IDENTIFICATION OF DUAL INHIBITORS FOR EGFR AND SRC PROTEIN KINASES: A MOLECULAR DOCKING STUDY

A

Dissertation Report

Submitted in Partial Fulfilment of the Requirements

For the Award of Degree of

Masters of Science

In

Biotechnology

Submitted by:

Rahul Gupta

302101039

M.Sc. Biotechnology-

External Guide:

Dr. Subhash M. Agarwal

Scientist F, Head of Department,

Bioinformatics Division,

ICMR-NICPR, Noida

Internal guide:

Dr. Siddharth Sharma

Associate Professor

Department of Biotechnology,

Thapar Institute of Engineering and Technology, Patiala

icmr | NICPR

INDIAN COUNCIL OF MEDICAL RESEARCH | NATIONAL INSTITUTE OF CANCER PREVENTION AND RESEARCH

CERTIFICATE

This is to certify that **Mr. Rahul Gupta (302101039)** has worked on the dissertation entitled **"Identification of dual inhibitors for EGFR and SRC protein kinases: A molecular docking study"** under my supervision and guidance. The contents of the thesis, being submitted to the **Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala** for the partial fulfillment of the award of Masters of Science in Biotechnology are original. It has been carried out by the candidate himself during 2nd Jan – 30th June, 2023. This thesis has not been submitted in full or part for the award of any other degree or diploma to this or any other university. Any part of this work cannot be published anywhere in any form without permission from undersigned.

Dr. Subhash Mohan Agarwal

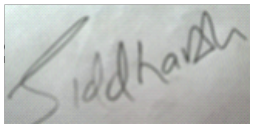
Scientist 'F' and Head

Bioinformatics Division

CERTIFICATE

On the basis of declaration submitted by **Rahul Gupta**, student of **M.Sc. Biotechnology**, I hereby certify that the project titled "**Identification of Dual Inhibitors for EGFR And SRC Protein Kinases: A Molecular Docking Study**" which is submitted to **Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala**, in partial fulfilment of the requirement for the award of the degree of M.Sc. Biotechnology, is an original contribution with existing knowledge and faithful record of work carried out by them with my guidance.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.



Dr. Siddharth Sharma
Associate Professor
Department of Biotechnology,
Thapar Institute of Engineering and
Technology, Patiala

DECLARATION

I hereby declare that the work presented in the dissertation entitled “Identification of Dual Inhibitors for EGFR And SRC Protein Kinases: A Molecular Docking Study” in partial fulfilment of the requirement for the award of the degree of Master of Science in Biotechnology, Department of Biotechnology, Thapar Institute of Engineering and Technology (TIET), Patiala, is a genuine record of my own work during the period from January 2023 to June 2023, under the supervision of Dr. Subhash M. Agarwal, Scientist ‘F’, Head of Department, Bioinformatics Division, ICMR-NICPR, Noida. The matter incorporated in this thesis has not been submitted to any other university or institute for the award of any degree in India or abroad.

Rahul Gupta

(302101039)

ACKNOWLEDGMENT

I would like to express my gratitude to my supervisor Dr. Subhash M. Agarwal, Scientist 'F', Head of Department, Bioinformatics Division, ICMR-NICPR, Noida. His valuable feedback and immense encouragement always pushed me to do more. This project has been a great learning opportunity for me and it would not have been possible without his expertise and guidance in the field. I could not have undertaken this journey without his unwavering support and uplifting spirit. I will always be thankful for this opportunity to work under him.

I would especially like to thank my internal guide, Dr. Siddharth Sharma, Associate Professor, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala for helping me to carry out this dissertation efficiently by giving his precious time and guiding me during the duration of this project.

I would like to extend my sincere thanks to our Head of Department, Dr. M.S. Reddy, Professor, for allowing me to pursue this dissertation work and to every teacher of the Department of Biotechnology for their support and guidance.

Last but not the least, I convey a hearty thanks to my family, friends and God Almighty for their constant support, wishes and blessings, which has constantly motivated me to follow the path towards the completion of this project.

Rahul Gupta

Table of Contents

List of Figures

List of tables

Abbreviations

SCLC	Small-cell lung cancer
NSCLC	Non-Small-cell lung cancer
EGFR	Epidermal Growth Factor Receptor
SFK	SRC Family Kinases
RTK	Receptor Tyrosine Kinases
TKI	Tyrosine Kinase Inhibitor
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
STAT	Signal transducer and activator of transcription
sdf	Structure data file
GUI	Graphical User Interface
NIH	National Institutes of Health
PMC	PubMed Central
NLM	National Library of Medicine
NCBI	National Centre for Biotechnology Information
G score	Glide score
HIA	Human Intestinal Absorption
PMID	PubMed ID
HTVS	High throughput virtual screening
SP	Standard precision
XP	Extra precision
BBB	Blood Brain Barrier
SIB	Swiss Institute of Bioinformatics

Abstract

Background: The human Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein that has been found to cause a variety of cancers, especially lung cancer. SRC is a non-receptor tyrosine kinase whose expression correlates with cancer progression and metastatic disease by facilitating the actions of other signalling proteins like EGFR. It has been studied that interactions between EGFR and SRC may contribute to enhanced cancer development and resistance to already present inhibitors. So, there is a need to identify those molecules which can act as dual inhibitors for both EGFR and SRC protein kinases. In this study, molecular docking approach was used to identify such inhibitors.

Objective: To identify dual inhibitors for EGFR and SRC proteins using Molecular Docking approach.

Method and Results: Literature survey was performed in order to identify compounds that have the capacity to inhibit EGFR based on their IC_{50} value. 1433 compounds were identified from 121 research papers. The structures of these compounds were created using Marvin Sketch and PubChem and a library of these compounds was generated. Following that, we used the PDB database to search for 3D structure of the target protein. The SRC protein (PDB: 1Y57) was selected as the target protein. We then docked our target protein with the ligand library generated in the current study using Schrödinger's Glide in order to identify potential ligands with the highest docking score and affinity for the SRC protein. The docking process was performed in 3 modes: HTVS, SP and XP. After molecular docking step, 18 compounds were identified which had higher docking score than the reference cocrystal. ADMET study was then performed to determine the pharmacokinetic characteristics of these ligands. Finally, 3 compounds were selected which had highest docking scores and showed 0 violations to Lipinski's rule.

Conclusion: The current study has identified several compounds with higher docking score than the reference cocrystal as well as with 0 violations to Lipinski's Rule of Five. Among these the top 3 compounds with the highest docking scores and best ADMET properties have been discussed in detail and these can also be taken forward for future research as potential dual inhibitors of EGFR and SRC protein kinases.

Keywords: Lung cancer; Epidermal Growth Factor Receptor; epidermal growth factor; SRC protein; dual inhibitors; Molecular docking; ADMET

1. Introduction

Cancer is a disease which occurs when few of the body's cells grow out of control and spread to other body parts. The human body is comprised of trillions of cells, and cancer has the ability to appear in nearly any location within them. When cancer begins in one area and then spreads to another, it is known as metastatic cancer. This spread of cancer cells is referred to as metastasis. Cancer is a genetic disease that results from alterations to the genes that regulate the behavior of our cells, including growth, division, and function. The most frequently occurring types of cancer include prostate, breast, lung, bowel, and rectal tumors.

The field of cancer bioinformatics plays a crucial role in the practice of systems clinical medicine for cancer. It serves as a fundamental tool and approach in investigating cancer within this context. Cancer bioinformatics offers a specialized approach to applying bioinformatics methods to cancer research, focusing on the specific characteristics of disease metabolisms, signalling, communication, and proliferation. (Wu *et al*., 2012)

In 2020, lung cancer affected approximately 2.2 million individuals worldwide and caused 1.8 million deaths. It is a significant contributor to cancer-related fatalities among both males and females. Lung cancer comprises two primary types: Non-small cell lung cancer (NSCLC) and Small cell lung cancer (SCLC). NSCLC, which accounts for approximately 75% of lung cancers, has been challenging to treat due to poorly understood pathological mechanisms (Bethune *et al.*, 2010). NSCLCs generally consist of 3 types of cancer: squamous-cell carcinoma large-cell carcinoma and adenocarcinoma. Studies have demonstrated that the EGFR group of growth factors are a critical molecule that promote the development and spread of lung cancer.

EGFR belongs to a group of receptor tyrosine kinases known as the erbB family. This family comprises closely related members including erbB1, also known as EGFR, erbB2 (HER2), erbB3, and erbB4. In various pathological conditions, particularly in lung and breast cancer as

well as glioblastoma, EGFR functions as a driver of tumorigenesis. It has an extracellular ligand-binding domain, a transmembrane portion, and intracellular tyrosine kinase and regulatory domains. Upon binding of a specific ligand, a conformational change occurs in the normally functioning EGFR and phosphorylation of the intracellular domain occurs, leading to downstream signalling through various pathways.

Drugs or inhibitors that block EGFR proteins are being used in the treatment of several types of cancer. The common types of inhibitors are Reversible, Irreversible, Covalent, Non-covalent inhibitors (Bethune *et al*., 2010)

Small molecules known as tyrosine kinase inhibitors (TKIs) can either be reversible or irreversible. They work by binding to ATP binding pockets on the intracellular catalytic kinase domain of receptor tyrosine kinases (RTKs), which can prevent autophosphorylation and activation of various downstream signalling pathways (Seshacharyulu *et al.,* 2012). Gefitinib, erlotinib, lapatinib, and icotinib are considered first-generation EGFR inhibitors among the approved EGFR-TKIs, while afatinib, neratinib, and dacomitinib are classified as second-generation EGFR-TKIs. The third-generation EGFR inhibitors include almonertinib, olmutinib, and Osimertinib (Abourehab *et al.*, 2021). Alternative methods such as monoclonal antibodies targeting EGFR are specifically developed to target the extracellular portion of EGFR, resulting in a ligand competitive inhibition. This inhibits receptor dimerization, auto-phosphorylation, and downstream signalling.

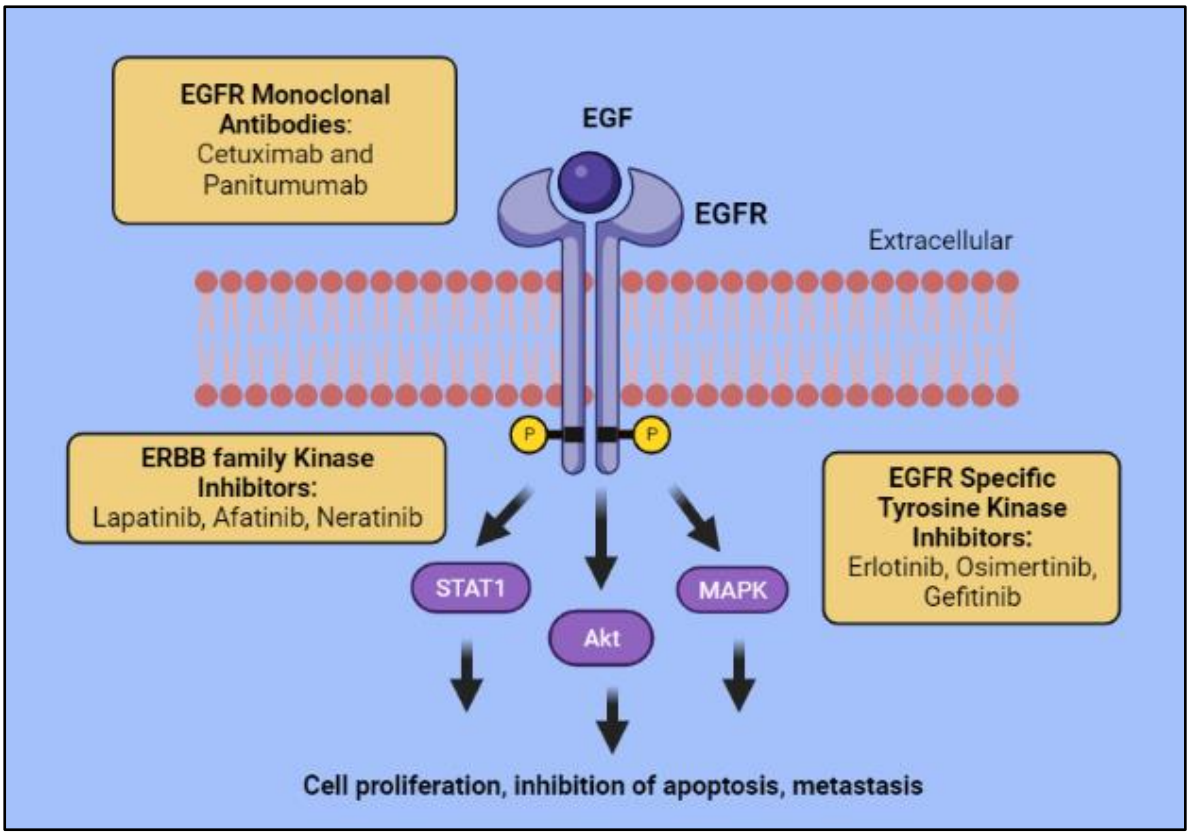


Fig 1: Diagram illustrating the EGFR activators, inhibitors, and outcomes of EGFR signalling

SRC family kinases (SFKs) are 60 kDa non-receptor tyrosine kinases, involved in the control of vital physiological processes like cell division, differentiation, apoptosis, migration, and metabolism. Structurally, c-SRC mainly consists of seven parts from N-terminus to C-terminus: SRC homology 4 (SH4) domain, unique domain, SH3 domain, SH2 domain, SH2 kinase binding domain, SH1 domain and C-terminal negative regulatory region.

The protein c-SRC has been found to be excessively expressed and significantly activated in many types of human cancers. The correlation between the activation of SRC and the advancement of cancer is noteworthy. c-SRC acts as a co-transducer of transmembrane signals produced from many polypeptide growth factor receptors, including the EGFR. Targeting SRC is being seen as a promising therapeutic approach for cancer treatment **(**Biscardi, J. S. *et al*., 1999)

Research has shown that the relationship between the epidermal growth factor receptor (EGFR) and the nonreceptor tyrosine kinase c-SRC could potentially play a role in the development of an aggressive phenotype in various types of human tumors. This increased tumorigenesis correlates with EGF-induced physical binding of the EGF receptor to c-SRC, tyrosine phosphorylation of the EGF receptor in two new sites, and increased tyrosine phosphorylation of two EGF receptor substrates: SRC homology 2-containing adapter protein (Shc) and phospholipase Cγ (Kloth *et al*., 1999**).**

A study has identified these c-SRC-dependent phosphorylation sites on the EGFR as Tyr845 and Tyr1101 It was found that cells expressing a Y845F variant of the EGFR were unable to synthesize DNA in response to EGF treatment, thus proving the importance of this phosphorylation. It was further shown that STAT5b is tyrosine phosphorylated in response to EGF in EGFR-overexpressing cells, and this tyrosine phosphorylation is partially dependent on c-SRC kinase activity and the presence of c-SRC-mediated phosphorylation at Tyr845 on EGFR. Thus, STAT5b activation is a necessary signal in the pathway resulting from EGFR and c-SRC overexpression, association and c-SRC mediated Tyr845 phosphorylation **(**Kloth *et al*., 1999**).**

Apart from these, it was also seen that overexpression of EGFR plus c-SRC induced the down-regulation of E-cadherin. Loss of E-cadherin mediated adhesion activity is associated with progression of non-invasive tumor to an invasive stage.

Biological interaction between overexpressed EGFR and c-SRC was further evident from the ability of doubly transduced MECs to undergo anchorage independent growth, a trait which is often used to measure the cellular oncogenicity in vitro. A clear increase in soft agar colony formation was also observed as result of EGFR and c-SRC co-overexpression **(**Dimri *et al*., 2007**)**.

Molecular docking is a computer modelling technique used to predict the structure of a receptor and ligand complex. The ligand is usually a small molecule or other protein, and the receptor is often a protein or a nucleic acid molecule such as DNA or RNA. There are two main methods for virtual screening (VS) - ligand-based and structure-based. Molecular docking is the most commonly used method in structure based virtual screening. Molecular docking can be broadly defined as the process by which a ligand is coupled to a target to form a stable complex.

The process of docking involves two main steps: figuring out where and how the ligand fits into the binding sites, and then determining how strong the bond will be. There are different types of docking, but protein-ligand docking is especially important in the pharmaceutical industry. (Meng *et al*., 2011).

The ADMET study is a crucial evaluation of a drug's pharmacokinetics. This study assesses the drug's absorption, distribution, metabolism, excretion, and toxicity. It helps to determine how much of a drug is absorbed when taken orally and how much of the drug is absorbed through the gastrointestinal tract. Poor absorption can lead to issues with distribution and metabolism, potentially causing neurotoxicity and nephrotoxicity (Flores-Holguín *et al.*, 2021). Several studies have shown that drugs have failed in clinical trials due to adverse side effects and toxicities caused by the drugs, which have proven extremely detrimental and costly in drug development. Since in vivo and in vitro evaluations are costly and time consuming, in silico methods are being widely used to estimate these properties (Srivastava *et al*., 2022). Therefore, ADMET study is a vital component of computational drug design

2. Review of Literature:

2.1. Cancer:

Cancer is a disease in which certain body cells grow out of control and spread to other parts of the body. The human body is comprised of trillions of cells, and cancer has the ability to appear in nearly any location within them. Normal cells multiply when the body needs them and die when the body does not need them. Cancer occurs when cell growth in the body is out of control and cells divide too quickly, or when the cells forget to die (Mathur *et al.,* 2015). These cells can form tumors, which are lumps of tissue.

Tumors can be benign or malignant; Benign tumors can usually be removed and will not spread to other body parts. While malignant tumors grow aggressively and invade other body tissues, allowing tumor cells to enter the bloodstream or the lymphatic system, which spreads the tumor to other places in the body This process of cancer spread is called metastasis and the areas of tumor growth at these distant places in body are called metastases (Pandi *et al*. 2016).

Causes of Cancer:

Cancer is a medical condition that arises from genetic or epigenetic modifications in somatic cells. Surprisingly, only a small percentage of cancer cases, about 5-10%, are attributed to genetic defects. On the other hand, the majority of cancer cases, approximately 90-95%, are linked to environmental and lifestyle factors. These factors may include habits such as smoking cigarettes, consuming fried foods and red meat, drinking alcohol, being exposed to the sun, and being exposed to environmental pollutants. In addition, other factors like infections, stress, being overweight, and leading an inactive lifestyle are also associated with the development of cancer.

In order to reduce the risk of developing cancer, it is recommended to abstain from smoking, increase the consumption of fruit and vegetables, reduce alcohol consumption, limit caloric

intake, exercise, abstain from direct sunlight exposure, reduce meat consumption, consume whole grains, receive vaccinations, and undergo regular medical examinations. Numerous studies have demonstrated that a diet that is rich in fruit and vegetables, spices and grains can have a positive impact on the prevention from cancer (Anand *et al*., 2008)

Types of Cancer:

Cancer can be mainly distributed into these 5 types:

Carcinomas: It begins in the tissues or the skin, covering the glands and the surface of internal organs. It develops into a solid tumor.

Sarcomas: The formation of sarcomas begin in the tissues that provide structural support to the body. These tissues may include nerves, tendons and joints, as well as fat and blood vessels. Sarcomas may also form in bone, lymph vessels, muscles or cartilage.

Leukemias: Leukemia is a type of cancer of the blood that originates from the uncontrolled growth and proliferation of healthy blood cells. It is classified into four distinct subtypes: acute myeloid, acute lymphocytic, chronic lymphocytic, and chronic myeloid leukemia.

Lymphomas: Lymphoma is a type of cancer that starts in the lymphoid system. The lymphoid system is a complex network of lymph nodes and blood vessels that helps fight infection. Lymphomas include Hodgkin and NonHodgkin lymphoma.

Melanomas: Melanoma is a type of cancer that begins in cells called melanocytes. Melanocytes are specialized cells that produce melanin, the pigment that colors the skin. Most melanomas start on the skin but can also form in other pigmented tissues such as the eye.

Apart from these, there are more than 100 kinds of cancers that affect humans but the most common among all are: bladder, breast, thyroid, colon cancer, lung cancer, kidney (renal cell and renal pelvis) cancer, prostate cancer and pancreatic cancer (Malik *et al*., 2020)

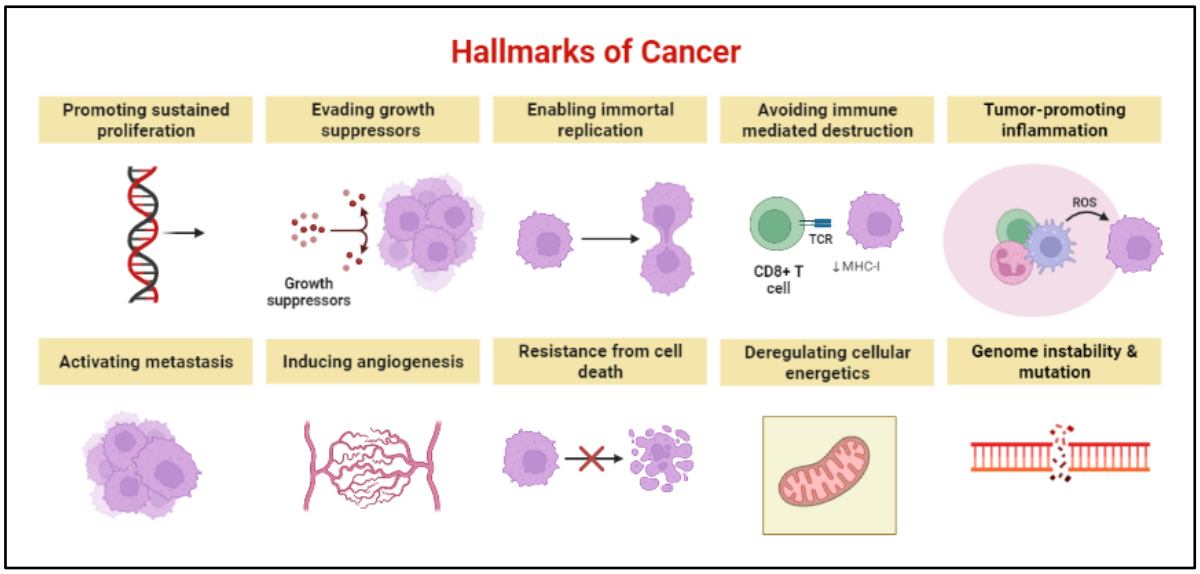


Fig 2: Hallmarks of Cancer

Cancer bioinformatics:

The field of cancer bioinformatics plays a crucial role in the practice of systems clinical medicine for cancer. It serves as a fundamental tool and approach in investigating cancer within this context. Cancer bioinformatics offers a specialized approach to applying bioinformatics methods to cancer research, focusing on the specific characteristics of disease metabolisms, signalling, communication, and proliferation. (Wu *et al*., 2012)

2.2. Lung Cancer:

Lung cancer, also known as lung carcinoma, is a malignant tumor of the lung, which is characterized by the uncontrolled cell growth in lung tissues. It is a significant contributor to cancer-related fatalities among both males and females. If this growth is not treated, it can spread outside the lungs by the process of metastasis to nearby tissues or other parts of the body (Pandi et al. 2016).

Lung cancer comprises two primary types: Non-small cell lung cancer (NSCLC) and Small cell lung cancer (SCLC). Non-small cell lung cancer, which accounts for the majority (approximately 75%) of lung cancers, has proven difficult to treat due to poorly understood

pathological mechanisms. NSCLCs generally consist of 3 types of cancer: squamous-cell carcinoma large-cell carcinoma and adenocarcinoma.

Squamous cell carcinoma: The development of squamous cell carcinoma of the lung typically begins in the middle of the bronchus. Process like necrosis and cavitation within the center of cancer is a frequent occurrence. Generally, squamous cell lung cancers that are well-differentiated tend to develop at a slower rate than other cancer types.

Adenocarcinoma: A significant proportion of lung cancers, 29.4%, are classified as Adenocarcinoma. This type of cancer typically originates in the peripheral lung tissue. Smoking is the primary contributor to the development of this type of lung cancer. However, it is also the most common form of lung cancer among individuals who have never smoked.

Large cell lung carcinoma: Large cell lung carcinoma accounts for 10.7%, of all lung cancers. This type of cancer is characterized by its rapid growth, which occurs near the surface of the lung. It is characterized by its lack of differentiation and its tendency to metastasize rapidly (Pandi *et al*. 2016).

Studies have demonstrated that the EGFR group of growth factors are a critical molecule that promote the development and spread of lung cancer. Mutations in EGFR have been discovered in relation with some lung cancers. Some studies have also shown that EGFR expression in NSCLC is associated with reduced survival, recurrent lymph node metastasis, and poor chemosensitivity.

2.3. EGFR and its role in Cancer:

Receptor Tyrosine Kinases (RTKs) are a class of transmembrane proteins that play a major role in cell growth and cell survival. RTKs have also been shown to play a role in the induction and progression of various types of malignancies in humans. The most well-characterized

RTKs are EGFR (Epidermal Growth Factor Receptor) and ERBB-2 (Her-2). The physiological function of EGFR is to regulate the development and homeostasis of epithelial tissue.

EGFR overexpression has been associated with the development or progression of a variety of malignancies, including head and neck cancer, lung cancer, pancreas cancer, bladder cancer and breast cancer. Specifically, high metastatic rate, short survival time and poor prognosis due to EGFR overexpression have been observed in patients with Squamous Cell Carcinoma of the Lung (Rusnak *et al*., 2010)

EGFR Structure:

The Epidermal growth factor receptor is a transmembrane glycoprotein with an extracellular epidermal growth factor-binding domain, a transmembrane portion and an intracellular tyrosine kinase domain that regulates signalling pathways to control cell proliferation. Upon binding of a specific ligand (eg, epidermal growth factor), the normally functioning EGFR undergoes a conformational change and phosphorylation of the intracellular domain, leading to downstream signal transduction through various pathways. Raf1-extracellular signal-regulated kinase, PI3K/Akt, and Signal transducer and activator of transcription (STAT) factors are some of these pathways (Bethune *et al*., 2010).

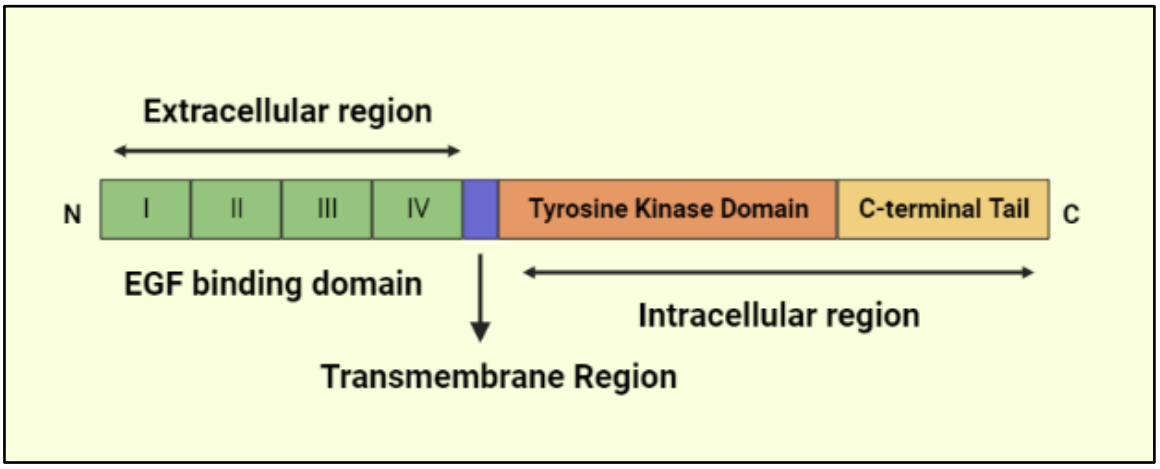


Fig 3: EGFR Structure

EGFR Signaling:

The EGFR mediated signal transduction pathway is highly intricate. It commences with the activation of the EGFR by a ligand, which results in receptor dimerization. Subsequently, the C- terminal tail is transphosphorylated, followed by the transmission of the signal through a series of complex signalling pathways to stimulate the expression of novel genes. EGFR is responsible for the largest number of distinct signalling pathways out of the four family members.

These pathways include the ERK MAPK, PI3K-AKT, SRC, PLC-γ1-PKC, JNK, and JAK-STAT pathways. As these pathways are interconnected, activating the EGFR results in the activation of an entire signalling network that is associated with a broad range of outcomes, including cell proliferation and growth, differentiation and migration, and the inhibition of apoptosis (Wee *et al*., 2017).

EGFR inhibitors:

Given EGFR's functional role in a variety of cellular activities, numerous techniques that target and interfere with EGFR-mediated effects have been developed. Currently, monoclonal antibodies and small molecule tyrosine kinase inhibitors are utilised as two distinct therapeutic approaches for targeting EGFR in a variety of human cancers. Each of these strategies employs a unique mechanism of action. Recent studies also indicate the use of various chemopreventive agents for EGFR gene-level downregulation. The monoclonal antibodies against EGFR were specifically designed to be directed against the extracellular region of EGFR which creates a ligand competitive inhibition, thus preventing receptor dimerization, auto-phosphorylation and downstream signalling. Apart from inhibiting EGFR signalling, these monoclonal antibodies will induce receptor internalization, ubiqutinization, degradation and prolonged

downregulation. Some of the common anti-EGFR monoclonal antibodies are Cetuximab and Panitumumab.

Small molecules known as tyrosine kinase inhibitors (TKIs) can either be reversible or irreversible. They work by binding to ATP binding pockets on the intracellular catalytic kinase domain of receptor tyrosine kinases (RTKs), which can prevent autophosphorylation and activation of various downstream signalling pathways (Seshacharyulu *et al.,* 2012).

First-generation EGFR inhibitors include Gefitinib, Erlotinib, Lapatinib, and icotinib among the approved EGFR-TKIs. This class of drugs bind reversibly to the PTK domain of the EGFR, which inhibits the binding of ATP to the EGFR, thereby inhibiting EGFR activation and cellular proliferation. Afatinib, neratinib, and dacomitinib, on the other hand, are classified as second-generation EGFR-TKIs. These three inhibitors contain a Michael acceptor site, allowing them to bind covalently to the EGFR and inhibit its kinase activity irreversibly. Almonertinib, olmutinib, and osimertinib are examples of third-generation EGFR inhibitors. They can form a covalent bond with the cysteine residue in the EGFR. On the other hand, Brigatinib, vandetanib, and pyrotinib are classified as multi-kinase inhibitors due to their inhibitory activities against kinases other than the EGFR (Abourehab *et al*., 2021)**.**

2.4. SRC and its role in Cancer:

The proto-oncogene c-SRC (SRC) is a non-receptor tyrosine kinase whose expression and activity correlates with advanced malignancy and poor prognosis in several human cancers. SRC family kinases (SFKs) refer to the nine enzymes with SRC homology. SFKs are the largest nonreceptor-tyrosine kinase family in the human body that interact directly with a variety of other tyrosine-tyrosine-kinases, including receptor tyrosinases, G-protein coupled receptors (GPCRs), steroid receptors, signal transducers and transcription activators, as well as molecules involved in cellular adhesion and translocation. These interactions result in a broad

range of biological functions, including proliferation and growth, differentiation and remodelling of cells, as well as motility and migration and angiogenesis and survival (Wheeler *et al.,* 2009)

SRC Protein Structure:

SFKs are structurally related and contain some structural elements that are conserved among family members. These elements include the N-terminal SRC Homology 4 domain (SH4), the SRC Homology 3 domain (SH3), the SRC Homology 2 domain (SH2), the linker sequence, the tyrosine kinase domain, and the C-terminal tail. The N-terminal domain of SH4 functions as a myristoylation site, thereby targeting SFK to the cytoplasmic membrane. The SH3 domain binds amino acid sequences rich in proline residues. This domain is essential for SRC activity, intracellular localization, substrate recruitment, and substrate binding. The SH2 domain binds to short motifs containing phosphotyrosines. The SH2 and SH3 domains work together to regulate the catalytic activity of SFKs.

Increased SRC protein levels and its tyrosine kinase activity have been linked with cancer progression and metastasis by facilitating the actions of other signalling proteins such as EGFR. This accumulating evidence suggests that SFKs may be a promising therapeutic target for the treatment of solid tumors (Wheeler *et al.,* 2009). Given the crucial role of SRC in tumor development and extensive pre-clinical evidence of metastasis suppression by SRC inhibition, several clinically applicable small molecule SRC inhibitors have been developed and are undergoing clinical testing, particularly for metastatic diseases (Zhang *et al*., 2012).

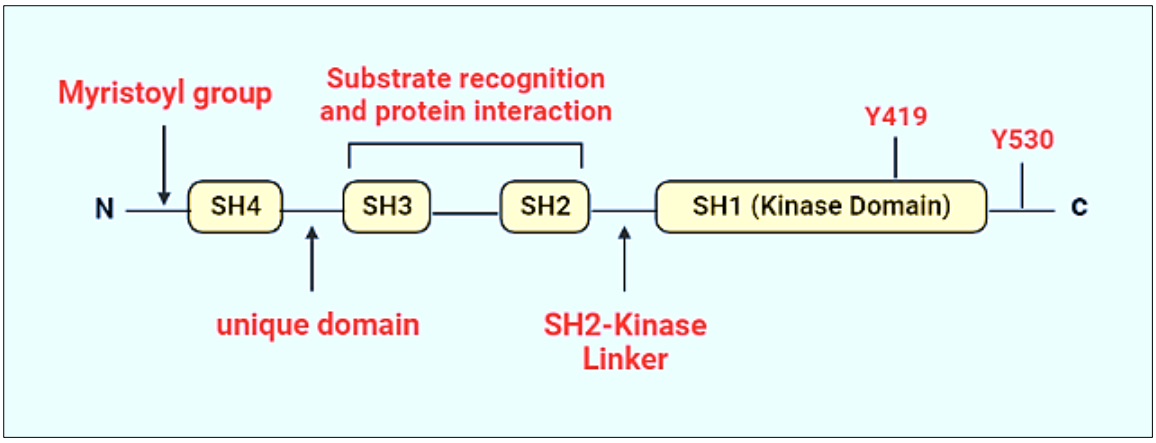


Fig 4: SRC protein structure

SRC Inhibitors:

Dasatinib (BMS-354825, Sprycel) is an oral small molecule SRC inhibitor that was originally approved by the FDA for potent inhibition of the BCR/ABL fusion protein and is currently used to treat chronic myeloid leukemia. Dasatinib has shown low inhibitory specificity and its potential targets include other SFK like Lck, Fyn and Yes. Saracatinib (AZD0530) and bosutinib are two other SFK inhibitors undergoing or having undergone clinical trials.

Saracatinib (AZD0530) is a bioavailable anilinequinazoline that is extremely selective for non-receptor tyrosine kinases, including SRC (IC50 = 2.7 nM), Yes (IC50 = 4 nM), and Lck (IC50 4 nM). Bosutinib (SKI-606) was generally well tolerated and its safety profile differed from that of dasatinib in a comparable patient population, with gastrointestinal side effects being the most common.

In addition, bosutinib had a more restricted inhibitory effect against SRC/Abl compared to dasatinib (Martellucci *et al*., 2020). However, the therapeutic efficacies of SRC inhibitors as a single agent in the treatment of different types of solid tumours have not been particularly encouraging.

Role of EGFR and SRC co-overexpression in tumor formation:

The mechanism of action of c-SRC is to potentiate the activation of EGFR (Epidermal Growth Factor Receptor) and to amplify its downstream oncogenic signals. It has been demonstrated that c-SRC, in conjunction with EGF, increases the pro-mitogenic signals upon EGF stimuli. c-SRC, together with activated EGF, are involved in the induction of cell transformation and the development of cancer. Furthermore, it has been observed that c-SRC binds to EGFR and phosphorylates the tyrosine residues in the C-terminal region of EGF, leading to a range of downstream effects (Belli, S *et al*., 2020**)**. In breast cancer cells expressing c-SRC and EGFR, control of proliferation and invasion of one of the two oncogenes is observed, while in other cancer cell types a synergistic interaction between them is necessary for tumor progression.

Model cell lines have helped to characterize the role of EGFR and c-SRC co-expression and association in the process of tumor formation. In a murine fibroblast cell model (C3H10T1/2), overexpression of both the EGF receptor and c-SRC results in a synergistic increase in mitogenesis, anchorage-independent growth, and tumorigenesis compared to EGFR or c-SRC overexpression in nude mice separately.

In a study, clones of the mouse fibroblast cell line C3H1OT1/2 overexpressing pp6Oc-SRC were generated. c-SRC encodes a 60-kilodalton phoshoprotein (pp6Oc-SRC) (Luttrell *et al*., 1998). Based on this study, it was found that c-SRC protein is involved in the cellular response to mitogenic stimulants such as epidermal growth factor (EGF). This increase in tumor progression correlates with EGF-induced physical binding of the EGF receptor to c-SRC, tyrosine phosphorylation of the EGF receptor in two new sites, and increased tyrosine phosphorylation of two EGF receptor substrates: SRC homology 2-containing adapter protein (Shc) and phospholipase Cγ. A study has identified these c-SRC dependent phosphorylation sites on the EGFR as Tyr845 and Tyr1101 of EGFR, which is itself not an autophosphorylation site, is among the known autophosphorylation sites in the C-terminal tail of EGFR. Tyr845 is

located in the activation loop of the catalytic domain (Kloth *et al*., 1999). Several studies have supported the notion of direct phosphorylation of the EGFR receptor by c-SRC. It has been seen that Tyr845 is homologous to Tyr416, which is autophosphorylation site in c-SRC.

It was found that cells expressing a Y845F variant of the EGFR were unable to synthesize DNA in response to EGF treatment, thus proving the importance of this phosphorylation. This finding also suggests that endogenous levels of c-SRC are able to mediate Tyr845 phosphorylation and that the Y845F form of the receptor acts in a dominant negative manner (Biscardi, J. S. *et al*., 1999). It was further shown that STAT5b is tyrosine phosphorylated in response to EGF in EGFR-overexpressing cells, and this tyrosine phosphorylation is partially dependent on c-SRC kinase activity and the presence of c-SRC-mediated phosphorylation at Tyr845 on EGFR. Thus, STAT5b activation is a necessary signal in the pathway resulting from EGFR and c-SRC overexpression, association and c-SRC mediated Tyr845 phosphorylation. It has been suggested that phosphorylation of tyrosine 845 can act directly as a docking site for STAT5b or it may act more indirectly by promoting phosphorylation at a different site of EGFR or even other signalling molecules in the complex.

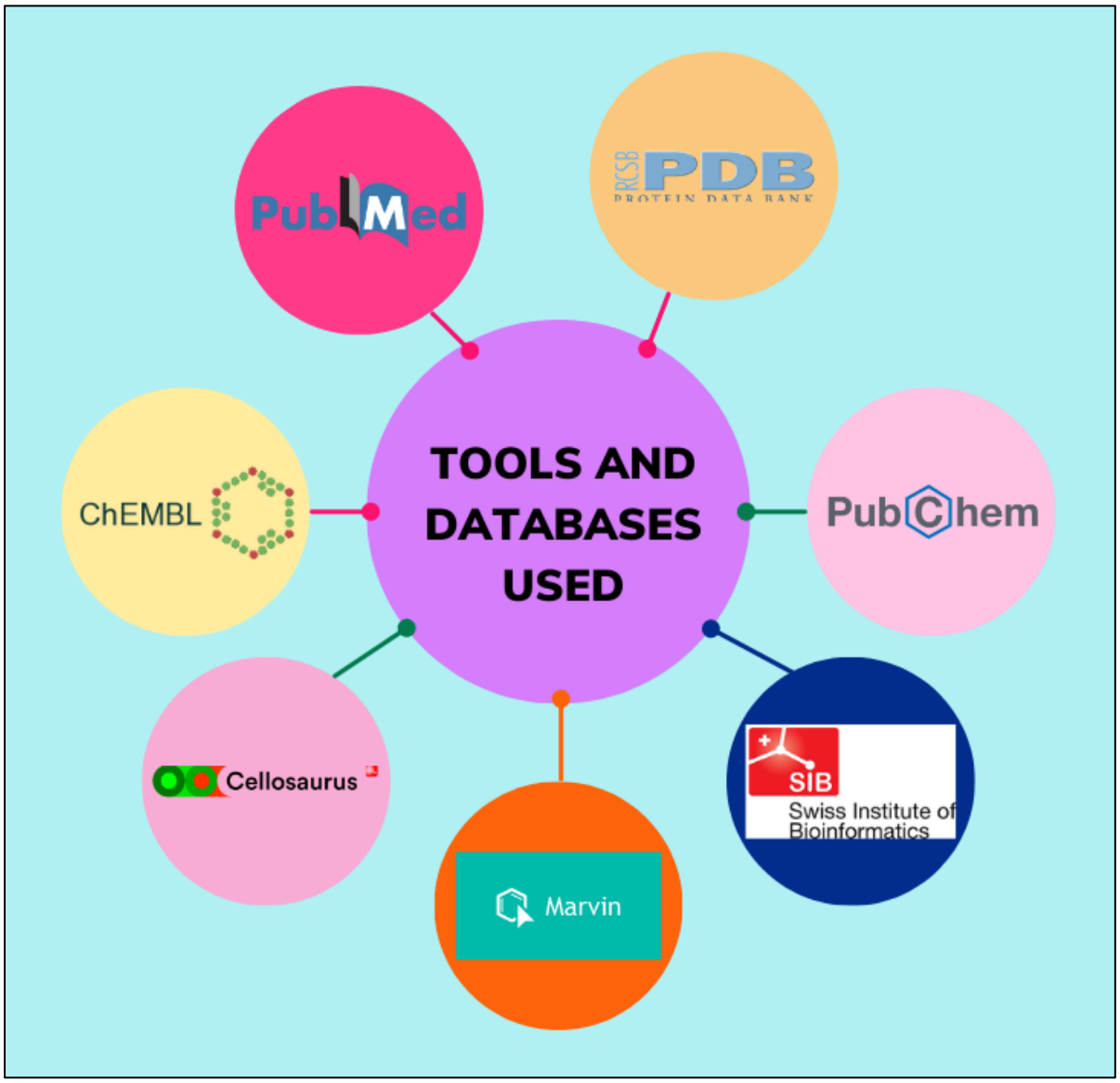
These studies provide evidence of the Cooperation between c-SRC and the EGF receptor in formation of tumors. This relationship between EGFR and c-SRC has been described in various human breast cancer cell lines. A Panel of 14 human breast tumor cell lines were characterized by their presence or absence of estrogen receptor, EGFR and c-SRC levels as well as the presence or absence of EGFR and c-SRC complexes. Five out of these 14 cell lines were found to overexpress both EGFR and c-SRC, and such overexpression highlights the relationship between these two proteins and phosphorylation of EGFR at the same two new sites that were found in the mouse fibroblast system (Kloth *et al*., 1999)

Apart from these, it was also seen that overexpression of EGFR plus c-SRC induced the down-regulation of E-cadherin. E-cadherin is the main adhesion protein at the adherens junction

expressed by epithelial cells and is known to play an important role in maintaining cell polarity. Loss of E-cadherin mediated adhesion activity is associated with progression of non-invasive tumor to an invasive stage. Analyses in epithelial cells show that SRC activation can enhance the down-regulation of E-cadherin, often in relation with EGFR stimuli, through ubiquitin-dependent endocytosis and lysosomal degradation**.**

In a subsequent study, co-overexpression of EGFR and c-SRC were simulated in well-characterized human MECs in order to investigate the possibility of c-SRC modulating EGFR-mediated MEC oncogenesis. Human MECs that are immortal but non-tumorigenic were genetically modified to overexpress either EGFR alone or in combination with c-SRC. This study was conducted to investigate the biological implications of such manipulations in 2D and 3D culture systems. C-SRC overexpression has been demonstrated to significantly enhance the capacity of overexpressing EGFR to produce a phenotype characterized by branched acini and irregular acini formation in three-dimensional Matrigel cultures, as well as to improve the invasion and anchorage independent growth in soft agar. Thus, establishing that c-SRC can modify EGFR-mediated mammary oncogenesis (Dimri *et al*., 2007).

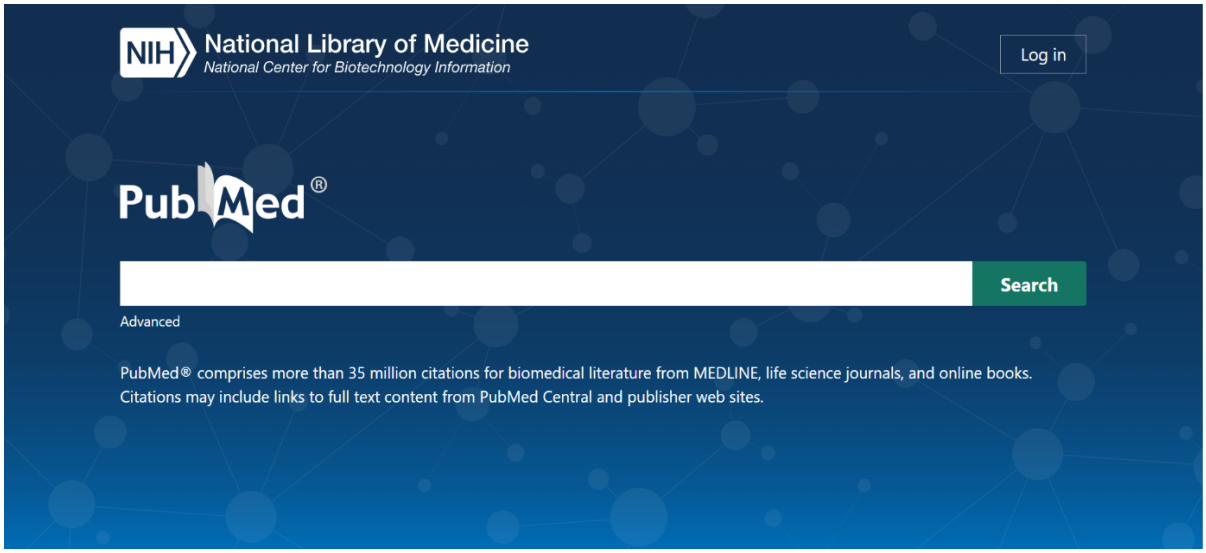
Tools and databases used:



Pubmed:

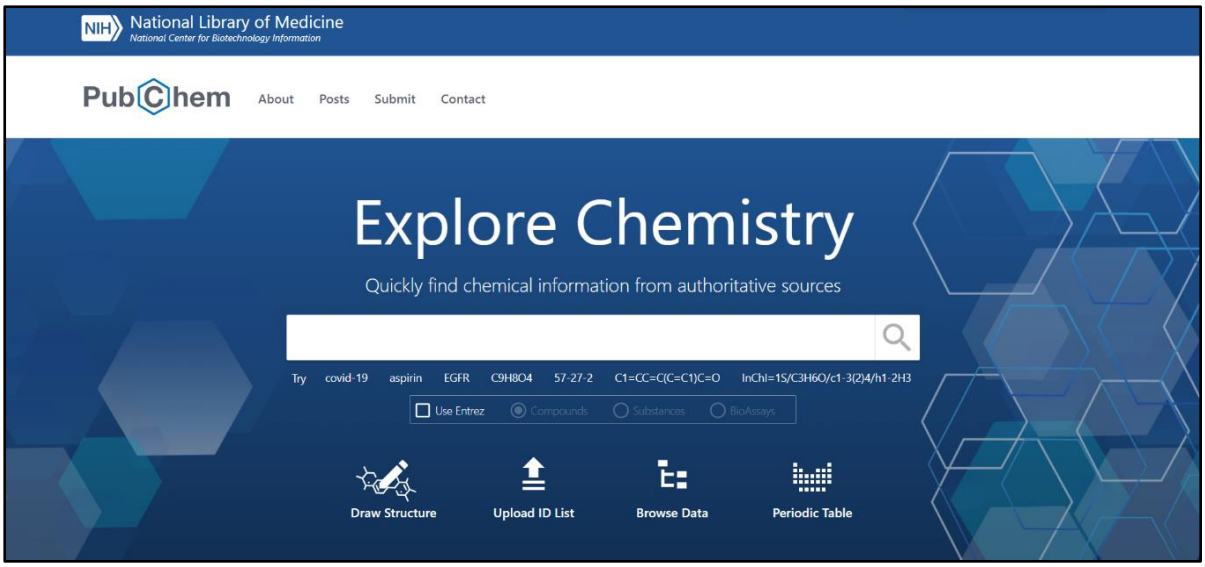
PubMed (https://pubmed.ncbi.nlm.nih.gov/) is an online database that provides a free service to facilitate the search for and retrieval of literature in the biomedical and life sciences fields, with the purpose of promoting health, both globally and personally. It contains over 35 million citations of biomedical literature, as well as abstracts of articles. Although the database does not contain full text journal articles, links to the article are often available when accessed from other sources, including the publisher's website, or from the National Library of Medicine's

(NLM) database, known as PubMed Central. Established in 1996, the database is hosted by the NCBI, a branch of the NIH, and is accessible to the public online.



Pubchem:

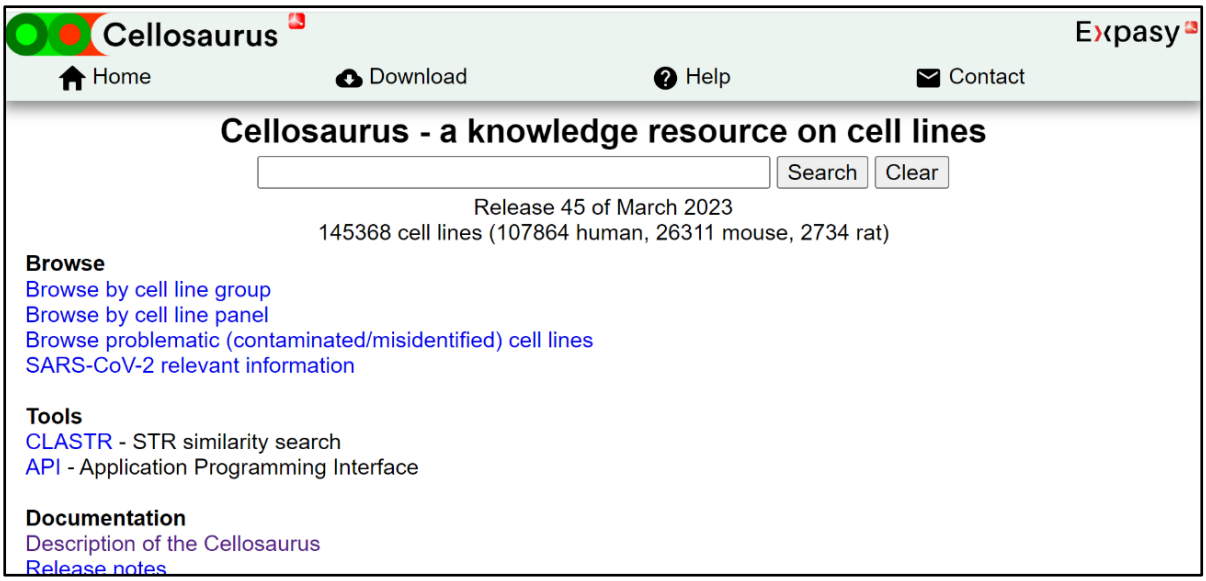
PubChem (https://pubchem.ncbi.nlm.nih.gov/) is a database of open-access chemical information hosted at the National Institutes of Health (NIH). Established in 2004, it has become a fundamental source of chemical information for researchers, academicians, and the public at large. It is composed primarily of small molecules, as well as larger molecules, such as nucleotide, carbohydrate, lipid, peptide, and chemically modified macromolecule. Among other things, PubChem collects data on chemical structure, identification, chemical and physicochemical characteristics, biological activity, patents, safety, and toxicity.



Cellosaurus:

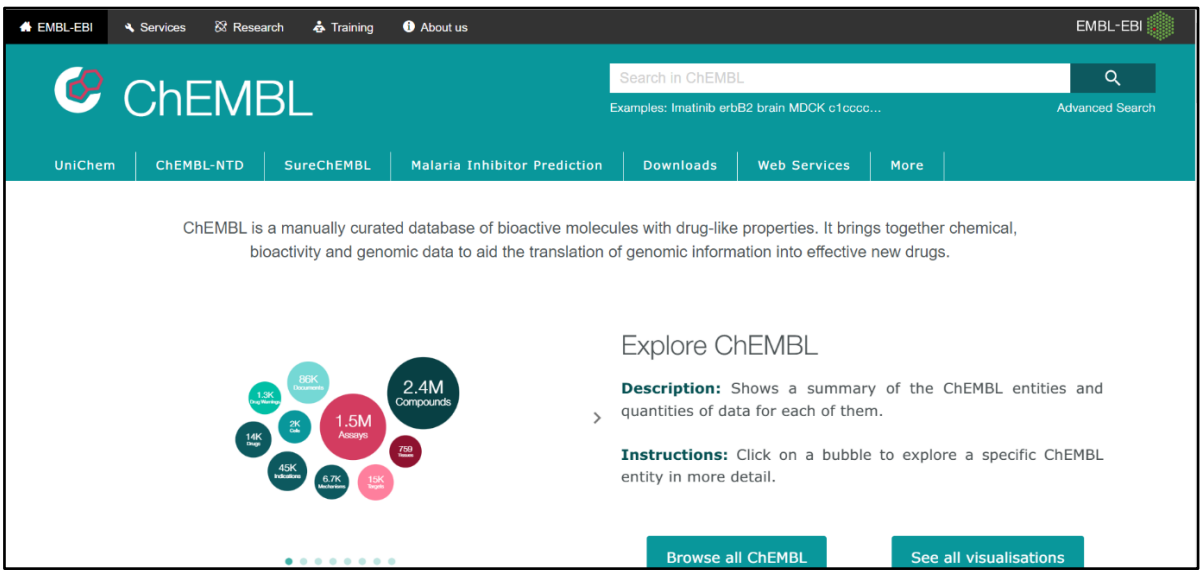
Cellosaurus (https://www.cellosaurus.org/search) is an online knowledge resource that stores information about cell lines. It aims to provide a comprehensive description of all cell lines utilized in biomedical research. The Swiss Institute of Bioinformatics (SIB) provides it. It contains information for over 144,000 cell lines as of December 2022. For each cell it provides the following information:

- A recommended name
- A list of synonyms
- A unique accession number
- The species of origin
- Structured comments
- The category a cell line belongs to
- Publication references



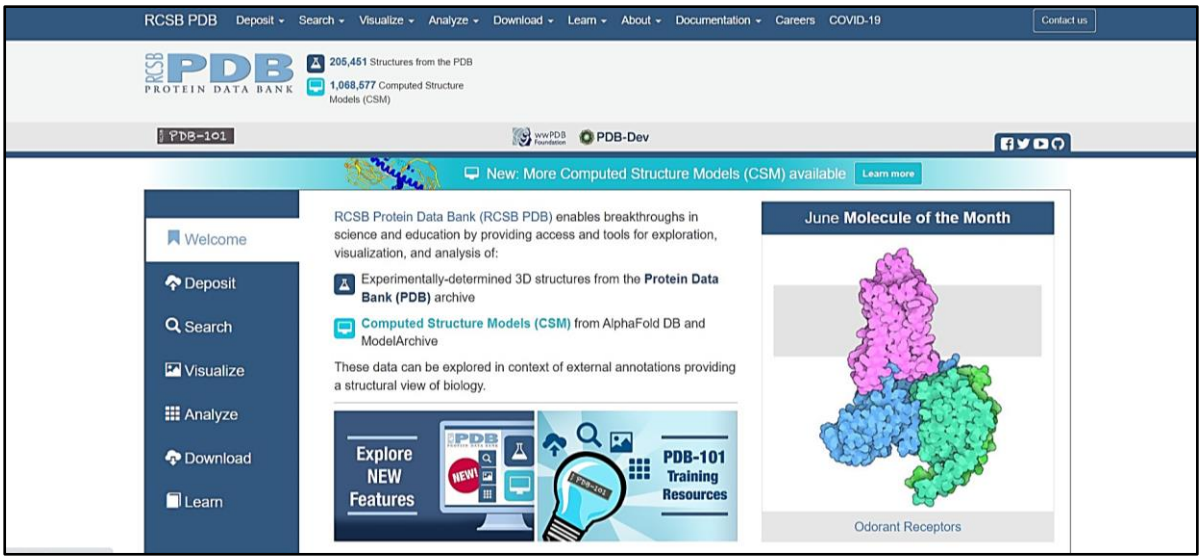
ChEMBL:

The ChEMBL database is a huge, open-source database that collects data on small molecules and how they work. This data is extracted from the full-text articles of the main Medicinal chemistry journals and links to approved drug and clinical development candidate data, like how the molecule works and what it's used for. It also links to Pubchem BioAssay and BindingDB databases, so users can get even more info (Mendez *et al*., 2019)



Protein Data Bank:

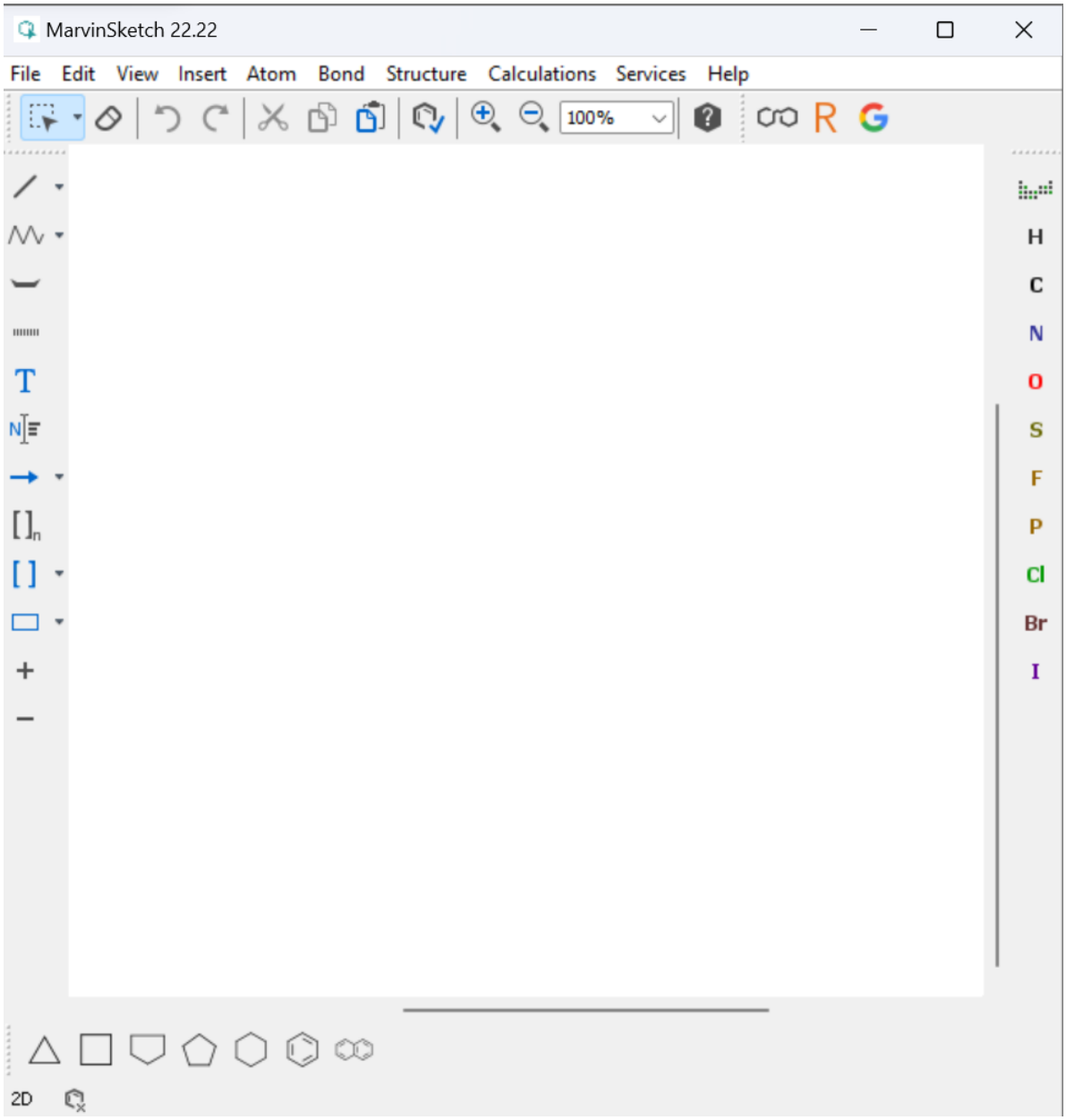
The Protein Data Bank (PDB) is a database containing the three-dimensional structural information of large biological entities, such as proteins and nucleic acids. The Protein Data Bank (PDB) library of 3D structure data for large biological molecules (proteins, DNA, and RNA) is a vital resource for study and education in fundamental biology, health, energy, and biotechnology.



The data, which was contributed by scientists and biochemistry researchers from all over the world and frequently gathered by X-ray crystallography, NMR spectroscopy, or, increasingly, cryo-electron microscopy, is available to the public at https://www.rcsb.org/. The Worldwide Protein Data Bank, or wwPDB, is the entity in charge of managing the PDB. The PDB's original file format is known simply as the PDB file format. PDB IDs are unique alphanumeric identifiers that are assigned to every structure that is added to the PDB database.

Marvin:

Marvin is a programme that allows you to sketch and visualize chemical structures and substructures. It includes Marvin Sketch and Marvin View. Marvin Sketch is a tool for drawing chemical structures. It supports checking valency, querying atoms and bonds, stereochemistry and user-defined models. Once a structure drawing is complete, it can also be exported to other databases. Marvin can handle molecule files in many formats including MDL mol, Compressed mol, unique SMILES, SMARTS, CML, XYZ etc. We can check the accuracy of our structure using Marvin's built-in structure checker. This function identifies and highlights any structural errors. Some special features include multi-level undo/redo, visual fragment placement, easy chain drawing, branching with one click. Marvin View is a 2D or 3D viewer that can display a single or several molecules, structures and compounds. This tool is perfect for users who don't need to perform hands-on chemical drawing, and want to display any number of structures simultaneously. Using Marvin View, we can also generate the chemical name of a structure that we have drawn (Csizmadia *et al*., 1999)



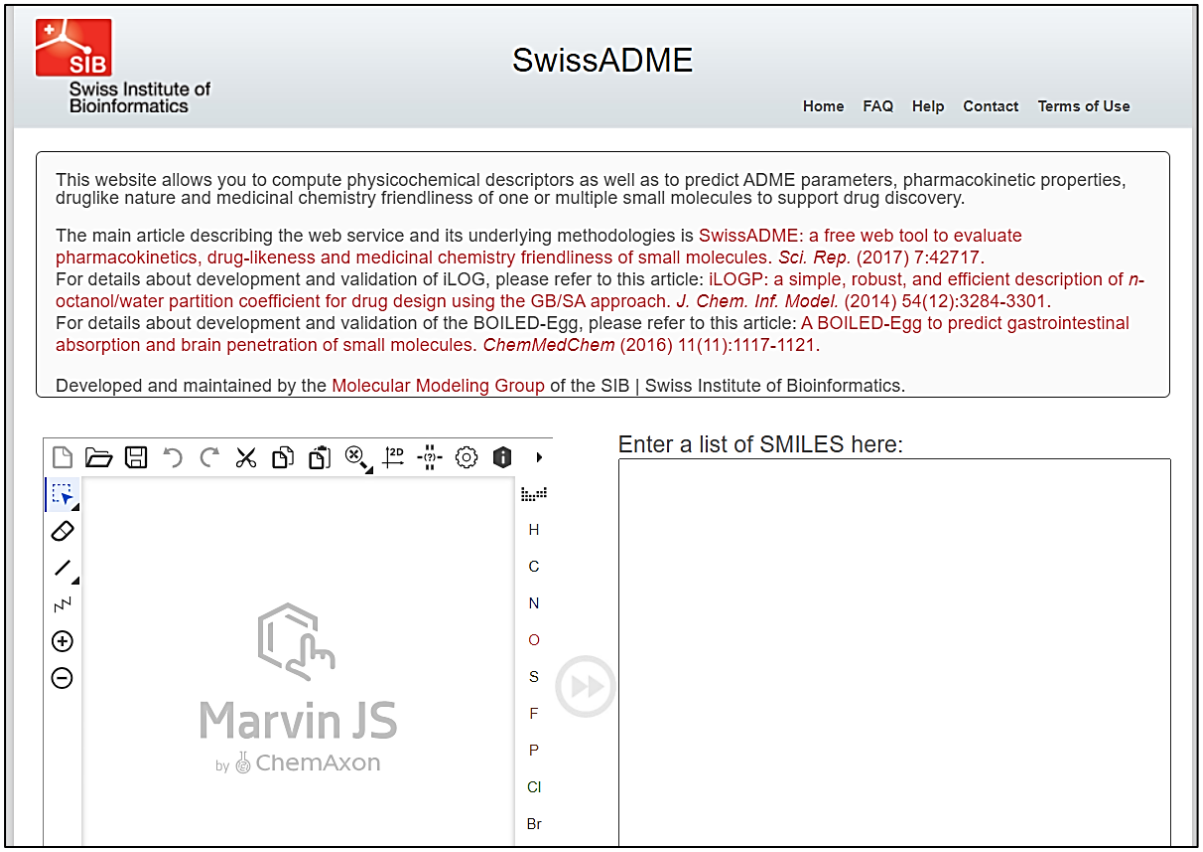
Maestro:

Maestro provides a graphical user interface for the creation, display, and alteration of chemical structures, the organization, loading, and storage of these structures and their associated data, and the creation, monitoring, and visualization of the results of calculations performed on those structures. All Schrodinger tools, including CombiGlide; Epik; Glide; Impact; Liaison;

Ligprep; MacroModel; Phase; Prime; QikProp; Qsite; and Strike, utilize Maestro as their graphical user interface.

Swiss ADME:

Swiss ADME is a free web tool that can be used to analyse pharmacokinetics and drug-likeness of small molecules, and how well they work with medicinal chemistry. It is developed and maintained by the Molecular Modelling Group of the SIB (Swiss Institute of Bioinformatics). To facilitate drug discovery, it can compute physicochemical descriptors and predict ADME parameters, pharmacokinetic properties, druglike nature, and medicinal chemistry affinity of a single or multiple small molecules. SMILES file format is generally accepted as an input format in Swiss ADME.



2.5. Molecular Docking:

Molecular docking has revolutionized the field of drug discovery, enabling researchers to identify promising drug candidates quickly and efficiently. Molecular docking has numerous uses and applications in drug discovery, such as structure–activity studies, lead optimisation, discovering potential leads by virtual screening, and providing binding hypotheses to facilitate predictions for mutagenesis studies. (Morris *et al*., 2008).

There are two main methods for virtual screening (VS) - ligand-based and structure-based. Ligand-based methods are used when we have a set of active ligand molecules but little to no structural information about the targets. In this case, we can use methods like pharmacophore modelling and quantitative structure activity relationship (QSAR) to help us. On the other hand, structure-based drug design is used when we have more information about the target's structure. Molecular docking is the most commonly used method for this, and it has been around since the early 1980s. So, depending on what information we have available, we can use either ligand-based or structure-based methods for virtual screening (Meng *et al*., 2011)

Ligand-based computational methods include two key components: an effective similarity measure and a robust scoring approach. Furthermore, the computational method must be able to evaluate many ligands with sufficient precision and speed. The scoring method employed in Ligand based screening should be capable of distinguishing active compounds from inactive compounds during the ranking stage (Hamza A. *et al*., 2012)

What is molecular docking

Molecular docking study is a way to model how a small molecule, also known as a ligand, interacts with a protein at a tiny, atomic level. This helps us to understand how ligands behave at the binding site of target proteins and to learn more about basic biochemical processes. There are different types of docking, but protein-ligand docking is especially important in the

pharmaceutical industry. This involves looking for the right shape of the ligand within the target protein when we already know what the protein structure looks like (Meng *et al*., 2011)

How docking is done:

The first docking methods were derived from the “lock-and-key” hypothesis proposed by Fischer. Fischer proposed that both ligands and receptors could be thought of as rigid bodies, and their affinity was directly related to a geometrical fit between their respective shapes. Subsequently, Koshland proposed the "induced-fit" theory, according to which ligands and receptors should be thought of as being flexible during docking. They proposed that each movement of a backbone affects several side chains, as opposed to relatively independent movements of a single side chain. As a result, the sampling process in a Fully Flexible Receptor/ligand Docking is of a greater order of magnitude than in a Flexible Docking with a Rigid Receptor. These flexible docking algorithms can predict not only a molecule's binding mode more precisely than a rigid body algorithm, but also the binding affinity of a molecule to other compounds (Pagadala *et al*., 2017)

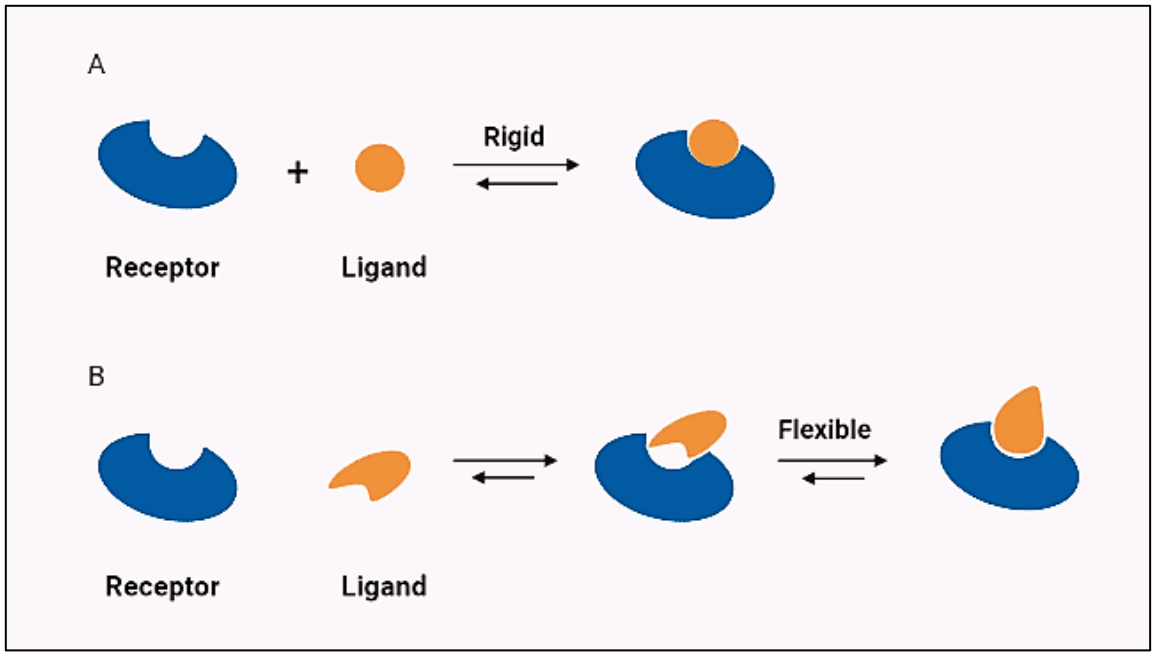


Fig5: Two models of molecular docking: (A) lock-and-key model (B) Induced fit model

The process of docking involves two main steps: figuring out where and how the ligand fits into the binding sites, and then determining how strong the bond will be. To determine if the ligand is active or not, a scoring system is used. This system predicts how well the ligand and protein will bond together. It's best to know the location of the binding site before docking, as this can improve the efficiency of the process. Sometimes, the binding site is already known before the ligands are tested. Other times, information about the sites can be found by comparing the target protein to similar proteins or to proteins that have already been tested with other ligands (Meng *et al*., 2011)

Algorithms used in docking:

A docking procedure is essentially a combination of a search algorithm and a scoring function. A search algorithm predicts the orientation and conformations of ligand binding, commonly known as posing. Common search algorithms include:

Incremental Construction (IC):

In this method the ligand is broken down into several fragments by breaking the rotatable bonds of the ligand, Then, one of the fragments is chosen to be the first to dock into the active site. The anchor is typically the largest fragment, or the piece that may have a major functional role or interact with protein. Other fragments may be added gradually. Various orientations are created to fit in the active site. This way, the algorithm can figure out how flexible the ligand is. The IC algorithm, also known as anchor and grow method, has been used in a variety of docking programs, like DOCK, FLEXX, FLOG and Surflex (Meng *et al*., 2011).

Fast Shape Matching (SM):

Shape matching algorithms work by taking into account the geometric overlap of two molecules. Various algorithms are used to make several alignments between ligands and receptors. This approach can be used to find potential binding sites of a protein by searching for macromolecules on the surface (Dias *et al*., 2008).

Monte Carlo (MC) Simulations:

This method creates ligand poses by bond rotation or rigid-body translation, or rotation of the ligand. The resulting conformation is evaluated using an energy-based selection criterion. If the conformation meets the criterion, it is saved and further used to generate the next conformation. The iterations continue until the predetermined number of conformations is obtained. Applications of the Monte Carlo methods include an earlier version of AutoDock, ICM, QXP and Affinity (Meng *et al.*, 2011)

Simulated Annealing (SA):

In SA, a specific type of dynamic simulation is used to simulate a biological system. Each docking conformation is transferred into a simulation, where the temperature is reduced gradually over time during each cycle of simulation. This type of simulation may provide a higher precision result compared to MC, as it takes a more in-depth look at the conformational status and flexibility of the protein and ligands in different thermodynamic states over an extended period of time. When SA is used in combination with other techniques, it enhances precision and efficiency.

Distance Geometry (DG):

In this search method, information can be expressed as intra or intermolecular distances between molecules. These distances can be combined, allowing for the calculation of structures or conformations using these distances. Compared to other algorithms, the DG algorithm has a smaller number of distance constraints.

Evolutionary Programming (EP):

EP algorithms are used to solve problems by using computer models of evolutionary natural processes. Genetic operators, or computational approximations, are used to simulate the lifetime of the most advantageous features. In the sample space, where a problem or search routine exists and many possible solutions are available, each solution is ranked according to a set of parameters. Only the top-ranked solutions are retained for the following calculation loop. The process is iterated until an optimal solution is found (Dias *et al*., 2008).

Genetic Algorithm (GA):

The concept of the Genetic Algorithm (GA) is derived from Darwin's theory of Evolution. The degrees of freedom of a ligand are expressed as binary strings, known as genes. The genes make up a chromosome, which is the pose of a ligand. Mutations and crossover are the two types of genetic operators found in GA. Mutations cause random changes in the genes, while crossover changes the genes by exchanging them between two different chromosomes. When these genetic operators modify the genes, it results in a novel ligand structure (Meng *et al*., 2011) Genetic algorithms have been applied in programs like AutoDock, GOLD, DIVALI and DARWIN.

Tabu Search (TS):

Tabu search is a method of iterative search to solve optimization issues. The TS docking algorithm has proven to be very precise, being able to avoid trapping the simulation in local minima and avoid visiting known minimum energy conformations.

The primary goal of a docking algorithm should be to generate a rapid approach that can identify new lead compounds (in Virtual screening) or reproducing experimental conformations (for validation using experimental data) with the highest possible precision (Dias *et al.,* 2008).

Scoring Functions:

The goal of a scoring function is to distinguish the correct poses from the incorrect poses or binders from the inactive compounds within a reasonable calculation time. However, a scoring function involves an estimation, rather than a calculation, of the binding affinity of the protein to the ligand, and through these functions different assumptions and simplifications are adopted. The scoring functions can be classified into force-field based, empirical and knowledge-based scoring functions.

Table 1: Algorithms and scoring functions used by 7 different softwares (Fan *et al*., 2019)

Name	**Search algorithm**	**Evaluation method**	**Speed**	**Features & Application areas**
Flex X	Fragmentation Algorithm	Semi-empirical calculation on free energy	Fast	Flexible rigid docking. It Utilizes incremental construction strategy to virtually screen small molecule databases.

Gold	GA (genetic algorithm)	Semi-empirical calculation on free energy	Fast	Flexible docking. This is a GA-based docking program. This software has been highly evaluated for its reliability and precision.
Glide	Exhaustive systematic search	Semi-empirical calculation on free energy	Medium	Flexible docking. This software reduces the search range using domain knowledge and has 3 modes: HTVS (high throughout virtual screen), SP (Standard Precision) and XP (Extra Precision).
AutoDock	GA (genetic algorithm) LGA (lamarckian genetic algorithm)	Semi-empirical calculation on free energy	Medium	Flexible-rigid docking. This software is often used along with Autodock-tools and it is free of charge for academic purposes.
LeDOCK	Simulated annealing (SA), Genetic algorithm (GA)	Molecular force field	Fast	Flexible docking. LeDock is a new development in the field of molecular docking. Due to its speed and high precision, it is highly recommended for virtual screening tasks.
Dock	Fragmentation algorithm	Molecular force field	Fast	Flexible docking. It is used in a wide range of applications and is often used in the docking of

				flexible ligands with flexible proteins.
Autodock Vina	GA (genetic algorithm)	Semi-empirical calculation on free energy	Fast	Flexible-rigid docking. Autodock vina makes use of an iterated local search global optimizer and has been found to be faster than autodock 4.

Softwares used in Docking:

In the past twenty years, many different docking tools and programs have been developed, both for academic and commercial purposes. These include DOCK, AutoDock, FlexX, Surflex, GOLD, ICM, Glide, Cdocker, LigandFit, MCDock, FRED, MOE-Dock, LeDock, AutoDock Vina, rDock and UCSF Dock. including DOCK, AUTODOCK, GOLD, FLEXX, ZDOCK, and others. All of these softwares are based on a particular search algorithm (Pagadala *et al*., 2017).

Software used here for docking:

Schrodinger's Glide:

Glide (Friesner *et al*., 2004) developed by Schrödinger, Inc. is a software used for docking small molecules into proteins and other biopolymers. In contrast to other approaches used for the docking of ligands to a fixed 3D structural model of a well-known protein receptor, the Glide approach performs a comprehensive systematic analysis of the conformation, orientation, and position space of the ligand being docked.

Why Glide:

Glide differs from other docking programs in that it performs a full systematic analysis of the conformation, orientation, and position space of the dock ligand with OPLS-AA, a force field

optimized for liquid simulation. The optimal conformation is refined with Monte Carlo sampling.

Different Softwares were tested to determine their ability to generate the correct ligand binding mode to its target and to identify the known compounds with the highest scores in the virtual screening trials. It was found that Glide consistently predicted the ligand poses with 90.0% precision. Glide also correctly detected the crystallographic pose in 61% of cases, compared to GOLD (48%) and ICM (45%) (Pagadala *et al*., 2017)

Glide was found to be almost twice as precise as GOLD and more than twice as precise as FlexX when it comes to ligands with up to 20 rotatable bonds. It was also found that Glide was more precise than the recently introduced Surflex method.

Methodology used by Glide:

Glide employs a hierarchy of filters to identify ligand positions in the active site region of the receptor. On a grid, the shape and characteristics of the receptor are represented by various sets of fields, which provide progressively more precise scoring of ligand pose. These fields are created as a part of the pre-processing step of the calculation and thus only need to be calculated once per receptor.

The subsequent step results in the formation of a set of initial ligand conformations. These ligand conformations are chosen from an exhaustive list of minima within the ligand's torsion angle space and are presented in a compact, combinatorial form. As a result, initial screens are conducted throughout the phase space available to the ligands to identify potential ligand poses, which significantly decreases the region of phase space.

From the initial screening poses, ligands are reduced to a minimum in the receptor field using a standard molecular mechanical energy function (in this case OPLS - AA force field) in combination with a distance dependent dielectric model.

Finally, three to six of the lowest-energy poses achieved in this manner are subjected to the Monte Carlo procedure, which looks at nearby torsional minima. In some cases, this procedure is necessary to correctly align peripheral groups and sometimes changes internal torsion angle. In order to select the correct docked pose, we employ a composite scoring function, which we refer to as Emodel, which combines GlideScore with the ligand receptor molecule mechanics interaction energy and ligand strain energy. This composite scoring function is significantly more effective at selecting the correct pose than either the molecular mechanics energy or GlideScore individually (Friesner *et al*., 2004)

Features:

Complete solution:

Glide has all kinds of speed vs. accuracy options, like the HTVS (high-throughput virtual screening) mode for efficiently enriching million-compound libraries, the SP (standard precision) mode for reliably docking tens to hundreds of thousands of ligands with high accuracy, and the XP (extra precision) mode for getting rid of false positives by using more extensive sampling and more advanced scoring, which leads to even higher enrichment.

Virtual screening:

Glide provides a rational workflow for virtual screening from HTVS to SP to XP, enriching the data at every level such that only an order of magnitude fewer compounds need to be studied at the next higher accuracy level.

Accurate binding mode prediction:

Glide finds the right binding modes for many test cases. It works better than other docking programmes because it can get lower RMS variations from native co-crystallized structures.

Universal applicability:

Glide has universal applicability due to its superior docking precision and high enrichment across a wide variety of receptor types.

2.6. ADMET analysis:

The ADMET study is an extremely important part of the pharmacokinetics assessment of a drug. In this study, the toxicity of the drug is evaluated, along with its absorption, distribution, metabolism, and excretion. It is crucial in predicting a drug's fate and the effects it may have inside the body. For example, it helps determine how much of a drug is absorbed when taken orally and how much of the drug is absorbed through the gastrointestinal tract. Poor absorption can lead to issues with distribution and metabolism, potentially causing neurotoxicity and nephrotoxicity. Ultimately, the study seeks to understand how a drug molecule is processed within an organism. Therefore, ADMET study is a vital component of computational drug design (Flores-Holguín *et al*., 2021).

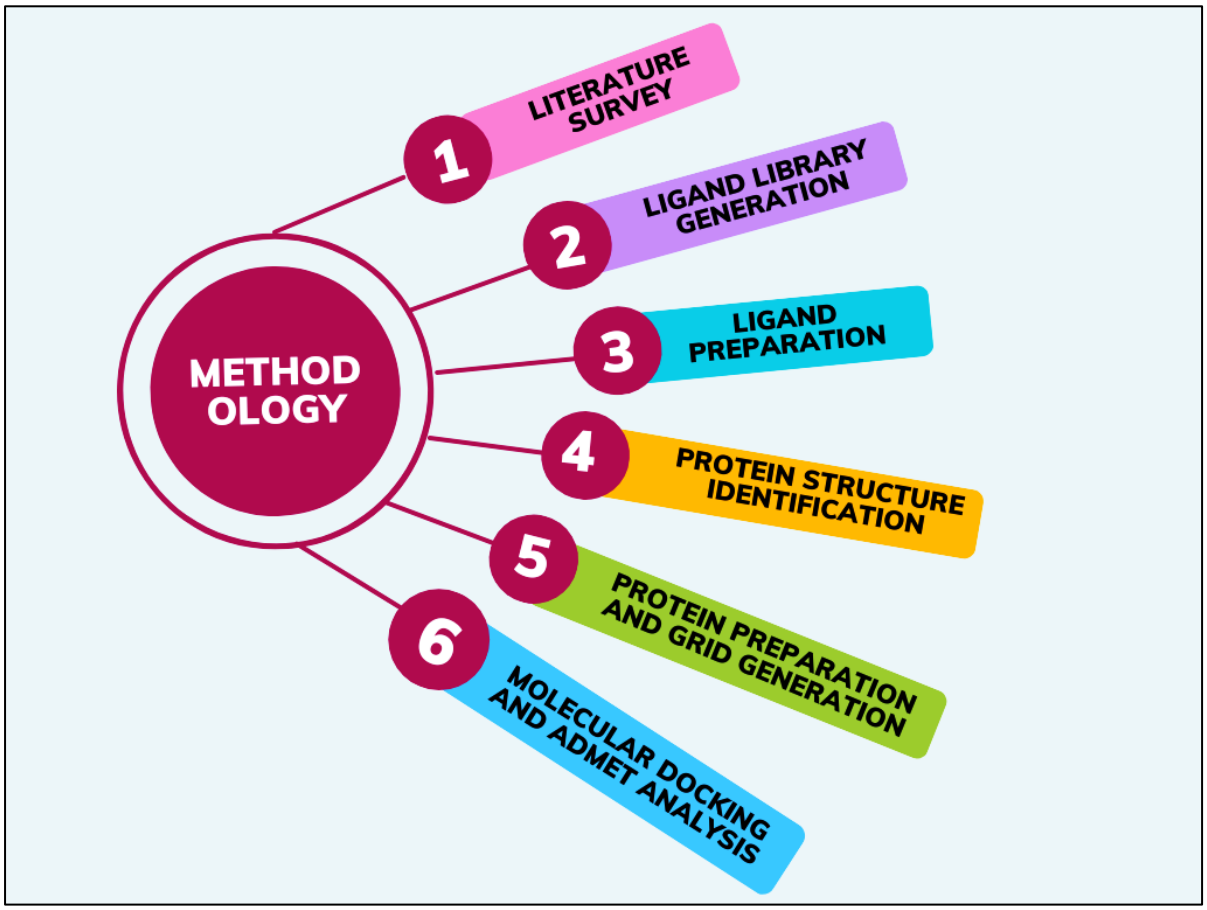
At the early stages of drug discovery, drug-likeness is a useful concept. The "rule of five" was first proposed in 1997 by Lipinski and his colleagues. This was the first and most widely used rule-based filter for drug-likeness, distinguishing whether or not a molecule is orally active by: Molecular weight (MW) ≤ 500, Octanol / water partition coefficient (A'log'P) ≤ 5, Number of Hydrogen Bond Donors ≤ 5 and Hydrogen Bond Acceptors ≤ 10. Under the rule of five, a molecule is not orally active if it fails to meet two or more of these four rules. These rules are not applicable for complex natural products because they are derived from relatively small molecules (Guan *et al*., 2019)

Several studies have shown that drugs have failed in clinical trials due to adverse side effects and toxicities caused by the drugs, which have proven extremely detrimental and costly in drug development. Since in vivo and in vitro evaluations are costly and time consuming, in silico methods are being widely used to estimate these properties. Performing in silico ADMET and

drug similarity predictions can help in discovering new targets and compounds with expected biological activity.

The blood-brain barrier (BBB) is a property which is used to predict various features of CNS vasculature. The lack of endothelial cell pores in the CNS vasculature and the variety of properties that regulate the transport of ions, molecules and cells make the CNS vasculature very restrictive in nature. Human intestinal absorption (HIA) is another important property in ADMET analysis. HIA is one of the key steps in transporting chemical compounds (drugs) to their intended destination in the body (Srivastava *et al*., 2022)

3. Methodology:



3.1. Literature survey:

A literature survey was conducted using PubMed (PubMed.nlm.ncbi.nih.gov). Research articles containing molecules that demonstrated inhibition against EGFR were selected. Structures were identified through a literature survey. The criteria for selection of the structures was the IC_{50} value of the molecules against EGFR. IC50 is the minimum inhibitory concentration of a drug by which a biological process can be inhibited by half.

Cellosaurus (Cellosaurus.org / Search) database was then used to verify the cell lines used for the experiments, the sequence variation and mutation in the cell lines, disease in the cell line and other details. For each compound, a record was maintained, PMID was obtained from the paper, and a bioactivity example was provided.

3.2. Ligand Library Generation:

The Ligand library was prepared manually by sketching the potential EGFR inhibitors identified through the literature survey using Marvin Sketch. Some of the structures were created by importing the chemical name of the structures from the research papers through Marvin Sketch and cross-checking the structures with those given in the papers. Those compounds whose chemical names were not mentioned in the research paper were sketched manually by drawing a central scaffold first and then attaching the side chains to it.

These structures were then validated through the Marvin Sketch structure checker and saved in a folder in different file formats: Structure data file (SDF), SMILES and png file formats for further use and docking. Structures of all ligands were finally validated using the PubChem database.

3.3. Ligand Preparation:

After generating the ligand library, the next step was to prepare the ligands for docking. Maestro Schrodinger interface was used for ligand preparation. Maestro is the Graphical user interface (GUI) for all of Schrodinger's products, including Glide and Ligprep. The ligands were first imported in Maestro. Then ligand preparation was performed using LigPrep. With the help of LigPrep, all the ligands were first converted from 2D to 3D conformation and the ligands were then neutralized by adding or removing protons. By using the desalter function in LigPrep, all the molecules in each ligand structure were eliminated except the molecule with the largest number of atoms. Multiple conformations for each input structure with different combinations of ionized states were produced with the help of ionizer function in LigPrep. The valency of these structures was also satisfied with the help of LigPrep.

3.4. Protein Structure Identification:

SRC protein was searched in PDB. A record of the PDB ID, title, resolution, experimental method, co-crystallized ligand, mutations, and missing residues present in the PDB structures was noted. Selection of the final protein structure was done based on the following criterias:

- Selected protein structure should be bound to a ligand molecule
- Absence of mutations
- Resolution should be less than 2Å
- Lesser number of missing residues
- Absence of missing residues at the binding site.
- Only kinase domain should be there in the protein structure

3.5. Protein Preparation:

The crystal structure of SRC protein was imported from PDB (protein data bank) (www.pdb.org).Protein preparation was done using Protein preparation workflow. While preparing a protein, missing hydrogen atoms, metals in the protein, bound ligands were fixed, and gaps due to missing amino acids were filled. With the help of the "Optimize H-bond assignments" option, the orientation of hydrogen atoms that were part of water molecules and also protein amino acids were adjusted to maximize the H-bond network. The pKa of different amino acids was also calculated in order to assign the correct protonation states for each of the amino acids in the protein. The Clean Up section then minimized the whole complex to a local energy minimum and the water molecules that were far away from the ligand were deleted since they are not involved in the interaction between protein and ligand.

3.6. Grid Generation:

After the protein is prepared, the next step was to identify the binding pocket and generate a grid around it. A receptor grid defines the area of interaction between the protein and the ligand. Grid preparation was done using Glide. To prepare the grid, we first identified the binding pocket in the prepared protein structure. After identifying the binding pocket, it was highlighted and then a grid was generated around this binding pocket. Grid generation was further used in docking step.

3.7. Molecular Docking:

After preparing the ligands and the protein, and generating the grid, Molecular Docking was performed using Glide. It was done in three steps: High throughput virtual screening (HTVS), Standard-precision (SP), and extra precision (XP). This was done in order to achieve accurate results. HTVS provides better sampling than SP, and XP provides more extensive sampling than SP. During each step of the docking process, only 50% of ligands with the highest docking scores were considered, and the remaining 50% of the ligands were eliminated. At each step of docking, the generated grid file was loaded, followed by the top 50% of the ligands. The first step of docking was performed in HTVS mode, then the second step of docking in SP mode, and the third and final step of docking in XP mode. Further, the docking score of the reference cocrytal was calculated using the XP docking step.

3.8. ADMET analysis:

All those ligands that had a glide score greater than the reference Cocrystal were further screened. ADMET properties like bioavailability, brain penetration properties, oral absorption and carcinogenicity properties, and other intestinal absorption properties of the

active compounds were determined using an online software Swiss ADME. SMILES were used to identify ADMET properties of these compounds. Ligands with 0 Violations of Lipinski's rule were selected. Boiled Egg Analysis was then performed on the selected ligands and those ligands that showed good Boiled Egg results were selected.

3.9. Docking Result Analysis:

The Ligands with the highest docking scores were selected and further analyzed. For this, the 2-D overlay was first deselected in order to hide the ligand 2-D view from the workplace. Docking results were analyzed by looking at properties like Ligand Interaction diagrams, Interaction complementarity, Geometric complementarity, Docking score. Now, we will look at the 2-D ligand interaction diagram to study the interactions between the ligand and protein. The ligand interaction diagram was visualized by opening the Ligand Interaction diagram panel in maestro. Then, the 3-D ligand interactions were visualized. For this, the ligand and the binding pocket were brought into focus and the interactions between them were represented as dashed lines. Here, different bond colours represent 9 different kinds of interactions. In the next step, the geometric complementarity was checked. For this the representation of protein was changed to surface. Finally, the docking scores were noted from the table option from the top right corner of the workplace.

4. Results:

4.1. Literature Survey and Ligand Library Generation:

The literature survey resulted in the identification of 1433 ligands from 121 research papers that showed inhibitory effect on Epidermal Growth Factor Receptor (EGFR). These ligands were thus used in the present study. The library of ligands was created manually by sketching potential EGFR inhibitors previously identified using Marvin Sketch, as

illustrated in Figure. A few structures were imported into Marvin Sketch by chemical name, and those structures whose names were not given were manually sketched. In total, 1433 compounds were identified and sketched and then saved in Structure Data File (sdf) format for docking**.**

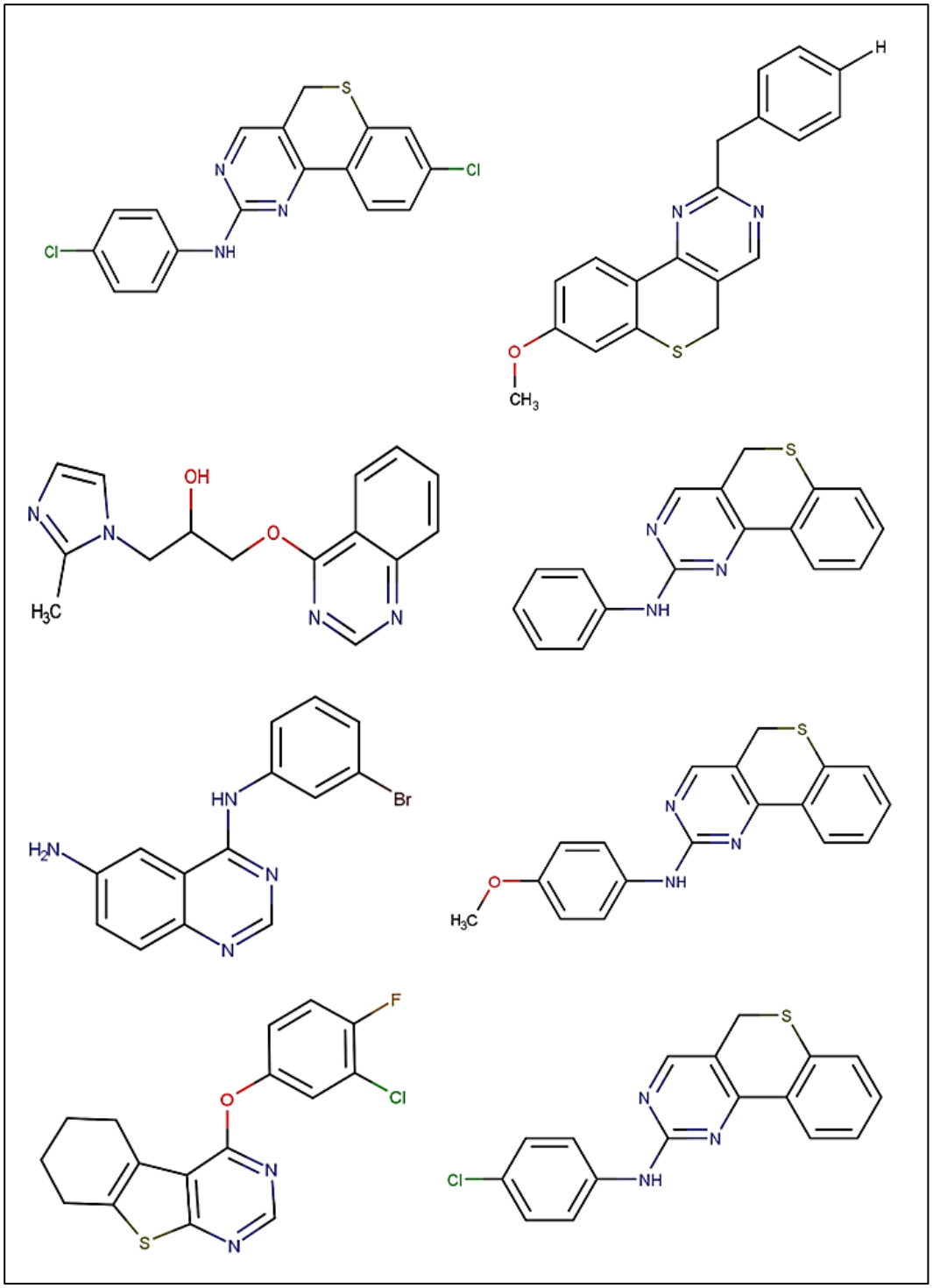


Fig6: Few Sketched Ligands

4.2. Protein Structure Identification:

The crystal structure of SRC (PDB: 1Y57) taken in this study was retrieved from PDB (protein data bank) (www.pdb.org). PDB: 1Y57 was selected because its resolution was 1.91 Å. Resolution is a measure of the level of details present in the diffraction pattern and the level of detail that will be seen when the electron density map is calculated. A resolution of less than 2Å is generally considered because the resolution affects sidechain locations, and so, accurate sidechain insertion is correlated with resolution. While NMR is more sensitive to the atomic level details of a molecular structure, X-ray diffraction is more sensitive to the overall shape of the molecule. So, structure estimated through X-RAY diffraction is considered over structure estimated through NMR. This protein structure was selected because it had no mutations in its structure. A Network of interactions between amino acids gives proteins their shape and stability, while mutations can disrupt these networks and lead to protein misfolding or instability. In order to get accurate docking results, it is important to use a protein structure that has a co-crystallized ligand that is chemically similar to our ligand. This protein structure had no missing residues and so there would be no interference with the ligand binding domain during docking process. Finally, A chain of PDB: 1Y57 was selected for docking.

4.3. Ligand Preparation:

Ligand preparation was performed using LigPrep. All the 1433 Ligands were first converted from 2D to 3D conformation with the help of LigPrep. The ligands were then neutralized by adding or removing protons. All the molecules in each ligand structure were eliminated except the molecule with the largest number of atoms. Multiple conformations for each input structure with different combinations of ionized states were produced with the help of LigPrep. The valency of these structures was also satisfied.

4.4. Protein Preparation and Grid Generation:

Protein preparation was done using Protein preparation workflow. Protein preparation was carried out by adding missing hydrogen atoms and filling the missing amino acids if any in the protein structure. With the help of the “Optimize H-bond assignments” option, the orientation of hydrogen atoms that are part of water molecules and also protein amino acids was adjusted in order to maximize the H-bond network. The pKa of different amino acids was also calculated to assign the correct protonation states for each of the amino acids in the protein. The whole complex was then minimized to a local energy minimum, and the water molecules that were far away from the ligand were deleted since they will not be involved in the interaction between protein and ligand.

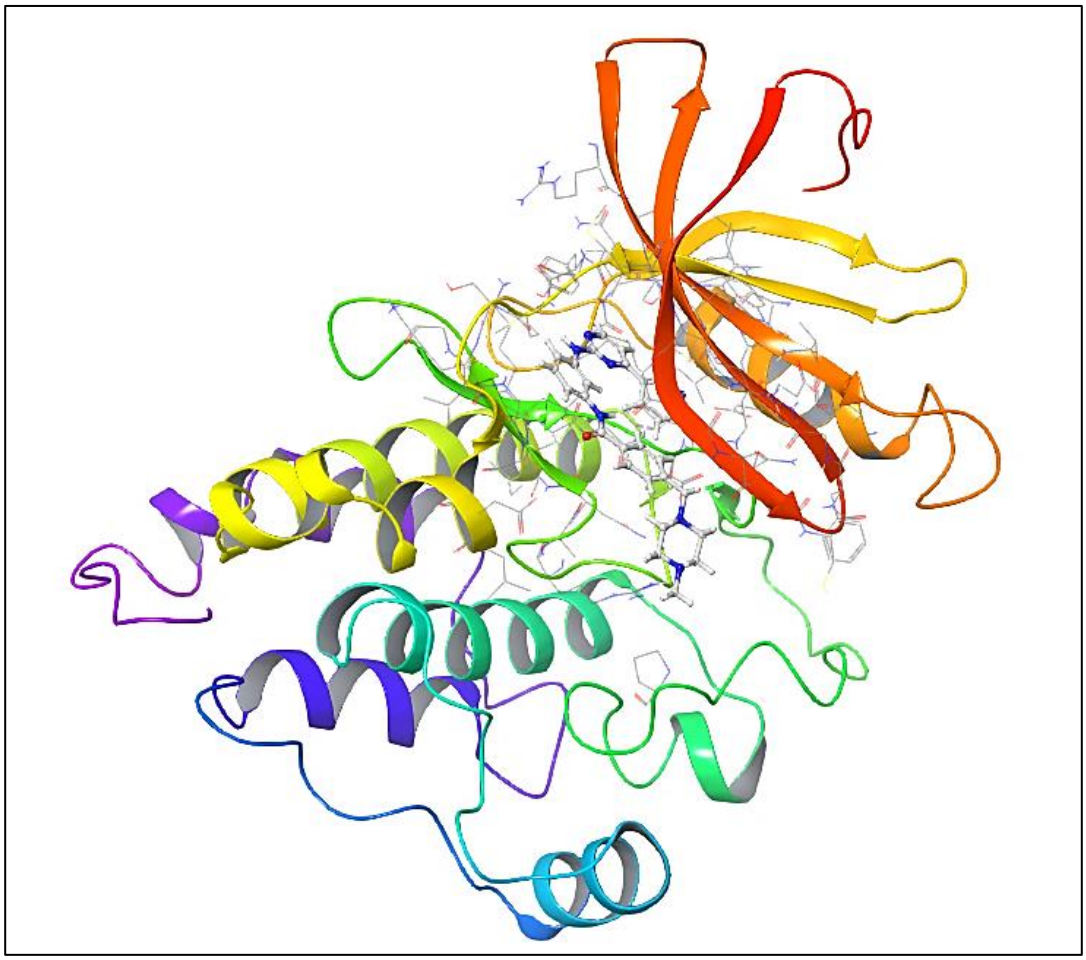


Fig7: Prepared protein structure of SRC protein (PDB: 1Y57)

After protein preparation, a grid was generated around the binding pocket of the prepared protein structure (PDB: 1Y57). The Receptor grid defines the area of interaction between the protein and the ligand. This was carried out with the receptor grid generation tool in maestro 12.7 which defines the area around the active site in terms of coordinates x, y, and z. A pink colored grid is generated around the binding site in Figure 8.

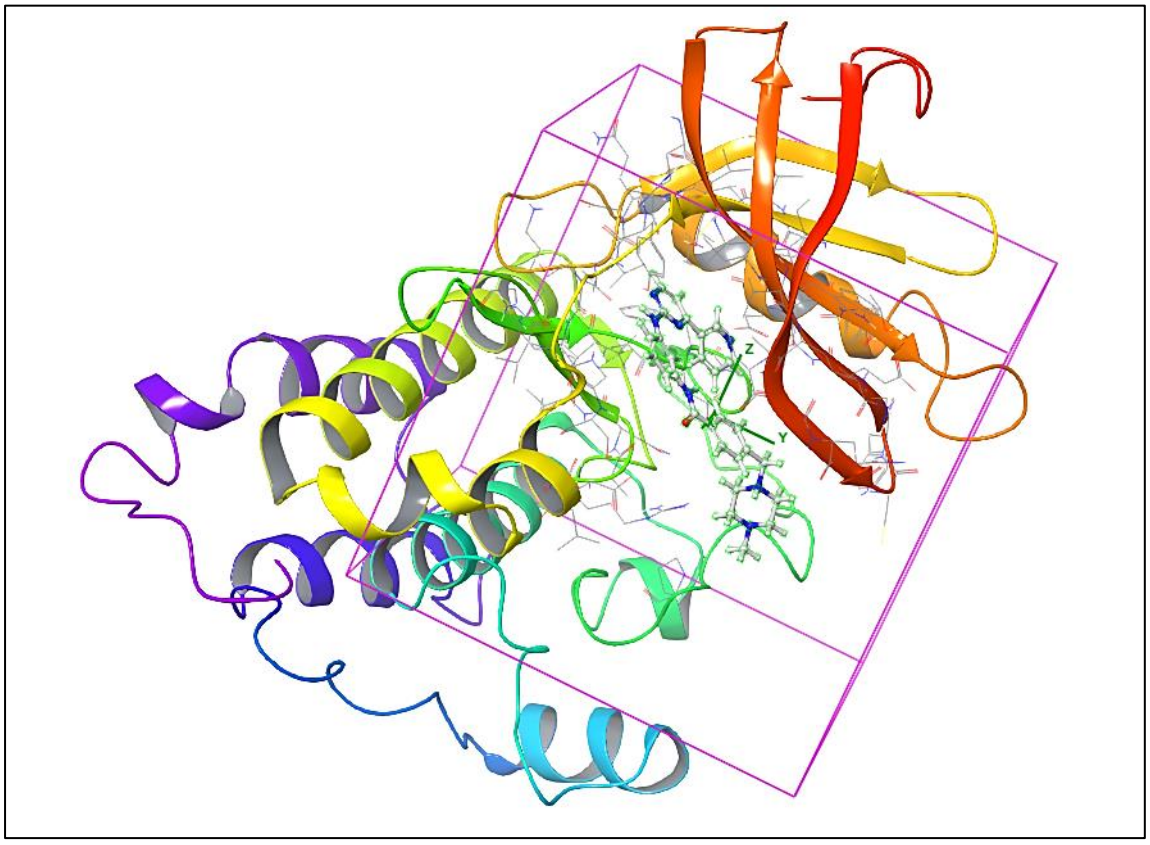


Fig8: Grid generation around binding site of SRC protein

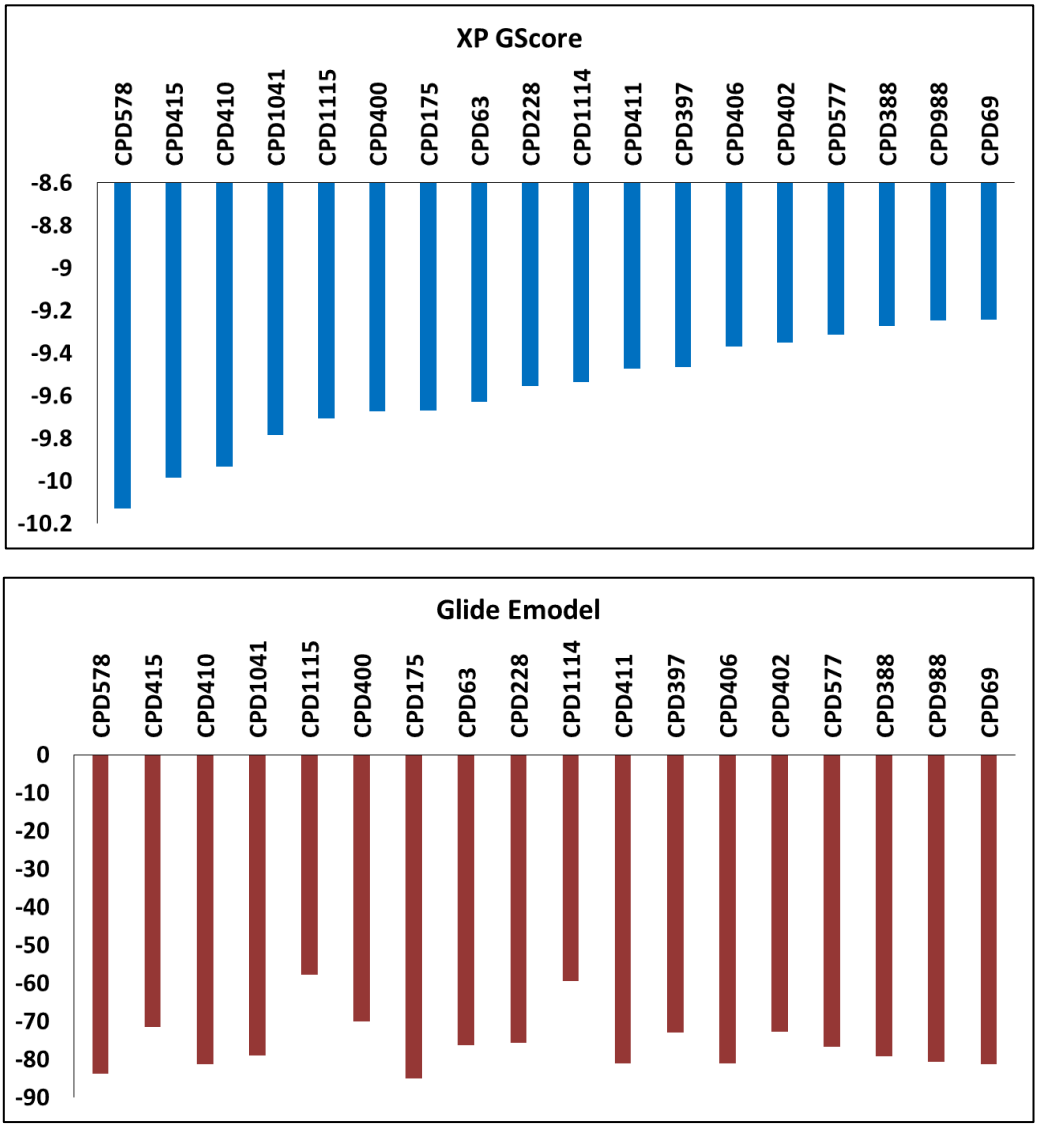
4.5. Molecular Docking and ADMET Analysis:

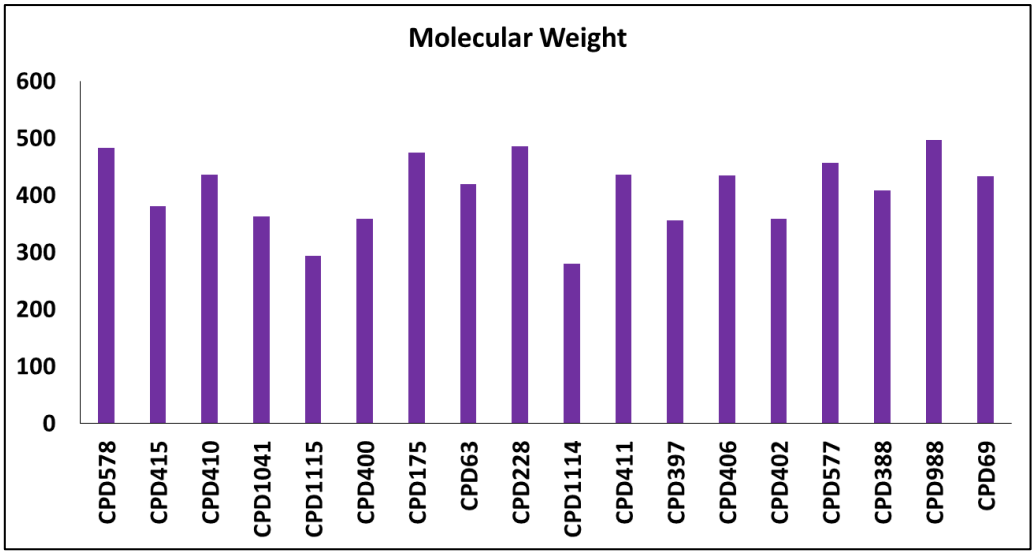
The prepared ligands were docked with the prepared SRC protein structure (PDB: 1Y57) using the Ligand Docking panel of Glide. 1433 ligands were considered for the HTVS docking process. Out of these 1433 ligands, HTVS resulted in 1421 docked ligands. 50% of these ligands, approximately 700 with higher glide scores were taken for the second step of docking, that is, SP. 50% of the ligands, approximately 300 with higher scores were again selected for the third step, XP and XP docking process was performed. The docking score of the reference cocrystal was then compared with the docking scores of these ligands. The XP docking score of the reference cocrystal was -9.199. 25 ligands were found to have a docking score higher than that of the reference cocrystal. ADMET Analysis of these 25 ligands was performed. Out of these, 18 ligands were selected that showed 0 violations to Lipinski's rule. Finally, 3 compounds with the highest docking scores were selected out of these 18 ligands. The docking scores of 18 ligands along with their details of Lipinski's Rule are listed in the given table:

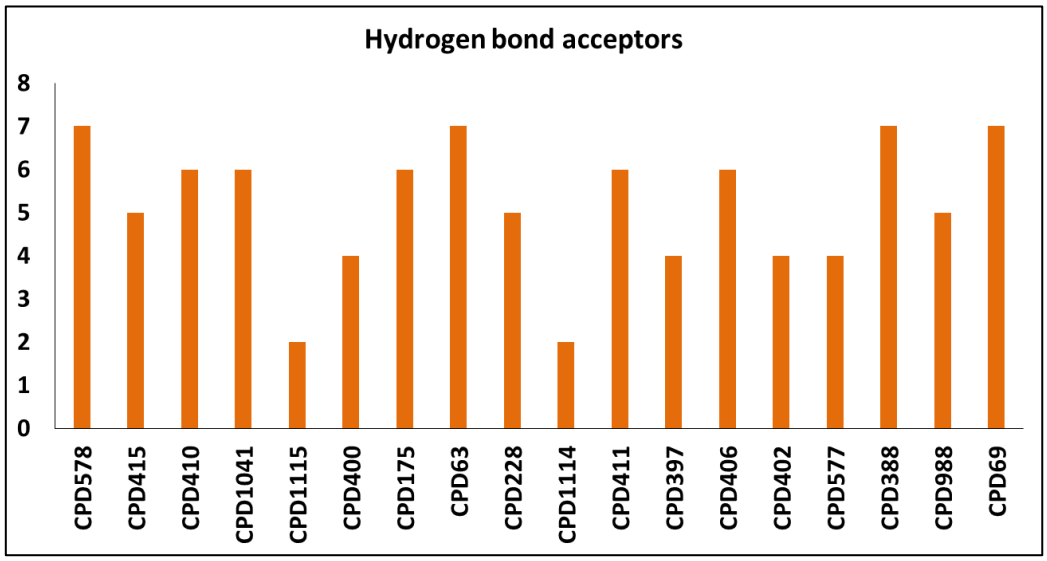
Table 2: Docking Scores and other details of Lipinski's rule of 18 ligands

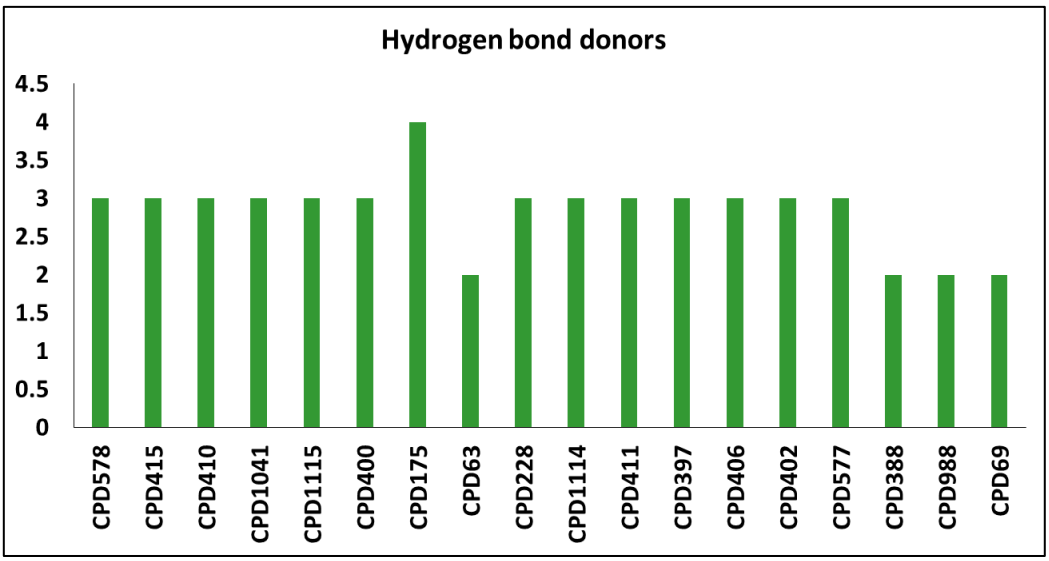
Title	XP GScore	Glide emodel	Molecular Weight	H-bond acceptors	H-bond donors	Partition coefficient (Log P)	Lipinski violations
CPD578	-10.13	-83.699	482.45	7	3	4.66	0
CPD415	-9.983	-71.499	380.45	5	3	1.71	0
CPD410	-9.934	-81.338	436.51	6	3	2.1	0
CPD1041	-9.785	-79.043	362.39	6	3	2.1	0
CPD1115	-9.705	-57.776	294.35	2	3	2.95	0
CPD400	-9.673	-70.068	358.4	4	3	3.05	0
CPD175	-9.669	-84.949	474.96	6	4	2.9	0
CPD63	-9.629	-76.286	418.85	7	2	3.09	0
CPD228	-9.556	-75.589	485.58	5	3	3.09	0
CPD1114	-9.537	-59.437	280.32	2	3	2.62	0
CPD411	-9.474	-81.058	436.51	6	3	2.14	0
CPD397	-9.466	-72.887	356.38	4	3	2.96	0
CPD406	-9.368	-81.134	434.49	6	3	2.22	0
CPD402	-9.351	-72.824	358.4	4	3	2.99	0
CPD577	-9.313	-76.624	456.54	4	3	4.61	0
CPD388	-9.271	-79.103	408.41	7	2	2.81	0
CPD988	-9.248	-80.584	496.6	5	2	4.06	0
CPD69	-9.243	-81.241	432.88	7	2	3.45	0

The range of XP docking scores was found to be from -9.243 to -10.13 for the 18 ligands. The range of Binding energy was from -59.437 to -84.949. Molecular Weight was from 280.32 to 496.6 for these 18 ligands. The maximum number of Hydrogen bond acceptors was found to be 7 while the minimum number was 2. Similarly, the maximum number of Hydrogen bond donors was found to be 4 while the minimum number was 2. The Partition coefficient (Log P) was found to be from 1.71 to 4.66 for these 18 ligands. These statistics have been represented in the form of bar graphs in Figure 9




Molecular Weight
600
500
400
300
200
100
0
CPD578
CPD415
CPD410
CPD1041
CPD1115
CPD400
CPD175
CPD63
CPD228
CPD1114
CPD411
CPD397
CPD406
CPD402
CPD577
CPD388
CPD988
CPD69


Hydrogen bond acceptors
8
7
6
5
4
3
2
1
0
CPD578
CPD415
CPD410
CPD1041
CPD1115
CPD400
CPD175
CPD63
CPD228
CPD1114
CPD411
CPD397
CPD406
CPD402
CPD577
CPD388
CPD988
CPD69


Hydrogen bond donors
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
CPD578
CPD415
CPD410
CPD1041
CPD1115
CPD400
CPD175
CPD63
CPD228
CPD1114
CPD411
CPD397
CPD406
CPD402
CPD577
CPD388
CPD988
CPD69

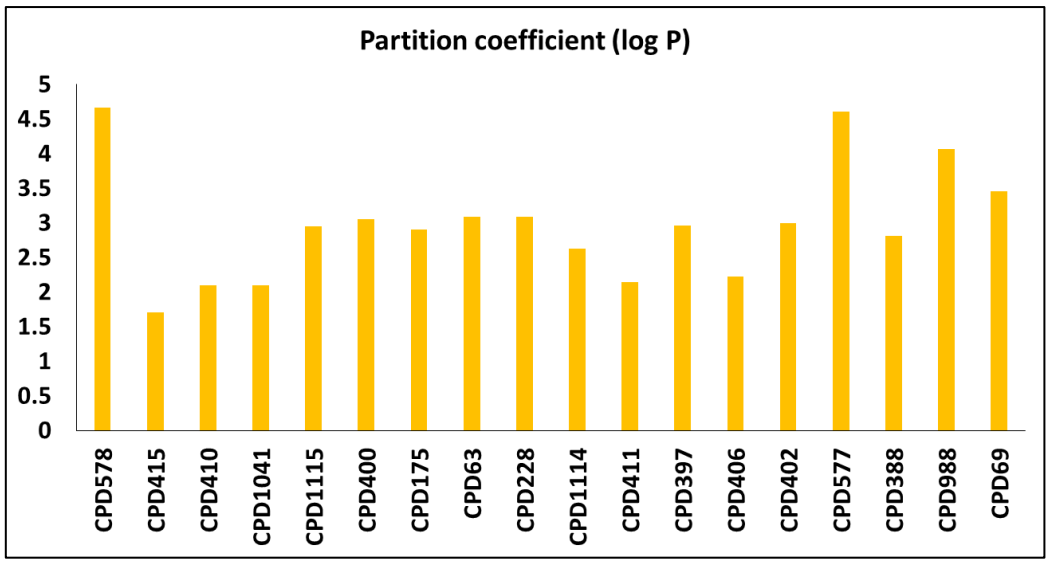


Fig9: Graphical Representation of the Docking Scores and other details of Lipinski's rule of 18 ligands

5. Discussion:

CPD578:

The analysis of CPD578 was done to study its interactions with the protein structure and the interacting amino acids as well as the groove where this ligand fits. It had the highest docking score of -10.13 kcal/mol and binding energy of -83.699 with 0 violations of Lipinski's Rule. The molecular weight was found to be 482.45 Da. It has 7 H-bond acceptors and 3 H-bond donors. Partition coefficient for this ligand was 4.66. The residues in the binding pocket of SRC (PDB: 1Y57) were Leu 273, Gly 274, Gln 275, Gly 276, Cys 277, Phe 278, Gly 279, Glu 280, Val 281, Ala 293, Leu 295, Met 314, Val 323, Thr 338, Glu 339, Tyr 340, Met 341, Ser 342, Lys 343, Gly 344, Leu 393, Ala 403, Asp 404 and Phe 405. CPD578 interacted with the protein by forming 3 hydrogen bonds, one with Thr at 338 position and 2 hydrogen bonds with Met at position 341. This has been displayed in the figure. Further, it was found that the ligand fits well in the binding pocket of SRC protein.

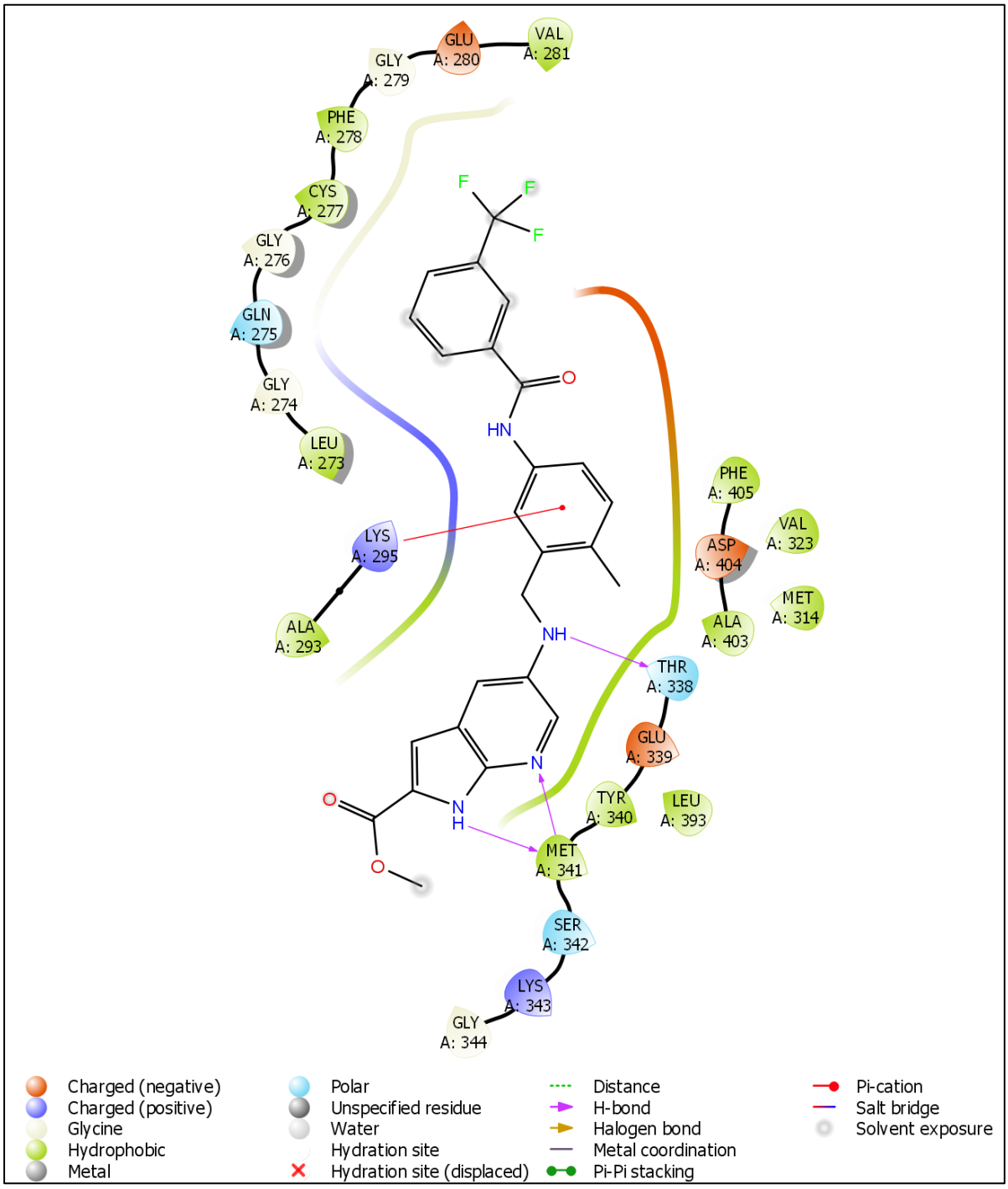


Fig10: 2D interaction of CPD578 with amino acid residue within the active site of SRC

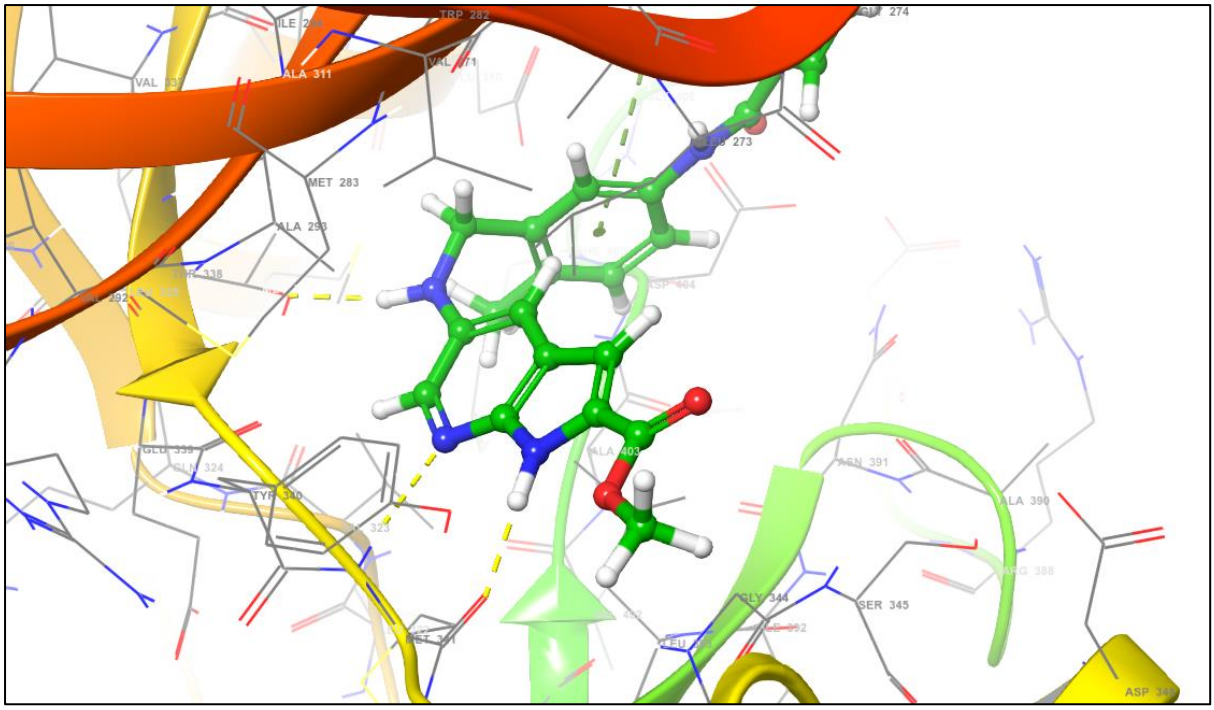


Fig11: 3D interaction of CPD578 with amino acid residue within the active site of SRC.

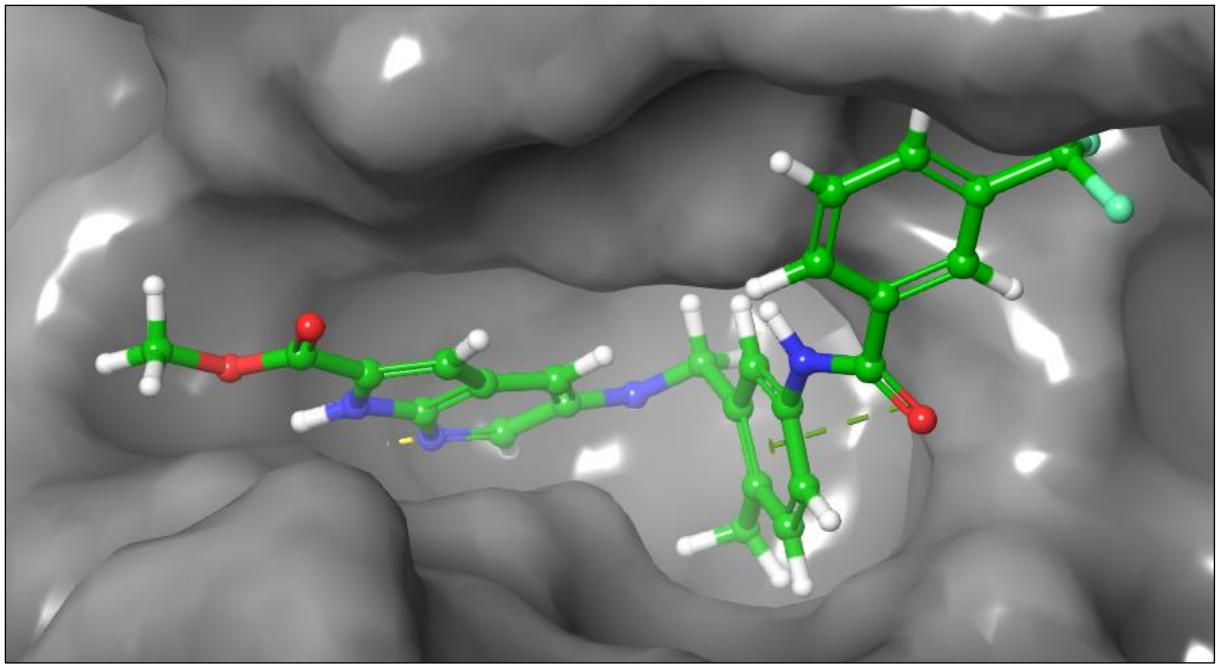


Fig12: CPD578 fits well in the binding site of SRC

CPD415:

The analysis of this ligand revealed its interactions with the protein as well as the interacting amino acids and the groove wherein it fits. CPD415 had the second best docking score of -9.983 kcal/mol and binding energy of -71.499 with 0 violations of Lipinski's Rule. The molecular weight was found to be 380.45 Da. It has 5 H-bond acceptors and 3 H-bond donors. Partition coefficient for this ligand was 1.71. The residues in the binding pocket of SRC (PDB: 1Y57) were Leu 273, Val 281, Ala 293, Lys 295, Val 323, Thr 338, Glu 339, Tyr 340, Met 341, Ser 342, Lys 343, Gly 344, Ser 345, Ala 390, Asn 391, Leu 393, Asp 404. CPD578 interacted with the protein by forming 4 hydrogen bonds, one bond each with Glu at 339 position and with Ala 390 and 2 hydrogen bonds with Met at position 341. This has been displayed in the figure. Further, it was found that ligand fits well in the binding pocket of SRC protein.

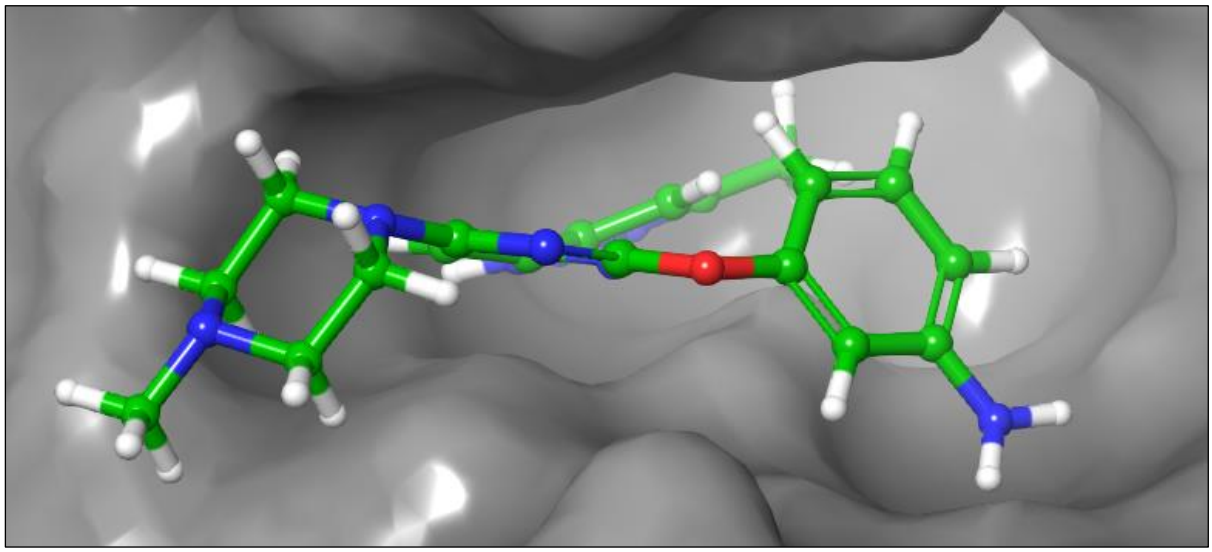


Fig13: CPD415 fits well in the binding site of SRC

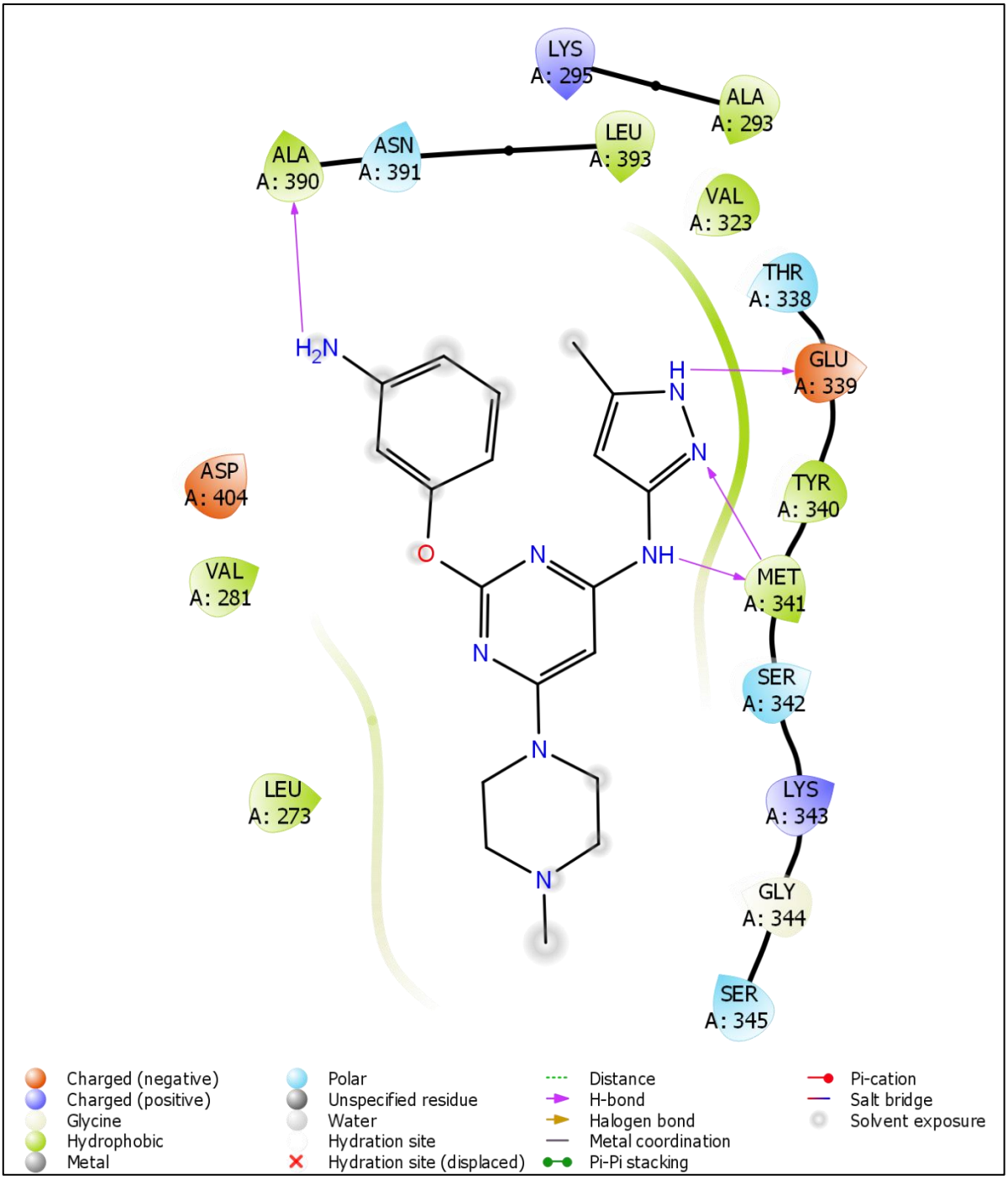


Fig14: 2D interaction of CPD415 with amino acid residue within the active site of SRC

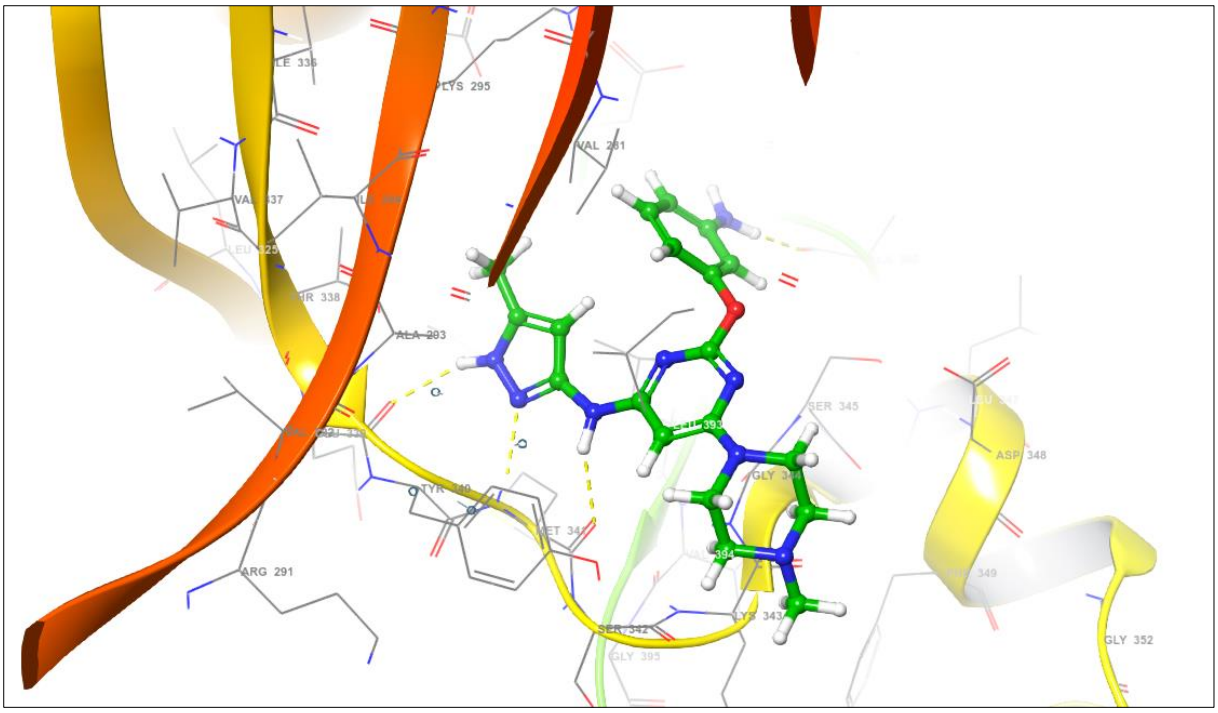


Fig15: 3D interaction of CPD415 with amino acid residue within the active site of SRC.

CPD410:

The analysis of CPD410 was done to study its interactions with the protein structure and the interacting amino acids as well as the groove where this ligand fits. It had the third best docking score of -9.934 kcal/mol and binding energy of -81.338 with 0 violations of Lipinski's Rule. The molecular weight was found to be 436.51 Da. It has 6 H-bond acceptors and 3 H-bond donors. Partition coefficient for this ligand was 2.1. The residues in the binding pocket of SRC (PDB: 1Y57) were Leu 273, Val 281, Ala 293, Val 323, Thr 338, Glu 339, Tyr 340, Met 341, Ser 342, Lys 343, Gly 344, Ser 345, Asp 348, Ala 390, Asn 391, Leu 393, Ala 403, Asp 404 and Phe 405. CPD410 interacted with the protein by forming 4 hydrogen bonds, one each with Gly at 338 position and Ala at 390 position and 2 hydrogen bonds with Met at position 341. This has been displayed in the figure. Further, it was found that the ligand fits well in the binding pocket of SRC protein.

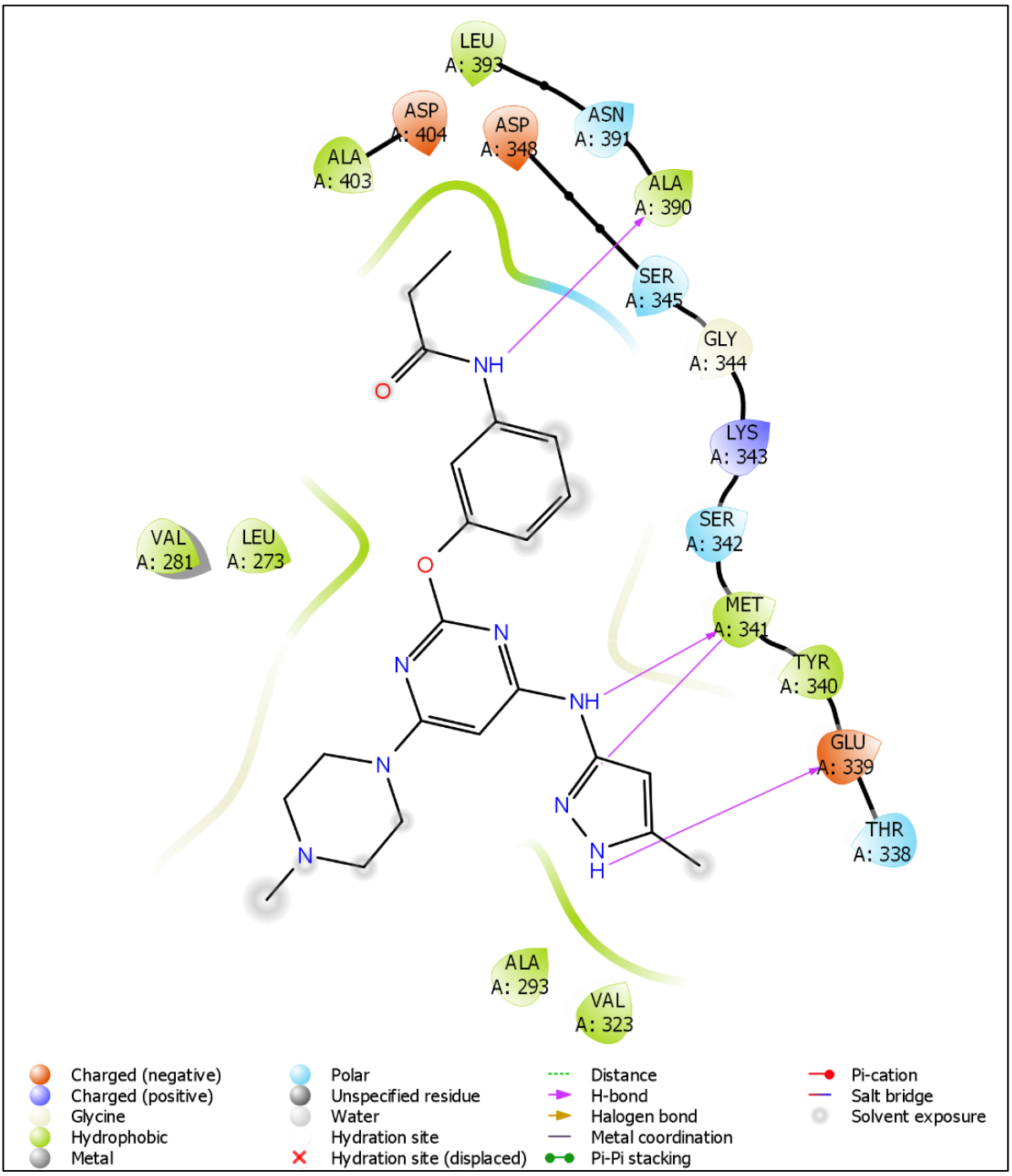


Fig16: 2D interaction of CPD410 with amino acid residue within the active site of SRC

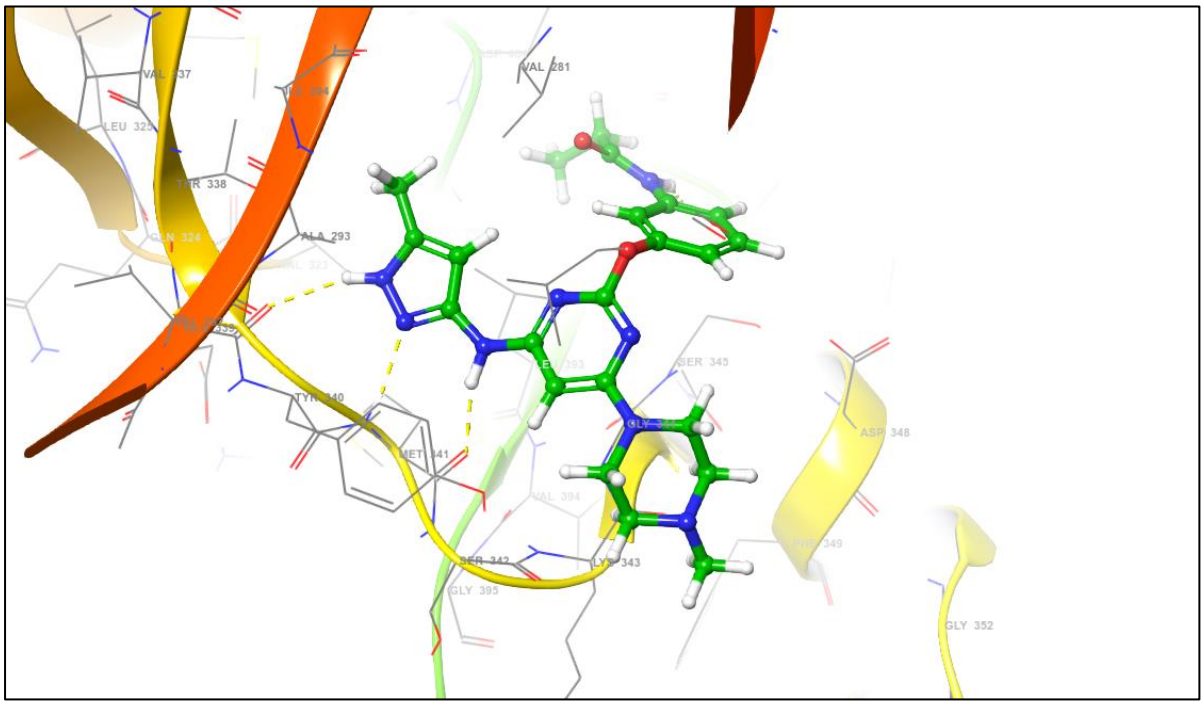


Fig17: 3D interaction of CPD410 with amino acid residue within the active site of SRC.

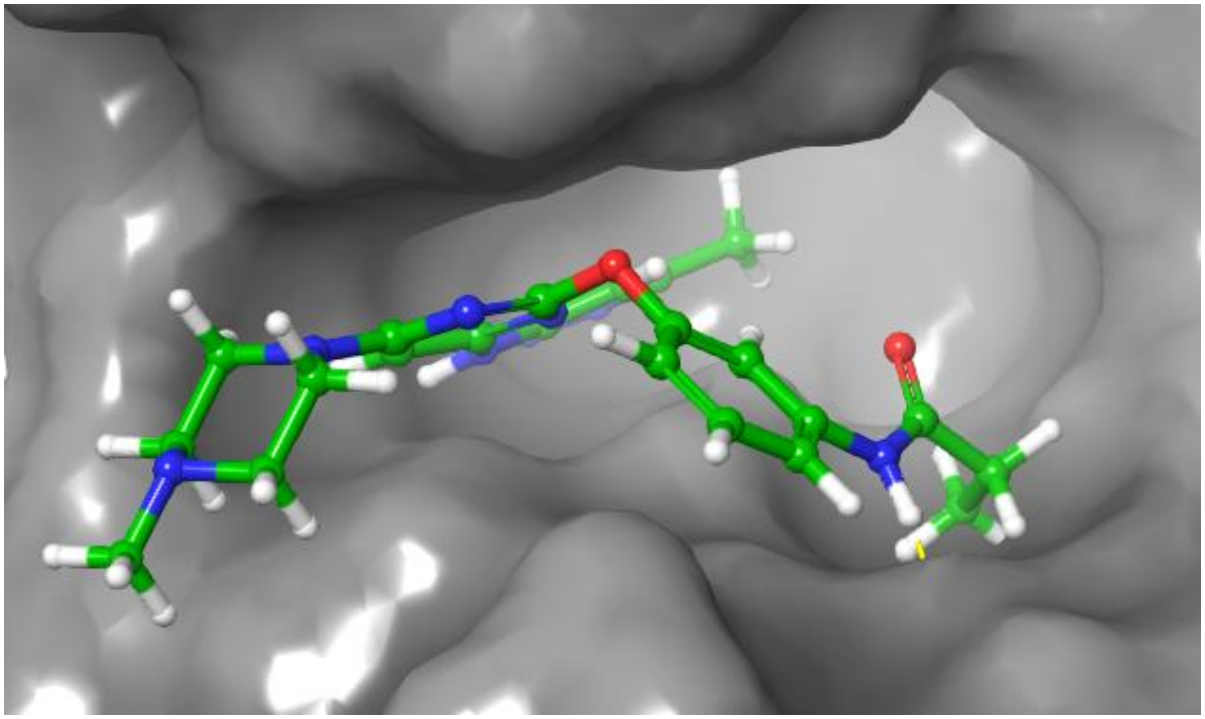


Fig18: CPD410 fits well in the binding site of SRC protein

Cancer is a leading cause of deaths worldwide. There are more than 100 types of cancer. Lung cancer, in particular, is responsible for numerous cancer-related fatalities in both men and women. Research has identified the EGFR group of growth factors as key molecules that promote lung cancer generation and propagation. The Epidermal growth factor receptor is a transmembrane glycoprotein that regulates signaling pathways to control cell proliferation. Several approaches have been developed to target and interfere with EGFR-mediated effects, including the use of monoclonal antibodies and small molecule Tyrosine kinase inhibitors.

SRC is a non-receptor tyrosine kinase that has been linked to advanced malignancy and poor prognosis in several human cancers. Increased SRC protein levels and its tyrosine kinase activity are involved in cancer progression and metastatic disease by facilitating the actions of other signaling proteins like EGFR. It has been studied that interactions between the epidermal growth factor receptor (EGFR) and the nonreceptor tyrosine kinase c-SRC may contribute to an aggressive phenotype in multiple human tumors. Hence, there is a need to identify inhibitors that can target both SRC and EGFR protein kinases.

6. Conclusion:

The goal of this investigation was to identify such molecules that have the potential to inhibit both EGFR and SRC protein kinases simultaneously. To achieve this goal, a ligand library was generated and docked with SRC protein (PDB: 1Y57) using a molecular docking approach. Through this process, three compounds were identified that had higher docking scores than the reference co-crystal and exhibited high binding affinity towards SRC protein (PDB: 1Y57). Subsequently, ADMET analysis was conducted, which showed that these three compounds did not violate Lipinski's rule of five and had desirable pharmacokinetic characteristics. These findings suggest that further research could help to develop these compounds as potent dual inhibitors of EGFR and SRC protein kinases, which could play a crucial role in preventing cancer development.

References:

1. Bethune, G., Bethune, D., Ridgway, N., & Xu, Z. (2010). Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. Journal of Thoracic Disease, 2(1), 48–51.
2. Anand, P., Kunnumakkara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., Sung, B., & Aggarwal, B. B. (2008). Cancer is a preventable disease that requires major lifestyle changes. Pharmaceutical Research, 25(9), 2097–2116.
3. Wu, D., Rice, C. M., & Wang, X. (2012). Cancer bioinformatics: a new approach to systems clinical medicine. BMC Bioinformatics, 13(1), 71.
4. Rusnak, D. W., Affleck, K., Cockerill, S. G., Stubberfield, C., Harris, R., Page, M., Smith, K. J., Guntrip, S. B., Carter, M. C., Shaw, R. J., Jowett, A., Stables, J., Topley, P., Wood, E. R., Brignola, P. S., Kadwell, S. H., Reep, B. R., Mullin, R. J., Alligood, K. J., … Lackey, K. (2001). The characterization of novel, dual ErbB-2/EGFR, tyrosine kinase inhibitors: potential therapy for cancer. Cancer Research, 61(19), 7196–7203.
5. Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The next Generation. *Cell* **2011**, *144* (5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013.
6. Biesalski, H. K.; de Mesquita, B. B.; Chesson, A.; Chytil, F.; Grimble, R.; Hermus, R. J.; Kohrle, J.; Lotan, R.; Norpoth, K.; Pastorino, U.; Thurnham, D. European Consensus Statement on Lung Cancer: Risk Factors and Prevention. Lung Cancer Panel. CA. Cancer J. Clin. 1998, 48 (3), 167–176. https://doi.org/10.3322/canjclin.48.3.167.
7. Singh, N.; Agrawal, S.; Jiwnani, S.; Khosla, D.; Malik, P. S.; Mohan, A.; Penumadu, P.; Prasad, K. T. Lung Cancer in India. *J. Thorac. Oncol.* **2021**, *16* (8), 1250–1266. https://doi.org/10.1016/j.jtho.2021.02.004.

8. Niu, Z.; Jin, R.; Zhang, Y.; Li, H. Signaling Pathways and Targeted Therapies in Lung Squamous Cell Carcinoma: Mechanisms and Clinical Trials. Signal Transduct. Target. Ther. 2022, 7 (1). https://doi.org/10.1038/s41392-022-01200-x.
9. Seshacharyulu, P., Ponnusamy, M. P., Haridas, D., Jain, M., Ganti, A. K., & Batra, S. K. (2012). Targeting the EGFR signalling pathway in cancer therapy. Expert Opinion on Therapeutic Targets, 16(1)
10. Citri, A.; Yarden, Y. EGF-ERBB Signalling: Towards the Systems Level. Nat. Rev. Mol. Cell Biol. 2006, 7 (7), 505–516. https://doi.org/10.1038/nrm1962
11. Abourehab, M. A. S., Alqahtani, A. M., Youssif, B. G. M., & Gouda, A. M. (2021). Globally approved EGFR inhibitors: Insights into their syntheses, target kinases, biological activities, receptor interactions, and metabolism. Molecules (Basel, Switzerland), 26(21), 6677.
12. Ayati, A.; Moghimi, S.; Salarinejad, S.; Safavi, M.; Pouramiri, B.; Foroumadi, A. A Review on Progression of Epidermal Growth Factor Receptor (EGFR) Inhibitors as an Efficient Approach in Cancer Targeted Therapy. Bioorg. Chem. 2020, 99 (March), 103811. https://doi.org/10.1016/j.bioorg.2020.103811.
13. Wee, P., & Wang, Z. (2017). Epidermal growth factor receptor cell proliferation signalling pathways. Cancers, 9(5), 52.
14. Yewale, C.; Baradia, D.; Vhora, I.; Patil, S.; Misra, A. Epidermal Growth Factor Receptor Targeting in Cancer: A Review of Trends and Strategies. Biomaterials 2013, 34 (34), 8690–8707. https://doi.org/10.1016/j.biomaterials.2013.07.100.
15. Wheeler, D. L., Iida, M., & Dunn, E. F. (2009). The role of SRC in solid tumors. The Oncologist, 14(7), 667–678. https://doi.org/10.1634/theoncologist.2009-0009
16. Zhang, S., & Yu, D. (2012). Targeting SRC family kinases in anti-cancer therapies: turning promise into triumph. Trends in Pharmacological Sciences, 33(3), 122–128.

17. Martellucci, S., Clementi, L., Sabetta, S., Mattei, V., Botta, L., & Angelucci, A. (2020). SRC Family Kinases as therapeutic targets in advanced solid tumors: What we have learned so far. Cancers, 12(6), 1448. https://doi.org/10.3390/cancers12061448
18. Belli, S., Esposito, D., Servetto, A., Pesapane, A., Formisano, L., & Bianco, R. (2020). C-SRC and EGFR inhibition in molecular cancer therapy: What else can we improve? Cancers, 12(6), 1489.
19. Biscardi, J. S., Maa, M. C., Tice, D. A., Cox, M. E., Leu, T. H., & Parsons, S. J. (1999). c-SRC-mediated phosphorylation of the epidermal growth factor receptor on Tyr845 and Tyr1101 is associated with modulation of receptor function. The Journal of Biological Chemistry, 274(12), 8335–8343.
20. Luttrell, D. K., Luttrell, L. M., & Parsons, S. J. (1988). Augmented mitogenic responsiveness to epidermal growth factor in murine fibroblasts that overexpress pp60c-SRC. Molecular and Cellular Biology, 8(1), 497–501.
21. Kloth, M. T., Laughlin, K. K., Biscardi, J. S., Boerner, J. L., Parsons, S. J., & Silva, C. M. (2003). STAT5b, a mediator of synergism between c-SRC and the epidermal growth factor receptor. The Journal of Biological Chemistry, 278(3), 1671–1679.
22. Dimri, M., Naramura, M., Duan, L., Chen, J., Ortega-Cava, C., Chen, G., Goswami, R., Fernandes, N., Gao, Q., Dimri, G. P., Band, V., & Band, H. (2007). Modeling breast cancer-associated c-SRC and EGFR overexpression in human MECs: c-SRC and EGFR cooperatively promote aberrant three-dimensional acinar structure and invasive behavior. Cancer Research, 67(9), 4164–4172.
23. Csizmadia, P. (1999). MarvinSketch and MarvinView: Molecule applets for the world wide web. Proceedings of the 3rd International Electronic Conference on Synthetic Organic Chemistry.

24. Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., Magariños, M. P., Mosquera, J. F., Mutowo, P., Nowotka, M., Gordillo-Marañón, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C. J., Segura-Cabrera, A., … Leach, A. R. (2019). ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Research, 47(D1), D930–D940. https://doi.org/10.1093/nar/gky1075

25. Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. Methods in Molecular Biology (Clifton, N.J.), 443, 365–382. https://doi.org/10.1007/978-1-59745-177-2_19

26. Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. Current Computer-Aided Drug Design, 7(2), 146–157. https://doi.org/10.2174/157340911795677602

27. Hamza, A., Wei, N.-N., & Zhan, C.-G. (2012). Ligand-based virtual screening approach using a new scoring function. Journal of Chemical Information and Modeling, 52(4), 963–974. https://doi.org/10.1021/ci200617d

28. Astalakshmi, D. .; T, G.; K B, G. S.; M, N.; M R, H. H. S.; S, G.; Latha, D. S. .; Kumar, D. M. S. Over View on Molecular Docking: A Powerful Approach for Structure Based Drug Discovery. Int. J. Pharm. Sci. Rev. Res. 2022, 77 (2), 146–157. https://doi.org/10.47583/ijpsrr.2022.v77i02.029

29. Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. Biophysical Reviews, 9(2), 91–102. https://doi.org/10.1007/s12551-016-0247-1

30. Dias, R., & de Azevedo, W. F., Jr. (2008). Molecular docking algorithms. Current Drug Targets, 9(12), 1040–1047. https://doi.org/10.2174/138945008786949432

31. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1.

Method and assessment of docking accuracy. Journal of Medicinal Chemistry, 47(7), 1739–1749. https://doi.org/10.1021/jm0306430

32. Flores-Holguín, N., Frau, J., & Glossman-Mitnik, D. (2021). Computational pharmacokinetics report, ADMET study and conceptual DFT-based estimation of the chemical reactivity properties of marine cyclopeptides. ChemistryOpen, 10(11), 1142–1149. https://doi.org/10.1002/open.202100178
33. Guan, L., Yang, H., Cai, Y., Sun, L., Di, P., Li, W., Liu, G., & Tang, Y. (2019). ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness. MedChemComm, 10(1), 148–157. https://doi.org/10.1039/c8md00472b
34. Srivastava, V., Yadav, A., & Sarkar, P. (2022). Molecular docking and ADMET study of bioactive compounds of Glycyrrhiza glabra against main protease of SARS-CoV2. Materials Today: Proceedings, 49, 2999–3007.

RE-2022-144826-plag-report

ORIGINALITY REPORT

18 %	12%	15%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.jbc.org Internet Source	1 %
2	academic.oup.com Internet Source	1 %
3	Submitted to University of Greenwich Student Paper	1 %
4	Xiang Wang, Yujia Cao, Siyu Chen, Jiachen Lin, Xin Yang, Dejian Huang. "Structure-Activity Relationship (SAR) of Flavones on Their Anti- Inflammatory Activity in Murine Macrophages in Culture through the NF-κB Pathway and c- Src Kinase Receptor", Journal of Agricultural and Food Chemistry, 2022 Publication	1 %