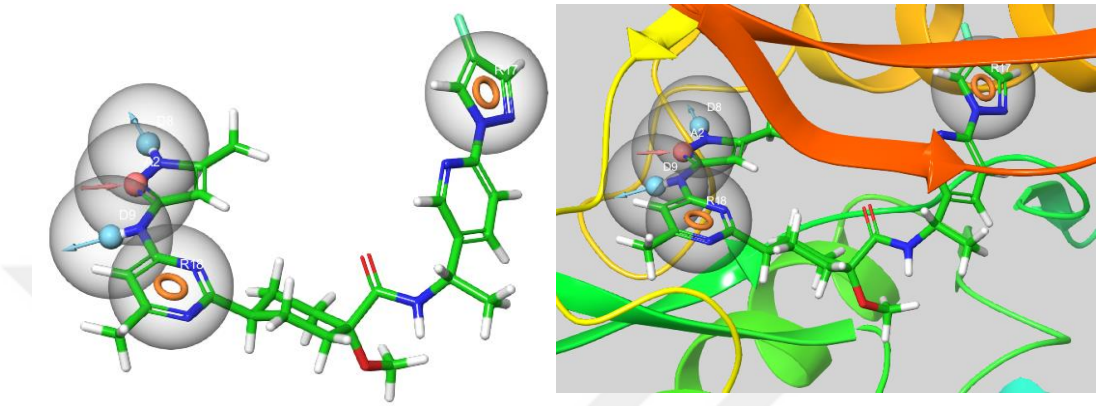


Figure 11. Matching ADDRR hypothesis on Pralsetinib in complex with RET protein.

5.2.2 Ligand Filtering Results. To implement this approach, four additional libraries were incorporated from the Chemdiv database, namely Anticancer, Indole, CMET, and FDA-approved libraries, along with the libraries obtained from text mining in the previous method. Therefore, a total of 14 libraries were screened based on the selected hypothesis. Table 21. displays the total number of compounds in each library.

Table 21.

Number of Molecules in Each Library.

Library name	Number of compounds	Library name	Number of compounds
Anticancer_Chemdiv	61538	Quinazolin	1943971
Indole_chemdiv	9753	Trifluoromethyl	13135584
CMET_chemdiv	15080	Oxazol	18873912
FDA2022_chemdiv	2452	Piperazin	23438212
[(1r,4r)-4-hydroxycyclohexyl] amino	2450	Pyrazol	59598058
Fluorophenol	87084	Pyridin	82890693
Pyrazin-2-carboxamide	808826	Piperidin	88836619

In order to screen the ligands using the hypothesis, it is necessary to convert their structures into three-dimensional representations using the ligprep tool in Maestro. However, due to the large number of molecules in eight libraries (Pyrazin-2-carboxamid, Quinazolin, Trifluoromethyl, Oxazol, Piperidin, Pyrazol, Pyridin, and Piperidin), it is important to implement filtering methods to reduce their numbers in a meaningful way. To accomplish this, a script was developed to filter these eight libraries based on Lipinski's rule of five and extract the 2D pharmacophore features of each compound. The selected molecules for screening are those that have at least one hydrogen bond acceptor (A), two hydrogen bond donors (D), and two aromatic rings (R). The resulting number of compounds after applying the filters in each library is shown in Table 22.

Table 22. *Table of Compounds Number, Before and After Applying the Filters.*

Library Name	**Total number of compounds**	**NO. of compound after Lipinski filtering**	**No. of compounds after 2D pharmacophore filtering**
Pyrazin-2-carboxamide	808826	808409	265254
Quinazolin	1943971	1943147	725334
Trifluoromethyl	13135584	13133987	1861924
Oxazol	18873912	18836970	3675265
Piperazin	23438212	23376816	2017852
Pyrazol	59598058	59315668	17328653
Pyridin	82890693	82773428	19475110
Piperidin	88836619	88355043	9750272

5.2.3 Ligand Screening and METACORE™/METADRUG ™ Results.

After converting all compounds into three-dimensional structures using Ligprep, the Ligand and Database Screening tool of PHASE was employed to fit the selected hypothesis onto the ligands. The screening process took place in two different ways:

- Screening ligands where four out of the five features of the reference hypothesis matched the ligands.

- Screening ligands where all five features of the reference hypothesis matched the ligands.

Following the fitting process, a fitness score was calculated for each ligand to assess how well the reference hypothesis aligned with each ligand. Subsequently, a normal distribution curve was generated for all the molecules in each library, and the z-score values of the fitness scores were calculated using a Python script. A cutoff value of 2 was set to select molecules, and only those with z-score values of 2 or above were further analyzed. However, if the number of compounds with z-score values above 2 was less than 20, the top 20 molecules with high fitness scores were selected instead. Then their anticancer activity scores predicted using the METACORE^{TM} platform and ligands with scores over the cutoff value, 0.5, are selected. Table 23. displays the number of compounds in each library where four reference features matched and the number of compounds in that library with z-score values above 2 and the number of compounds with anticancer activity higher than 0.5. Additionally, Table 24. illustrates the number of compounds in which all five reference features matched and had z-score values above 2 and anticancer activity scores higher than 0.5.

Table 23.

Number of Compounds in Each Library That Contain 4 Features of Total 5 with Z-Score Value Above 2 and Anticancer Activity Higher Than 0.5.

Library Name	**4 Feature Fitting (Number of compounds)**	**Z-score values ≥ 2**	**Anticancer Activity ≥ 0.5**
Anticancer_Chemdiv	3389	23	22
Indole_chemdiv	365	Top 20	14
CMET_chemdiv	1091	Top 20	18
FDA2022_chemdiv	294	Top 20	13
[(1r,4r)-4-hydroxycyclohexyl] amino	229	Top 20	19
Fluorophenol	13674	Top 20	20
Pyrazin-2-carboxamide	75199	444	444
Quinazolin	181859	521	420
Trifluoromethyl	375685	670	562
Oxazol	742089	9193	7103
Piperazin	156068	1279	853
Pyrazol	4169229	8380	5787
Pyridin	3213502	15923	13373
Piperidin	1005469	13342	9481

Table 24.

Number of Compounds in Each Library Contain All 5 Features with Z-Score Value Above 2 and Anticancer Activity Higher Than 0.5.

Library Name	5 Feature Fitting (number of compounds)	Z-score values ≥ 2	Anticancer Activity ≥ 0.5
Anticancer_Chemdiv	7	7	7
Indole_chemdiv	2	2	2
CMET_chemdiv	7	7	7
FDA2022_chemdiv	1	-	-
[(1r,4r)-4-hydroxycyclo hexyl] amino	0	-	-
Fluorophenol	0	-	-
Pyrazin-2-carboxamide	47	Top 20	19
Quinazolin	299	Top 20	20
Trifluoromethyl	36	Top 20	12
Oxazol	340	Top 20	20
Piperazin	1033	Top 20	19
Pyrazol	8332	109	89
Pyridin	5696	149	144
Piperidin	2233	74	57

5.1.1 Molecular Docking, Molecular Dynamics Simulations, and MM/GBSA Results. For the compounds in each library with anticancer activity scores higher than 0.5, glide/SP docking was performed. From each library, the top 10 molecules with the lowest docking scores were selected to proceed with molecular dynamics (MD) simulations. Additionally, for the reference molecule Pralsetinib, glide/SP docking was used to redock the ligand and compare the resulting pose with the co-crystal structure. As a result, two poses were selected as reference pose, first pose with a docking score of -9.491 (kcal/mol) and an RMSD value of 2.842 Å chosen as the reference pose for docking, and the other pose with the docking score of -8.911 (kcal/mol) and an RMSD value of 2.542 Å selected as a first pose for MD simulations because of lower RMSD value between that pose and co-crystal pose. The disparity between the redocking poses and the co-crystal pose is depicted in Figure 12. and the docking results are presented in Table 25.

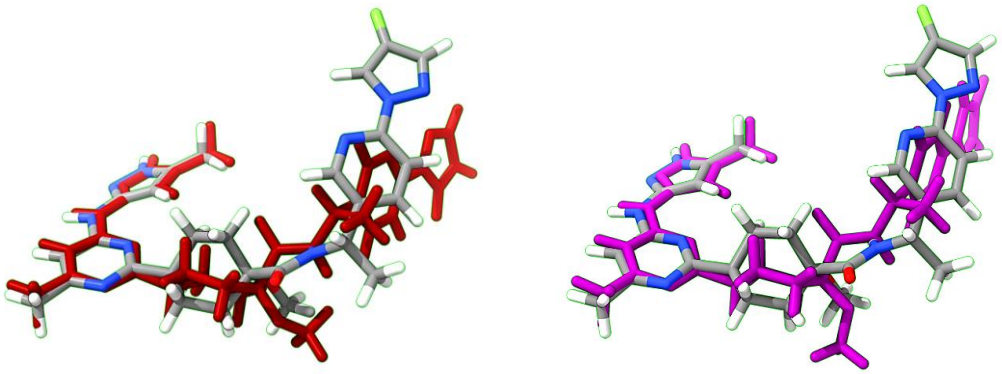


Figure 12. The three-dimensional structure of the co-crystal structure of PRALSETINIB aligns on the redocking pose with better docking score (shown in red) and with low RMSD value (shown in purple).

Table 25.

Docking Score and RMSD Values of Pralsetinib Glide/SP Redocking Poses.

	Docking score (kcal/mol)	**RMSD (Å) from cocrystal**
PRALSETINIB redocking pose	-9.491	2.842
PRALSETINIB redocking pose	-8.911	2.542

Docking scores for all ligands in each library were recorded, and three repeats of 10 ns MD simulations were conducted for the top 10 molecules from each library. From each MD simulation, 100 protein-ligand complexes were extracted, and the MM/GBSA free energy was calculated. The average MM/GBSA values were determined for each MD simulation, and the overall average was computed across the three repeats. This process was repeated for each ligand in each library, and the total averages were compiled for further analysis and selection of hit molecules. The redock pose of PRALSETINIB, which exhibited a low RMSD value, was also subjected to the same processes, and its total MM/GBSA average was set as the reference score (-71.275 kcal/mol). Using a threshold value of -65 kcal/mol for the total MM/GBSA averages, a total of 12 molecules were identified as hit molecules with values below the threshold. In Table 26. all 12 molecules are named as Hit1-Hit12 for better use in Tables and merged with their ZINC id, Then the results for these 12 hit molecules are presented in Table 27. Also, the 2D structures of 12 hit molecules and their 3D structures with fitting pharmacophores on them are illustrated in Figures 13 and 14 respectively.

Table 26.

All 12 Hit Molecules, Their ZINC ID, and Their IUPAC Names.

	Zinc ID	**IUPAC Name**
Hit1	ZINC767808756	(((1-phenyl-1H-pyrazol-4-yl)methyl)carbamoyl)methyl 3-(1H-1,2,4-triazol-3-yl)benzoate
Hit2	ZINC364815311	2-[4-(pyridin-2-yloxy)phenoxy]-N-[2-(1H-1,2,4-triazol-3-yl)phenyl] acetamide
Hit3	ZINC476943771	3-(5-chlorothiophen-2-yl)-N-{3-[(2-hydroxyethyl)sulfanyl]-1H-1,2,4-triazol-5-yl}-1-methyl-1H-pyrazole-4-carboxamide
Hit4	ZINC513956406	3-oxo-N-[(1R)-1-{4-[(phenylcarbamoyl)amino]phenyl}ethyl]-2H,3H-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide
Hit5	ZINC513768922	3-oxo-N-{3-[2-(pyridin-2-yl)ethynyl]phenyl}-2H,3H-[1,2,4] triazolo [4,3-a]pyridine-6-carboxamide
Hit6	ZINC96309479	(S)-1-(4-methylbenzyl)-3-(1-(3-(pyridin-4-yl)-1H-pyrazol-5-yl) piperidin-3-yl) urea
Hit7	ZINC96309483	(S)-1-(4-fluorobenzyl)-3-(1-(3-(pyridin-4-yl)-1H-pyrazol-5-yl) piperidin-3-yl) urea
Hit8	ZINC257289132	1-(4-{6-[(5-methyl-1H-pyrazol-3-yl) amino]pyridin-2-yl}piperidin-1-yl)-3-phenylpropan-1-one
Hit9	ZINC9677876	5-amino-N-(3,4-dimethoxyphenethyl)-1-(2-((2-ethoxyphenyl) amino)-2-oxoethyl)-1H-1,2,3-triazole-4-carboxamide
Hit10	ZINC792285195	N'-{3-[(1R)-1,2-dihydroxyethyl] phenyl}-N-[1-(4-methylphenyl)-5-(propan-2-yl)-1H-pyrazol-4-yl]ethanediamide
Hit11	ZINC44892020	(S)-2-(2,5-dioxo-2,3,4,5-tetrahydro-1H-benzo[e][1,4] diazepin-3-yl)-N-(5-(4-methoxybenzyl)-1,3,4-thiadiazol-2-yl)acetamide
Hit12	ZINC64449182	4-((6,7-dimethoxy-2,4-dioxo-1,4-dihydroquinazolin-3(2H)-yl) methyl)-N-(4-fluorobenzyl) benzamide

Table 27.

Results Determined For 12 Hit Molecules, Docking Scores, MM/GBSA Averages, Anticancer Activity Scores, and Fitness Scores.

Name	ZINC ID	Docking Score (kcal/mol)	Total MM/GBSA average (kcal/mol)	Anticancer activity score	Fitness Score
PRALSETINIB	-	-9.491	-71.276 ± 4.44	0.75	-
Hit1	ZINC767808756	-9.924	-68.368 ± 3.25	0.57	1.838
Hit2	ZINC364815311	-9.71	-69.106 ± 3.05	0.72	1.830
Hit3	ZINC476943771	-10.085	-65.675 ± 3.22	0.68	1.752
Hit4	ZINC513956406	-9.861	-75.633 ± 4.0	0.69	1.699
Hit5	ZINC513768922	-10.095	-65.365 ± 3.20	0.7	1.659
Hit6	ZINC96309479	-6.931	-68.721 ± 4.46	0.78	1.656
Hit7	ZINC96309483	-6.587	-66.030 ± 4.36	0.77	1.656
Hit8	ZINC257289132	-9.449	-65.425 ± 2.81	0.63	1.729
Hit9	ZINC9677876	-5.964	-65.444 ± 4.28	0.67	0.296
Hit10	ZINC792285195	-9.051	-66.867 ± 6.12	0.62	1.610
Hit11	ZINC44892020	-6.777	-70.844 ± 4.64	0.69	1.674
Hit12	ZINC64449182	-7.323	-65.163 ± 7.22	0.68	1.581

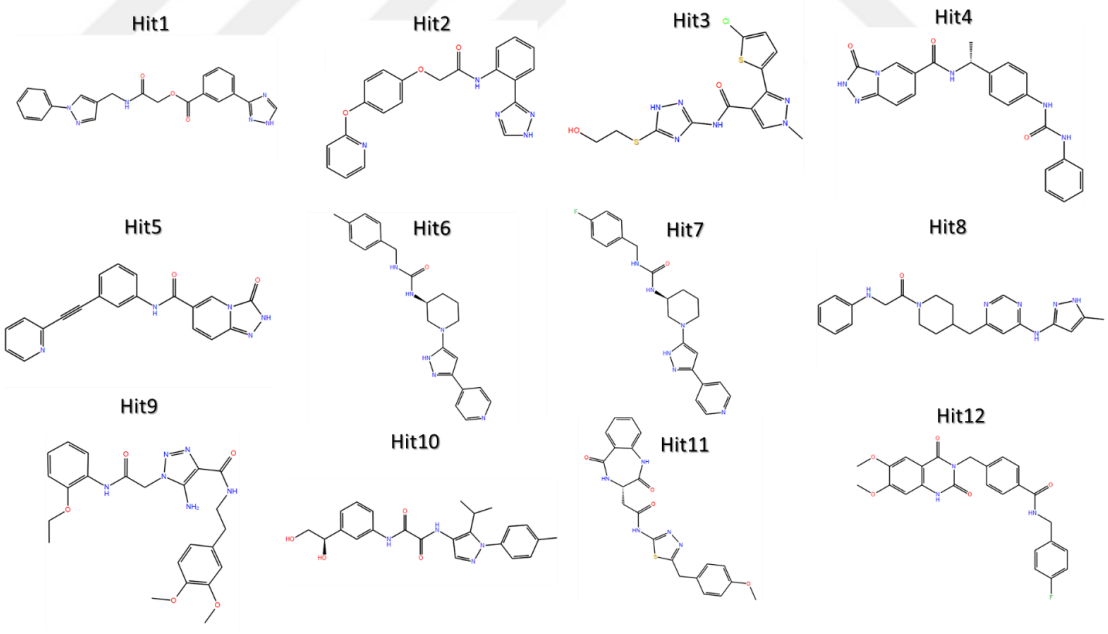


Figure 13. Two-dimensional structures of 12 hit molecules.

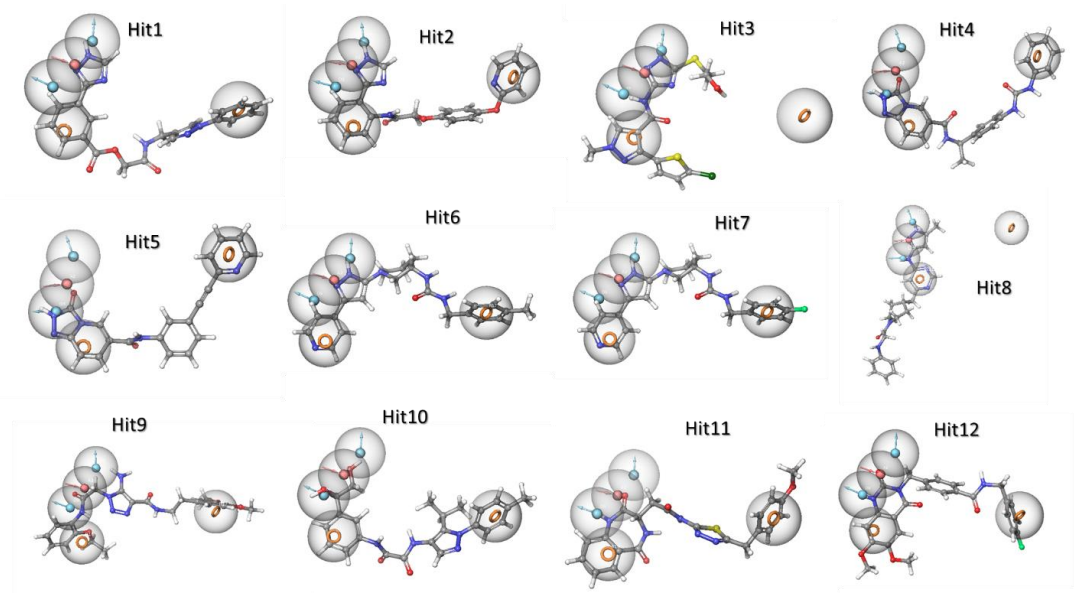


Figure 14. Three-dimensional structures of 12 hit molecules with pharmacophore features fit on them.

For further analysis, 12 hit molecules and the reference molecule underwent 200 ns MD simulations, collecting a total of 2000 frames. The purpose was to investigate the behavior of the ligands and their interactions over an extended simulation time. The results are presented in Table 28. indicate that the MM/GBSA average for the Hit10 ligand significantly increased, the MM/GBSA average of this ligand was -66.867 kcal/mol during the 10 ns MD simulation but it increases to -48.861 kcal/mol by increasing the simulation time to 200 ns. This change in MM/GBSA average suggested weaker protein-ligand interactions and less favorable binding affinity as the simulation time increased. This ligand may encounter challenges in establishing stable interactions within the protein's binding site. Conversely, there were Hit9 ligands in which the MM/GBSA average decreased from -65.444 kcal/mol to -79.447 kcal/mol, indicating more favorable interactions and potentially stronger binding affinity.

Table 28.

Total MM/GBSA Average of 12 Hit Molecules and PRALSETINIB During 10 ns and 200 ns MD Simulations.

Name	Total MM/GBSA average (kcal/mol) (10 ns)	MM/GBSA average (kcal/mol) (200 ns)
PRALSETINIB	-71.276 ± 4.44	-75.112 ± 3.25
Hit1	-68.368 ± 3.25	-67.845 ± 3.28
Hit2	-69.106 ± 3.05	-70.020 ± 3.49
Hit3	-65.675 ± 3.22	-61.818 ± 4.10
Hit4	-75.633 ± 4.0	-71.679 ± 3.79
Hit5	-65.365 ± 3.20	-65.700 ± 3.14
Hit6	-68.721 ± 4.46	-61.846 ± 4.73
Hit7	-66.030 ± 4.36	-71.732 ± 3.35
Hit8	-65.425 ± 2.81	-62.482 ± 2.86
Hit9	-65.444 ± 4.28	-79.447 ± 6.37
Hit10	-66.867 ± 6.12	-48.861 ± 6.67
Hit11	-70.844 ± 4.64	-63.103 ± 6.41
Hit12	-65.163 ± 7.22	-52.210 ± 8.00

5.2 Comparative Analysis of Hit Molecules Obtained from Two Methods

We present an analysis of selected molecules obtained from two methods and their interactions with mutations on the RET protein. Following a stringent threshold of -65 kcal/mol for the MM/GBSA average, none of the molecules from the first method met the criteria. However, from the second method, 12 hit molecules exhibited an MM/GBSA average above the threshold. Detailed analysis of the molecular interactions during MD simulations was conducted, and the findings are illustrated through interaction diagrams and histograms (Figure 16). According to Figure 15. Pralsetinib exhibits strong hydrogen bonding interactions with Ala 807 and Glu 805 residues. It also demonstrates favorable hydrophobic interactions with Leu 730, Val 738, Ala 756, Lys 758, Val 804, and Leu 881 residues. Additionally, Figure 17. provides information about the RMSD (Root Mean Square Deviation) of the ligand while it is fitted onto the protein and presents probability distributions using normal curves for the 12 hit molecules and PRALSETINIB. Analysis of the Figure reveals that certain molecules have low RMSD probabilities, indicating a wide range of RMSD values and instability during the simulation.

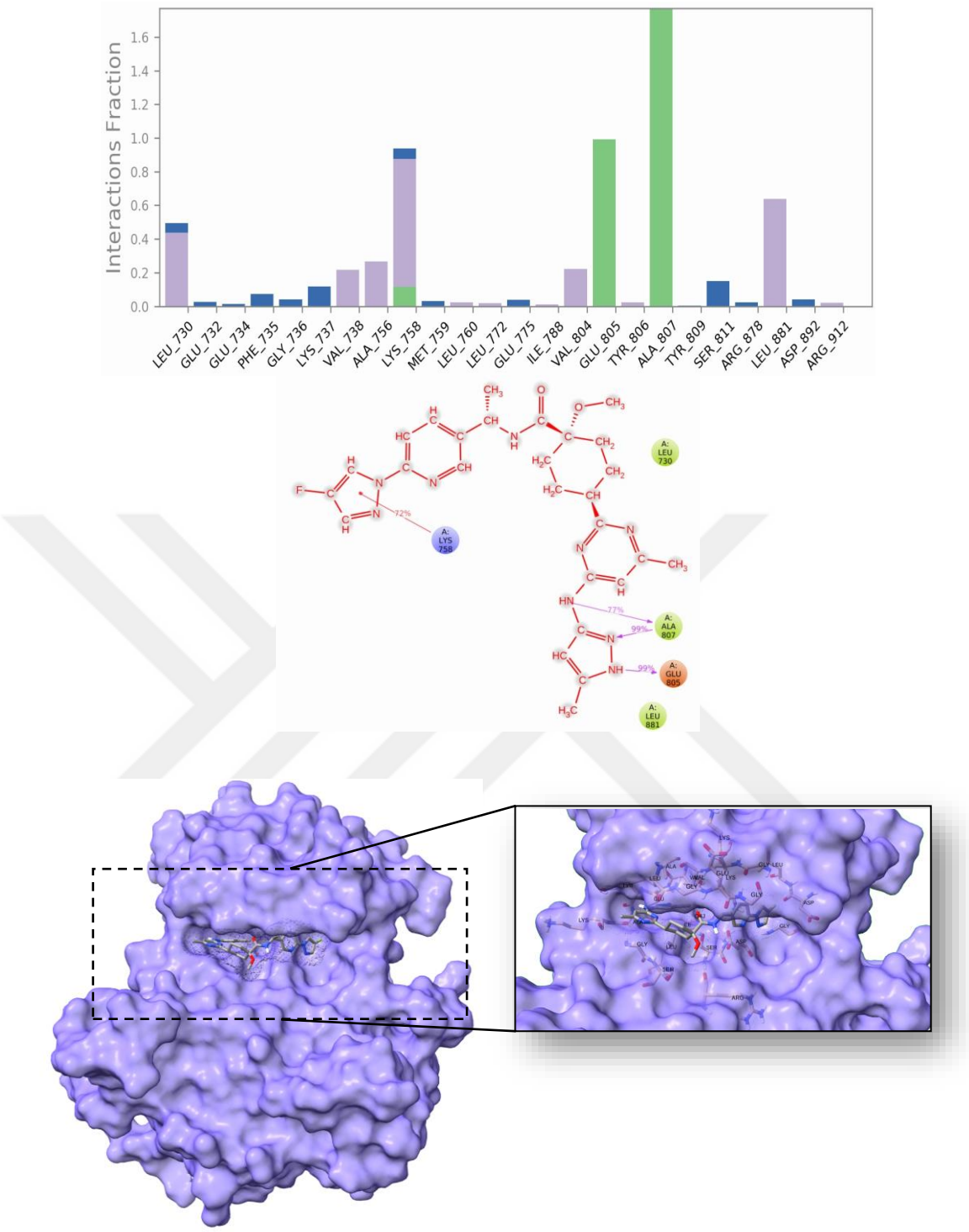


Figure 15. Interaction diagram and histogram of PRALSETINIB during 200 ns MD simulation

(Green, blue, and purple colors represent the H-bonds, water bridges, and hydrophobic interactions respectively). The surface view of the wildtype RET protein in complex with Pralsetinib is depicted at the bottom. The residues within the binding site of the protein have been labeled for clarity.

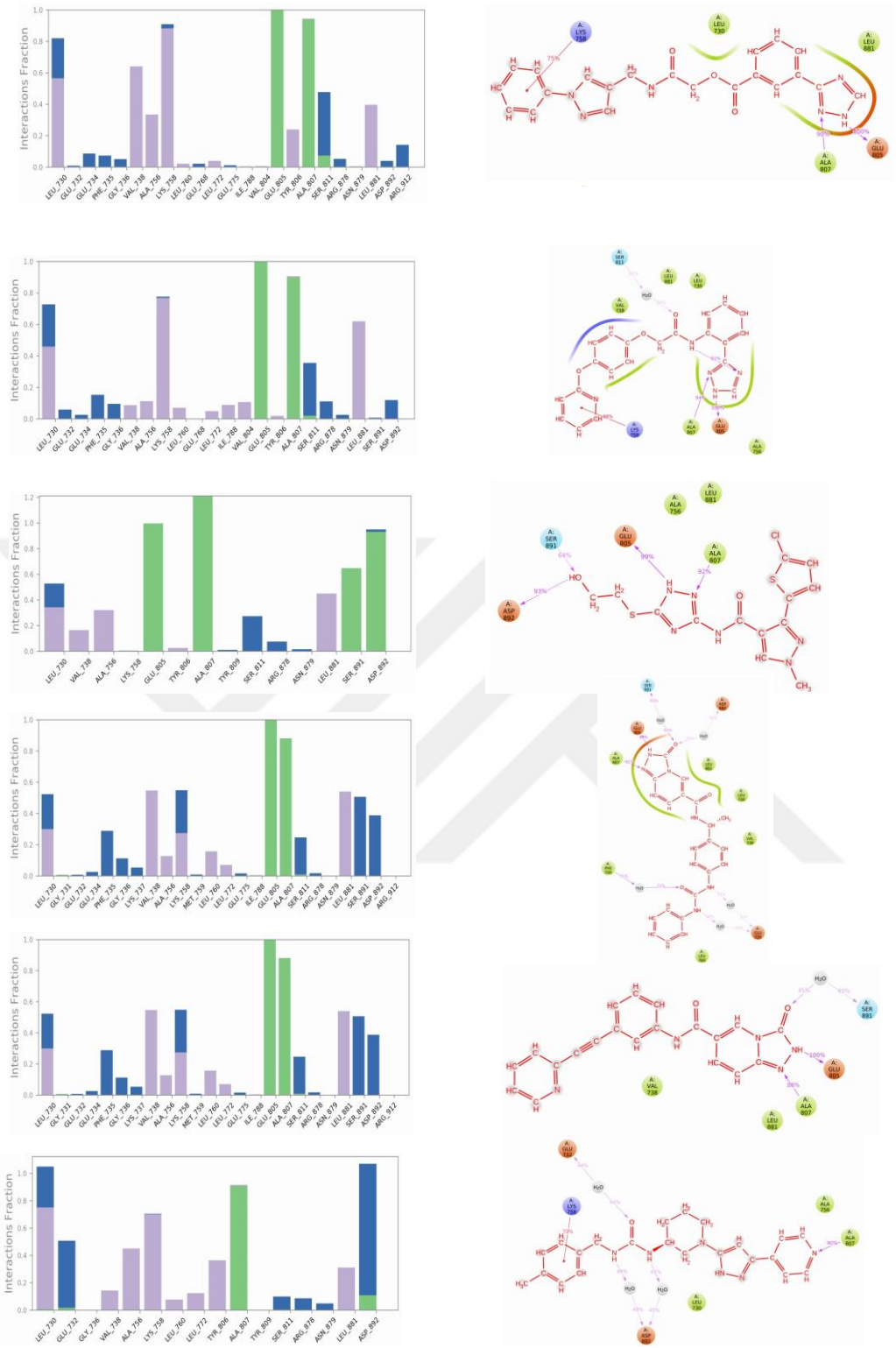


Figure 16. Interaction diagram and histogram of 12 hit molecules during 200 ns MD simulation
(Hit1-Hit12 from top to bottom respectively), (green, blue, purple, and pink colors represent the H-bonds, water bridges, and hydrophobic and ionic interactions respectively).

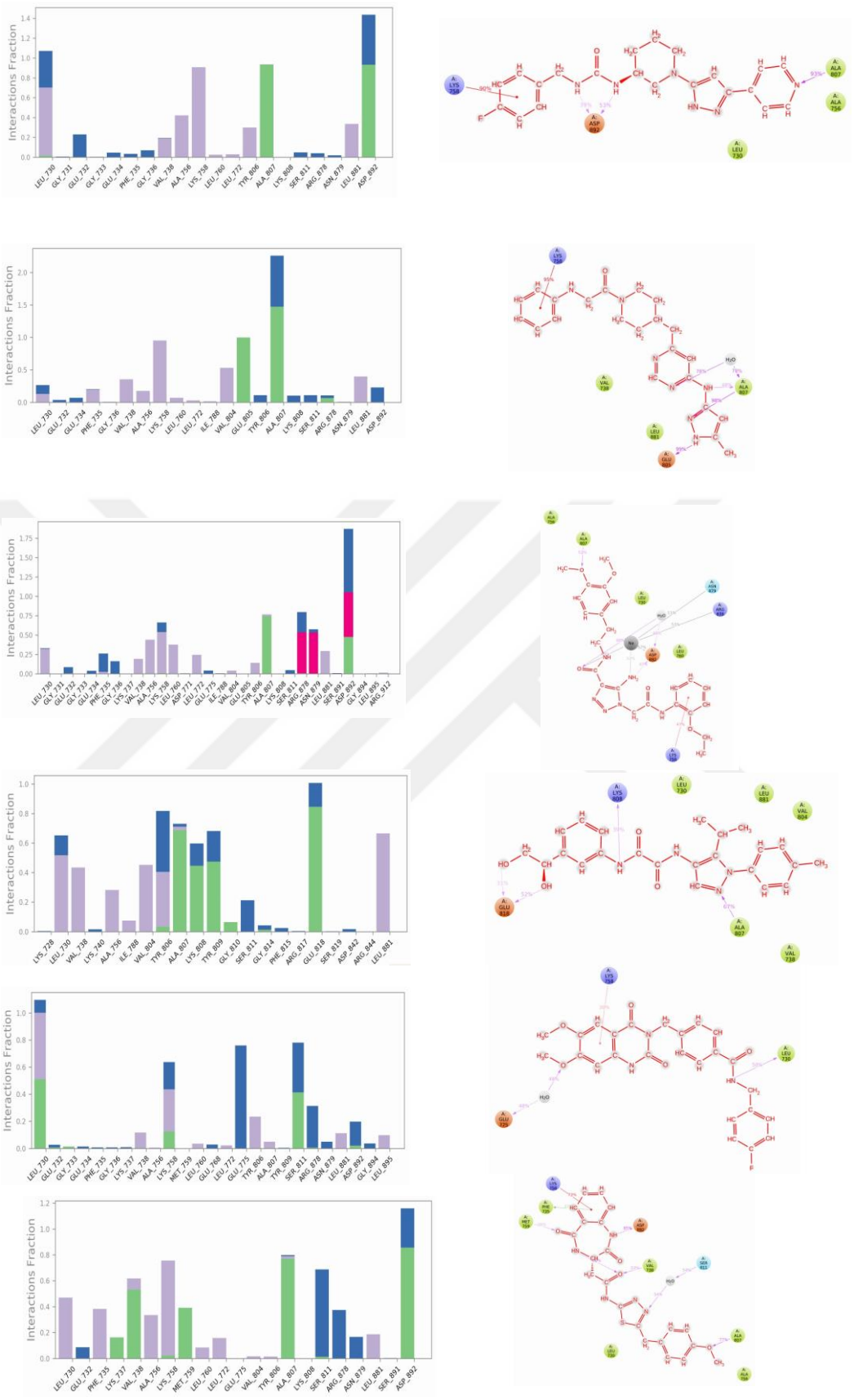


Figure 16 *(Cont.d)*

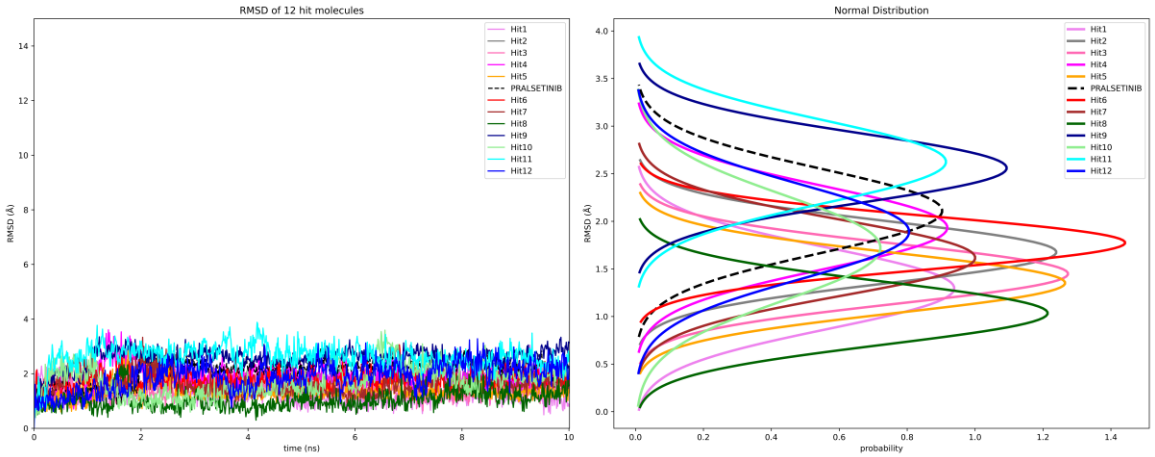


Figure 17. Lig fit prot RMSD and normal distribution probability plots of Pralsetinib and 12 hit molecules during 10 ns MD simulation in wildtype RET protein.

5.2.1 Molecular Docking, MD Simulations and MM/GBSA Results for Mutated Proteins.

The 12 hit molecules were individually combined with eleven mutated RET proteins, and the glide/SP docking method was performed for each combination. The docking poses obtained were used as initial poses for 10 ns molecular dynamics (MD) simulations, and the MM/GBSA free energy values were calculated. The differences in MM/GBSA average values (ΔΔG) between the wildtype RET and each mutated protein were calculated for each hit molecule. If the MM/GBSA average value of a hit molecule was higher than -60 kcal/mol in a mutated protein, it was considered as showing resistance to that mutation. Eight single mutations and three double mutations were identified as important based on the literature. During the analysis of single mutations, it was found that Hit1 and Hit9 molecules had MM/GBSA values lower than -60 kcal/mol, indicating that they did not show resistance to these mutations. Similarly, in the analysis of double mutations, Hit4, Hit9, and Hit12 molecules had MM/GBSA averages lower than -60 kcal/mol, suggesting that they may not show resistance to double mutations. Notably, the Hit9 molecule was found to be common in both the single mutations and double mutations categories. Graph depicting the RMSD of the ligand during protein fitting and probability distributions using normal curves (Figure 18.) supported these findings. The Hit9 molecule exhibited stable RMSD values throughout the MD simulations, particularly in simulations with mutated proteins. Moreover, the MM/GBSA averages and ΔΔG values for hit molecules are given in Tables 29-39.

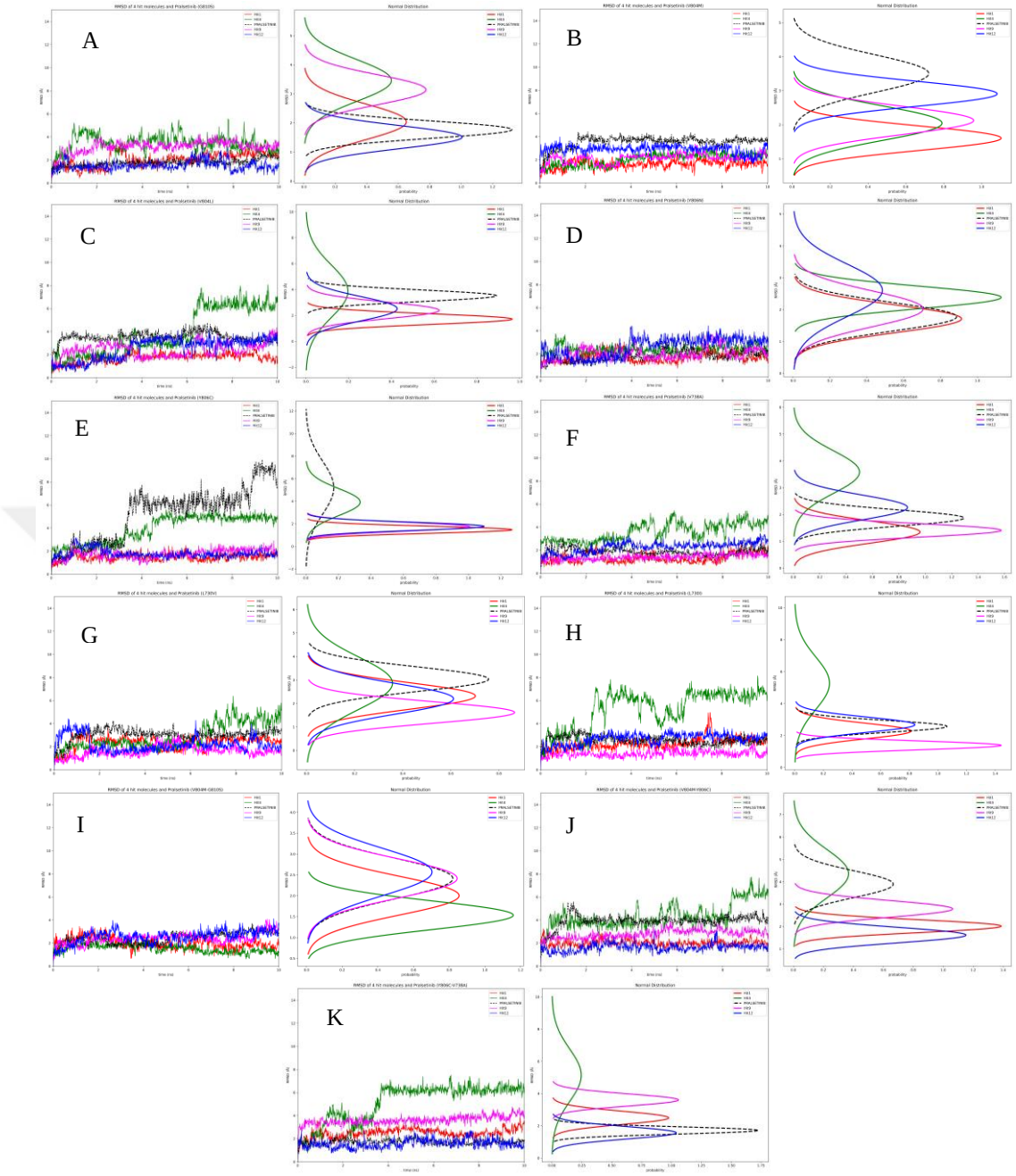


Figure 18. Lig Fit Prot graph of RMSD (Å)/time (ns) for 10 ns simulations and normal distribution probability plots

Graphs for PRALSETINIB and Hit1, Hit4, Hit9, and Hit12 compounds on RET protein with (A) G810S, (B) V804M, (C) V804L, (D) Y806N, (E) Y806C, (F) V738A, (G) L730V, (H) L730I, (I) V804M-G810S, (J) V804M-Y806C, (K) Y806C-V738A mutations. (green, blue, black, red, and magenta colors represent Hit4, Hit12, PRALSETINIB, Hit1, and Hit9 respectively)

From the data presented in Table 29. it can be observed that the mutation of residue 810 from Glycine to Serine leads to an increase in the docking score of Pralsetinib from -9.491 kcal/mol to -5.952 kcal/mol. This indicates that the G810S

mutation decreases the binding affinity of Pralsetinib to the RET protein. However, the MM/GBSA average value does not show a significant change, suggesting that the mutation may have a limited impact on the overall binding energy and stability of the complex. Figure 19. which displays the interaction histograms, reveals that only Glu 775 residue forms a hydrogen bond with Pralsetinib, while Leu 730 and Val 738 residues contribute to hydrophobic interactions. Similar results are observed for other mutations such as V804L/M, Y806C/N, L730V, V804M/Y806C, and Y806C/V738A, indicating low interactions during the MD simulations as shown in Figure 19. On the other hand, mutations such as V738A, L730I, and V804M/G810S result in an increase in the docking score but a decrease in the MM/GBSA average value. Furthermore, it is evident that during the mutations, Pralsetinib forms hydrogen bonds with different residues compared to those in the wildtype structure. Interestingly, in some cases, there are no hydrogen bonds observed between Pralsetinib and the mutated RET protein during MD simulations. It has been observed that the positioning of Pralsetinib alters notably when it is docked onto the majority of mutated RET proteins. Appendix A.1 provides a visual representation of the various docking positions of Pralsetinib in RET with different mutations.

Table 29.

Docking Scores, MM/GBSA Averages, And ΔΔG Results of PRALSETINIB And 12 Hit Molecules for RET Protein with G810S Mutation.

	G810S			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-5.952	-71.276 ± 4.44	-70.422 ± 5.88	0.853
Hit1	-9.381	-68.368 ± 3.25	-66.735 ± 4.97	1.633
Hit2	-6.909	-69.106 ± 3.05	-65.591 ± 3.69	3.515
Hit3	-5.837	-65.675 ± 3.22	-50.653 ± 3.94	15.022
Hit4	-8.233	-75.633 ± 4.0	-65.409 ± 4.74	10.224
Hit5	-8.616	-65.365 ± 3.20	-65.181 ± 3.27	0.184
Hit6	-7.404	-68.721 ± 4.46	-66.453 ± 4.64	2.269
Hit7	-7.173	-66.030 ± 4.36	-48.914 ± 7.01	17.116
Hit8	-8.962	-65.425 ± 2.81	-65.990 ± 4.19	-0.565
Hit9	-6.769	-65.444 ± 4.28	-70.107 ± 4.27	-4.663
Hit10	-6.746	-66.867 ± 6.12	-59.493 ± 3.59	7.374
Hit11	-7.237	-70.844 ± 4.64	-69.080 ± 2.84	1.763
Hit12	-7.986	-65.163 ± 7.22	-70.565 ± 4.45	-5.401

Table 30.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with V804L Mutation.

	V804L			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-6.442	-71.276 ± 4.44	-67.371 ± 5.28	3.905
Hit1	-8.872	-68.368 ± 3.25	-68.290 ± 3.35	0.078
Hit2	-8.336	-69.106 ± 3.05	-70.621 ± 3.30	-1.515
Hit3	-8.153	-65.675 ± 3.22	-61.651 ± 4.04	4.025
Hit4	-7.701	-75.633 ± 4.0	-68.167 ± 5.71	7.467
Hit5	-8.277	-65.365 ± 3.20	-65.523 ± 3.11	-0.158
Hit6	-7.156	-68.721 ± 4.46	-63.568 ± 5.67	5.153
Hit7	-6.419	-66.030 ± 4.36	-63.060 ± 4.20	2.970
Hit8	-9.019	-65.425 ± 2.81	-65.581 ± 3.14	-0.156
Hit9	-7.276	-65.444 ± 4.28	-67.536 ± 7.30	-2.092
Hit10	-7.054	-66.867 ± 6.12	-55.696 ± 4.48	11.171
Hit11	-6.845	-70.844 ± 4.64	-64.617 ± 4.87	6.227
Hit12	-6.852	-65.163 ± 7.22	-57.444 ± 6.11	7.719

Table 31.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with V804M Mutation.

	V804M			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-6.386	-71.276 ± 4.44	-64.583 ± 3.32	6.693
Hit1	-9.384	-68.368 ± 3.25	-69.399 ± 3.55	-1.032
Hit2	-6.793	-69.106 ± 3.05	-64.775 ± 4.44	4.331
Hit3	-7.718	-65.675 ± 3.22	-66.271 ± 3.76	-0.596
Hit4	-8.775	-75.633 ± 4.0	-69.430 ± 3.85	6.203
Hit5	-7.994	-65.365 ± 3.20	-64.623 ± 2.79	0.742
Hit6	-7.058	-68.721 ± 4.46	-59.632 ± 3.11	9.089
Hit7	-7.527	-66.030 ± 4.36	-70.322 ± 5.54	-4.292
Hit8	-7.404	-65.425 ± 2.81	-64.803 ± 3.31	0.622
Hit9	-7.189	-65.444 ± 4.28	-71.029 ± 6.37	-5.585
Hit10	-8.19	-66.867 ± 6.12	-59.358 ± 8.94	7.509
Hit11	-6.764	-70.844 ± 4.64	-65.557 ± 4.45	5.286
Hit12	-6.784	-65.163 ± 7.22	-58.197 ± 4.36	6.966

Table 32.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with Y806N Mutation.

	Y806N			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-9.172	-71.276 ± 4.44	-70.527 ± 3.27	0.749
Hit1	-9.816	-68.368 ± 3.25	-70.163 ± 2.99	-1.796
Hit2	-9.154	-69.106 ± 3.05	-67.011 ± 3.19	2.094
Hit3	-8.361	-65.675 ± 3.22	-59.176 ± 4.83	6.499
Hit4	-9.045	-75.633 ± 4.0	-64.184 ± 3.59	11.449
Hit5	-8.692	-65.365 ± 3.20	-59.488 ± 3.63	5.877
Hit6	-7.29	-68.721 ± 4.46	-57.580 ± 4.77	11.141
Hit7	-7.595	-66.030 ± 4.36	-59.660 ± 4.91	6.370
Hit8	-6.366	-65.425 ± 2.81	-59.499 ± 4.20	5.926
Hit9	-7.393	-65.444 ± 4.28	-65.937 ± 6.03	-0.493
Hit10	-6.603	-66.867 ± 6.12	-47.763 ± 4.31	19.104
Hit11	-6.817	-70.844 ± 4.64	-56.535 ± 7.18	14.309
Hit12	-7.002	-65.163 ± 7.22	-59.317 ± 5.91	5.846

Table 33.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with Y806C Mutation.

	Y806C			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-7.846	-71.276 ± 4.44	-55.282 ± 7.17	15.994
Hit1	-8.333	-68.368 ± 3.25	-68.106 ± 3.18	0.262
Hit2	-8.415	-69.106 ± 3.05	-68.405 ± 2.93	0.700
Hit3	-5.552	-65.675 ± 3.22	-54.615 ± 5.84	11.061
Hit4	-8.024	-75.633 ± 4.0	-66.074 ± 4.29	9.560
Hit5	-6.211	-65.365 ± 3.20	-61.694 ± 3.93	3.671
Hit6	-6.538	-68.721 ± 4.46	-44.831 ± 5.16	23.890
Hit7	-6.579	-66.030 ± 4.36	-50.145 ± 6.71	15.885
Hit8	-6.711	-65.425 ± 2.81	-62.577 ± 3.27	2.848
Hit9	-6.55	-65.444 ± 4.28	-69.954 ± 4.81	-4.510
Hit10	-6.294	-66.867 ± 6.12	-58.217 ± 3.95	8.650
Hit11	-6.39	-70.844 ± 4.64	-56.191 ± 5.52	14.653
Hit12	-7.094	-65.163 ± 7.22	-69.849 ± 4.58	-4.686

Table 34.

Docking Scores, MM/GBSA Averages and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with V738A Mutation.

	V738A			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-8.557	-71.276 ± 4.44	-74.830 ± 3.47	-3.555
Hit1	-9.108	-68.368 ± 3.25	-67.224 ± 3.97	1.144
Hit2	-7.698	-69.106 ± 3.05	-66.344 ± 3.97	2.761
Hit3	-6.134	-65.675 ± 3.22	-53.235 ± 2.98	12.440
Hit4	-8.246	-75.633 ± 4.0	-57.834 ± 8.41	17.799
Hit5	-9.112	-65.365 ± 3.20	-62.706 ± 2.51	2.659
Hit6	-6.994	-68.721 ± 4.46	-67.042 ± 4.24	1.679
Hit7	-7.412	-66.030 ± 4.36	-68.318 ± 4.22	-2.288
Hit8	-8.66	-65.425 ± 2.81	-62.871 ± 2.96	2.554
Hit9	-6.199	-65.444 ± 4.28	-89.540 ± 6.13	-24.096
Hit10	-5.843	-66.867 ± 6.12	-61.954 ± 6.70	4.913
Hit11	-6.062	-70.844 ± 4.64	-60.636 ± 6.75	10.208
Hit12	-6.171	-65.163 ± 7.22	-67.527 ± 5.25	-2.364

Table 35.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with L730V Mutation.

	L730V			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-6.432	-71.276 ± 4.44	-66.568 ± 3.46	4.708
Hit1	-9.396	-68.368 ± 3.25	-72.185 ± 3.68	-3.817
Hit2	-8.139	-69.106 ± 3.05	-72.727 ± 3.98	-3.621
Hit3	-7.162	-65.675 ± 3.22	-61.621 ± 3.85	4.054
Hit4	-7.541	-75.633 ± 4.0	-60.969 ± 5.32	14.664
Hit5	-8.076	-65.365 ± 3.20	-65.049 ± 2.95	0.316
Hit6	-7.136	-68.721 ± 4.46	-58.776 ± 4.56	9.946
Hit7	-6.011	-66.030 ± 4.36	-50.923 ± 4.20	15.106
Hit8	-7.343	-65.425 ± 2.81	-65.154 ± 3.03	0.271
Hit9	-6.693	-65.444 ± 4.28	-66.019 ± 5.02	-0.575
Hit10	-6.325	-66.867 ± 6.12	-43.242 ± 6.87	23.625
Hit11	-6.058	-70.844 ± 4.64	-62.419 ± 7.00	8.425
Hit12	-7.196	-65.163 ± 7.22	-64.220 ± 5.39	0.944

Table 36.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with L730I Mutation.

	L730I			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average kcal/mol**	**ΔΔG**
PRALSETINIB	-5.789	-71.276 ± 4.44	-74.972 ± 4.42	-3.696
Hit1	-7.164	-68.368 ± 3.25	-64.213 ± 4.51	4.155
Hit2	-5.452	-69.106 ± 3.05	-43.788 ± 5.10	25.317
Hit3	-5.711	-65.675 ± 3.22	-55.680 ± 4.56	9.996
Hit4	-6.927	-75.633 ± 4.0	-57.996 ± 3.65	17.637
Hit5	-6.738	-65.365 ± 3.20	-62.848 ± 2.86	2.517
Hit6	-7.165	-68.721 ± 4.46	-60.419 ± 3.26	8.303
Hit7	-6.136	-66.030 ± 4.36	-71.463 ± 3.79	-5.434
Hit8	-6.444	-65.425 ± 2.81	-51.096 ± 6.34	14.329
Hit9	-5.108	-65.444 ± 4.28	-68.327 ± 5.64	-2.882
Hit10	-5.672	-66.867 ± 6.12	-56.270 ± 4.13	10.597
Hit11	-5.451	-70.844 ± 4.64	-50.242 ± 5.97	20.601
Hit12	-5.494	-65.163 ± 7.22	-64.355 ± 5.04	0.808

Table 37.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with Y806C-V738A Mutations.

	Y806C-V738A			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average (kcal/mol)**	**ΔΔG**
PRALSETINIB	-9.341	-71.276 ± 4.44	-69.866 ± 3.58	1.410
Hit1	-6.038	-68.368 ± 3.25	-52.464 ± 4.19	15.904
Hit2	-6.93	-69.106 ± 3.05	-58.725 ± 7.23	10.381
Hit3	-6.204	-65.675 ± 3.22	-64.416 ± 3.87	1.259
Hit4	-7.739	-75.633 ± 4.0	-65.313 ± 3.77	10.320
Hit5	-9.642	-65.365 ± 3.20	-61.160 ± 2.58	4.205
Hit6	-5.985	-68.721 ± 4.46	-71.835 ± 4.58	-3.113
Hit7	-5.996	-66.030 ± 4.36	-49.785 ± 4.58	16.245
Hit8	-8.751	-65.425 ± 2.81	-63.732 ± 3.33	1.693
Hit9	-5.593	-65.444 ± 4.28	-73.889 ± 5.63	-8.445
Hit10	-6.144	-66.867 ± 6.12	-48.664 ± 6.30	18.203
Hit11	-6.275	-70.844 ± 4.64	-60.156 ± 6.30	10.688
Hit12	-7.31	-65.163 ± 7.22	-62.607 ± 4.65	2.557

Table 38.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with V804M-Y806C Mutations.

	V804M-Y806C			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average (kcal/mol)**	**ΔΔG**
PRALSETINIB	-6.33	-71.276 ± 4.44	-60.055 ± 3.29	11.220
Hit1	-7.581	-68.368 ± 3.25	-61.190 ± 3.02	7.178
Hit2	-5.372	-69.106 ± 3.05	-44.059 ± 4.37	25.047
Hit3	-7.292	-65.675 ± 3.22	-57.153 ± 4.9	8.522
Hit4	-8.266	-75.633 ± 4.0	-60.286 ± 3.60	15.348
Hit5	-6.823	-65.365 ± 3.20	-59.493 ± 3.07	5.872
Hit6	-7.3	-68.721 ± 4.46	-66.299 ± 5.75	2.422
Hit7	-7.657	-66.030 ± 4.36	-69.789 ± 3.80	-3.760
Hit8	-6.739	-65.425 ± 2.81	-56.402 ± 5.64	9.023
Hit9	-6.5	-65.444 ± 4.28	-65.264 ± 5.32	0.180
Hit10	-6.27	-66.867 ± 6.12	-60.695 ± 4.15	6.172
Hit11	-6.783	-70.844 ± 4.64	-54.689 ± 6.01	16.155
Hit12	-7.368	-65.163 ± 7.22	-64.793 ± 3.51	0.370

Table 39.

Docking Scores, MM/GBSA Averages, and ΔΔG Results of PRALSETINIB and 12 Hit Molecules for RET Protein with V804M- G810S Mutations.

	V804M-G810S			
Name	**Docking Score kcal/mol**	**Wildtype MM/GBSA average (kcal/mol)**	**MM/GBSA average (kcal/mol)**	**ΔΔG**
PRALSETINIB	-6.231	-71.276 ± 4.44	-73.890 ± 4.63	-2.614
Hit1	-7.398	-68.368 ± 3.25	-65.635 ± 5.44	2.733
Hit2	-6.867	-69.106 ± 3.05	-63.212 ± 4.97	5.894
Hit3	-7.524	-65.675 ± 3.22	-61.878 ± 4.74	3.797
Hit4	-7.861	-75.633 ± 4.0	-73.985 ± 3.36	1.649
Hit5	-7.063	-65.365 ± 3.20	-61.412 ± 3.50	3.953
Hit6	-6.6	-68.721 ± 4.46	-52.713 ± 8.08	16.008
Hit7	-6.558	-66.030 ± 4.36	-63.063 ± 4.98	2.967
Hit8	-7.285	-65.425 ± 2.81	-56.958 ± 4.16	8.467
Hit9	-6.628	-65.444 ± 4.28	-64.061 ± 6.55	1.383
Hit10	-6.54	-66.867 ± 6.12	-56.792 ± 8.07	10.075
Hit11	-6.395	-70.844 ± 4.64	-61.382 ± 11.63	9.461
Hit12	-5.971	-65.163 ± 7.22	-62.987 ± 5.94	2.177

Based on the aforementioned tables, it can be observed that the Hit9 molecule, selected as a promising compound capable of forming favorable interactions with both wildtype and mutated RET, exhibits a docking score of -5.96 kcal/mol and an MM/GBSA average value of -65.163 kcal/mol in the wildtype RET. The results shown in Figures 19 and 20 indicate that during various mutations, both the docking scores

and MM/GBSA average values of Hit9 decrease, and ΔΔG value of that is mostly negative during all mutations. Furthermore, Figure 21 illustrates the interaction histogram of the Hit9 molecule during different mutations. In the wildtype RET, Hit9 forms hydrogen bonds with Ala 807 and Asp 892, engages in hydrophobic interactions with Leu 730, Ala 756, Lys 756, and Leu 760 residues, and participates in ionic interactions with Arg 878, Asn 879, and Asp 892 residues. Interestingly, during RET mutations, the Hit9 molecule maintains hydrogen bond formation with Ala 807, Asp 892 residues, or both, similar to its interaction pattern in the wildtype. Additionally, hydrophobic interactions with Leu 730 and Lys 756 residues persist in most of the mutations. The stability of these interactions in both wildtype and mutated RET suggests that Hit9 may not exhibit resistance during these mutations. Also, as shown in appendix A.2, the docking pose of the Hit9 molecule while docking on mutated RET doesn't show significant differences in comparison with the pose of it on wildtype RET. Moreover, it appears to be more favorable than Pralsetinib in inhibiting the RET signaling pathway during various mutations. This is supported by the fact that Pralsetinib, in some mutations, fails to form hydrogen bonds and exhibits completely different binding poses compared to the wildtype.

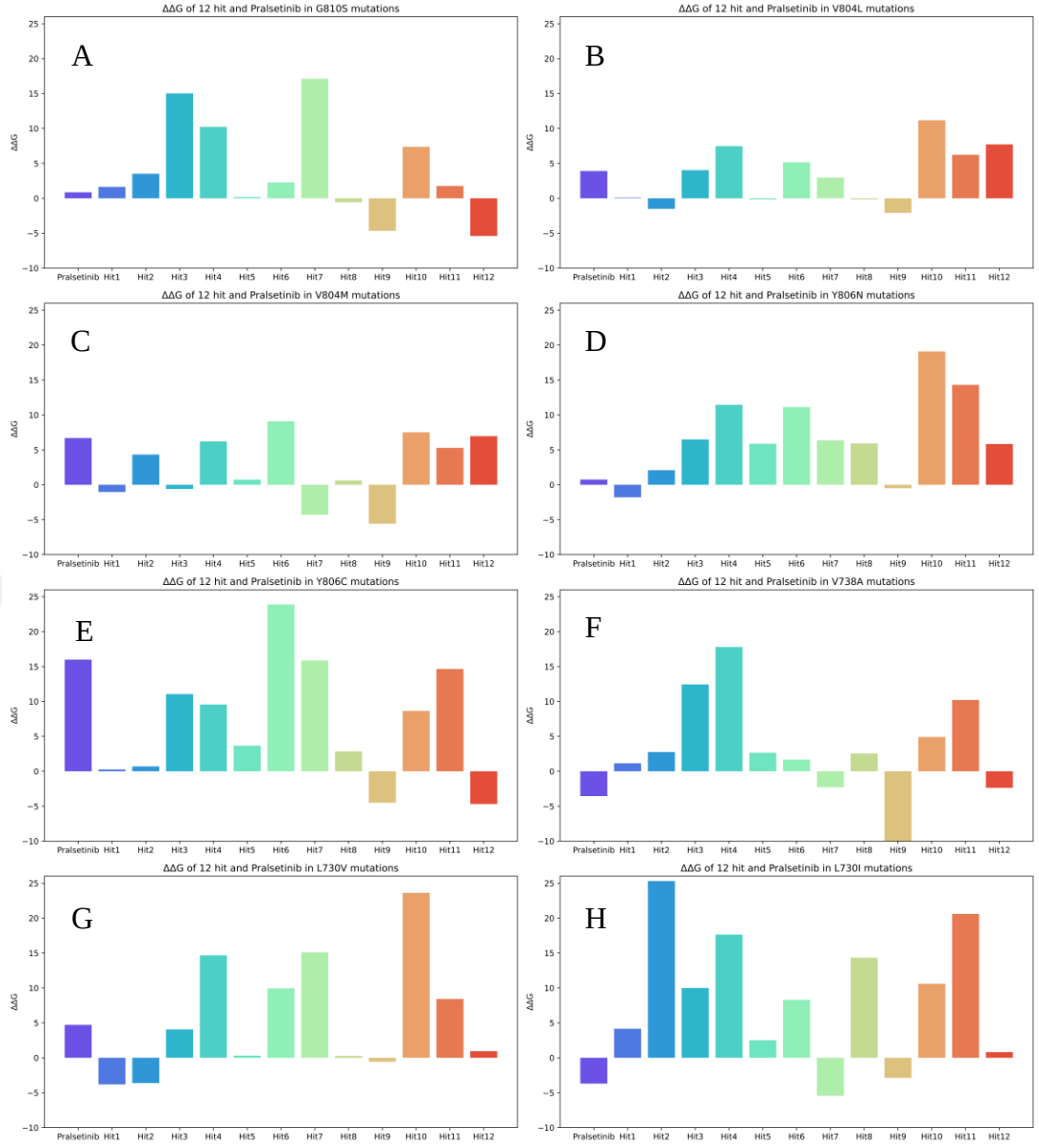


Figure 19. MM/GBSA average difference between wildtype and mutated RET in 12 hit molecules and Pralsetinib.
(A) G810S, (B) V804L, (C) V804M, (D) Y806N, (E) Y806C, (F) V738A, (G) L730V, (H) L730I.

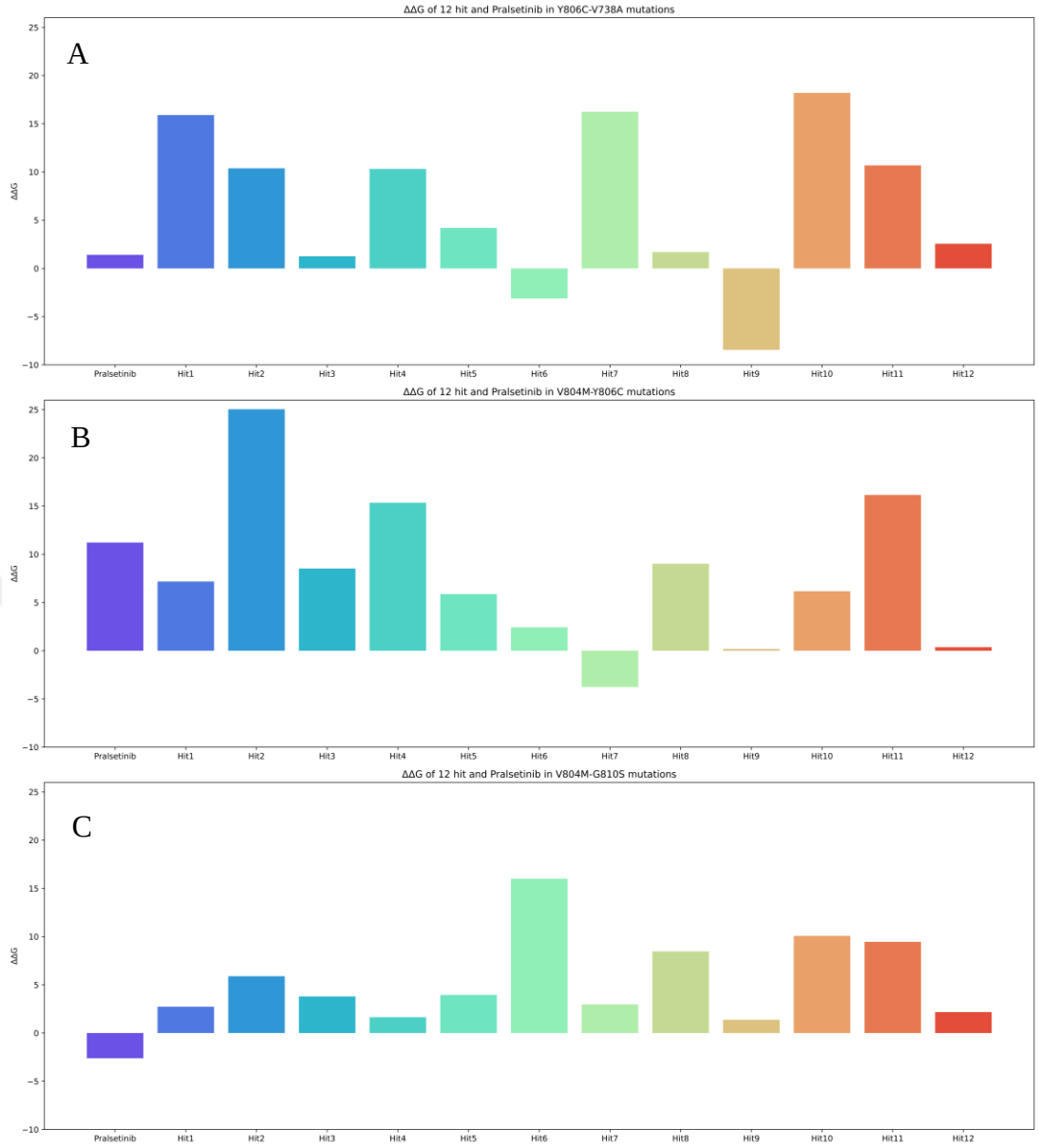


Figure 20. MM/GBSA average difference between wildtype and mutated RET in 12 hit molecules and Pralsetinib.
(A) Y806C/V738A, (B) V804M/Y806C, (C) V804M/G810S.

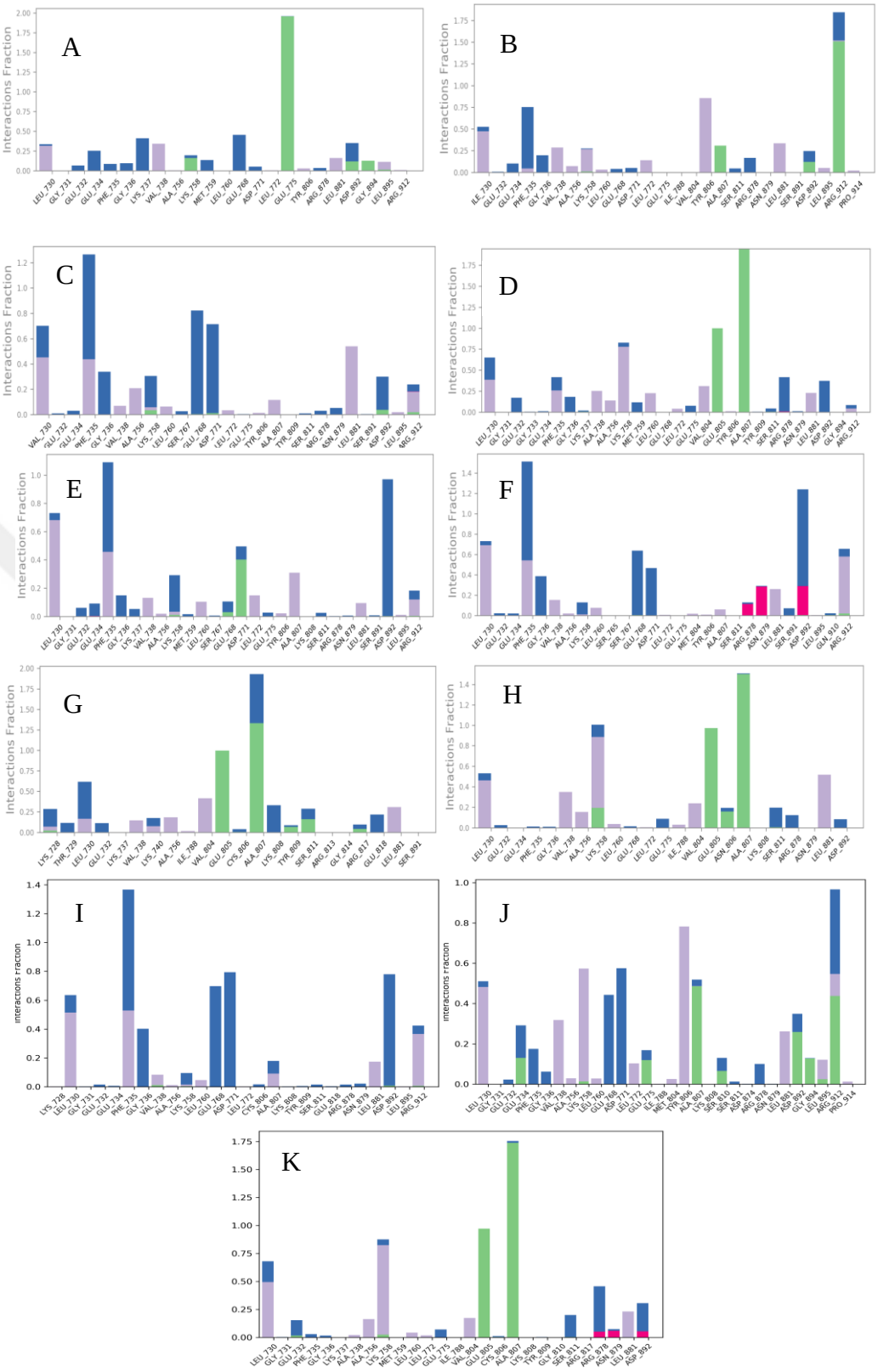


Figure 21. Interaction histogram of Pralsetinib during 10 ns MD simulation (A) G810S, (B) L730I, (C) L730V, (D) V738A, (E) V804L, (F) V804M, (G) Y806C, (H) Y806N, (I) V804M/Y806C, (J) V804M/G810S and (K) Y806C/V738A mutations.

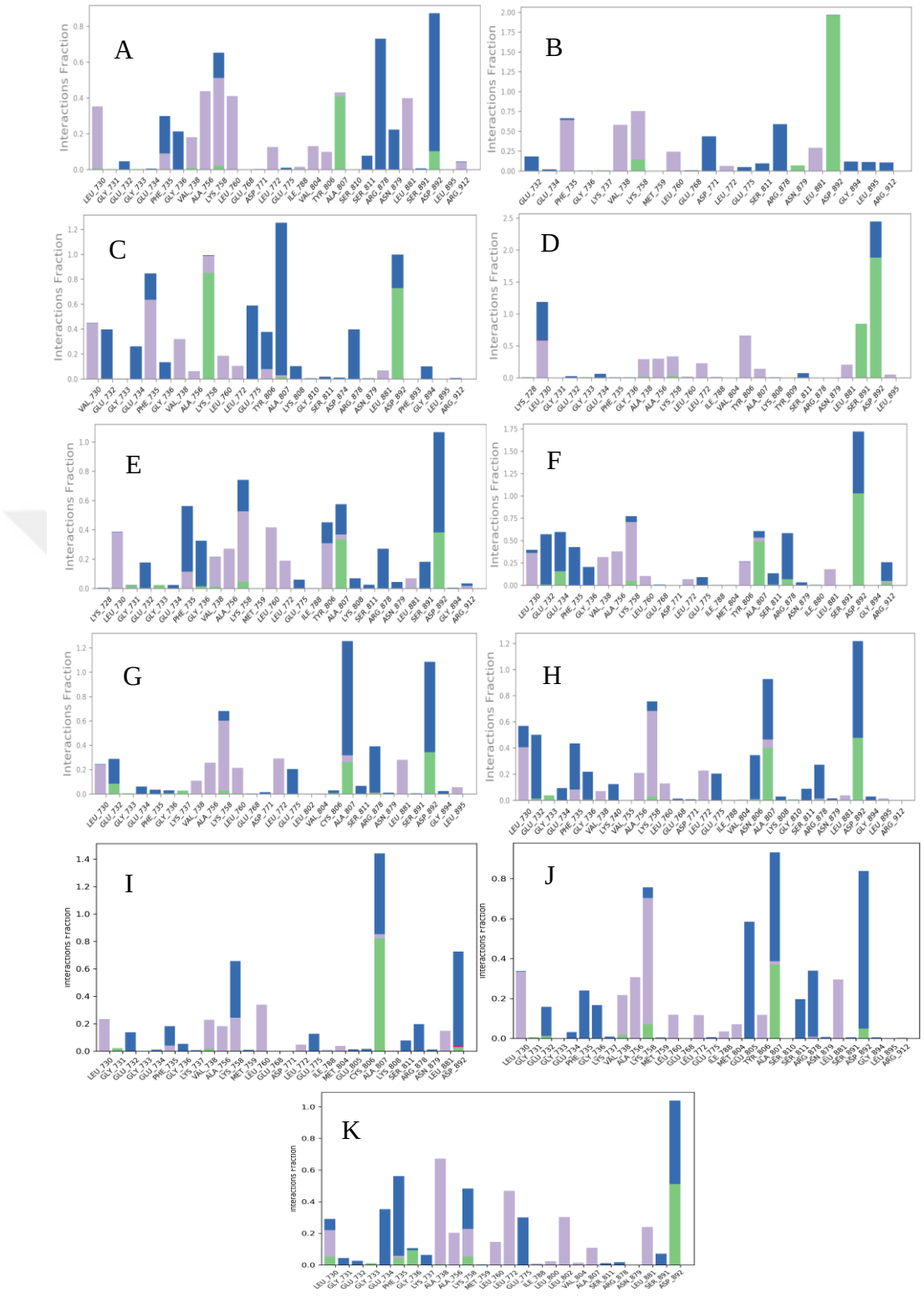


Figure 22. Interaction histogram of Hit9 molecule during 10 ns MD simulation (A) G810S, (B) L730I, (C) L730V, (D) V738A, (E) V804L, (F) V804M, (G) Y806C, (H) Y806N, (I) V804M/Y806C, (J) V804M/G810S and (K) Y806C/V738A mutations.

To assess the stability of Hit9 under various mutations, a 50 ns molecular dynamics (MD) simulation was conducted, and 2000 frames were collected from each simulation. The findings are presented in Table 40. revealing no significant changes during the 50 ns MD simulation. Additionally, the three-dimensional structure of RET bound to the Hit9 compound, with labeled residues within the binding site, is depicted in Figure 23. Furthermore, Appendix A.2 showcases the docking poses of the Hit9 compound in all eleven mutated RET structures.

Table 40.

MM/GBSA Average of Hit9 Molecule During 50 Ns MD Simulations.

Hit9	**G810S**	**V804L**	**V804M**	**Y806N**	**Y806C**	**V738A**	**L730V**	**L730I**
MM/GBSA (kcal/mol) (10 ns)	-70.107 ± 4.27	-67.536 ± 7.30	-71.029 ± 6.37	-65.937 ± 6.03	-69.954 ± 4.81	-89.540 ± 6.13	-66.019 ± 5.02	-68.327 ± 5.64
MM/GBSA (kcal/mol) (50 ns)	-69.639 ± 4.66	-69.460 ± 7.07	-70.393 ± 6.23	-59.897 ± 5.09	-70.578 ± 5.73	-88.613 ± 5.47	-69.054 ± 4.93	-62.274 ± 4.38

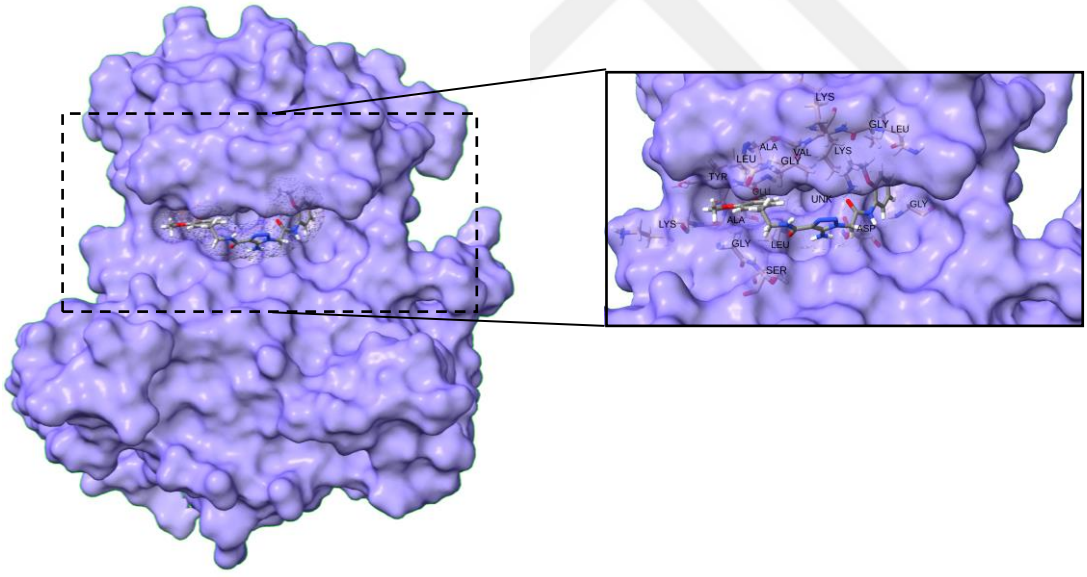


Figure 23. The surface view of the wildtype RET protein in complex with Hit9. The residues within the binding site of the protein have been labeled.

To predict the toxicity scores of the compounds Hit1, Hit4, Hit9, and Hit12, QSAR models in the METACORE^{TM}/METADRUG^{TM} platform were utilized. Typically, a cutoff value of 0.5 is used, where scores exceeding 0.5 indicate potential

toxicity. However, for the cytotoxicity model, a cutoff value of 6 is preferred, with lower scores being more desirable. Similarly, the skin sensitization model employs a cutoff value of 10, and scores higher than 10 indicate weak to moderate sensitization. Based on the data presented in Table 5.41, the hit molecules demonstrate a lower potential for toxicity upon general analysis. Notably, Hit9, which has been identified as a compound lacking resistance to the studied mutations, exhibits a significantly low potential for toxicity. As per the AMES and Cytotoxicity models, their predicted scores are 0.55 and 6.57, respectively.

Table 41.

Toxicity Prediction Scores (TP) of 4 hit molecules (Scores Don't Pass the Cutoff Represented by Red).

Name	Hit1	Hit4	Hit9	Hit12
AMES (TP)	0.36	0.44	**0.55**	0.46
Cytotoxicity model, -log GI50 (M) (TP)	4.52	4.64	**6.57**	5.08
MRTD (TP)	0.15	0.1	0.28	0.31
Carcinogenicity Mouse Female (TP)	0.33	0.34	0.47	0.05
Carcinogenicity Mouse Male (TP)	0.44	0.16	0.21	0.1
Carcinogenicity Rat Female (TP)	0.3	0.29	0.22	0.02
Carcinogenicity Rat Male (TP)	0.22	0.21	0.15	0.01
Anemia (TP)	0.18	0.27	0.15	0.13
Carcinogenicity (TP)	0.21	0.26	0.17	0.03
Cardiotoxicity (TP)	0.26	0.27	0.15	0.41
Genotoxicity (TP)	0.35	0.28	0.15	0.2
Hepatotoxicity (TP)	0.15	0.19	0.21	0.12
Nephrotoxicity (TP)	0.21	0.08	0.13	0.12
Neurotoxicity (TP)	0.15	0.11	0.1	0.01
Liver Cholestasis (TP)	0.33	0.13	0.07	0.08
Liver Lipid Accumulation (TP)	0.21	0.34	0.39	0.27
Liver Necrosis (TP)	**0.56**	0.25	0.43	**0.64**
Liver Weight Gain (TP)	0.25	0.09	0.09	0.37
Kidney Necrosis (TP)	0.06	0.11	0.11	0.08
Kidney Weight Gain (TP)	0.11	0.03	0.06	0.04
Nephron Injury (TP)	0.29	0.07	0.09	0.31
SkinSens, EC3 (TP)	31.85	42.07	59.62	31.7
Nasal pathology (TP)	0.15	0.04	0.07	0.11
Testicular toxicity (TP)	0.3	0.32	0.27	0.02
Pulmonary toxicity (TP)	0.12	0.04	0.05	0.05
Epididymis toxicity (TP)	0.21	0.14	0.13	0.09
Reactive	OK	R	OK	OK

Chapter 6
Discussion

This thesis aims to use new methodologies to screen chemical databases to find hit molecules that can bind to the RET protein with favorable interactions to inhibit the signaling pathway of the RET protein even during several mutations.

Through text mining, we generated 10 chemical libraries for analysis. Additionally, we selected 4 libraries from the ChemDiv database for further investigation. Collectively, these libraries encompass a total of 290 million compounds that will undergo analysis to identify potential hit molecules capable of inhibiting the RET proto-oncogene. The focus is on identifying compounds that remain effective against various point mutations, ensuring their potential as inhibitors is unaffected.

Two methods were employed and described in detail in the methods section. In the first method of this thesis, we aim to create QSAR models based on their numerical properties using all known compounds that targeted RET proteins instead of creating 3D QSAR based on the pharmacophore model for a few numbers of compounds which are obtained in previous studies (Kuchana et al. 2022). So, the pIC_{50} values for the huge number of compounds can be predicted in a shorter time duration. In this method, a QSAR model was developed using 793 molecules from the ChEMBL database that directly targets the RET protein. The dataset was divided into a 70% training set and a 30% test set, resulting in an internal set with an R-square value of 0.876. The QSAR model was further tested on an external set of 92 molecules, yielding an R-square value of 0.652. These R-square values suggest that the model has a strong correlation between the input structural features of compounds and their activity and properties, indicating its reliability in predicting the biological activity and property of new compounds. This established QSAR model was then applied to predict the pIC_{50} values of all 10 libraries identified through text mining. Next, a series of steps were undertaken, including the elimination of compounds based on z-score calculations and consideration of their anticancer activity. Molecular docking was performed using a virtual screen workflow. Among the top 10 molecules with low docking scores in each library, molecular dynamics (MD) simulations were conducted with 5 repeats. The MM/GBSA energy calculation was performed for these molecules.

The results revealed 4 molecules with low MM/GBSA scores across all libraries. However, when compared to the reference molecules, Pralsetinib and Selpercatinib, these 4 molecules did not exhibit favorable MM/GBSA energy values. Selpercatinib had a glide/XP value of -9.164 kcal/mol and an MM/GBSA value of -75.67 kcal/mol, while Pralsetinib had a glide/XP value of -9.463 kcal/mol and an MM/GBSA value of -75.543 kcal/mol. On the other hand, the top molecule from a library containing the pyrazin-2-carboxamid fragment had a glide/XP value of -9.053 kcal/mol, along with an MM/GBSA value of -61.673 kcal/mol. In the second method, we introduced a novel approach to add innovation to the study. Previous research, such as the study conducted by Ouassaf et al. (2022), has commonly employed structure-based pharmacophore modeling to determine the pharmacophore hypothesis of a ligand-based on its cocrystal structure. This hypothesis is then utilized in virtual screening. In this thesis, we propose a different approach by utilizing e-pharmacophore modeling throughout all frames of molecular dynamics (MD) simulations. This allows us to determine which pharmacophore hypotheses are favorable during the entire MD simulations and whether they are generated in a low RMSD pose of the ligand. By employing this method, we aim to gain insights into the most relevant pharmacophore hypotheses that emerge throughout the entire dynamics of the ligand in the simulations. So, in the second method, Structure-based pharmacophore models were generated using e-pharmacophore algorithms from 5000 frames of a 500ns molecular dynamics (MD) simulation. Among the generated models, the ADDRR hypothesis was observed frequently across the frames. Notably, the ADDRR hypothesis was identified in frame 4291, which exhibited a low root mean square deviation (RMSD) value of 1.098 Å, making it a promising candidate. This hypothesis from frame 4291 was selected for ligand screening across all fourteen libraries. Following the filtering of libraries based on Lipinski rules and the inclusion of a minimum ADDRR 2D pharmacophore, the selected compounds underwent z-score calculations of their phase screen scores and assessment of their anticancer activity using binary QSAR. Lipinski filtering is applied to reduce the number of compounds analyzed by ensuring they contain a minimum number of ADDRR features. However, it's important to note that 2D pharmacophore analysis can still be performed without Lipinski filtering. In such cases, numerous compounds that do not adhere to the Lipinski rules may be identified as hit molecules. Subsequently, Glide/SP docking was performed on these compounds, and 10ns MD

simulations were conducted in three repeats and their MM/GBSA energy values are calculated.

The reference molecule, Pralsetinib, exhibited a Glide/SP value of -9.491 kcal/mol and an MM/GBSA value of -71.276 kcal/mol. Among all the libraries, twelve hit molecules were identified with MM/GBSA values exceeding -65 kcal/mol. These twelve molecules underwent 200ns MD simulations, revealing that the MM/GBSA value for Pralsetinib decreased to -75.111 kcal/mol. Notably, only the Hit9 molecule exhibited an MM/GBSA value superior to Pralsetinib. Hit9 had a Glide/SP value of -5.964 kcal/mol, indicating its potential to bind to the ATP binding site of RET. Furthermore, its MM/GBSA value decreased from -65.444 to -79.446 kcal/mol. All 12 molecules were subsequently docked into eight mutated RET and three double-mutated RET proteins. Subsequent 10ns MD simulations were performed, and their MM/GBSA values were calculated. Through these calculations, it was observed that the Hit9 molecule showed the potential to bind to the mutated RET, and its MM/GBSA value decreased during the mutation process. In comparison, the binding pose of Pralsetinib in the mutated RET proteins showed significant changes, potentially influenced by the docking algorithms employed. The observed significant change in the docking pose of Pralsetinib could potentially be attributed to its resistance to the mutations under investigation. In clinical cases, Pralsetinib has been shown to exhibit resistance to these mutations. While docking Pralsetinib to the binding site of RET, it is bind to that region even if the interactions are not optimal. Consequently, this substantial change in the docking pose may indicate that Pralsetinib cannot effectively penetrate the binding pocket of the mutated RET, which could contribute to its resistance. Therefore, considering the potential challenges posed by the mutated residues in the binding pocket, it is plausible that employing an alternative docking method such as induce fit docking could yield better results when docking Pralsetinib to the ATP binding pocket. It is also worth noting that the presence of these mutations may induce conformational changes in the structure of the protein binding pocket. Consequently, Pralsetinib may not be able to bind to the mutated binding pocket in the same pose as it does with the wildtype RET. These conformational changes might explain the discrepancies observed between the results obtained from various in vivo and xenograft studies and the findings of our study. Furthermore, it's important to note that conducting 10 ns MD simulations provides a relatively short timeframe for

observing the detailed interactions between molecules and the residues within the binding pocket. Longer simulation times may be necessary to gain a more comprehensive understanding of how the molecules interact with the binding pocket residues.

Chapter 7
Conclusion

In conclusion, our study aimed to explore novel hit molecules capable of inhibiting the signaling pathway of the RET proto-oncogene to reduce tumor size in medullary thyroid cancer and NSCLC. We employed two methods to achieve this objective. The first method involved the analysis of hit molecules obtained from QSAR models based on the numerical properties of compounds targeting RET proteins.

The second method utilized a unique approach that involved generating structure-based pharmacophores from molecular dynamics simulation trajectories. We selected the most favorable pharmacophore hypothesis based on a low RMSD pose. From this method, we obtained 12 hit molecules. Subsequently, we analyzed the activity of these 12 hit molecules in RET proteins with mutations in their binding pocket. We compared their results with Pralsetinib, which served as the reference molecule. Remarkably, our findings revealed that the Hit9 molecule exhibited the potential to bind to the ATP binding pocket of the RET protein and demonstrated a favorable MM/GBSA score compared to Pralsetinib.

Overall, our study provides valuable insights into potential hit molecules that can effectively target the ATP binding site of the RET protein, offering promise for further exploration in the development of therapeutics for medullary thyroid cancer and NSCLC.

REFERENCES

Agrawal, Nishant, Rehan Akbani, B. Arman Aksoy, Adrian Ally, Harindra Arachchi, Sylvia L. Asa, J. Todd Auman, Miruna Balasundaram, Saianand Balu, Stephen B. Baylin, Madhusmita Behera, Brady Bernard, Rameen Beroukhim, Justin A. Bishop, Aaron D. Black, Tom Bodenheimer, Lori Boice, Moiz S. Bootwalla, Jay Bowen, Reanne Bowlby, Christopher A. Bristow, Robin Brookens, Denise Brooks, Robert Bryant, Elizabeth Buda, Yaron S. N. Butterfield, Tobias Carling, Rebecca Carlsen, Scott L. Carter, Sally E. Carty, Timothy A. Chan, Amy Y. Chen, Andrew D. Cherniack, Dorothy Cheung, Lynda Chin, Juok Cho, Andy Chu, Eric Chuah, Kristian Cibulskis, Giovanni Ciriello, Amanda Clarke, Gary L. Clayman, Leslie Cope, John A. Copland, Kyle Covington, Ludmila Danilova, Tanja Davidsen, John A. Demchok, Daniel DiCara, Noreen Dhalla, Rajiv Dhir, Sheliann S. Dookran, Gideon Dresdner, Jonathan Eldridge, Greg Eley, Adel K. El-Naggar, Stephanie Eng, James A. Fagin, Timothy Fennell, Robert L. Ferris, Sheila Fisher, Scott Frazer, Jessica Frick, Stacey B. Gabriel, Ian Ganly, Jianjiong Gao, Levi A. Garraway, Julie M. Gastier-Foster, Gad Getz, Nils Gehlenborg, Ronald Ghossein, Richard A. Gibbs, Thomas J. Giordano, Karen Gomez-Hernandez, Jonna Grimsby, Benjamin Gross, Ranabir Guin, Angela Hadjipanayis, Hollie A. Harper, D. Neil Hayes, David I. Heiman, James G. Herman, Katherine A. Hoadley, Matan Hofree, Robert A. Holt, Alan P. Hoyle, Franklin W. Huang, Mei Huang, Carolyn M. Hutter, Trey Ideker, Lisa Iype, Anders Jacobsen, Stuart R. Jefferys, Corbin D. Jones, Steven J. M. Jones, Katayoon Kasaian, Electron Kebebew, Fadlo R. Khuri, Jaegil Kim, Roger Kramer, Richard Kreisberg, Raju Kucherlapati, David J. Kwiatkowski, Marc Ladanyi, Phillip H. Lai, Peter W. Laird, Eric Lander, Michael S. Lawrence, Darlene Lee, Eunjung Lee, Semin Lee, William Lee, Kristen M. Leraas, Tara M. Lichtenberg, Lee Lichtenstein, Pei Lin, Shiyun Ling, Jinze Liu, Wenbin Liu, Yingchun Liu, Virginia A. LiVolsi, Yiling Lu, Yussanne Ma, Harshad S. Mahadeshwar, Marco A. Marra, Michael Mayo, David G. McFadden, Shaowu Meng, Matthew Meyerson, Piotr A. Mieczkowski, Michael Miller, Gordon Mills, Richard A. Moore, Lisle E. Mose, Andrew J. Mungall, Bradley A.

Murray, Yuri E. Nikiforov, Michael S. Noble, Akinyemi I. Ojesina, Taofeek K. Owonikoko, Bradley A. Ozenberger, Angeliki Pantazi, Michael Parfenov, Peter J. Park, Joel S. Parker, Evan O. Paull, Chandra Sekhar Pedamallu, Charles M. Perou, Jan F. Prins, Alexei Protopopov, Suresh S. Ramalingam, Nilsa C. Ramirez, Ricardo Ramirez, Benjamin J. Raphael, W. Kimryn Rathmell, Xiaojia Ren, Sheila M. Reynolds, Esther Rheinbay, Matthew D. Ringel, Michael Rivera, Jeffrey Roach, A. Gordon Robertson, Mara W. Rosenberg, Matthew Rosenthal, Sara Sadeghi, Gordon Saksena, Chris Sander, Netty Santoso, Jacqueline E. Schein, Nikolaus Schultz, Steven E. Schumacher, Raja R. Seethala, Jonathan Seidman, Yasin Senbabaoglu, Sahil Seth, Samantha Sharpe, Kenna R. Mill. Shaw, John P. Shen, Ronglai Shen, Steven Sherman, Margi Sheth, Yan Shi, Ilya Shmulevich, Gabriel L. Sica, Janae V. Simons, Rileen Sinha, Payal Sipahimalani, Robert C. Smallridge, Heidi J. Sofia, Matthew G. Soloway, Xingzhi Song, Carrie Sougnez, Chip Stewart, Petar Stojanov, Joshua M. Stuart, S. Onur Sumer, Yichao Sun, Barbara Tabak, Angela Tam, Donghui Tan, Jiabin Tang, Roy Tarnuzzer, Barry S. Taylor, Nina Thiessen, Leigh Thorne, Vésteinn Thorsson, R. Michael Tuttle, Christopher B. Umbricht, David J. Van Den Berg, Fabio Vandin, Umadevi Veluvolu, Roel G. W. Verhaak, Michelle Vinco, Doug Voet, Vonn Walter, Zhining Wang, Scot Waring, Paul M. Weinberger, Nils Weinhold, John N. Weinstein, Daniel J. Weisenberger, David Wheeler, Matthew D. Wilkerson, Jocelyn Wilson, Michelle Williams, Daniel A. Winer, Lisa Wise, Junyuan Wu, Liu Xi, Andrew W. Xu, Liming Yang, Lixing Yang, Travis I. Zack, Martha A. Zeiger, Dong Zeng, Jean Claude Zenklusen, Ni Zhao, Hailei Zhang, Jianhua Zhang, Jiashan Zhang, Wei Zhang, Erik Zmuda, and Lihua Zou. 2014. “Integrated Genomic Characterization of Papillary Thyroid Carcinoma.” *Cell* 159(3):676–90. doi: 10.1016/J.CELL.2014.09.050.

Alder, B. J., and T. E. Wainwright. 1959. “Studies in Molecular Dynamics. I. General Method.” *The Journal of Chemical Physics* 31(2):459–66. doi: 10.1063/1.1730376.

Andrade, Chittaranjan. 2021. “Z Scores, Standard Scores, and Composite Test Scores.” *Indian Journal of Psychological Medicine* 43(6):555. doi: 10.1177/02537176211046525.

Arthur, David Ebuka, Stephen Ejeh, and Adamu Uzairu. 2020. “Quantitative Structure-Activity Relationship (QSAR) and Design of Novel Ligands That Demonstrate High Potency and Target Selectivity as Protein Tyrosine Phosphatase 1B (PTP 1B) Inhibitors as an Effective Strategy Used to Model Anti-Diabetic Agents.” *Journal of Receptor and Signal Transduction Research* 40(6):501–20. doi: 10.1080/10799893.2020.1759092.

Ashorobi, Damilola, and Peter P. Lopez. 2023. “Follicular Thyroid Cancer.” *StatPearls*.

Blume-Jensen, Peter, and Tony Hunter. 2001. “Oncogenic Kinase Signalling.” *Nature* 411(6835):355–65. doi: 10.1038/35077225.

Bowers, Kevin J., Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw. 2006. “Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters.” *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC’06*. doi: 10.1145/1188455.1188544.

Brogi, Simone, Teodorico Castro Ramalho, Kamil Kuca, José L. Medina-Franco, and Marian Valko. 2020. “Editorial: In Silico Methods for Drug Design and Discovery.” *Frontiers in Chemistry* 8:612. doi: 10.3389/FCHEM.2020.00612/BIBTEX.

Brown, A. C., and T. R. Fraser. 1868. “On the Connection between Chemical Constitution and Physiological Action; with Special Reference to the Physiological Action of the Salts of the Ammonium Bases Derived from

Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia - PubMed." Retrieved May 26, 2023 (https://pubmed.ncbi.nlm.nih.gov/17230757/).

Casals, Eudald, Muriel F. Gusta, Macarena Cobaleda-Siles, Ana Garcia-Sanz, and Victor F. Puntes. 2017. "Cancer Resistance to Treatment and Antiresistance Tools Offered by Multimodal Multifunctional Nanoparticles." *Cancer Nanotechnology* 8(1). doi: 10.1186/S12645-017-0030-4.

Chang, L., and M. Karin. 2001. "Mammalian MAP Kinase Signalling Cascades." *Nature* 410(6824):37–40. doi: 10.1038/35065000.

Chapman, Paul B., Axel Hauschild, Caroline Robert, John B. Haanen, Paolo Ascierto, James Larkin, Reinhard Dummer, Claus Garbe, Alessandro Testori, Michele Maio, David Hogg, Paul Lorigan, Celeste Lebbe, Thomas Jouary, Dirk Schadendorf, Antoni Ribas, Steven J. O'Day, Jeffrey A. Sosman, John M. Kirkwood, Alexander M. M. Eggermont, Brigitte Dreno, Keith Nolop, Jiang Li, Betty Nelson, Jeannie Hou, Richard J. Lee, Keith T. Flaherty, and Grant A. McArthur. 2011. "Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation." *The New England Journal of Medicine* 364(26):2507–16. doi: 10.1056/NEJMOA1103782.

Chen, I. Jen, and Nicolas Foloppe. 2010. "Drug-like Bioactive Structures and Conformational Coverage with the LigPrep/ConfGen Suite: Comparison to Programs MOE and Catalyst." *Journal of Chemical Information and Modeling* 50(5):822–39. doi: 10.1021/CI100026X.

Chen, Ying Nan P., Matthew J. Lamarche, Ho Man Chan, Peter Fekkes, Jorge Garcia-Fortanet, Michael G. Acker, Brandon Antonakos, Christine Hiu Tung Chen, Zhouliang Chen, Vesselina G. Cooke, Jason R. Dobson, Zhan Deng, Feng Fei, Brant Firestone, Michelle Fodor, Cary Fridrich, Hui Gao, Denise Grunenfelder, Huai Xiang Hao, Jaison Jacob, Samuel Ho, Kathy Hsiao, Zhao B. Kang, Rajesh Karki, Mitsunori Kato, Jay Larrow, Laura R. La Bonte, Francois Lenoir, Gang Liu, Shumei Liu, Dyuti Majumdar, Matthew J. Meyer, Mark Palermo, Lawrence

Perez, Minying Pu, Edmund Price, Christopher Quinn, Subarna Shakya, Michael D. Shultz, Joanna Slisz, Kavitha Venkatesan, Ping Wang, Markus Warmuth, Sarah Williams, Guizhi Yang, Jing Yuan, Ji Hu Zhang, Ping Zhu, Timothy Ramsey, Nicholas J. Keen, William R. Sellers, Travis Stams, and Pascal D. Fortin. 2016. "Allosteric Inhibition of SHP2 Phosphatase Inhibits Cancers Driven by Receptor Tyrosine Kinases." *Nature 2016 535:7610* 535(7610):148–52. doi: 10.1038/nature18621.

Christopher J., Cramer. 2004. "Essentials of Computational Chemistry." *John Wiley & Sons* 1–624.

Cibas, Edmund S., and Syed Z. Ali. 2009. "The Bethesda System for Reporting Thyroid Cytopathology." *Thyroid : Official Journal of the American Thyroid Association* 19(11):1159–65. doi: 10.1089/THY.2009.0274.

Cortes, J. E., D. W. Kim, J. Pinilla-Ibarz, P. le Coutre, R. Paquette, C. Chuah, F. E. Nicolini, J. F. Apperley, H. J. Khoury, M. Talpaz, J. DiPersio, D. J. DeAngelo, E. Abruzzese, D. Rea, M. Baccarani, M. C. Müller, C. Gambacorti-Passerini, S. Wong, S. Lustgarten, V. M. Rivera, T. Clackson, C. D. Turner, F. G. Haluska, F. Guilhot, M. W. Deininger, A. Hochhaus, T. Hughes, J. M. Goldman, N. P. Shah, and H. Kantarjian. 2013. "A Phase 2 Trial of Ponatinib in Philadelphia Chromosome–Positive Leukemias." *New England Journal of Medicine* 369(19):1783–96. doi: 10.1056/NEJMOA1306494/SUPPL_FILE/NEJMOA1306494_APPENDIX.PDF.

Crnčić, Tatjana Bogović, Maja Ilić Tomaš, Neva Girotto, and Svjetlana Grbac Ivanković. 2020. "Risk Factors for Thyroid Cancer: What Do We Know So Far?" *Acta Clinica Croatica* 59.(Supplement 1):66–72. doi: 10.20471/ACC.2020.59.S1.08.

Davies, Louise, and H. Gilbert Welch. 2014. “Current Thyroid Cancer Trends in the United States.” *JAMA Otolaryngology-- Head & Neck Surgery* 140(4):317–22. doi: 10.1001/JAMAOTO.2014.1.

DeLano, Warren L., Mark H. Ultsch, Abraham M. De Vos, and James A. Wells. 2000. “Convergent Solutions to Binding at a Protein-Protein Interface.” *Science* 287(5456):1279–83. doi: 10.1126/SCIENCE.287.5456.1279/SUPPL_FILE/1044724S1_THUMB.GIF.

Detterbeck, Frank C., Daniel J. Boffa, Anthony W. Kim, and Lynn T. Tanoue. 2017. “The Eighth Edition Lung Cancer Stage Classification.” *Chest* 151(1):193–203. doi: 10.1016/j.chest.2016.10.010.

Dogan, Berna, and Serdar Durdagi. 2021. “Drug Re-Positioning Studies for Novel HIV-1 Inhibitors Using Binary QSAR Models and Multi-Target-Driven In Silico Studies.” *Molecular Informatics* 40(2). doi: 10.1002/MINF.202000012.

Doll, Richard, and Richard Peto. 1976. “Mortality in Relation to Smoking: 20 Years’ Observations on Male British Doctors.” *British Medical Journal* 2(6051):1525. doi: 10.1136/BMJ.2.6051.1525.

Dr. BİRİNCİ, Şuayip, M. Mahir Dr.ÜLGÜ, Berrak BORA Dr. BAŞARA, İrem SOYTUTAN ÇAĞLAR, Asiye AYGÜN, Tuğcan Adem ÖZDEMİR, and Banu KULALİ. 2020. *THE MINISTRY of HEALTH of TÜRKİYE HEALTH STATISTICS YEARBOOK*.

Drilon, A., G. Oxnard, L. Wirth, B. Besse, O. Gautschi, S. W. D. Tan, H. Loong, T. Bauer, Y. J. Kim, A. Horiike, K. Park, M. Shah, C. McCoach, L. Bazhenova, T. Seto, M. Brose, N. Pennell, J. Weiss, I. Matos, N. Peled, B. C. Cho, Y. Ohe, K. Reckamp, V. Boni, M. Satouchi, G. Falchook, W. Akerley, H. Daga, T. Sakamoto, J. Patel, N. Lakhani, F. Barlesi, M. Burkard, V. Zhu, V. Moreno Garcia, J. Medioni, M. Matrana, C. Rolfo, D. H. Lee, H. Nechushtan, M. Johnson, V. Velcheti, M. Nishio, R. Toyozawa, K. Ohashi, L. Song, J. Han, A. Spira, F. De

Braud, K. Staal Rohrberg, S. Takeuchi, J. Sakakibara, S. Waqar, H. Kenmotsu, F. Wilson, B.Nair, E. Olek, J. Kherani, K. Ebata, E. Zhu, M. Nguyen, L. Yang, X. Huang, S. Cruickshank, S. Rothenberg, B. Solomon, K. Goto, and V. Subbiah. 2019. "PL02.08 Registrational Results of LIBRETTO-001: A Phase 1/2 Trial of LOXO-292 in Patients with RET Fusion-Positive Lung Cancers." *Journal of Thoracic Oncology* 14(10):S6–7. doi: 10.1016/J.JTHO.2019.08.059.

Drilon, Alexander, Zishuo I. Hu, Gillianne G. Y. Lai, and Daniel S. W. Tan. 2017. "Targeting RET-Driven Cancers: Lessons from Evolving Preclinical and Clinical Landscapes." *Nature Reviews Clinical Oncology 2017 15:3* 15(3):151–67. doi: 10.1038/nrclinonc.2017.175.

Drilon, Alexander, Zishuo I. Hu, Gillianne G. Y. Lai, and Daniel S. W. Tan. 2018a. "Targeting RET-Driven Cancers: Lessons from Evolving Preclinical and Clinical Landscapes." *Nature Reviews. Clinical Oncology* 15(3):151–67. doi: 10.1038/NRCLINONC.2017.175.

Drilon, Alexander, Zishuo I. Hu, Gillianne G. Y. Lai, and Daniel S. W. Tan. 2018b. "Targeting RET-Driven Cancers: Lessons from Evolving Preclinical and Clinical Landscapes." *Nature Reviews. Clinical Oncology* 15(3):151–67. doi: 10.1038/NRCLINONC.2017.175.

Drilon, Alexander, Zishuo I. Hu, Gillianne G. Y. Lai, and Daniel S. W. Tan. 2018c. "Targeting RET-Driven Cancers: Lessons from Evolving Preclinical and Clinical Landscapes." *Nature Reviews. Clinical Oncology* 15(3):151–67. doi: 10.1038/NRCLINONC.2017.175.

Drilon, Alexander, Geoffrey R. Oxnard, Daniel S. W. Tan, Herbert H. F. Loong, Melissa Johnson, Justin Gainor, Caroline E. McCoach, Oliver Gautschi, Benjamin Besse, Byoung C. Cho, Nir Peled, Jared Weiss, Yu-Jung Kim, Yuichiro Ohe, Makoto Nishio, Keunchil Park, Jyoti Patel, Takashi Seto, Tomohiro Sakamoto, Ezra Rosen, Manisha H. Shah, Fabrice Barlesi, Philippe A. Cassier, Lyudmila Bazhenova, Filippo De Braud, Elena Garralda, Vamsidhar Velcheti, Miyako

Satouchi, Kadoaki Ohashi, Nathan A. Pennell, Karen L. Reckamp, Grace K. Dy, Jürgen Wolf, Benjamin Solomon, Gerald Falchook, Kevin Ebata, Michele Nguyen, Binoj Nair, Edward Y. Zhu, Luxi Yang, Xin Huang, Elizabeth Olek, S. Michael Rothenberg, Koichi Goto, and Vivek Subbiah. 2020. " Efficacy of Selpercatinib in RET Fusion–Positive Non–Small-Cell Lung Cancer ." *New England Journal of Medicine* 383(9):813–24. doi: 10.1056/NEJMOA2005653/SUPPL_FILE/NEJMOA2005653_DATA-SHARING.PDF.

Drilon, Alexander, Natasha Rekhtman, Maria Arcila, Lu Wang, Andy Ni, Melanie Albano, Martine Van Voorthuysen, Romel Somwar, Roger S. Smith, Joseph Montecalvo, Andrew Plodkowski, Michelle S. Ginsberg, Gregory J. Riely, Charles M. Rudin, Marc Ladanyi, and Mark G. Kris. 2016. "Cabozantinib in Patients with Advanced RET-Rearranged Non-Small-Cell Lung Cancer: An Open-Label, Single-Centre, Phase 2, Single-Arm Trial." *The Lancet Oncology* 17(12):1653–60. doi: 10.1016/S1470-2045(16)30562-9.

Drilon, Alexander, Lu Wang, Adnan Hasanovic, Yoshiyuki Suehara, Doron Lipson, Phil Stephens, Jeffrey Ross, Vincent Miller, Michelle Ginsberg, Maureen F. Zakowski, Mark G. Kris, Marc Ladanyi, and Naiyer Rizvi. 2013. "Response to Cabozantinib in Patients with RET Fusion-Positive Lung Adenocarcinomas." *Cancer Discovery* 3(6):630–35. doi: 10.1158/2159-8290.CD-13-0035.

Dror, Ron O., Robert M. Dirks, J. P. Grossman, Huafeng Xu, and David E. Shaw. 2012. "Biomolecular Simulation: A Computational Microscope for Molecular Biology." *Annual Review of Biophysics* 41(1):429–52. doi: 10.1146/ANNUREV-BIOPHYS-042910-155245.

Drosten, Matthias, and Brigitte M. Pützer. 2006. "Mechanisms of Disease: Cancer Targeting and the Impact of Oncogenic RET for Medullary Thyroid Carcinoma Therapy." *Nature Clinical Practice Oncology 2006 3:10* 3(10):564–74. doi: 10.1038/ncponc0610.

Druker, Brian J., Charles L. Sawyers, Hagop Kantarjian, Debra J. Resta, Sofia Fernandes Reese, John M. Ford, Renaud Capdeville, and Moshe Talpaz. 2001. "Activity of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in the Blast Crisis of Chronic Myeloid Leukemia and Acute Lymphoblastic Leukemia with the Philadelphia Chromosome." *The New England Journal of Medicine* 344(14):1038–42. doi: 10.1056/NEJM200104053441402.

Ekins, S., A. Bugrim, L. Brovold, E. Kirillov, Y. Nikolsky, E. Rakhmatulin, S. Sorokina, A. Ryabov, T. Serebryiskaya, A. Melnikov, J. Metz, and T. Nikolskaya. 2006. "Algorithms for Network Analysis in Systems-ADME/Tox Using the MetaCore and MetaDrug Platforms." *Xenobiotica; the Fate of Foreign Compounds in Biological Systems* 36(10–11):877–901. doi: 10.1080/00498250600861660.

Essmann, Ulrich, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. 1995. "A Smooth Particle Mesh Ewald Method." *The Journal of Chemical Physics* 103(19):8577–93. doi: 10.1063/1.470117.

De Falco, Valentina, Francesca Carlomagno, Hong yu Li, and Massimo Santoro. 2017. "The Molecular Basis for RET Tyrosine-Kinase Inhibitors in Thyroid Cancer." *Best Practice & Research Clinical Endocrinology & Metabolism* 31(3):307–18. doi: 10.1016/J.BEEM.2017.04.013.

FDA. 2020a. "FDA Approves Pralsetinib for Lung Cancer with RET Gene Fusions." Retrieved April 11, 2023 (https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-pralsetinib-lung-cancer-ret-gene-fusions).

FDA. 2020b. "FDA Approves Selpercatinib for Lung and Thyroid Cancers with RET Gene Mutations or Fusions." Retrieved April 10, 2023 (https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-selpercatinib-lung-and-thyroid-cancers-ret-gene-mutations-or-fusions).

Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. 2020. “Global Cancer Observatory: Cancer Today.” *International Agency for Research on Cancer*. Retrieved (https://gco.iarc.fr/today).

Friesner, Richard A., Robert B. Murphy, Matthew P. Repasky, Leah L. Frye, Jeremy R. Greenwood, Thomas A. Halgren, Paul C. Sanschagrin, and Daniel T. Mainz. 2006. “Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes.” *Journal of Medicinal Chemistry* 49(21):6177–96. doi: 10.1021/JM051256O/SUPPL_FILE/JM051256OSI20060602_023733.PDF.

Fujita, Toshio. 1995. “QSAR and Drug Design : New Developments and Applications.” 493.

Gainor, Justin F., Giuseppe Curigliano, Dong Wan Kim, Dae Ho Lee, Benjamin Besse, Christina S. Baik, Robert C. Doebele, Philippe A. Cassier, Gilberto Lopes, Daniel S. W. Tan, Elena Garralda, Luis G. Paz-Ares, Byoung Chul Cho, Shirish M. Gadgeel, Michael Thomas, Stephen V. Liu, Matthew H. Taylor, Aaron S. Mansfield, Viola W. Zhu, Corinne Clifford, Hui Zhang, Michael Palmer, Jennifer Green, Christopher D. Turner, and Vivek Subbiah. 2021. “Pralsetinib for RET Fusion-Positive Non-Small-Cell Lung Cancer (ARROW): A Multi-Cohort, Open-Label, Phase 1/2 Study.” *The Lancet Oncology* 22(7):959–69. doi: 10.1016/S1470-2045(21)00247-3.

Gainor, Justin F., and Alice T. Shaw. 2013. “Novel Targets in Non-Small Cell Lung Cancer: ROS1 and RET Fusions.” *The Oncologist* 18(7):865–75. doi: 10.1634/THEONCOLOGIST.2013-0095.

GAVRETO™ (pralsetinib) Prescribing Information (U.S.). 2020. “Blueprint Medicines Announces FDA Approval of GAVRETO™ (Pralsetinib) for the Treatment of Adults with Metastatic RET Fusion-Positive Non-Small Cell Lung Cancer.” Retrieved April 10, 2023 (https://ir.blueprintmedicines.com/news-

releases/news-release-details/blueprint-medicines-announces-fda-approval-gavretotm-pralsetinib).

Gazdar, Adi F., Paul A. Bunn, and John D. Minna. 2017. “Small-Cell Lung Cancer: What We Know, What We Need to Know and the Path Forward.” *Nature Reviews Cancer 2017 17:12* 17(12):725–37. doi: 10.1038/nrc.2017.87.

Genheden, Samuel, and Ulf Ryde. 2015. “The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities.” *Expert Opinion on Drug Discovery* 10(5):449–61. doi: 10.1517/17460441.2015.1032936.

George, Julie, Jing Shan Lim, Se Jin Jang, Yupeng Cun, Luka Ozretia, Gu Kong, Frauke Leenders, Xin Lu, Lynnette Fernández-Cuesta, Graziella Bosco, Christian Müller, Ilona Dahmen, Nadine S. Jahchan, Kwon Sik Park, Dian Yang, Anthony N. Karnezis, Dedeepya Vaka, Angela Torres, Maia Segura Wang, Jan O. Korbel, Roopika Menon, Sung Min Chun, Deokhoon Kim, Matt Wilkerson, Neil Hayes, David Engelmann, Brigitte Pützer, Marc Bos, Sebastian Michels, Ignacija Vlasic, Danila Seidel, Berit Pinther, Philipp Schaub, Christian Becker, Janine Altmüller, Jun Yokota, Takashi Kohno, Reika Iwakawa, Koji Tsuta, Masayuki Noguchi, Thomas Muley, Hans Hoffmann, Philipp A. Schnabel, Iver Petersen, Yuan Chen, Alex Soltermann, Verena Tischler, Chang Min Choi, Yong Hee Kim, Pierre P. Massion, Yong Zou, Dragana Jovanovic, Milica Kontic, Gavin M. Wright, Prudence A. Russell, Benjamin Solomon, Ina Koch, Michael Lindner, Lucia A. Muscarella, Annamaria La Torre, John K. Field, Marko Jakopovic, Jelena Knezevic, Esmeralda Castaños-Vélez, Luca Roz, Ugo Pastorino, Odd Terje Brustugun, Marius Lund-Iversen, Erik Thunnissen, Jens Köhler, Martin Schuler, Johan Botling, Martin Sandelin, Montserrat Sanchez-Cespedes, Helga B. Salvesen, Viktor Achter, Ulrich Lang, Magdalena Bogus, Peter M. Schneider, Thomas Zander, Sascha Ansén, Michael Hallek, Jürgen Wolf, Martin Vingron, Yasushi Yatabe, William D. Travis, Peter Nürnberg, Christian Reinhardt, Sven Perner, Lukas Heukamp, Reinhard Büttner, Stefan A. Haas, Elisabeth Brambilla, Martin Peifer, Julien Sage, and Roman K. Thomas. 2015. “Comprehensive

Genomic Profiles of Small Cell Lung Cancer." *Nature 2015 524:7563* 524(7563):47–53. doi: 10.1038/nature14664.

Griesinger, F., G. Curigliano, M. Thomas, V. Subbiah, C. S. Baik, D. S. W. Tan, D. H. Lee, D. Misch, E. Garralda, D. W. Kim, A. J. van der Wekken, J. F. Gainor, L. Paz-Ares, S. V. Liu, G. P. Kalemkerian, Y. Houvras, D. W. Bowles, A. S. Mansfield, J. J. Lin, V. Smoljanovic, A. Rahman, S. Kong, A. Zalutskaya, M. Louie-Gao, A. L. Boral, and J. Mazières. 2022. "Safety and Efficacy of Pralsetinib in RET Fusion-Positive Non-Small-Cell Lung Cancer Including as First-Line Therapy: Update from the ARROW Trial." *Annals of Oncology : Official Journal of the European Society for Medical Oncology* 33(11):1168–78. doi: 10.1016/J.ANNONC.2022.08.002.

Haddad, Robert I., Lindsay Bischoff, Douglas Ball, Victor Bernet, Erik Blomain, Naifa Lamki Busaidy, Michael Campbell, Paxton Dickson, Quan Yang Duh, Hormoz Ehya, Whitney S. Goldner, Theresa Guo, Megan Haymart, Shelby Holt, Jason P. Hunt, Andrei Iagaru, Fouad Kandeel, Dominick M. Lamonica, Susan Mandel, Stephanie Markovina, Bryan McIver, Christopher D. Raeburn, Rod Rezaee, John A. Ridge, Mara Y. Roth, Randall P. Scheri, Jatin P. Shah, Jennifer A. Sipos, Rebecca Sippel, Cord Sturgeon, Thomas N. Wang, Lori J. Wirth, Richard J. Wong, Michael Yeh, Carly J. Cassara, and Susan Darlow. 2022. "Thyroid Carcinoma, Version 2.2022 NCCN CLINICAL PRACTICE GUIDELINES IN ONCOLOGY." *JNCCN Journal of the National Comprehensive Cancer Network* 20(8):925–51. doi: 10.6004/JNCCN.2022.0040.

Halgren, Thomas A., Robert B. Murphy, Richard A. Friesner, Hege S. Beard, Leah L. Frye, W. Thomas Pollard, and Jay L. Banks. 2004. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening." *Journal of Medicinal Chemistry* 47(7):1750–59. doi: 10.1021/JM030644S/SUPPL_FILE/JM030644S_S.PDF.

Harder, Edward, Wolfgang Damm, Jon Maple, Chuanjie Wu, Mark Reboul, Jin Yu Xiang, Lingle Wang, Dmitry Lupyan, Markus K. Dahlgren, Jennifer L. Knight,

Joseph W. Kaus, David S. Cerutti, Goran Krilov, William L. Jorgensen, Robert Abel, and Richard A. Friesner. 2016. “OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins.” *Journal of Chemical Theory and Computation* 12(1):281–96. doi: 10.1021/ACS.JCTC.5B00864.

Haugen, Bryan R., Erik K. Alexander, Keith C. Bible, Gerard M. Doherty, Susan J. Mandel, Yuri E. Nikiforov, Furio Pacini, Gregory W. Randolph, Anna M. Sawka, Martin Schlumberger, Kathryn G. Schuff, Steven I. Sherman, Julie Ann Sosa, David L. Steward, R. Michael Tuttle, and Leonard Wartofsky. 2016. “2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer.” *Thyroid* 26(1):1–133. doi: 10.1089/THY.2015.0020/ASSET/IMAGES/LARGE/FIGURE 8.JPEG.

Heinrich, Michael C., Christopher L. Corless, George D. Demetri, Charles D. Blanke, Margaret Von Mehren, Heikki Joensuu, Laura S. McGreevey, Chang Jie Chen, Annick D. Van Den Abbeele, Brian J. Druker, Beate Kiese, Burton Eisenberg, Peter J. Roberts, Samuel Singer, Christopher D. M. Fletcher, Sandra Silberman, Sasa Dimitrijevic, and Jonathan A. Fletcher. 2003. “Kinase Mutations and Imatinib Response in Patients with Metastatic Gastrointestinal Stromal Tumor.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 21(23):4342–49. doi: 10.1200/JCO.2003.04.190.

Herbst. 2013. “Lung Cancer Lung Cancer Lung Cancer.” *Conn’s Current Therapy 2020* 2030(November):133–41.

Horn, Leora, Aaron S. Mansfield, Aleksandra Szczęsna, Libor Havel, Maciej Krzakowski, Maximilian J. Hochmair, Florian Huemer, György Losonczy, Melissa L. Johnson, Makoto Nishio, Martin Reck, Tony Mok, Sivuonthanh Lam, David S. Shames, Juan Liu, Beiying Ding, Ariel Lopez-Chavez, Fairooz Kabbinavar, Wei Lin, Alan Sandler, and Stephen V. Liu. 2018. “First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer.”

New England Journal of Medicine 379(23):2220–29. doi: 10.1056/NEJMOA1809064/SUPPL_FILE/NEJMOA1809064_DATA-SHARING.PDF.

Hubbard, Stevan R. 1999. "Structural Analysis of Receptor Tyrosine Kinases." *Progress in Biophysics and Molecular Biology* 71(3–4):343–58. doi: 10.1016/S0079-6107(98)00047-9.

Jacobson, Matthew P., David L. Pincus, Chaya S. Rapp, Tyler J. F. Day, Barry Honig, David E. Shaw, and Richard A. Friesner. 2004. "A Hierarchical Approach to All-Atom Protein Loop Prediction." *Proteins* 55(2):351–67. doi: 10.1002/PROT.10613.

Jain, Ajay N. 1996. "Scoring Noncovalent Protein-Ligand Interactions: A Continuous Differentiable Function Tuned to Compute Binding Affinities." *Journal of Computer-Aided Molecular Design* 10(5):427–40. doi: 10.1007/BF00124474/METRICS.

Jiménez-Luna, José, Alberto Cuzzolin, Giovanni Bolcato, Mattia Sturlese, and Stefano Moro. 2020. "A Deep-Learning Approach toward Rational Molecular Docking Protocol Selection." *Molecules 2020, Vol. 25, Page 2487* 25(11):2487. doi: 10.3390/MOLECULES25112487.

Jing, Shuqian, Duanzhi Wen, Yanbin Yu, Paige L. Holst, Yi Luo, Mei Fang, Rami Tamir, Laarni Antonio, Zheng Hu, Rod Cupples, Jean Claude Louis, Sylvia Hu, Bruce W. Altrock, and Gary M. Fox. 1996. "GDNF-Induced Activation of the Ret Protein Tyrosine Kinase Is Mediated by GDNFR-Alpha, a Novel Receptor for GDNF." *Cell* 85(7):1113–24. doi: 10.1016/S0092-8674(00)81311-2.

Joensuu, H. 2006. "Gastrointestinal Stromal Tumor (GIST)." *Annals of Oncology* 17(SUPPL. 10). doi: 10.1093/annonc/mdl274.

Johnson, Douglas B., Monica V. Estrada, Roberto Salgado, Violeta Sanchez, Deon B. Doxie, Susan R. Opalenik, Anna E. Vilgelm, Emily Feld, Adam S. Johnson, Allison R. Greenplate, Melinda E. Sanders, Christine M. Lovly, Dennie T. Frederick, Mark C. Kelley, Ann Richmond, Jonathan M. Irish, Yu Shyr, Ryan J. Sullivan, Igor Puzanov, Jeffrey A. Sosman, and Justin M. Balko. 2016. "Melanoma-Specific MHC-II Expression Represents a Tumour-Autonomous Phenotype and Predicts Response to Anti-PD-1/PD-L1 Therapy." *Nature Communications 2016 7:1* 7(1):1–10. doi: 10.1038/ncomms10582.

Joo, Won Duk, Irene Visintin, and Gil Mor. 2013. "Targeted Cancer Therapy – Are the Days of Systemic Chemotherapy Numbered?" *Maturitas* 76(4):308–14. doi: 10.1016/J.MATURITAS.2013.09.008.

Jorgensen, William L., Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. 1983. "Comparison of Simple Potential Functions for Simulating Liquid Water." *The Journal of Chemical Physics* 79(2):926–35. doi: 10.1063/1.445869.

Kantarjian, Hagop, Francis Giles, Lydia Wunderle, Kapil Bhalla, Susan O'Brien, Barbara Wassmann, Chiaki Tanaka, Paul Manley, Patricia Rae, William Mietlowski, Kathy Bochinski, Andreas Hochhaus, James D. Griffin, Dieter Hoelzer, Maher Albitar, Margaret Dugan, Jorge Cortes, Leila Alland, and Oliver G. Ottmann. 2006. "Nilotinib in Imatinib-Resistant CML and Philadelphia Chromosome–Positive ALL." *New England Journal of Medicine* 354(24):2542–51. doi: 10.1056/NEJMOA055104/SUPPL_FILE/NEJM_KANTARJIAN_2542SA1.PDF.

Karplus, Martin, and J. Andrew McCammon. 2002. "Molecular Dynamics Simulations of Biomolecules." *Nature Structural Biology* 9(9):646–52. doi: 10.1038/NSB0902-646.

Knowles, Phillip P., Judith Murray-Rust, Svend Kjær, Rizaldy P. Scott, Sarah Hanrahan, Massimo Santoro, Carlos F. Ibáñez, and Neil Q. McDonald. 2006. "Structure and Chemical Inhibition of the RET Tyrosine Kinase Domain." *Journal of Biological Chemistry* 281(44):33577–87. doi: 10.1074/JBC.M605604200.

Kramer, Edgar R., and Birgit Liss. 2015. "GDNF–Ret Signaling in Midbrain Dopaminergic Neurons and Its Implication for Parkinson Disease." *FEBS Letters* 589(24PartA):3760–72. doi: 10.1016/J.FEBSLET.2015.11.006.

Kuchana, Vinutha, Vaeshnavi Kashetti, Sai Kiran Reddy Peddi, Sreekanth Sivan, and Vijjulatha Manga. 2022. "Integrated Computational Approach for in Silico Design of New Purinyl Pyridine Derivatives as B-Raf Kinase Inhibitors." *Journal of Receptors and Signal Transduction* 42(5):439–53. doi: 10.1080/10799893.2021.1999472/SUPPL_FILE/IRST_A_1999472_SM9172.PDF.

Kumi Kawai, and Naoya Asai Masahide Takahashi. 2020. "Roles of the RET Proto-Oncogene in Cancer and Development." *JMA Journal* 3(3):175–81. doi: 10.31662/JMAJ.2020-0021.

Lanzi, Cinzia, Giuliana Cassinelli, Valentina Nicolini, and Franco Zunino. 2009. "Targeting RET for Thyroid Cancer Therapy." *Biochemical Pharmacology* 77(3):297–309. doi: 10.1016/J.BCP.2008.10.033.

Leach, Andrew R. 1996. "Molecular Modelling : Principles and Applications." 595.

Lemmon, Mark A., and Joseph Schlessinger. 2010. "Cell Signaling by Receptor Tyrosine Kinases." *Cell* 141(7):1117–34. doi: 10.1016/J.CELL.2010.06.011.

Li, Jianing, Severin T. Schneebeli, Joseph Bylund, Ramy Farid, and Richard A. Friesner. 2011. "IDSite: An Accurate Approach to Predict P450-Mediated Drug Metabolism." *Journal of Chemical Theory and Computation* 7(11):3829–45. doi: 10.1021/CT200462Q/SUPPL_FILE/CT200462Q_SI_001.PDF.

Li, Tao, Wen Yu Yang, Ting Ting Liu, Yao Li, Lu Liu, Xuan Zheng, Lei Zhao, Fan Zhang, and Yi Hu. 2023. “Advances in the Diagnosis and Treatment of a Driving Target: RET Rearrangements in Non-Small-Cell Lung Cancer (NSCLC) Especially in China.” *Technology in Cancer Research and Treatment* 22. doi: 10.1177/15330338221148802/ASSET/IMAGES/LARGE/10.1177_15330338221148802-FIG1.JPEG.

Lichtenfels, Ana Julia De F. C., Diana A. Van Der Plaat, Kim De Jong, Cleo C. Van Diemen, Dirkje S. Postma, Ivana Nedeljkovic, Cornelia M. Van Duijn, Najaf Amin, Sacha La Bastide-Van Gemert, Maaike De Vries, Cavin K. Ward-Caviness, Kathrin Wolf, Melanie Waldenberger, Annette Peters, Ronald P. Stolk, Bert Brunekreef, H. Marike Boezen, and Judith M. Vonk. 2018. “Long-Term Air Pollution Exposure, Genome-Wide DNA Methylation and Lung Function in the LifeLines Cohort Study.” *Environmental Health Perspectives* 126(2). doi: 10.1289/EHP2045.

Lin, Jessica J., and Alice T. Shaw. 2016. “Resisting Resistance: Targeted Therapies in Lung Cancer.” *Trends in Cancer* 2(7):350–64. doi: 10.1016/J.TRECAN.2016.05.010.

Lipinski, Christopher A., Franco Lombardo, Beryl W. Dominy, and Paul J. Feeney. 2001. “Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings.” *Advanced Drug Delivery Reviews* 46(1–3):3–26. doi: 10.1016/S0169-409X(00)00129-0.

Liu, Xuan, Xueqing Hu, Tao Shen, Qi Li, Blaine H. M. Mooers, and Jie Wu. 2020. “RET Kinase Alterations in Targeted Cancer Therapy.” *Cancer Drug Resistance* 3(3):472. doi: 10.20517/CDR.2020.15.

Lu, Yue C., Olha V. Nazarko, Richard Sando, Gabriel S. Salzman, Thomas C. Südhof, and Demet Araç. 2015. “Structural Basis of Latrophilin-FLRT-UNC5 Interaction in Cell Adhesion.” *Structure (London, England : 1993)* 23(9):1678–91. doi: 10.1016/J.STR.2015.06.024.

Lukeman, John M. 2015. “What Is Lung Cancer?” *Perspectives in Lung Cancer* 30–40. doi: 10.1159/000400400.

Luo, Zhibo, Lingli Wang, Zhifei Fu, Bin Shuai, Miaorong Luo, Guoping Hu, Jian Chen, Jikui Sun, Jiansong Wang, Jian Li, Shuhui Chen, and Yang Zhang. 2021. “Discovery and Optimization of Selective RET Inhibitors via Scaffold Hopping.” *Bioorganic & Medicinal Chemistry Letters* 47:128149. doi: 10.1016/J.BMCL.2021.128149.

Madhavi Sastry, G., Matvey Adzhigirey, Tyler Day, Ramakrishna Annabhimoju, and Woody Sherman. 2013. “Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments.” *Journal of Computer-Aided Molecular Design* 27(3):221–34. doi: 10.1007/S10822-013-9644-8.

Maemondo, Makoto, Akira Inoue, Kunihiko Kobayashi, Shunichi Sugawara, Satoshi Oizumi, Hiroshi Isobe, Akihiko Gemma, Masao Harada, Hirohisa Yoshizawa, Ichiro Kinoshita, Yuka Fujita, Shoji Okinaga, Haruto Hirano, Kozo Yoshimori, Toshiyuki Harada, Takashi Ogura, Masahiro Ando, Hitoshi Miyazawa, Tomoaki Tanaka, Yasuo Saijo, Koichi Hagiwara, Satoshi Morita, and Toshihiro Nukiwa. 2010. “Gefitinib or Chemotherapy for Non–Small-Cell Lung Cancer with Mutated EGFR.” *New England Journal of Medicine* 362(25):2380–88. doi: 10.1056/NEJMOA0909530/SUPPL_FILE/NEJMOA0909530_DISCLOSURES.PDF.

Manning, Brendan D., and Lewis C. Cantley. 2007. “AKT/PKB Signaling: Navigating Downstream.” *Cell* 129(7):1261–74. doi: 10.1016/J.CELL.2007.06.009.

Marsh, D. J., D. L. Learoyd, and B. G. Robinson. 1995. “Medullary Thyroid Carcinoma: Recent Advances and Management Update.” *Thyroid : Official Journal of the American Thyroid Association* 5(5):407–24. doi: 10.1089/THY.1995.5.407.

Martyna, Glenn J., Michael L. Klein, and Mark Tuckerman. 1992. “Nosé–Hoover Chains: The Canonical Ensemble via Continuous Dynamics.” *The Journal of Chemical Physics* 97(4):2635–43. doi: 10.1063/1.463940.

McCammon, J. Andrew, Bruce R. Gelin, and Martin Karplus. 1977. “Dynamics of Folded Proteins.” *Nature* 267(5612):585–90. doi: 10.1038/267585A0.

McKinney, Wes. 2010. “Data Structures for Statistical Computing in Python.” *Proceedings of the 9th Python in Science Conference* 56–61. doi: 10.25080/MAJORA-92BF1922-00A.

Mendelsohn, John, and Jose Baselga. 2000. “The EGF Receptor Family as Targets for Cancer Therapy.” *Oncogene 2000 19:56* 19(56):6550–65. doi: 10.1038/sj.onc.1204082.

Mendez, David, Anna Gaulton, A. Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula Magariños, Juan F. Mosquera, Prudence Mutowo, Michał Nowotka, María Gordillo-Marañón, Fiona Hunter, Laura Junco, Grace Mugumbate, Milagros Rodriguez-Lopez, Francis Atkinson, Nicolas Bosc, Chris J. Radoux, Aldo Segura-Cabrera, Anne Hersey, and Andrew R. Leach. 2019. “ChEMBL: Towards Direct Deposition of Bioassay Data.” *Nucleic Acids Research* 47(D1):D930–40. doi: 10.1093/NAR/GKY1075.

Mok, Tony S., Yi-Long Wu, Myung-Ju Ahn, Marina C. Garassino, Hye R. Kim, Suresh S. Ramalingam, Frances A. Shepherd, Yong He, Hiroaki Akamatsu, Willemijn S. M. E. Theelen, Chee K. Lee, Martin Sebastian, Alison Templeton, Helen Mann, Marcelo Marotti, Serban Ghiorghiu, and Vassiliki A. Papadimitrakopoulou. 2017. “Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer.” *The New England Journal of Medicine* 376(7):629–40. doi: 10.1056/NEJMOA1612674.

Molina, Julian R., Ping Yang, Stephen D. Cassivi, Steven E. Schild, and Alex A. Adjei. 2008. “Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship.” *Mayo Clinic Proceedings* 83(5):584–94. doi: 10.4065/83.5.584.

Morris, Garrett M., David S. Goodsell, Robert S. Halliday, Ruth Huey, William E. Hart, Richard K. Belew, and Arthur J. Olson. 1639. "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function." *Journal of Computational Chemistry* 19(14):16391662. doi: 10.1002/(SICI)1096-987X(19981115)19:14.

Morris, Luc G. T., Andrew G. Sikora, Tor D. Tosteson, and Louise Davies. 2013. "The Increasing Incidence of Thyroid Cancer: The Influence of Access to Care." *Thyroid : Official Journal of the American Thyroid Association* 23(7):885–91. doi: 10.1089/THY.2013.0045.

Motzer, Robert J., Thomas E. Hutson, Piotr Tomczak, M. Dror Michaelson, Ronald M. Bukowski, Olivier Rixe, Stéphane Oudard, Sylvie Negrier, Cezary Szczylik, Sindy T. Kim, Isan Chen, Paul W. Bycott, Charles M. Baum, and Robert A. Figlin. 2007. "Sunitinib versus Interferon Alfa in Metastatic Renal-Cell Carcinoma." *The New England Journal of Medicine* 356(2):115–24. doi: 10.1056/NEJMOA065044.

Moumbock, Aurélien F. A., Jianyu Li, Hoai T. T. Tran, Rahel Hinkelmann, Evelyn Lamy, Henning J. Jessen, and Stefan Günther. 2021. "EPharmaLib: A Versatile Library of e-Pharmacophores to Address Small-Molecule (Poly-)Pharmacology." *Journal of Chemical Information and Modeling* 61(7):3659–66. doi: 10.1021/ACS.JCIM.1C00135/ASSET/IMAGES/LARGE/CI1C00135_0007.JPEG.

Moura, Margarida M., Branca M. Cavaco, António E. Pinto, and Valeriano Leite. 2011. "High Prevalence of RAS Mutations in RET-Negative Sporadic Medullary Thyroid Carcinomas." *The Journal of Clinical Endocrinology & Metabolism* 96(5):E863–68. doi: 10.1210/JC.2010-1921.

Mulligan, Lois M. 2014. "RET Revisited: Expanding the Oncogenic Portfolio." *Nature Reviews Cancer 2014 14:3* 14(3):173–86. doi: 10.1038/nrc3680.

Mulligan, Lois M. 2018. “GDNF and the RET Receptor in Cancer: New Insights and Therapeutic Potential.” *Frontiers in Physiology* 9(JAN). doi: 10.3389/FPHYS.2018.01873.

Mulligan, Lois M. 2019. “GDNF and the RET Receptor in Cancer: New Insights and Therapeutic Potential.” *Frontiers in Physiology* 10(JAN):1873. doi: 10.3389/FPHYS.2018.01873/BIBTEX.

Myshkin, Eugene, Richard Brennan, Tatiana Khasanova, Tatiana Sitnik, Tatiana Serebriyskaya, Elena Litvinova, Alexey Guryanov, Yuri Nikolsky, Tatiana Nikolskaya, and Svetlana Bureeva. 2012. “Prediction of Organ Toxicity Endpoints by QSAR Modeling Based on Precise Chemical-Histopathology Annotations.” *Chemical Biology & Drug Design* 80(3):406–16. doi: 10.1111/J.1747-0285.2012.01411.X.

Nikiforov, Yuri E., and Marina N. Nikiforova. 2011. “Molecular Genetics and Diagnosis of Thyroid Cancer.” *Nature Reviews. Endocrinology* 7(10):569–80. doi: 10.1038/NRENDO.2011.142.

Östman, Arne, and Frank D. Böhmer. 2001. “Regulation of Receptor Tyrosine Kinase Signaling by Protein Tyrosine Phosphatases.” *Trends in Cell Biology* 11(6):258–66. doi: 10.1016/S0962-8924(01)01990-0.

Ouassaf, Mebarka, Ossama Daoui, Sarfaraz Alam, Souad Elkhattabi, Salah Belaidi, and Samir Chtita. 2022. “Pharmacophore-Based Virtual Screening, Molecular Docking, and Molecular Dynamics Studies for the Discovery of Novel FLT3 Inhibitors.” *Journal of Biomolecular Structure and Dynamics*. doi: 10.1080/07391102.2022.2123403/SUPPL_FILE/TBSD_A_2123403_SM1349.DOCX.

Paik, Paul K., Maria E. Arcila, Michael Fara, Camelia S. Sima, Vincent A. Miller, Mark G. Kris, Marc Ladanyi, and Gregory J. Riely. 2011. “Clinical Characteristics

of Patients with Lung Adenocarcinomas Harboring BRAF Mutations." *Journal of Clinical Oncology* 29(15):2046–51. doi: 10.1200/JCO.2010.33.1280.

Pan, Yue, Chao Deng, Zhenhua Qiu, Chenghui Cao, and Fang Wu. 2021. "The Resistance Mechanisms and Treatment Strategies for ALK-Rearranged Non-Small Cell Lung Cancer." *Frontiers in Oncology* 11:3945. doi: 10.3389/FONC.2021.713530/BIBTEX.

Parkash, Vimal, Veli Matti Leppänen, Heidi Virtanen, Jaana M. Jurvansuu, Maxim M. Bespalov, Yulia A. Sidorova, Pia Runeberg-Roos, Mart Saarma, and Adrian Goldman. 2008. "The Structure of the Glial Cell Line-Derived Neurotrophic Factor-Coreceptor Complex: INSIGHTS INTO RET SIGNALING AND HEPARIN BINDING." *Journal of Biological Chemistry* 283(50):35164–72. doi: 10.1074/JBC.M802543200.

PDQ Adult Treatment Editorial Board. 2002. "Small Cell Lung Cancer Treatment (PDQ®): Patient Version." *PDQ Cancer Information Summaries*.

Plaza-Menacho, I., L. Mologni, and N. Q. McDonald. 2014. "Mechanisms of RET Signaling in Cancer: Current and Future Implications for Targeted Therapy." *Cellular Signalling* 26(8):1743–52. doi: 10.1016/J.CELLSIG.2014.03.032.

Plaza-Menacho, Iván. 2018. "Structure and Function of RET in Multiple Endocrine Neoplasia Type 2." *Endocrine-Related Cancer* 25(2):T79–90. doi: 10.1530/ERC-17-0354.

Plaza-Menacho, Iván, Luca Mologni, Elisa Sala, Carlo Gambacorti-Passerini, Anthony I. Magee, Thera P. Links, Robert M. W. Hofstra, David Barford, and Clare M. Isacke. 2007. "Sorafenib Functions to Potently Suppress RET Tyrosine Kinase Activity by Direct Enzymatic Inhibition and Promoting RET Lysosomal Degradation Independent of Proteasomal Targeting." *Journal of Biological Chemistry* 282(40):29230–40. doi: 10.1074/JBC.M703461200.

Rini, Brian I., Bernard Escudier, Piotr Tomczak, Andrey Kaprin, Cezary Szczylik, Thomas E. Hutson, M. Dror Michaelson, Vera A. Gorbunova, Martin E. Gore, Igor G. Rusakov, Sylvie Negrier, Yen Chuan Ou, Daniel Castellano, Ho Yeong Lim, Hirotsugu Uemura, Jamal Tarazi, David Cella, Connie Chen, Brad Rosbrook, Sinil Kim, and Robert J. Motzer. 2011. "Comparative Effectiveness of Axitinib versus Sorafenib in Advanced Renal Cell Carcinoma (AXIS): A Randomised Phase 3 Trial." *The Lancet* 378(9807):1931–39. doi: 10.1016/S0140-6736(11)61613-9.

Romei, Cristina, Raffaele Ciampi, and Rossella Elisei. 2016. "A Comprehensive Overview of the Role of the RET Proto-Oncogene in Thyroid Carcinoma." *Nature Reviews Endocrinology 2016 12:4* 12(4):192–202. doi: 10.1038/nrendo.2016.11.

Roos, Katarina, Chuanjie Wu, Wolfgang Damm, Mark Reboul, James M. Stevenson, Chao Lu, Markus K. Dahlgren, Sayan Mondal, Wei Chen, Lingle Wang, Robert Abel, Richard A. Friesner, and Edward D. Harder. 2019. "OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules." *Journal of Chemical Theory and Computation* 15(3):1863–74. doi: 10.1021/ACS.JCTC.8B01026.

Rosen, Ezra Y., Melissa L. Johnson, Sarah E. Clifford, Romel Somwar, Jennifer F. Kherani, Jieun Son, Arrien A. Bertram, Monika A. Davare, Eric Gladstone, Elena V. Ivanova, Dahlia N. Henry, Elaine M. Kelley, Mika Lin, Marina S. D. Milan, Binoj C. Nair, Elizabeth A. Olek, Jenna E. Scanlon, Morana Vojnic, Kevin Ebata, Jaclyn F. Hechtman, Bob T. Li, Lynette M. Sholl, Barry S. Taylor, Marc Ladanyi, Pasi A. Janne, S. Michael Rothenberg, Alexander Drilon, and Geoffrey R. Oxnard. 2021. "Overcoming MET-Dependent Resistance to Selective RET Inhibition in Patients with RET Fusion-Positive Lung Cancer by Combining Selpercatinib with Crizotinib." *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 27(1):34–42. doi: 10.1158/1078-0432.CCR-20-2278.

Roskoski, Robert. 2016. “Classification of Small Molecule Protein Kinase Inhibitors Based upon the Structures of Their Drug-Enzyme Complexes.” *Pharmacological Research* 103:26–48. doi: 10.1016/J.PHRS.2015.10.021.

Roskoski, Robert. 2021. “Properties of FDA-Approved Small Molecule Protein Kinase Inhibitors: A 2021 Update.” *Pharmacological Research* 165. doi: 10.1016/J.PHRS.2021.105463.

Rudin, Charles M., and John T. Poirier. 2016. “Shining Light on Novel Targets and Therapies.” *Nature Reviews Clinical Oncology 2016 14:2* 14(2):75–76. doi: 10.1038/nrclinonc.2016.203.

Salvatore, Domenico, Massimo Santoro, and Martin Schlumberger. 2021. “The Importance of the RET Gene in Thyroid Cancer and Therapeutic Implications.” *Nature Reviews Endocrinology 2021 17:5* 17(5):296–306. doi: 10.1038/s41574-021-00470-9.

Santoro, Massimo, Marialuisa Moccia, Giorgia Federico, and Francesca Carlomagno. 2020. “RET Gene Fusions in Malignancies of the Thyroid and Other Tissues.” *Genes 2020, Vol. 11, Page 424* 11(4):424. doi: 10.3390/GENES11040424.

Shabbir, Arafat, Arsenije Kojadinovic, Tabinda Shafiq, and Prabhjot S. Mundi. 2023. “Targeting RET Alterations in Cancer: Recent Progress and Future Directions.” *Critical Reviews in Oncology/Hematology* 181. doi: 10.1016/J.CRITREVONC.2022.103882.

Shaw, Alice T., Dong-Wan Kim, Kazuhiko Nakagawa, Takashi Seto, Lucio Crinó, Myung-Ju Ahn, Tommaso De Pas, Benjamin Besse, Benjamin J. Solomon, Fiona Blackhall, Yi-Long Wu, Michael Thomas, Kenneth J. O’Byrne, Denis Moro-Sibilot, D. Ross Camidge, Tony Mok, Vera Hirsh, Gregory J. Riely, Shrividya Iyer, Vanessa Tassell, Anna Polli, Keith D. Wilner, and Pasi A. Jänne. 2013. “ Crizotinib versus Chemotherapy in Advanced ALK -Positive Lung Cancer .” *New England Journal of Medicine* 368(25):2385–94. doi:

10.1056/NEJMOA1214886/SUPPL_FILE/NEJMOA1214486_APPENDIX.PDF
.

Shelley, John C., Anuradha Cholleti, Leah L. Frye, Jeremy R. Greenwood, Mathew R. Timlin, and Makoto Uchimaya. 2007. "Epik: A Software Program for PK(a) Prediction and Protonation State Generation for Drug-like Molecules." *Journal of Computer-Aided Molecular Design* 21(12):681–91. doi: 10.1007/S10822-007-9133-Z.

Siddhartha Mukherjee. 2011. "The Emperor of All Maladies." P. 57(4) 460 in *A biography of cancer*. Simon and Schuster Canada.

Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. 2020. "Cancer Statistics, 2020." *CA: A Cancer Journal for Clinicians* 70(1):7–30. doi: 10.3322/CAAC.21590.

Siveen, Kodappully S., Kirti S. Prabhu, Iman W. Achkar, Shilpa Kuttikrishnan, Sunitha Shyam, Abdul Q. Khan, Maysaloun Merhi, Said Dermime, and Shahab Uddin. 2018. "Role of Non Receptor Tyrosine Kinases in Hematological Malignances and Its Targeting by Natural Products." *Molecular Cancer 2018 17:1* 17(1):1–21. doi: 10.1186/S12943-018-0788-Y.

Smallridge, Robert C., Kenneth B. Ain, Sylvia L. Asa, Keith C. Bible, James D. Brierley, Kenneth D. Burman, Electron Kebebew, Nancy Y. Lee, Yuri E. Nikiforov, M. Sara Rosenthal, Manisha H. Shah, Ashok R. Shaha, and R. Michael Tuttle. 2012. "American Thyroid Association Guidelines for Management of Patients with Anaplastic Thyroid Cancer." *Thyroid : Official Journal of the American Thyroid Association* 22(11):1104–39. doi: 10.1089/THY.2012.0302.

Sterling, Teague, and John J. Irwin. 2015. "ZINC 15 - Ligand Discovery for Everyone." *Journal of Chemical Information and Modeling* 55(11):2324–37. doi: 10.1021/ACS.JCIM.5B00559/ASSET/IMAGES/LARGE/CI-2015-00559J_0005.JPEG.

Subbiah, V., T. Shen, S. S. Terzyan, X. Liu, X. Hu, K. P. Patel, M. Hu, M. Cabanillas, A. Behrang, F. Meric-Bernstam, P. T. T. Vo, B. H. M. Mooers, and J. Wu. 2021. "Structural Basis of Acquired Resistance to Selpercatinib and Pralsetinib Mediated by Non-Gatekeeper RET Mutations." *Annals of Oncology : Official Journal of the European Society for Medical Oncology* 32(2):261–68. doi: 10.1016/J.ANNONC.2020.10.599.

Subbiah, V., V. Velcheti, B. B. Tuch, K. Ebata, N. L. Busaidy, M. E. Cabanillas, L. J. Wirth, S. Stock, S. Smith, V. Lauriault, S. Corsi-Travali, D. Henry, M. Burkard, R. Hamor, K. Bouhana, S. Winski, R. D. Wallace, D. Hartley, S. Rhodes, M. Reddy, B. J. Brandhuber, S. Andrews, S. M. Rothenberg, and A. Drilon. 2018. "Selective RET Kinase Inhibition for Patients with RET-Altered Cancers." *Annals of Oncology : Official Journal of the European Society for Medical Oncology* 29(8):1869–76. doi: 10.1093/ANNONC/MDY137.

Subbiah, Vivek, Justin F. Gainor, Rami Rahal, Jason D. Brubaker, Joseph L. Kim, Michelle Maynard, Wei Hu, Qiongfang Cao, Michael P. Sheets, Douglas Wilson, Kevin J. Wilson, Lucian Dipietro, Paul Fleming, Michael Palmer, Mimi I. Hu, Lori Wirth, Marcia S. Brose, Sai Hong Ignatius Ou, Matthew Taylor, Elena Garralda, Stephen Miller, Beni Wolf, Christoph Lengauer, Timothy Guzi, and Erica K. Evans. 2018. "Precision Targeted Therapy with BLU-667 for RET-Driven Cancers." *Cancer Discovery* 8(7):836–49. doi: 10.1158/2159-8290.CD-18-0338/42851/AM/PRECISION-TARGETED-THERAPY-WITH-BLU-667-FOR-RET.

Suhandi, Cecep, Ersa Fadhilah, Nurfianti Silvia, Annisa Atusholihah, Randy Rassi Prayoga, Sandra Megantara, and Muchtaridi Muchtaridi. 2021. "Molecular Docking Study of Mangosteen (Garcinia Mangostana L.) Xanthone-Derived Isolates as Anti Androgen." *Indonesian Journal of Cancer Chemoprevention 2021, Vol. 12, Pages 11-20* 12(1):11–20. doi: 10.14499/INDONESIANJCANCHEMOPREV12ISS1PP11-20.

Takahashi, Masahide, Jerome Ritz, and Geoffrey M. Cooper. 1985. “Activation of a Novel Human Transforming Gene, Ret, by DNA Rearrangement.” *Cell* 42(2):581–88. doi: 10.1016/0092-8674(85)90115-1.

Taylor, Matthew H., Justin F. Gainor, Mimi I. Nan Hu, Viola Weijia Zhu, Gilberto Lopes, Sophie Leboulleux, Marcia S. Brose, Martin H. Schuler, Daniel W. Bowles, Dong-Wan Kim, Christina S. Baik, Elena Garralda, Chia-Chi Lin, Douglas Adkins, Debashis Sarker, Giuseppe Curigliano, Hui Zhang, Corinne Clifford, Christopher D. Turner, and Vivek Subbiah. 2019. “Activity and Tolerability of BLU-667, a Highly Potent and Selective RET Inhibitor, in Patients with Advanced RET-Altered Thyroid Cancers.” *Https://Doi.Org/10.1200/JCO.2019.37.15_suppl.6018* 37(15_suppl):6018–6018. doi: 10.1200/JCO.2019.37.15_SUPPL.6018.

The American Cancer Society. 2021. “The History of Cancer.” *The American Cancer Society*. Retrieved (https://www.cancer.org/cancer/understanding-cancer/history-of-cancer.html).

The American Cancer Society medical and editorial content team. 2019. “Lung Cancer Causes, Risk Factors, and Prevention.” *The American Cancer Society* 1–14.

Travis, William D., Elisabeth Brambilla, Andrew G. Nicholson, Yasushi Yatabe, John H. M. Austin, Mary Beth Beasley, Lucian R. Chirieac, Sanja Dacic, Edwina Duhig, Douglas B. Flieder, Kim Geisinger, Fred R. Hirsch, Yuichi Ishikawa, Keith M. Kerr, Masayuki Noguchi, Giuseppe Pelosi, Charles A. Powell, Ming Sound Tsao, and Ignacio Wistuba. 2015. “The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification.” *Journal of Thoracic Oncology : Official Publication of the International Association for the Study of Lung Cancer* 10(9):1243–60. doi: 10.1097/JTO.0000000000000630.

Trenker, Raphael, and Natalia Jura. 2020. “Receptor Tyrosine Kinase Activation: From the Ligand Perspective.” *Current Opinion in Cell Biology* 63:174–85. doi: 10.1016/J.CEB.2020.01.016.

Vidal, Marcos, Samuel Wells, Anderson Ryan, and Ross Cagan. 2005. “ZD6474 Suppresses Oncogenic RET Isoforms in a Drosophila Model for Type 2 Multiple Endocrine Neoplasia Syndromes and Papillary Thyroid Carcinoma.” *Cancer Research* 65(9):3538–41. doi: 10.1158/0008-5472.CAN-04-4561.

Vodopivec, Danica M., and Mimi I. Hu. 2022. “RET Kinase Inhibitors for RET-Altered Thyroid Cancers.” *Therapeutic Advances in Medical Oncology* 14. doi: 10.1177/17588359221101691/ASSET/IMAGES/LARGE/10.1177_17588359221101691-FIG3.JPEG.

Wang, Ercheng, Huiyong Sun, Junmei Wang, Zhe Wang, Hui Liu, John Z. H. Zhang, and Tingjun Hou. 2019. “End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design.” *Chemical Reviews* 119(16):9478–9508. doi: 10.1021/ACS.CHEMREV.9B00055.

Wang, Renxiao, Ying Fu, and Luhua Lai. 1997. “A New Atom-Additive Method for Calculating Partition Coefficients.” *Journal of Chemical Information and Computer Sciences* 37(3):615–21. doi: 10.1021/CI960169P/SUPPL_FILE/CI615.PDF.

Wang, Xinquan. 2013. “Structural Studies of GDNF Family Ligands with Their Receptors-Insights into Ligand Recognition and Activation of Receptor Tyrosine Kinase RET.” *Biochimica et Biophysica Acta* 1834(10):2205–12. doi: 10.1016/J.BBAPAP.2012.10.008.

Weinstein, I. Bernard, and Andrew Joe. 2008. “Oncogene Addiction.” *Cancer Research* 68(9):3077–80. doi: 10.1158/0008-5472.CAN-07-3293.

Wells, Samuel A., Sylvia L. Asa, Henning Dralle, Rossella Elisei, Douglas B. Evans, Robert F. Gagel, Nancy Lee, Andreas MacHens, Jeffrey F. Moley, Furio Pacini,

Friedhelm Raue, Karin Frank-Raue, Bruce Robinson, M. Sara Rosenthal, Massimo Santoro, Martin Schlumberger, Manisha Shah, and Steven G. Waguespack. 2015a. “Revised American Thyroid Association Guidelines for the Management of Medullary Thyroid Carcinoma.” *Thyroid* 25(6):567–610. doi: 10.1089/THY.2014.0335/ASSET/IMAGES/LARGE/FIGURE 3.JPEG.

Wells, Samuel A., Sylvia L. Asa, Henning Dralle, Rossella Elisei, Douglas B. Evans, Robert F. Gagel, Nancy Lee, Andreas MacHens, Jeffrey F. Moley, Furio Pacini, Friedhelm Raue, Karin Frank-Raue, Bruce Robinson, M. Sara Rosenthal, Massimo Santoro, Martin Schlumberger, Manisha Shah, and Steven G. Waguespack. 2015b. “Revised American Thyroid Association Guidelines for the Management of Medullary Thyroid Carcinoma.” *Thyroid* 25(6):567–610. doi: 10.1089/THY.2014.0335/ASSET/IMAGES/LARGE/FIGURE 3.JPEG.

Wells, Samuel A., and Massimo Santoro. 2009. “Targeting the RET Pathway in Thyroid Cancer.” *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 15(23):7119–23. doi: 10.1158/1078-0432.CCR-08-2742.

Wolber, Gerhard, and Thierry Langer. 2005. “LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters.” *Journal of Chemical Information and Modeling* 45(1):160–69. doi: 10.1021/CI049885E.

Wong, Chi Heem, Kien Wei Siah, and Andrew W. Lo. 2019. “Estimation of Clinical Trial Success Rates and Related Parameters.” *Biostatistics* 20(2):273–86. doi: 10.1093/BIOSTATISTICS/KXX069.

Wu, David., David Chandler, and David Chandler. 1988. “Solutions Manual for Introduction to Modern Statistical Mechanics.” 91.

Xia, Bing, and Sai Hong Ignatius Ou. 2020. “Simultaneous RET Solvent-Front and Gatekeeper Resistance Mutations In Trans: A Rare TKI-Specific Therapeutic

Challenge?" *Journal of Thoracic Oncology : Official Publication of the International Association for the Study of Lung Cancer* 15(4):479–81. doi: 10.1016/J.JTHO.2020.02.001.

Xing, Mingzhao. 2013. "Molecular Pathogenesis and Mechanisms of Thyroid Cancer." *Nature Reviews. Cancer* 13(3):184. doi: 10.1038/NRC3431.

Yarden, Yosef, and Mark X. Sliwkowski. 2001. "Untangling the ErbB Signalling Network." *Nature Reviews. Molecular Cell Biology* 2(2):127–37. doi: 10.1038/35052073.

Zhang, Ji Hu, Thomas D. Y. Chung, and Kevin R. Oldenburg. 1999. "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays." *Journal of Biomolecular Screening* 4(2):67–73. doi: 10.1177/108705719900400206.

Zhang, Yujie, Zhichao Xing, Tianyou Liu, Minghai Tang, Li Mi, Jingqiang Zhu, Wenshuang Wu, and Tao Wei. 2022. "Targeted Therapy and Drug Resistance in Thyroid Cancer." *European Journal of Medicinal Chemistry* 238. doi: 10.1016/J.EJMECH.2022.114500.