***KENYATTA UNIVERSITY***

***SCHOOL OF PHARMACY***

***IDENTIFICATION OF NOVEL EPIDERMAL GROWTH FACTOR RECEPTOR TYROSINE KINASE INHIBITOR CANDIDATES USING COMPUTER AIDED DRUG DESIGN***

***IAN NDEU MACHARIA***

***BACHELOR OF PHARMACY***

***P110S/17863/2016***

***A RESEARCH DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF BACHELOR OF PHARMACY, IN THE SCHOOL OF PHARMACY***

***APRIL 2022***

# **DECLARATION**

**Student:**

I declare that this research proposal is my original work and has not been presented in any other University.

Signed: ………………………………………. Date: ……………………

Ian N. Macharia

P110S/17863/2016

**Supervisor:**

This research proposal has been submitted for examination with my approval as the University supervisor.

Signed: ………………………………………… Date: ………………………

Dr. Evelyene Wanzala

Department of Chemistry and Industrial Pharmacy School of Pharmacy

# **ABSTRACT**

Computer aided drug design (CADD) is widely used in the pharmaceutical industry as a method of identifying potential novel drug molecules by performing virtual screening of compound libraries containing millions of drug-like molecules. CADD techniques can be classified as structure or ligand based where the former requires knowing the structure of the target macromolecule (protein, receptor, enzyme, nucleic acid) whereas the latter is used when there is no 3D macromolecule structure and relies on knowledge of compounds that bind to the target site on the macromolecule. The Epidermal Growth Factor Receptor (EGFR) belongs to the receptor tyrosine kinase family of receptors and plays a crucial role in cell proliferation thus aberrant mutations result in uncontrollable growth of cells leading to the development of cancer. EGFR inhibitors have served as a means of targeted therapy in cancers that are highly dependent on the EGFR signaling pathways such as Non-small Cell Lung Cancer (NSCLC), metastatic colorectal cancer, pancreatic cancer and breast cancer. The main aim of this project is to use structure-based pharmacophore modelling, virtual screening and molecular docking software (AutoDock Vina) to identify potential EGFR inhibitor candidates that may be used as lead compounds in drug synthesis and development.

TABLE OF CONTENTS

[DECLARATION 2](#_Toc101414191)

[ABSTRACT 3](#_Toc101414192)

[CHAPTER ONE 6](#_Toc101414193)

[INTRODUCTION 6](#_Toc101414194)

[1.1 Background 6](#_Toc101414195)

[1.2 Problem statement 8](#_Toc101414196)

[1.3 Rationale of the study 9](#_Toc101414197)

[1.4 Hypothesis 10](#_Toc101414198)

[1.5 Objectives 10](#_Toc101414199)

[**1.5.1** **General objective** 10](#_Toc101414200)

[**1.5.2** **Specific objective** 10](#_Toc101414201)

[1.6 Limitations of the study 11](#_Toc101414202)

[1.7 Delimitations of the study 11](#_Toc101414203)

[1.8 Assumptions of the study 11](#_Toc101414204)

[CHAPTER TWO 12](#_Toc101414205)

[LITERATURE REVIEW 12](#_Toc101414206)

[2.1. Introduction 12](#_Toc101414207)

[2.2. Epidermal growth factor receptors 12](#_Toc101414208)

[2.3. Epidermal growth factor receptor tyrosine kinase inhibitors 13](#_Toc101414209)

[2.4. Genetic mutations and mechanisms of resistance to EGFR 14](#_Toc101414210)

[2.4. Pharmacophore modeling 16](#_Toc101414211)

[2.5. Molecular Docking and Virtual Screening 17](#_Toc101414212)

[CHAPTER THREE 19](#_Toc101414213)

[MATERIALS AND METHODOLOGY 19](#_Toc101414214)

[3.1 Study design 19](#_Toc101414215)

[3.2 Study setting 19](#_Toc101414216)

[3.3 Input files and software used 20](#_Toc101414217)

[3.4 Equipment 20](#_Toc101414218)

[3.5 Methods 21](#_Toc101414219)

[**3.5.1** **Generation of Structure-Based Pharmacophore Model** 21](#_Toc101414220)

[**3.5.2** **Virtual Screening using Pharmacophore Model** 21](#_Toc101414221)

[**3.5.3** **Receptor and Ligand Preparation for docking** 21](#_Toc101414222)

[**3.5.4** **Molecular Docking Simulation** 22](#_Toc101414223)

[**3.5.5** **Pharmacokinetic analysis using Swiss ADME** 22](#_Toc101414224)

[**3.5.6** **Toxicity Testing using ProTox-II** 22](#_Toc101414225)

[**3.5.7** **Analysis of ligand-receptor interactions** 22](#_Toc101414226)

[**3.5.8** **Summary of the process** 23](#_Toc101414227)

[3.6 Data management & analysis 23](#_Toc101414228)

[CHAPTER FOUR 24](#_Toc101414229)

[RESULTS 24](#_Toc101414230)

[4.1 Structure-based pharmacophore modeling and virtual screening 24](#_Toc101414232)

[4.2 Molecular docking of chemical ligands using AutoDock Vina 25](#_Toc101414233)

[4.3 Pharmacokinetic analysis of ligands using Swiss ADME 26](#_Toc101414234)

[4.4 Toxicity testing using ProTox II 27](#_Toc101414235)

[4.5 Protein-ligand interaction analysis 28](#_Toc101414236)

[CHAPTER FIVE 30](#_Toc101414237)

[DISCUSSION AND CONCLUSION 30](#_Toc101414238)

[5.1 Introduction 30](#_Toc101414240)

[5.2 Discussion 30](#_Toc101414241)

[5.3 Conclusion 30](#_Toc101414242)

[5.4 Limitations 31](#_Toc101414243)

[5.5 Recommendations 31](#_Toc101414244)

[REFERENCES 31](#_Toc101414245)

**List of figures**

[*Figure 1 Pharmacophore features determined by Pharmit were used to perform virtual screening of the ZINC 15 database. Two hydrophobic groups, two hydrogen acceptors and a hydrogen donor were identified as interacting pharmacophore groups.* 21](file:////Users/admin/Desktop/untitled%20folder/RESEARCH%20RESULTS%20BY%20IAN%20MACHARIA.docx#_Toc101423041)

[*Figure 2: Pharmacophore features determined by Pharmit were used to perform virtual screening of the ZINC 15 database. Two hydrophobic groups, two hydrogen acceptors and a hydrogen donor were identified as interacting pharmacophore groups.* 21](#_Toc101423042)

[*Figure 3: Predicted physicochemical and pharmacokinetic properties of ZINC000009405320* 23](#_Toc101423043)

[*Figure 4: Predicted physicochemical and pharmacokinetic properties of ZINC000067291103* 23](#_Toc101423044)

[*Figure 5: Predicted physicochemical and pharmacokinetic properties ZINC000009405320* 23](#_Toc101423045)

[*Figure 6: Computational toxicity estimations of ZINC000064971506* 24](#_Toc101423046)

[*Figure 7: Computational toxicity estimations of ZINC000009405320* 24](#_Toc101423047)

[*Figure 8: Computational toxicity estimations of ZINC000009405320* 24](#_Toc101423048)

# **CHAPTER ONE**

# **INTRODUCTION**

## **Background**

Technological advancements made in the last decade have enabled great advances in the computational capabilities available for utilization in pharmacological research and this has had a great impact in computational chemistry as well. Computer aided drug design (CADD) stems from these developments and has been utilized significantly in the pharmaceutical industry in aiding drug development with techniques such as in virtual screening, protein-ligand docking simulations and molecular dynamics. CADD is also significantly more cost effective than traditional methods involved in drug development such as high throughput screening (HTS) since it requires minimal compound design or prior knowledge, making it an economical choice for many pharmaceutical industries and biotechnology companies alike.

The Epidermal Growth Factor Receptors (EGFR) are one of the most significant members of the receptor tyrosine kinase superfamily as they are involved in cellular proliferation. Mutations in EGFR result in occurrence of increased angiogenesis, uncontrollable growth and metastasis seen in malignant neoplasms. Tumors which rely on EGFR pathways are seen to be susceptible to EGFR inhibitors which can be classified either as anti-EGFR monoclonal antibodies or small molecule tyrosine kinase inhibitors (TKIs). Monoclonal antibodies function by selectively binding to the extracellular domain of EGFR thereby preventing other endogenous ligands (such as epidermal growth factor) from binding resulting in inhibition of downstream intracellular signal transduction and phosphorylation of intracellular substrates needed for cellular growth and differentiation. Examples of the most commonly utilized anti-EGFR monoclonal antibodies are Cetuximab and Necitumumab. Small molecule TKIs are relatively more cost effective and less complex to produce in comparison monoclonal antibodies. They function by binding to the intracellular tyrosine kinase domain of EGFR thus preventing phosphorylation of proteins involved in signal transduction. The first generation TKIs are reversible competitive inhibitors which include gefitinib and erlotinib. The major problem with these first-generation drugs is emergence of resistance after a median of one year due to point mutations, which renders them ineffective and unable to halt progression of disease. Second and third generation TKIs have been developed to address this concern although instances of acquired resistance have been reported. (Suda et al., 2009)

Currently, very few EGFR TKI inhibitors have passed clinical trials and are available in the market. Also, emergence of resistance occurring when using TKIs reduces the number of viable treatment options available thus there is need for new drugs to be developed and computer aided drug design techniques can be used as modern and cost-effective alternatives to aid drug development efforts. This study aims to use virtual screening of ligand libraries using molecular docking simulations to obtain suitable candidates that may have appreciable action as EGFR inhibitors to be used as leads in drug synthesis. Virtual screening involves searching large diverse compound databases to obtain novel structures that are likely to bind to a biological macromolecular target such as a receptor or enzyme. Many new compounds have been discovered using this technique and has proven to be indispensable in modern drug design efforts. Virtual screening is performed usually on pre-filtered databases using pharmacophore models to narrow down the chemical search space for more relevant search results (Sunseri & Koes, 2016). Chemical similarity searches may also be conducted to identify potential active ligands as well as 3D shape based virtual screening methods which serve to group compounds with similar topometry needed for binding to the biological receptor site. There are various ligand libraries available for virtual screening such as the NCBI PubChem, ZINC and eMolecules all of which are accessible online. Subsets of databases may be created from original larger libraries so as to reduce the computational effort required to perform the screening process. Ligand based virtual screening (LBVS) may require only milliseconds to perform structural comparisons. A computer with a single CPU may be sufficient enough to conduct LBVS screening session within hours whereas. Structural based virtual screening (SBVS) may require the use of parallel computing e.g. Linux clusters and batch processing to handle the workload.

Specific parameters and guidelines also may be applied during virtual screening of compound libraries to prune unwanted drug molecules. An example is the Lipinski’s rule of 5 which may be used to filter out compounds. ADMET (Absorption, distribution, metabolism, excretion and toxicity) analyses may also be virtually assessed by using freely available webservers such as the ProTox-II web server.

## **Problem statement**

Early phase drug development activities such as the hit to lead process and lead optimization using conventional methods are known to be quite expensive, require a great deal of training and are generally more time consuming. Computational efforts through a combination of structure-based pharmacophore models, virtual compound library screening and molecular docking simulations have been shown to yield considerable results in far less time and with less cost incurred in the process. There are currently less than 30 Food and Drug Administration (FDA) approved EGFR tyrosine kinase inhibitors and high rates of resistance developing especially in Non-Small Cell Lung Cancer (NSCLC) towards these drugs. Therefore, this research will focus on using computer aided drug design (CADD) methods to identify novel EGFR tyrosine kinase inhibitor candidates from compound libraries, which may be used as starting lead compounds in early phase drug development.

## **Rationale of the study**

EGFR TKIs have successfully been used as first line treatment in dealing with Non-Small Cell Lung Cancer (NSCLC) however, emergence of resistance to these drugs is usually inevitable during the course of treatment due to the T790M mutation which affects the ATP binding region of the tyrosine kinase domain of EGFR, rendering the drug ineffective in halting disease progression (Suda et al., 2009). Development of resistance, coupled with the few available FDA approved EGFR inhibitors, necessitates increased development of newer drug compounds to overcome such constraints. Third generation EGFR TKIs have been developed to avert the occurrence of resistance by covalently binding to the ATP binding domain of tyrosine kinase but currently, Osimertinib is the only FDA approved third generation EGFR TKI. The purpose of this study is to identify novel EGFR tyrosine kinase inhibitor candidates using computer aided drug design (CADD) techniques, which may be used as leads early phase drug development. The study is also of importance to other researchers and scientists seeking to develop EGFR TKIs using CADD and ultimately use them in biological assays for further investigation as lead compounds.

## **Hypothesis**

Structure based pharmacophore models, virtual screening and molecular docking simulations can be used to identify novel EGFR TKI inhibitor candidates.

These computer aided design methods are also considerably less expensive and time consuming compared to traditional ways of identifying lead compounds.

## **Objectives**

## **General objective**

To identify novel Epidermal Growth Factor Receptor tyrosine kinase inhibitor candidates which may be used as lead compounds in early phase drug development using structure-based pharmacophore models, virtual screening and molecular docking

kinase inhibitor using Pharmit which will be used as a filter in during virtual simulation.

### **Specific objective**

1. To create a structure-based pharmacophore model of Epidermal Growth Factor Receptor tyrosine kinase and perform virtual screening of compounds using Pharmit.
2. To perform molecular docking simulations with AutoDock Vina using compounds filtered from virtual screening as ligands to EGFR tyrosine kinase
3. To analyse the pharmacokinetic properties of the resultant lead compounds identified using SWISS ADME and ProTox-II.

## **Limitations of the study**

The computing resources utilized in this study are not powerful enough as may be desired thus the total number of compounds screened was limited to enable conclusion of the study within the time allocated. Furthermore, the molecular docking algorithms AutoDock Vina to performing docking simulations are stochastic (non-determninistic) and as such , multiple experiments will have to be conducted to achieve conclusive results.

## **Delimitations of the study**

The softwares used in conducting the study are freely available online thus do not incur any cost. Given the in-silico nature of the study, conducting the study is expected to take less time and have little financial constraints.

## **Assumptions of the study**

During molecular docking simulations, macromolecules such as receptors are assumed to have a rigid structure. This is done to save time and also save computational resources.

# **CHAPTER TWO**

# **LITERATURE REVIEW**

## 

## **2.1. Introduction**

This section serves to provide detailed knowledge regarding key concepts and areas that are covered during the course of the research project. Also included are accounts and results of studies performed by other researchers in their respective fields.

## 

## **2.2. Epidermal growth factor receptors**

The epidermal growth factor receptor (EGFR) family of tyrosine kinases is also referred to as the HER or ErbB family. They consist of four members namely HER1 (ErbB1), HER2 (ErbB2), HER3 (ErbB3) and HER4 (ErbB4). EGFR comprises of an extracellular binding domain containing 621 amino acids, a transmembrane domain with 23 amino acids and an intracellular/cytoplasmic domain made up of a protein tyrosine kinase domain with 542 amino acids and a C-terminal phosphorylation domain (Abourehab et al., 2021). Binding of endogenous ligands such as epidermal growth factor (EGF) ,transforming growth factor (TGF), amphiregulin or others which causes the receptor subunits to dimerize. In the process of dimerization, two members of the same EGFR family may undergo homodimerization while different members may also dimerize through heterodimerization, mostly with HER2. (Purba et al., 2017). Dimerization further leads to activation of the intrinsic tyrosine kinase activity, causing autophosphorylation of the intracellular tyrosine residues (found on the C-terminal domain of EGFR). (Gerber, 2022)

This results in activation of downstream signaling proteins which interact with the phosphorylated tyrosine residues, leading to generation of various signal transduction cascades. These cascades include the Ras-Raf-MEK (Mitogen-activated and extracellular-regulated kinase), P13K(Phosphatidylinositol-3-kinase)-Akt and STAT (Signal transducer and activator of transcription) pathways. These cascades eventually promote DNA synthesis and cellular proliferation.

## 

## **2.3. Epidermal growth factor receptor tyrosine kinase inhibitors**

The advent of targeted cancer therapy introduced a new class of drugs such as small molecule inhibitors which gave way to the discovery of novel and potent inhibitors. As of writing of this review, fourteen EGFR TKIs have been approved for treatment of various cancers. These drugs have proved to be highly efficacious but are now plagued by emergence of mutations which confer resistance to a fair number of these small molecule inhibitors. The first rationally designed kinase inhibitor was Imatinib which was approved in 2001 and since then, great attention has been directed towards the design of new kinase inhibitors. Since then, new EGFR TKIs have been approved for different types of cancers and can be divided into two groups. The first group consists of drugs approved by the FDA namely afatinib, brigatinib, dacomitinib, neratinib, Osimertinib, pyrotinib and vandetinib. The second group have been approved for treatment outside of the USA and include almonertinib, icotinib, simotinib and olmutinib. First generation EGFR TKIs are gefitinib, erlotinib, icotinib and lapatinib. They work by reversibly binding to tyrosine kinase domain of EGFR thereby outcompeting ATP for the binding pocket resulting in reduced activation of EGFR and cellular proliferation. A general property of first generation TKIs is that they have a basic quinazoline nucleus attached to an aniline group (Abourehab et al., 2021).

Second generation TKIs include afatinib, nefatinib and dacomitinib. The presence of a Michael acceptor site enables them to bind covalently to the receptor thus giving them and advantage against first generation inhibitors. Their chemical structure contains a quinazoline or quinoline nucleus with a crotonamide side chain (Michael acceptor) whereby its terminal carbon is substituted with a tert-amino group.

Third generation TKIs include Osimertinib, almonertinib and olmutinib. Chemically, they consist of a pyrimidine nucleus substituted with an aniline or phenoxy moiety. These moieties bear and acrylamide group which form a covalent bond with cysteine at amino acid position 797 on EGFR. A number of fourth generation inhibitors are currently undergoing clinical trials such as BBT-176 (Phase II/III trials for NSCLC) but are yet to receive any approval for clinical use. (Abourehab et al., 2021)

## 

## **2.4. Genetic mutations and mechanisms of resistance to EGFR**

The epidermal growth factor receptor is involved in a number of crucial cellular processes that are key to the survival of cells such as proliferation and differentiation thus unwanted activation of this receptor has been implicated in a variety of cancers.

Unfortunately, emergence of resistance to first and second generation TKIs is usually inevitable in most patients after 10 – 12 months on average (Abourehab et al., 2021). There exists a fair amount of mechanisms that evade the regulation of EGFR signaling which include enhanced production of ligands, EGFR protein overproduction, mutations that lead to auto-activation of the receptor, deficiency in EGFR downregulation or activation by heterologous regulation from other cell surface receptors (Purba et al., 2017). Mutations in the extracellular domain may lead to activation of the receptor without the binding of any ligand. Some EGFR mutants have missing extracellular domains which enable them to form the activated dimeric form of EGFR leading to auto-phosphorylation. However this active configuration differs from that of the wild EGFR as only a limited number of tyrosine residues are phosphorylated. This may cause defective downstream signaling seen in tumor cells. (Purba et al., 2017)

Exon 19 deletion (del 19) and Exon 21 substitution (L858R) account for about 85% of all EGFR mutations. The exon 19 mutation is seen to be more prevalent accounting for 45% of mutations. The exon 21 mutation involves substitution of leucine at position 858 with arginine. These mutations are referred to as activating mutations as they may lead to activation of receptors without the need for ligand binding. Others may lead to partial activation, requiring a secondary mutation for full activation. Mutations within exon 18, 19 and 21 are associated with increased sensitivity to EGFR TKIs such as erlotinib and gefitinib (Xu et al., 2020).

A secondary point mutation also occurs where threonine at position 790 is replaced by a methionine (T790M) produces a drug-resistant variant of the receptor and is found in more than half of patients with NSCLC. The T790M mutation increases the affinity of the receptor to adenosine triphosphate (ATP) relative to its affinity to tyrosine kinase inhibitors (TKI). This mutation has also been shown to harbor growth advantages to cells that express it thereby giving it a double role in the survival of lung cancer cells (Suda et al., 2009). Third generation inhibitors work differently compared to their first- and second-generation predecessors by binding covalently to the cysteine residue found at amino acid position 797. Additionally, the loss of this mutation has been noted as an early indicator of resistance to third generation TKIs as shown by Oxnard et al in their study (Leonetti et al., 2019).

Resistance to third generation EGFR TKIs has been noted to emerge from a tertiary C797S mutation which occurs at exon 20 and accounts for about 10 – 26% of resistance to Osimertinib treatment. This mutation leads to substitution of cysteine at amino acid position 797 for serine. As a consequence of this substitution, covalent bonding between the cysteine residue on EGFR and Osimertinib fails to take place. Consequently, the C797S mutation also leads to emergence of resistance to other third generation TKIs (Leonetti et al., 2019)

## 

## **2.4. Pharmacophore modeling**

Paul Erlich developed the concept of pharmacophores in the 1800’s where he believed the biological action of drugs was due to certain chemical groups present and molecules that elicited similar effects had the similar chemical groups. The IUPAC defines a pharmacophore to be an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target to trigger (or block) its biological response. (Voet et al., 2022). The notion of pharmacophores came about through observations whereby variation of certain chemical moieties lead to considerable changes in the activity of the compound whereas variations in of other parts of the compound resulted in only minor changes to activity. Pharmacophore elements are defined as an atom or collection of atoms common for active compounds relative to a receptor, and is essential to the activity of the compound. A pharmacophore model therefore consists of an ensemble of pharmacophore elements. When the spatial positions of these elements are elucidated, a 3D-pharmacophore model can be constructed. Pharmacophores have long been used in medicinal chemistry albeit limited to a topological context (2D). The emergence of computer aided drug design techniques introduced topographical (3D) derivations of the pharmacophore concept. Pharmacophores models can generally be divided into two categories based on how they are obtained. Structure based pharmacophore models are made by analyzing possible interactions between ligands and the target. These types of pharmacophore models are comprised of features such as hydrogen bonding, electrostatic charges, presence of aromatic groups, hydrophobic contact between ligand and target atoms. (Opo et al., 2021). A major challenge of structure-based pharmacophore models is the complexity that may arise due to numerous interaction points between the ligand and the target. Ligand based pharmacophore models are derived solely from structures and binding data of ligands without considering the three-dimensional structure of the target. This omission may lead to generation of more false positives as it disregards key important information about the binding site and ligand-target interactions.

## 

## **2.5. Molecular Docking and Virtual Screening**

Molecular docking is a computer aided drug design technique which is used to predict the binding modes of a ligand and a protein of known 3D structure. The availability of high-speed computers clusters of computers has allowed for larger virtual docking simulations to take place on entire compound libraries. (Cosconati et al., 2010). Molecular docking has been used in various applications such as structure-activity relationship studies, lead optimization and discovery of novel molecules through virtual screening. Computational docking comprises of two components, the first being a search method which is used to predict conformations by searching the available conformational space and a scoring function which is used to estimate the free binding energy of the ligand-protein complex and rank potential candidate molecules based on the free binding energy values (Morris & Lim-Wilby, 2008). There are different types of search methods used in molecular docking such as simulated annealing, genetic algorithms and local search. These methods are referred to as stochastic or non-deterministic meaning their results may slightly vary every time thus multiple docking simulations must be conducted to validate the results. Systematic/ deterministic search methods also exist which generate all possible ligand binding conformations by evaluating the degrees of freedom of the ligand but are usually more computationally expensive.

Scoring functions used in molecular docking are used to assess ligand-protein binding poses and give an estimate of the stability and favorability of each binding pose. (Cosconati et al., 2010). They also rank ligands based on the estimated binding affinity to the target protein whereby a lower and more negative value of free binding energy represents a more stable docking pose hence more favorable binding. Several scoring functions exist which can generally be classified as force field based, knowledge based, empirical based and machine learning based functions. The first three are referred to as classical scoring functions. Machine learning based scoring functions have been found to outperform the classical scoring functions and massive efforts are being taken to develop better scoring functions of these kind.

Virtual screening involves assessing the binding of compounds from a large database against a target protein of macromolecule to select the most favorable compounds to be used in further investigations as lead compounds in early drug design. Virtual screening can be performed by using pharmacophore models to query large ligand databases e.g. ZINC 15 database or by performing molecular docking on candidate ligands to select the those with the most favorable binding

# **CHAPTER THREE**

# **MATERIALS AND METHODOLOGY**

## **Study design**

The purpose of this study was to identify novel epidermal growth factor tyrosine kinase inhibitor candidates using computer aided drug design. This entailed generation of structure-based pharmacophore model using the crystal structure of Osimertinib in complex with EGFR (L858R/T790M/C797S) (PDB code 6LUD). This model was used to query a ligand database (ZINC 15) to identify compounds that fulfill parameters specified by the pharmacophore model. These ligands then underwent preparation to improve their suitability and compatibility with the docking software (AutoDock Vina). The EGFR- Osimertinib crystallized complex also underwent preparation using AutoDock Tools to improve their suitability in the docking simulation. Molecular docking was performed and the top three ranked ligands with the highest estimated binding affinity were subjected to further virtual pharmacokinetic and toxicity testing after conversion of the ranked ligands from PDBQT to the SMILES format.

## **Study setting**

The study was performed at my residence in South B, Nairobi Kenya on my personal computer with an internet connection. The operating system used was Mac OS Catalina 10.15.7 and the terminal program to be used is Iterm2. The shell program to be used is GNU bash, version 3.2.57(1)-release (x86\_64-apple-darwin19). Owing to the unavailability of Biovia Discovery Studio on Mac OS, a Microsoft Windows 10 installation was used to run the software on the same computer using Bootcamp which was meant to allow switching between the two operating systems when needed.

## **Input files and software used**

The study used a 3D representation of an EGFR-Osimertinib complex obtained using X-ray crystallography (PDB ID: 6LUD). The complex was available for download from the RCSB PDB (https://www.rcsb.org) databases. Ligand SDF files were generated and downloaded from Pharmit after performing a virtual screening of the ZINC 15 databases. Open Babel software was utilized in conversion of ligand files from the SDF file format to the PDBQT format which is compatible for molecular docking with AutoDock Vina.

Iterm2 terminal emulator running a GNU bash shell will be used to run the molecular docking simulations. AutoDock Vina will be used to perform the molecular docking simulation and generate results for the experiment. Biovia Discovery Studio will be used to analyze interactions between the ligands and receptor and also generate figures to visually display these interactions.

SwissADME and ProTox-II were used to perform virtual pharmacokinetic and toxicity tests on the top-ranking ligands obtained from molecular docking.

## **Equipment**

The study was performed on a MacBook Pro mid-2012. The processor used was a 2.5GHz dual-core Intel Core i5 processor (Turbo Boost up to 3.1GHz) with 3MB L3 cache. The RAM used will be 10GB of 1600MHz DDR3 memory. The storage disk of the device was 500GB 5400-rpm hard drive. The GPU used by the device was an Intel HD Graphics 4000

## **Methods**

### **Generation of Structure-Based Pharmacophore Model**

Pharmit (<https://pharmit.csb.pitt.edu/>) was used to identify the pharmacophore features that interact with the EGFR tyrosine kinase domain which consisted of two hydrogen acceptors, two hydrophobic groups and a hydrogen bond. The pharmacophore search was also configured to exclude compound conformations whose heavy atoms would intersect with the area occupied by the receptor (exclusive shape query). Other parameters set included selection of compounds with a molecular weight greater than 180 and less than or equal to 500, compounds with less than or equal to 10 rotational bonds, compounds with a LogP value of less than or equal to 5, compounds with less than or equal to 10 hydrogen bond acceptors and compounds with less than or equal to 5 hydrogen bond donors.

### **Virtual Screening using Pharmacophore Model**

The model generated was then used to perform a virtual screen of the ZINC 15 database which contained 13,190,370 molecules at the time of writing this report.

The initial search resulted in 6989 hit compounds. This number was then narrowed down by minimizing the energy of the hit compounds. Results with energy values greater than -8 kcal/mol were filtered out as unfavorable. Also, only one conformer per compound was to be considered to be a hit. The final result of the pharmacophore based virtual screening produced 101 compounds which were downloaded in the SDF file format to be used later in molecular docking. The screening session was then saved as a JSON file (named as pharmit.json) to allow repeating the process in the future with the same parameters.

### **Receptor and Ligand Preparation for docking**

Open babel (<https://openbabel.org/wiki/Main_Page>) was used to convert virtual screening hit results from the SDF format to the PDBQT format for compatibility with AutoDock Vina docking software. The hit ligand files were also renamed to match their ZINC 15 database ID using a simple GNU bash script (saved as extract\_ligands.sh) to save time as opposed to manually renaming each of the 101 files. Receptor preparation was performed using AutoDock Tools and entailed removal of heteroatoms which would interfere with the docking procedure such as water, solvent molecules and the Osimertinib ligand. Separation of Osimertinib from the tyrosine kinase domain was done to free the active site for binding of other candidate ligands during docking. Polar hydrogens as well as Kollman charges were added to the receptor since they contribute binding of ligands. The atom type representation format of the receptor was set to AD4 for compatibility with AutoDock. The receptor file was then be saved as a PDBQT file for compatibility with AutoDock Vina.

### **Molecular Docking Simulation**

A 3D grid box was placed around the receptor binding site previously occupied by Osimertinib using AutoDock Tools which specified the available space and coordinates for docking the receptor. A file containing the name of the receptor file (6lud.pdbqt) as well as values of the x, y and z dimensions (40 by 50 by 50 Angstroms) of the grid box was created (saved as conf.txt).

Molecular docking was conducted using AutoDock Vina software on the prepared receptor to assess the binding affinity of each ligand. Ranking of ligands was done according to their estimated free binding energy which translates to their respective binding affinity. Results from each ligand docking simulation was be compiled using a GNU bash script (named virtual\_screening.sh) to generate the final results in a single text file (named results.txt) found in separate folders. A total of four docking sessions were conducted and the result of each session stored separately in files denoted as results\_01, results\_02, results\_03 and results\_04

### **Pharmacokinetic analysis using Swiss ADME**

The top 3 ranked ligands obtained from molecular docking were subjected to virtual pharmacokinetic testing using Swiss ADME (<http://www.swissadme.ch>). The molecules were converted from the PDBQT format to the SMILES format using Open Babel then submitted to the Swiss ADME server.

### **Toxicity Testing using ProTox-II**

The top 5 ranked ligands obtained from molecular docking will be subjected to virtual toxicity testing using ProTox-II (<https://tox-new.charite.de/protox_II>). The molecules were converted from the PDBQT format to the SMILES format using Open Babel then submitted to the ProTox-II server.

### **Analysis of ligand-receptor interactions**

Further analysis of the interactions between the ligands and receptor was done using Biovia Discovery Studio which was used to generate figures to display these interactions.

### **Summary of the process**

## **Data management & analysis**

DUDE-Z EGFR benchmark database to will be used to validate the structure-based pharmacophore model generated by Pharmit. Re-docking of Osimetinib ligand to EGFR will be performed to obtain the estimated binding affinity to be compared against those of the candidate ligands after being docked. The files required for the simulation to occur and results will be stored both on the local hard drive of the computer and on a Git repository publicly accessible at <https://github.com/matteratomic/Virtual-docking-with-AutoDock.git>.

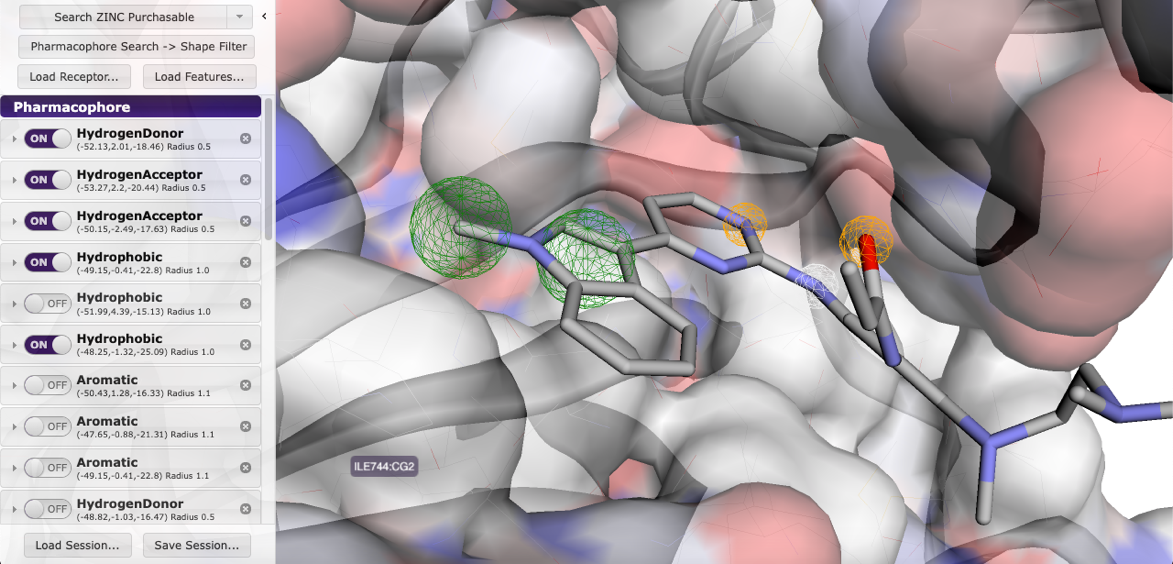
Results will be summarized as text files, charts, visual figures and tables showing the generated figures, free binding energy, molecular formula, chemical name and ZINC 15 database ID.

# **CHAPTER FOUR**

# **RESULTS**

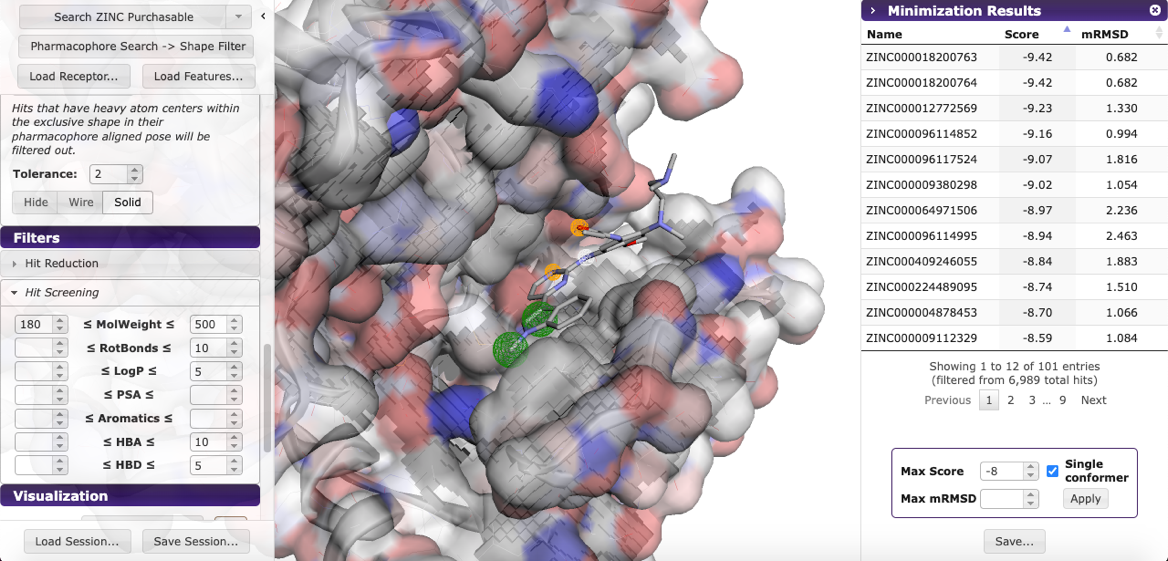


## **Structure-based pharmacophore modeling and virtual screening**



*Figure 1 Pharmacophore features determined by Pharmit were used to perform virtual screening of the ZINC 15 database. Two hydrophobic groups, two hydrogen acceptors and a hydrogen donor were identified as interacting pharmacophore groups.*

The pharmacophore model used in virtual screening of the compounds was generated using Pharmit (<https://pharmit.csb.pitt.edu/>)



*Figure 2: Pharmacophore features determined by Pharmit were used to perform virtual screening of the ZINC 15 database. Two hydrophobic groups, two hydrogen acceptors and a hydrogen donor were identified as interacting pharmacophore groups.*

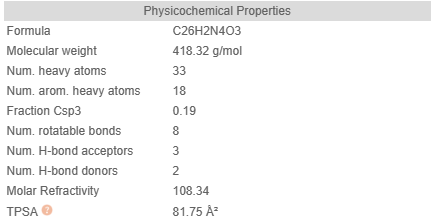
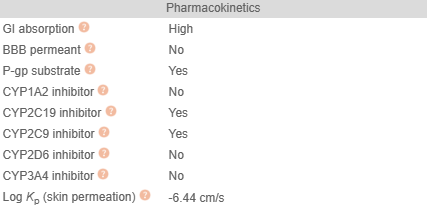
## **Molecular docking of chemical ligands using AutoDock Vina**

Table 4.1 Results of molecular docking using AutoDock Vina

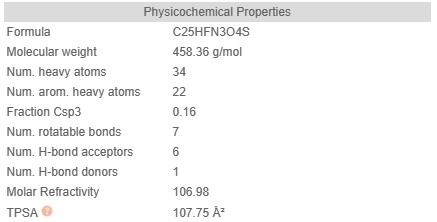
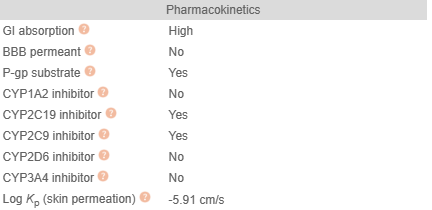
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Molecule ZINC ID | Chemical  Name | Molecular  Formula | Chemical  Structure | Binding Affinity |
| ZINC000064971506 | [4-[2-(o-tolylmethylamino)-2-oxo-ethyl]-3-oxo-N-(p-tolyl)-2H-quinoxaline-1-carboxamide](https://zinc12.docking.org/synonym/4-%5B2-%28o-tolylmethylamino%29-2-oxo-ethyl%5D-3-oxo-N-%28p-tolyl%29-2H-quinoxaline-1-carboxamide) | [C26H26N4O3](https://pubchem.ncbi.nlm.nih.gov/#query=C26H26N4O3) |  | -8.67 |
| ZINC000009405320 | [2-[3-(benzo[1,3]dioxol-5-ylmethyl)-4-oxo-pteridin-2-yl]sulfanyl-N-(3,4-dimethylphenyl)-acetamide](https://zinc12.docking.org/synonym/2-%5B3-%28benzo%5B1%2C3%5Ddioxol-5-ylmethyl%29-4-oxo-pteridin-2-yl%5Dsulfanyl-N-%283%2C4-dimethylphenyl%29-acetamide) | [C24H21N5O4S](https://pubchem.ncbi.nlm.nih.gov/#query=C24H21N5O4S) | C:\Users\LINK 1\Desktop\rank_2.png | -8.38 |
| ZINC000067291103 | 2-{[3-(1,3-benzodioxol-5-ylmethyl)-4-oxo-3,4-dihydin-2-yl]sulfanyl}-N-(3-fluoro-4-methylphenyl)acetamide | [[C25H20FN3O4S](https://pubchem.ncbi.nlm.nih.gov/#query=C25H20FN3O4S)](https://pubchem.ncbi.nlm.nih.gov/#query=C25H20FN3O4S) | C:\Users\LINK 1\Desktop\rank_3.png | -8.37 |
| ZINC000002970981 | [N-(3-chloro-4-methyl-phenyl)-2-[(4-keto-3-piperonyl-quinazolin-2-yl)thio]acetamide](https://zinc12.docking.org/synonym/N-%283-chloro-4-methyl-phenyl%29-2-%5B%284-keto-3-piperonyl-quinazolin-2-yl%29thio%5Dacetamide) | [C25H20ClN3O4S](https://pubchem.ncbi.nlm.nih.gov/#query=C25H20ClN3O4S) | C:\Users\LINK 1\Desktop\rank_4.png | -8.35 |
| ZINC000113940 | [3-(pyridin-2-ylthio)propanoic acid](https://zinc12.docking.org/synonym/3-(pyridin-2-ylthio)propanoic%20acid) | [C8H9NO2S](https://pubchem.ncbi.nlm.nih.gov/#query=C8H9NO2S) | C:\Users\LINK 1\Desktop\rank_5.png | -8.08 |

Molecular docking was performed using AutoDock Vina. The docking function used by AutoDock Vina was non-deterministic thus multiple docking simulations needed to be carried out to obtain compounds with higher average binding affinities. Docking of all 101 compounds in a single docking simulation took 1 hour 17 minutes on average.

## **Pharmacokinetic analysis of ligands using Swiss ADME**

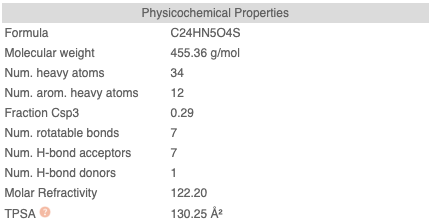
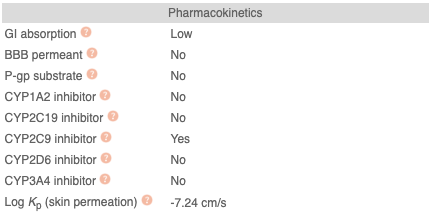


*Figure 3: Predicted physicochemical and pharmacokinetic properties of ZINC000009405320*



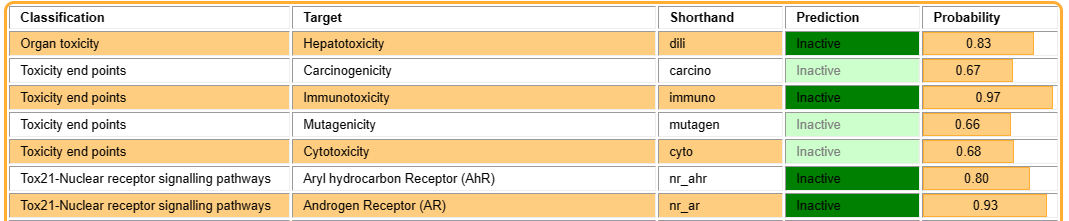
*Figure 4: Predicted physicochemical and pharmacokinetic properties of ZINC000067291103*

SWISS ADME was used to predict the physicochemical and pharmacokinetic properties of the top 3 ranked compounds with the highest estimated binding affinity.

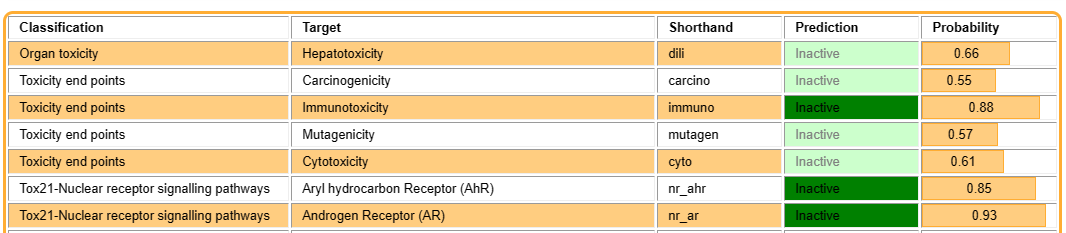


*Figure 5: Predicted physicochemical and pharmacokinetic properties ZINC000009405320*

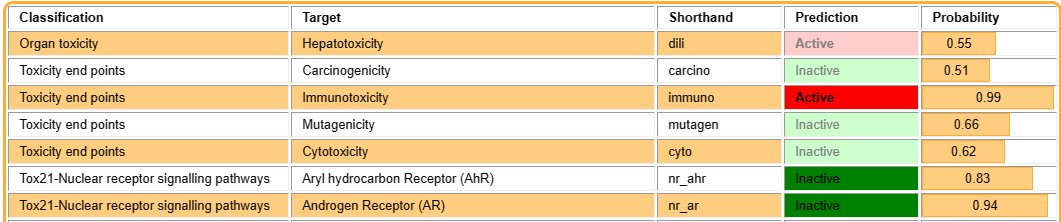
## **Toxicity testing using ProTox II**



*Figure 6: Computational toxicity estimations of ZINC000064971506*



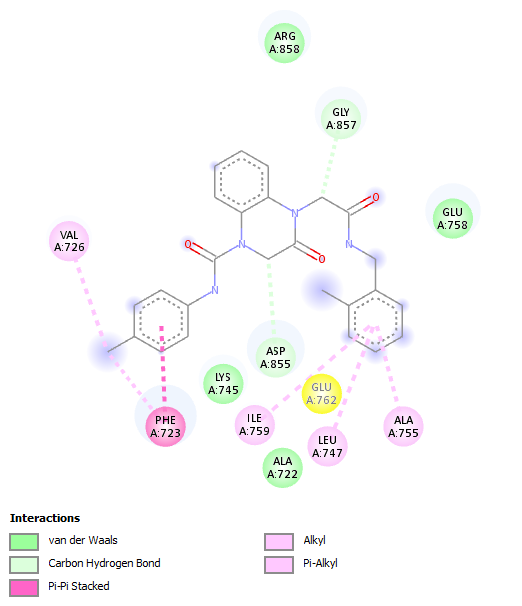
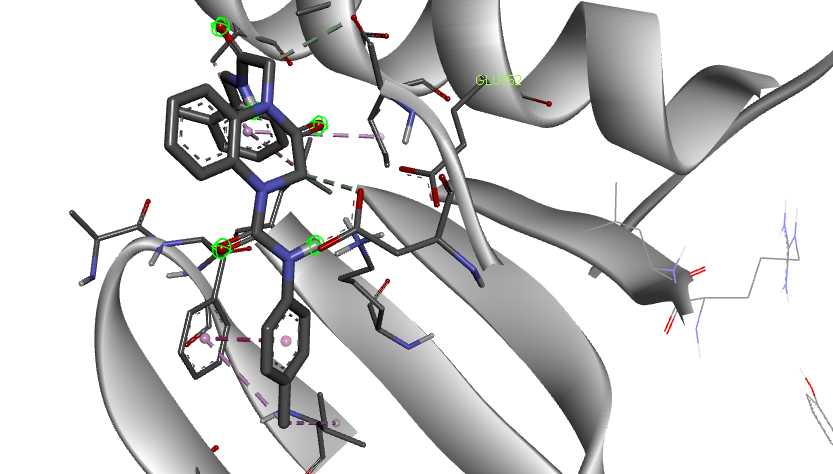
*Figure 7: Computational toxicity estimations of ZINC000009405320*



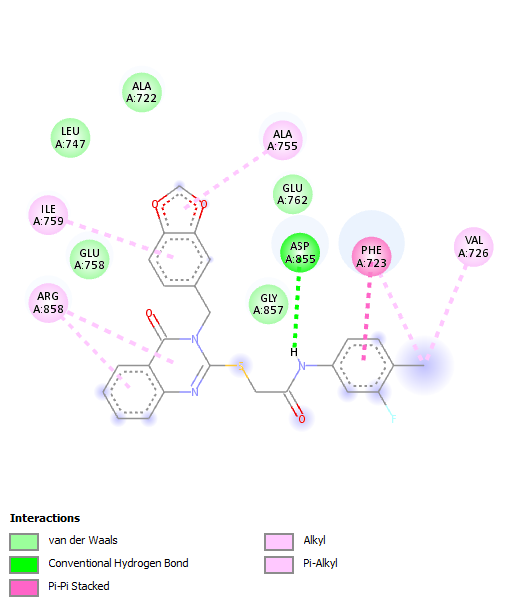
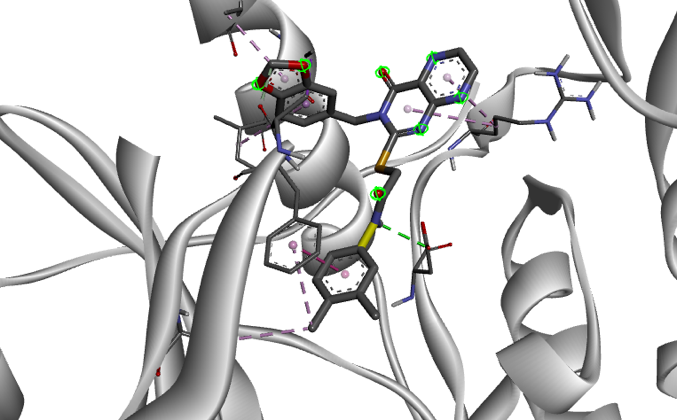
*Figure 8: Computational toxicity estimations of ZINC000009405320*

ProTox II was used to predict the physicochemical and pharmacokinetic properties of the top 3 ranked compounds with the highest estimated binding affinity.

## **Protein-ligand interaction analysis**

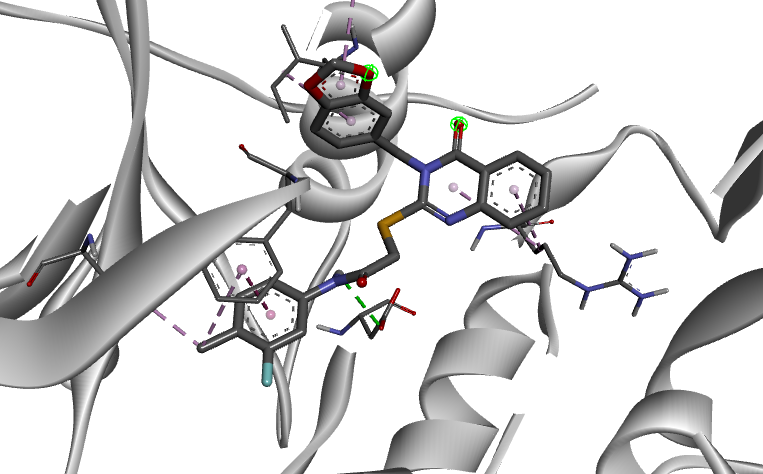
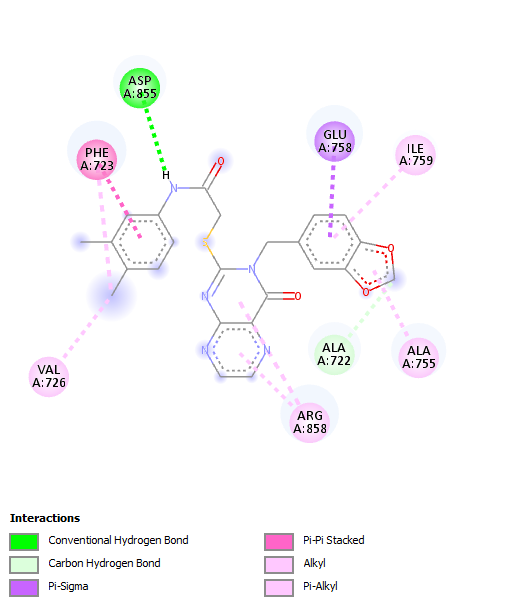


*Figure 8: Binding site interactions of* [*4-[2-(o-tolylmethylamino)-2-oxo-ethyl]-3-oxo-N-(p-tolyl)-2H-quinoxaline-1-carboxamide*](https://zinc12.docking.org/synonym/4-%5B2-%28o-tolylmethylamino%29-2-oxo-ethyl%5D-3-oxo-N-%28p-tolyl%29-2H-quinoxaline-1-carboxamide) *(ZINC Database ID: ZINC000064971506) with tyrosine kinase domain of EGFR.*



*Figure 9: Computational toxicity estimations of 2-[3-(benzo[1,3]dioxol-5-ylmethyl)-4-oxo-pteridin-2-yl]sulfanyl-N-(3,4-dimethylphenyl)-acetamide*

*(ZINC000009405320) with tyrosine kinase domain of EGFR*



*Figure 10: Computational toxicity estimations of 2-[3-(benzo[1,3]dioxol-5-ylmethyl)-4-oxo-pteridin-2-yl]sulfanyl-N-(3,4-dimethylphenyl)-acetamide*

*(ZINC000067291103) with tyrosine kinase domain of EGFR.*

# **CHAPTER FIVE**

# **DISCUSSION AND CONCLUSION**



## **Introduction**

This chapter gives a discussion and conclusion, limitations and recommendations based on the study conducted to identify novel epidermal growth factor receptor tyrosine kinase inhibitors

## **Discussion**

## **Conclusion**

The study was novel epidermal growth factor receptor candidates with less financial and temporal costs as opposed to the traditional able to show that computational drug design techniques could be used to identify methods of lead compound generation. The identified compounds also demonstrated a higher estimated binding affinity compared to Osimertinib (-6.8 kcal/mol), which is widely used as the first line drug in treating Non-Small Cell Lung Cancer with EGFR T790M mutations. Although it is important to note that the computational methods used are not proposed as an absolute substitute to wet laboratory testing since further studies may be required to reinforce or disprove viability of the selected novel ligands and also for the determination of biological activity.

## **Limitations**

The computational capability of the machine used to perform molecular docking was limited thus certain restrictions had to be made to achieve completion of the simulations at reasonable times. The restrictions involved reducing the number of compounds to be screened and also reducing the number of generated conformations which were used for docking.

## **Recommendations**

1. Due to the technical nature of the study, future researchers are encouraged to learn basic shell scripting and fundamentals of computer programming to automate tasks such as docking of multiple ligands and data aggregation. However, there are many programs that exist that automatically execute these tasks thus this recommendation is for better comprehension of the methods discussed, for those who wish to advance their knowledge in computational biology or bioinformatics and for the satisfaction of curiosity.

# **REFERENCES**

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, 26(21), 6677. https://doi.org/10.3390/molecules26216677

Cosconati, S., Forli, S., Perryman, A. L., Harris, R., Goodsell, D. S., & Olson, A. J. (2010). Virtual Screening with AutoDock: Theory and Practice. Expert Opinion on Drug Discovery, 5(6), 597–607. https://doi.org/10.1517/17460441.2010.484460

Leonetti, A., Sharma, S., Minari, R., Perego, P., Giovannetti, E., & Tiseo, M. (2019). Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. British Journal of Cancer, 121(9), 725–737. https://doi.org/10.1038/s41416-019-0573-8

Morris, G. M., & Lim-Wilby, M. (2008). Molecular Docking. In A. Kukol (Ed.), Molecular Modeling of Proteins (Vol. 443, pp. 365–382). Humana Press. https://doi.org/10.1007/978-1-59745-177-2\_19

Opo, F. A. D. M., Rahman, M. M., Ahammad, F., Ahmed, I., Bhuiyan, M. A., & Asiri, A. M. (2021). Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. Scientific Reports, 11(1), 4049. https://doi.org/10.1038/s41598-021-83626-x

Purba, E. R., Saita, E., & Maruyama, I. N. (2017). Activation of the EGF Receptor by Ligand Binding and Oncogenic Mutations: The “Rotation Model.” Cells, 6(2), 13. https://doi.org/10.3390/cells6020013

Solassol, I., Pinguet, F., & Quantin, X. (2019). FDA- and EMA-Approved Tyrosine Kinase Inhibitors in Advanced EGFR-Mutated Non-Small Cell Lung Cancer: Safety, Tolerability, Plasma Concentration Monitoring, and Management. Biomolecules, 9(11), 668. https://doi.org/10.3390/biom9110668

Suda, K., Onozato, R., Yatabe, Y., & Mitsudomi, T. (2009). EGFR T790M mutation: A double role in lung cancer cell survival? Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer, 4(1), 1–4. https://doi.org/10.1097/JTO.0b013e3181913c9f

Sunseri, J., & Koes, D. R. (2016). Pharmit: Interactive exploration of chemical space. Nucleic Acids Research, 44(W1), W442–W448. https://doi.org/10.1093/nar/gkw287

Xu, C., Lei, L., Wang, W., Lin, L., Zhu, Y., Wang, H., Miao, L., Wang, L., Zhuang, W., Fang, M., Lv, T., & Song, Y. (2020). Molecular Characteristics and Clinical Outcomes of EGFR Exon 19 C-Helix Deletion in Non–Small Cell Lung Cancer and Response to EGFR TKIs. Translational Oncology, 13(9), 100791. https://doi.org/10.1016/j.tranon.2020.100791