Identification of Novel Plant-Derived Inhibitors of the EGFR Kinase Domain using vHTS, QSAR and Molecular Docking Approaches

Abstract

The epidermal growth factor receptor (EGFR) protein tyrosine kinase (PTK) is a crucial target in the pursuit of anti-tumor drug discovery. This study investigates 305 phytochemicals from five known anticancer plants (*Anacardium occidentale*, *Annona muricata*, *Spondias mombin*, *Ocimum gratisimum*, and *Zingiber officinale*) for their potential as EGFR kinase domain inhibitors. Through Virtual High Throughput Screening (vHTS), lead compounds were identified and subjected to ADMET filtering. A Quantitative Structure-Activity Relationship (QSAR) model was developed using bioassay data from the ChEMBL database, exhibiting strong statistical robustness and external validation. Molecular docking studies revealed interactions of lead compounds with critical residues within the EGFR ATP kinase domain. Actinidine, berberine, and corydaline demonstrated adherence to Lipinski's rule of five, indicating drug-likeness. Notably, actinidine forms hydrophobic interactions with Phe-856, while berberine establishes hydrogen bonds with Asp-855. Corydaline engages in extensive hydrophobic and hydrogen bond interactions within the ATP pocket of the EGFR kinase domain. These findings underscore the potential of actinidine, berberine, and corydaline as EGFR kinase domain inhibitors, supported by a robust QSAR model, marking progress in the search for novel anticancer agents targeting EGFR inhibition.

Keywords: EGFR Kinase Domain; Anti-tumour; vHTS; QSAR

1.0. Introduction

The epidermal growth factor receptor (EGFR; also known as erbB1) is a member of the tyrosine kinase receptor family, which includes other members like HER2/neu (erbB2), erbB3, and erbB4 (Metibemu et al., 2019). Upon binding with ligands such as epidermal growth factor (EGF) and transforming growth factor-alpha (TGF-a), EGFR initiates a multitude of intracellular signal transduction pathways that regulate critical aspects of tumor cell behavior, including growth, proliferation, survival, metastasis, and angiogenesis (Rodriguez *et al*., 2023) Ligand binding leads to the formation of homo- or hetero-dimeric complexes, subsequently activating the tyrosine kinase domain (Paul and Hristova, 2019). This activation results in the phosphorylation and activation of various intracellular proteins, ultimately modulating gene transcription (Roux and Topisirovic, 2018).

The targeting of EGFR at various stages of cancer development is a promising strategy, with one approach focusing on the inhibition of the receptor's tyrosine kinase (RTK) domain (Metibemu et al., 2019). Receptor tyrosine kinase inhibitors, typically small molecules, compete with ATP for the intracellular orthosteric site of EGFR. An extensive array of cancer therapies developed thus far leverages plant-derived products, as plants have demonstrated significant anti-cancer properties (Metibemu et al., 2020a; Metibemu et al., 2020b; Akinloye et al., 2021; Metibemu et al., 2021). These plants serve as a valuable resource, potentially yielding new drugs due to their reservoir of natural chemicals with chemoprotective potential against cancer (Taneja and Qazi, 2007).

Despite the advances in EGFR-targeted cancer therapies, there remains a critical need for the development of innovative treatment strategies that effectively inhibit the EGFR tyrosine kinase domain (Halder *et al.,* 2023). This is particularly crucial given the role of EGFR in governing

various aspects of tumor development and progression. To address this gap, this study explores the potential of phytochemicals sourced from five widely recognized medicinal and antitumor plants: Anacardium occidentale, Annona muricata, Spondias mombin, Ocimum gratisimum, and Zingiber officinale (Cheng et al., 2021). By conducting a comprehensive analysis of these phytochemicals and their interactions with the EGFR kinase domain, this study aimed to contribute to the development of new and effective EGFR kinase domain inhibitors for potential integration into cancer treatment strategies. The research seeks to bridge the existing gap in knowledge regarding the utility of plant-derived compounds as targeted therapies for EGFR-associated malignancies.

2.0. Methodology

2.1. Data collection and preparation

Phytochemicals characterised from *Anacardium occidentale*, *Annona muricata*, *Spondias mombim*, *Oscimum gratisimum*, and *Zingiber officinale* were obtained from literature (Gajalakshmi et al., 2012). A total of three hundred and five (305) phytochemicals were downloaded in SDF (Structural data format) from the PubChem database (https://pubchem.ncbi.nlm.nih.gov). The SDF were converted to PDB format by Open Babel, and finally converted to PDBQT using ligprep command lines. The EGFR4 oncoprotein structure with PDB ID: 3BEL and crystallographic resolution of 2.30Aº was downloaded from the protein data bank (http://www.rcsb.org).

2.2. Virtual High Throughput Screening and molecular docking

Virtual High Throughput Screening, a computational screening technique was used to screen a pool of compounds library to probe the binding affinity of the target receptor with the library compounds (Sousa et al., 2024). The downloaded EGFR4 from the protein data bank (http://www.rcsb.org), was uploaded in Pymol, the grid coordinate was set as in the co-crystallized compound, X= 16.1 Y= 34.65 Z= 91.68. The phytochemicals were converted to PDB and PDBQT, using command lines. The protein-ligand docking was carried out using Autodock Vina (Le et al., 2024). The phytochemicals were docked into 3BEL catalytic site as occupied by the co-crystalized ligand (Rudrapal et al., 2022).

2.3. Validation of docking results

Validation of docking result was performed by alignments of EGFR4 kinase domain receptor sequences from the Pubmed repository against the ChemBL Database (www.ebiac.uk/chembl/) The eighty-five (85) EGFR4 (PDB ID: 3BEL) kinase inhibitors (compounds) obtained were downloaded in text format and converted to PDB format by Data Warrior version 2, and finally converted to PDBQT. The compounds obtained were docked into the kinase domain (X= 16.1 Y= 34.65 Z= 91.68). A correlation coefficient analysis of the relationship between the docking scores of the compounds and their corresponding pIC_{50} values was determined. The docking scores and the vHTS were also validated by determined the root mean square deviation (RMSD) of the co-crystallized compound within the catalytic domain. The co-crystallized compound was re-docked into the catalytic domain of 3BEL and RMSD was evaluated.

2.4. ADMET filtering

According to Lipinski *et al* (1997) the 'rule of five' depicts bioavailability of drugs. When there are more than 5 H-bond donors, 10 H-bond acceptors, molecular weight (MWT) greater than 500 and the calculated Log P (CLogP) greater than 5 (or MlogP>4.15), Poor absorption or permeation is more likely. The rule of five also describes molecular properties that are vital to the pharmacokinetics of drugs; these include the absorption, distribution, metabolism and excretion of compound (Ivanović et al., 2020). The Lipinski rule of five was used to filter our lead compounds. The Mavin Viewer software was used to establish the conformity of the lead compounds to the rule of five. The number of rotatable bonds and polar surface area, which are known to differentiate compounds that are orally active from those that are not (Caminero Gomes Soares et al., 2023). Compounds with 10 or fewer rotatable bonds and polar surface area equal or less than 140Å² have good oral bioavailability (Da Rocha et al., 2022).

2.5. Quantitative Structure Activity Relationship (QSAR).

2.5.1. Data collection and descriptor calculation

The bioassay IC_{50} data for EGFR4 kinase domain was downloaded from the chEMBL database (http://ebi.ac.uk), in excel format and converted to SDF using DataWarrior. The SDF structures were catenated and converted to 3-Dimensional structures using command lines. The Chemistry Development Kit (CDK) 1.4.6 was used to calculate the molecular descriptors (Willighagen,et al., 2017).

2.5.2. Data pre-treatment

The pre-treatment of the bioassay IC_{50} data from the chEMBL database (http://ebi.ac.uk) was

carried out with V-WSP algorithm **(**Metibemu, 2022) to remove co-linearity of descriptors.

2.5.3. Data set division: Training and test Sets

A dataset of 100 EGFR4 kinase domain inhibitors was obtained from the chEMBL database (http://ebi.ac.uk). The data set was split into training (70%) and test (30%) dataset with the Kennard Stone algorithm (Dataset Division GUI 1.2) (Olasupo et al., 2019).

2.5.4 Genetic algorithm and multiple linear regression

Genetic algorithm was used to perform the selection of significant variables (descriptors) (Hossain, 2022). The training set was used for the generation of the QSAR model. Multiple linear regression (MLR) was used for the generation of unbiased model.

3.0. Results and Discussion

3.1. Virtual High Throughput Screening

The docking score of the co-crystallized, POX, (4-amino-6- {[1-(3-fluorobenzyl)-1-Hindazol-5-yl] amino} pyrimidine-5-carbaldehyde O- (2-methoxyethyl) oxime) (National center for biotechnology information, 2024) against the EGFR4 kinase domain, 9.7 kcal/mol was used as the cut off for the selection of lead phytochemicals. Seven (7) phytochemicals with docking scores equal or greater than -9.7 kcal/mol were the lead phytochemicals (Table 1).

Table 1: The lead phytochemicals from the selected plants

Annona muricata	Docking Scores
Isoannonacin	-9.7
Spondias mombin	Docking Scores

Lupeol	-9.8
Anacardium occidentale	Docking Scores
Actinidine	-11.5
Chlorogenic Acid	-10.2
Corydaline	-10.1
Berberine	-9.9
Occimum gratissimum	Docking Scores
Rosmarinic acid	-9.7

3.2.

Validation of docking Scores

Analysis of the relationship between pIC_{50} and the corresponding docking scores of 85 compounds downloaded from Chembl (www.ebiac.uk/chembl/) gave a strong correlation of .577, significant at $p<0.01$ indicating a strong positive correlation (Table S1). This confirms that the Autodock algorithm used for Virtual Screening is reliable for predicting binding affinity of compounds and can be used to replicate wet experimental data. The redocked almost fit perfectly with the co-crystallized compound, and with RMSD of 0.09 Å. This further validated the reliability of the docking scores and vHTS.

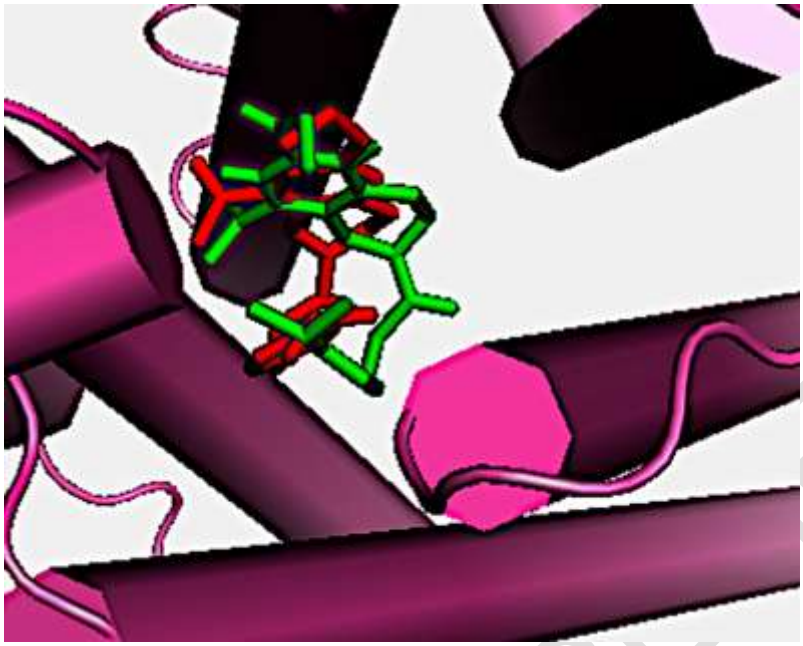


Figure 1: The redocked (green) and the co-crystallized compound (red) within the catalytic domain of 3BEL with RMSD of 0.09 Å

3.3. ADME and drug-likeness screening of lead compounds

ADME (Absorption, Distribution, Metabolism and Excretion) and drug-likeness of compounds can be determined using the Lipinski's rule of five (Lipinski *et al*., 1997). Using the docking score of POX, the co-crystalized inhibitor of the protein (-9.7 kcal/mol) as cut off for the selection of leads, only seven of all the phytochemicals docked have docking scores equal or greater than -9.7kcal/mol. The ADME determination was carried out with Marvin suite.

Table 2 shows the result of screening of the lead compounds using the Lipinski's rule, drugs with good oral bioavailability should violate not more than one of the rules of five. Chlorogenic acid, Lupeol, and Rosmarinic violated just one of the rules, however, Isoannonacin violated three of the rules of five hence it is filtered out. Actinidine, berberine, corydaline fit perfectly into the rule of five.

Table 2: Molecular properties of the lead compounds with respect to the Lipinski's rule of five.

Lead Compounds	Docking Scores	H.B.A (<=10)	H.B.D (<=5)	R.B (<=10)	XLog P (<=5)	M.M (<500)	PSA (<140 $Å^2$)
Actinidine	-11.5	1	0	0	2.4	147.221	12.89
Chlorogenic Acid*	-10.2	9	6	5	-0.4	354.311	164.75
Corydaline	-10.1	5	0	4	3.6	369.461	40.16
Berberine	-9.9	4	0	2	3.6	336.124	40.8
Lupeol*	-9.8	1	1	1	9.9	426.729	20.23
Isoannonacin***	-9.7	7	3	26	8.1	596.89	113.29
Rosmarinic Acid*	-9.7	8	5	7	2.4	360.318	144.52

Note: H.B.A: number of Hydrogen bond acceptors

H.B.D: number of Hydrogen bond donors

R.B: number of Rotatable bonds

XLogP: Octanol-water partition coefficient

M.W: Molecular weight

P.S.A: Polar surface area

- Indicates the number of violations of the Lipinski's rule of five

3.4. QSAR analysis

A QSAR predict the relationship that exists between the structure of compounds and biological

activity of a molecular system, geometric and chemical characteristics. Seventy EGFR kinase domain inhibitors obtained from ChemBL Database (www.ebiac.uk/chembl/) were used as the training set. Multiple linear regression and Genetic algorithm were used for the analysis.

3.4.1. Model summary

Table 3: Summary of the model table: The R, R squared, Adjusted R squared and the Durbin-Watson constant

R	R Square	Adjusted R Square
.988	.976	.941

Predictors: (Constant), XLogP, khs. sCH3, MOMI-XZ, WTPT-4, VC-5, Weta1.unity, PPSA-2, Lipinski Failures, Wnu2.unity, WT.unity

Dependent Variable: pIC_{50} (pIC_{50} = -log IC_{50})

Adjusted R square was calculated using the Stein's formula (Stein, 1972):

Adjusted R^2 = 1- $\left[\left(\frac{n-1}{n-k-1}\right)\left(\frac{n-2}{n-k-2}\right)\left(\frac{n+1}{n}\right)\right](1-R^2)$

Where:

R^2 = Measurement of the variability in the pIC_{50} accounted for by the descriptors in the model

n = Number of compounds in the training set

k = Number of descriptors in the model

The summary of the model is shown in Table 3, with R-value, Pearson correlation of .988. The model shows a very high positive Pearson Correlation. The R square value of .976 indicates the model can account for more than 97% of pIC_{50}. The adjusted R^2 is concerned with how well the model generalizes, that is, external validation of the model. The adjusted R^2 of .941 value depicts the external cross validation of the model is very good with a negligible 3.5% shrinkage in predicting external pIC_{50} and the model could be used for predicting pIC_{50} of potential EGFR kinase domain inhibitors with about 94% accuracy. A correlation plot (Figure 2) of the observed pIC50 versus the predicted pIC50 gives an R^2 value of 0.976, depicting a very strong corelation between the observed pIC50 and predicted pIC50. This shows the model can accurately predict pIC50 values (Figure 2) (Table S1). The closeness of the experimental and predicted pIC50 values of some of the training set in Table S1 further gives credence to the robustness of the model.

3.4.3. Generation of QSAR Model Equation

The equation of a straight line is given as

$$Y = MX + C \quad \text{i}$$

The equation of regression line is given as:

$$Y = B + B1X1! + B2X2! + B3X3! + \ldots + BnXn! \quad \text{ii}$$

$$pIC_{50} = Bo + B1X1! + B2X2! + B3X3! + \ldots BnXn! \quad \text{iii}$$

The equation of the model is given as:

Predicted pIC_{50} = 5.309 + (-0.535 * XLog) + (0.222 * MOMI-XZ) + (8.154 * VC-5) + (-0.34 * WTPT-4) + (-0.024 * Khs.sCH3) + (10.161 * Weta1.unity) + (-0.005 * WT.unity)

+ (4.994 * Wnu2.unity) + (0.001 * PPSA-2) + (-0.428 * Lipinski Failure) **iv**

$Y = pIC_{50}$ = dependent variable

b_0 is the constant

$b_1$is the regression coefficient

x_1 is the independent variable

Where,

WTPT-4: Sum of path lengths starting from oxygens

khs.sCH3 : Isothermal compressibility of the reference hard-sphere mixture

Weta1.unity: Directional WHIM, weighted by unit weights

WT.unity: Non-directional WHIM, weighted by unit weights

Wnu2.unity: Directional WHIM, weighted by unit weights

PPSA-2: Partial positive surface area * total positive charge on the molecule

LipinskiFailures: Number failures of the Lipinski's Rule Of 5

XLogP: Octanol-water partition Coefficient

MOMI-XZ: Moment of inertia along X/Z-axis

VC-5: Valence cluster, order

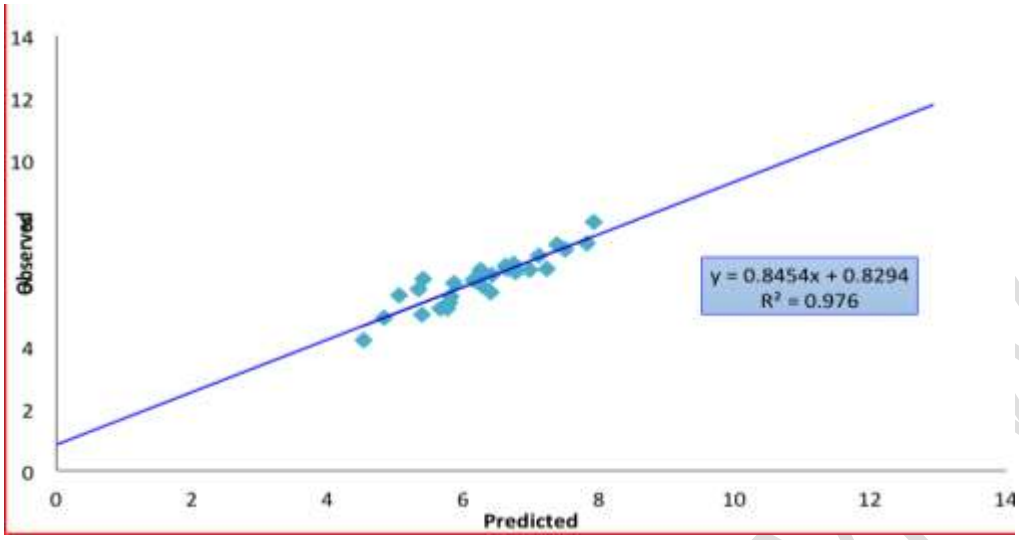


Figure 2: Scattered Plot of the observed pIC_{50} values against the predicted pIC_{50} Values of the training set

3.4.4. Contribution of the descriptors to the model

Figure 3 shows the contribution of the descriptors to the QSAR model. The Weta1.unity descriptor contributes the most to the model, followed by MOMI-XZ, VC-5, Wnu2.unity, PPSA-2 and WT.unity respectively.

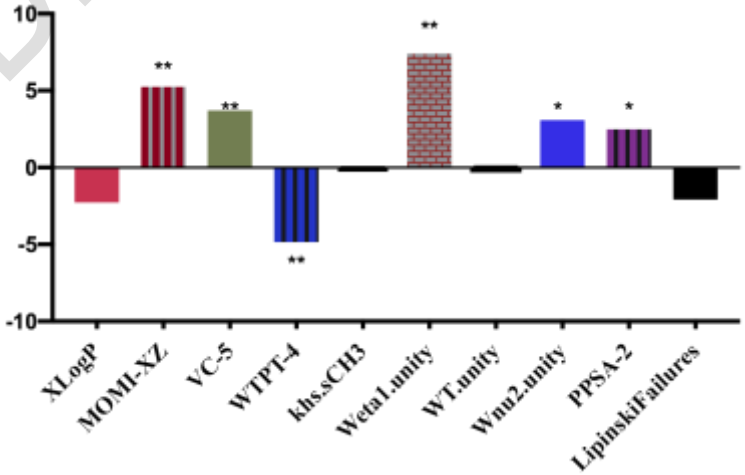
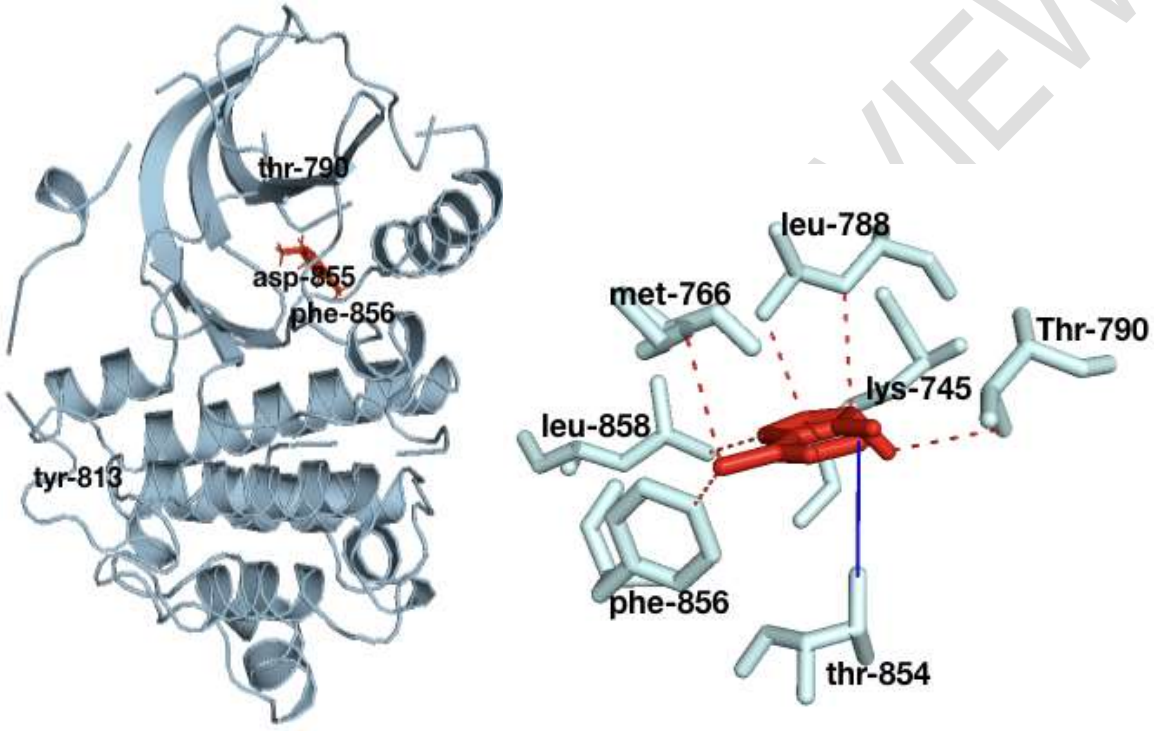


Figure 3: Contributions of the Descriptors to the QSAR Model

* Significant contribution at P<0.05 (95% confidence Interval)

** Significant contribution at P<0.01 (99% confidence Interval)

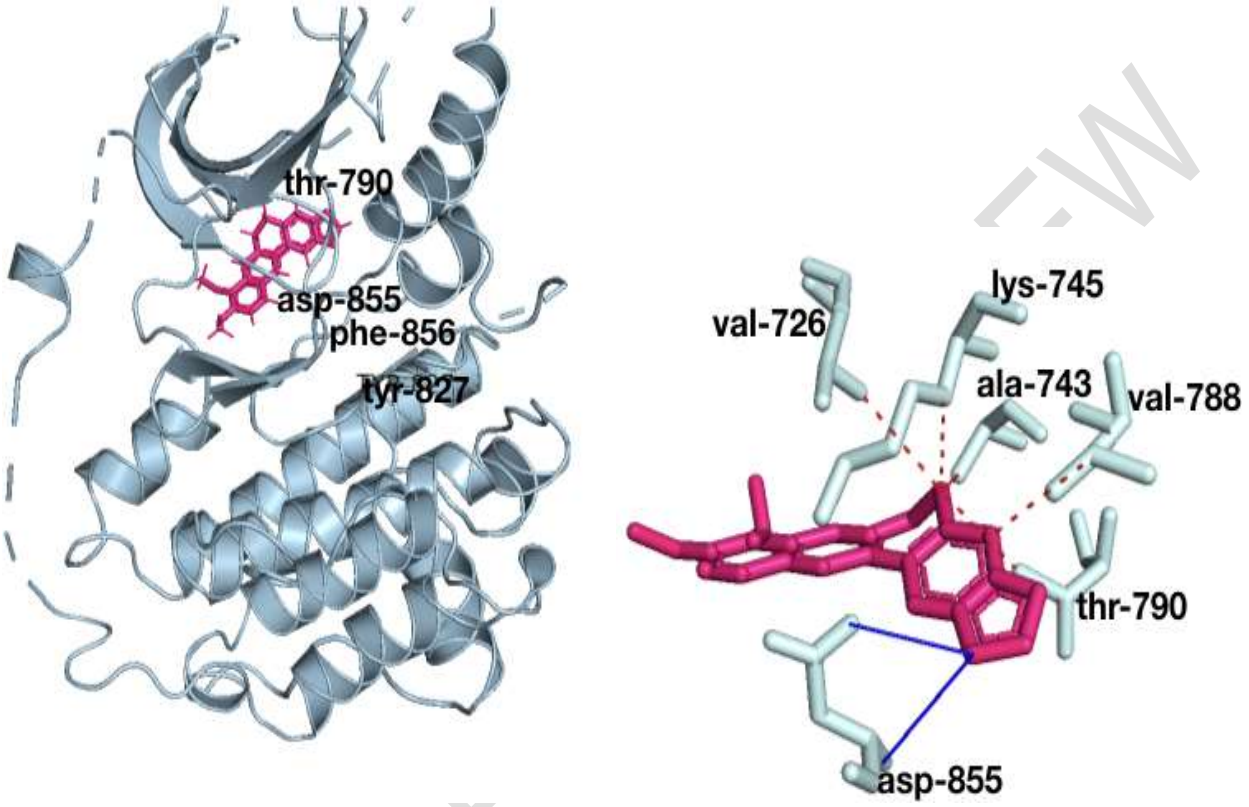
3.5. Molecular interactions of the lead compounds



A **B**

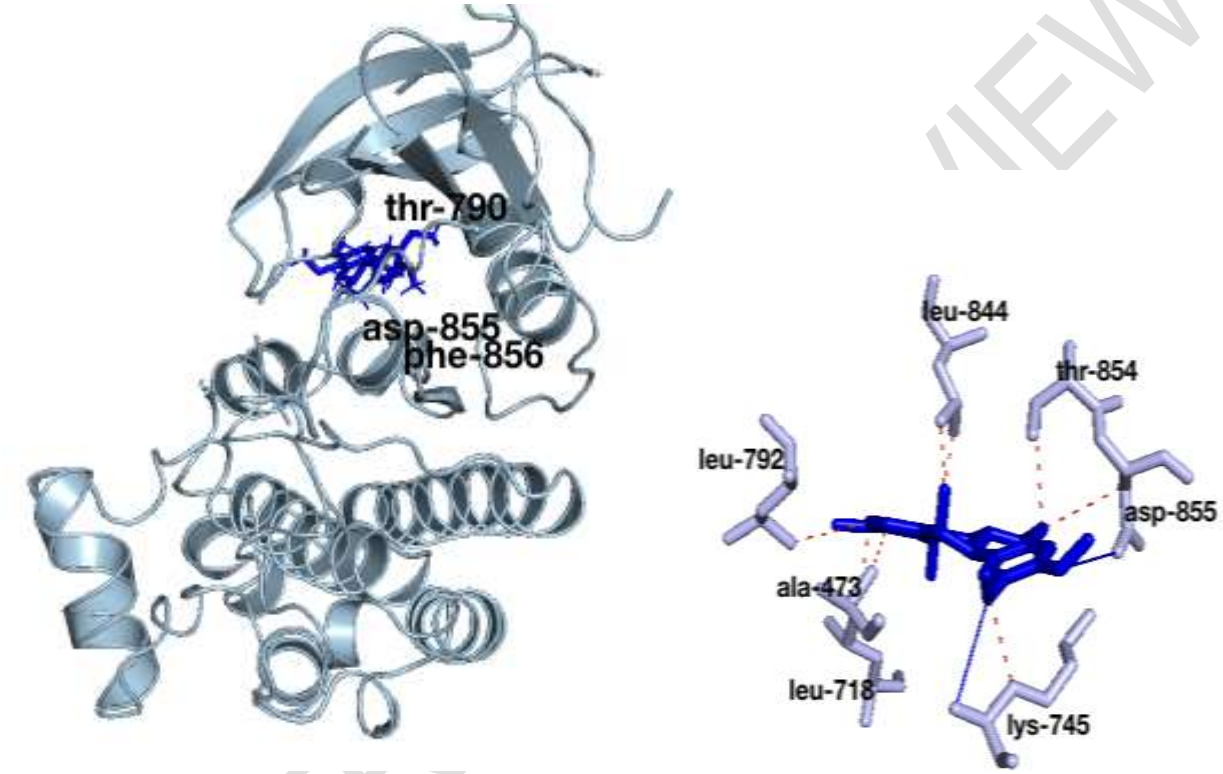
Figure 4: A). Crystal structure of the EGFR kinase domain in complex with Actinidine, (red), the inhibitor (actinidine) occupies the cleft between the N- and C-lobes of the kinase domain. The gatekeeper residue, thr-315, two residues of the DFG motif, asp-855 and phe-856 are revealed. Actinidine binds to the active state of the kinase with the activation loop in an "open" conformation and the DFG motif in the "out" conformation. B). Showing the specific

interactions of actinidine (red) within the ATP binding site, the blue lines represent hydrogen bonds while the dotted red lines represent hydrophobic interactions.



A **B**

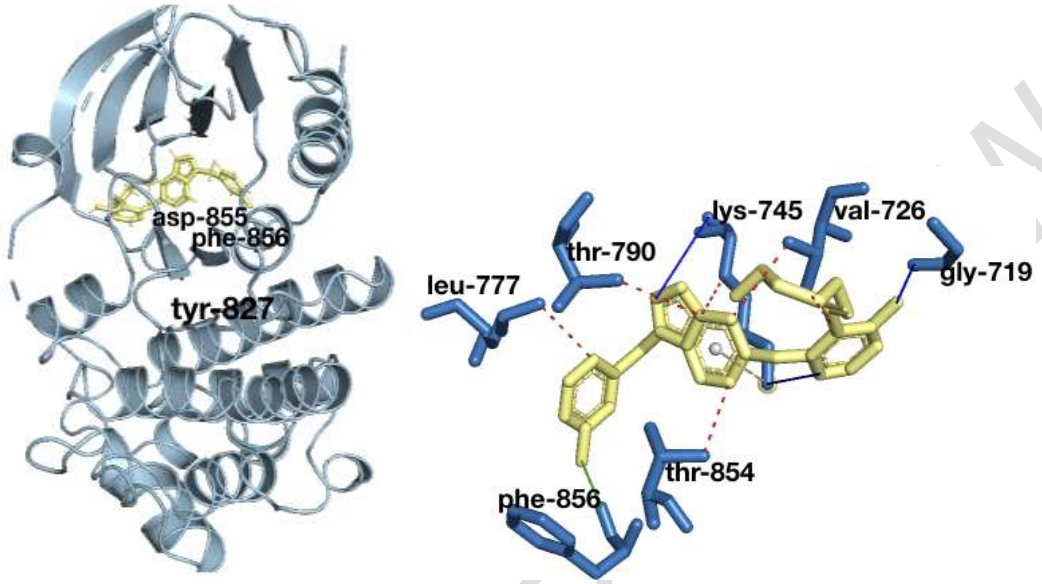
Figure 5: A). Crystal structure of the EGFR kinase domain in complex with Berberine (hot red), it occupies the cleft between the N- and C-lobes of the kinase domain. The gatekeeper residue, thr-790, two residues of the DFG motif, asp-855 and phe-856 and the activation loop tyr-827 are shown. Berberine binds to the active state of the kinase with the activation loop in an "open" conformation and the DFG motif in the "out" conformation. B). Showing the specific interactions of berberine (hot pink) within the ATP binding site, the blue lines represent hydrogen bonds while the dotted red lines represent hydrophobic interactions.



A B

Figure 6: A). Crystal structure of the EGFR kinase domain in complex with Corydaline (blue), it occupies the cleft between the N- and C-lobes of the kinase domain. The gatekeeper residue, thr-790, two residues of the DFG motif, asp-855 and phe-856 and the activation loop tyr-827 are shown. Corydaline binds to the active state of the kinase with the activation loop in an "open" conformation and the DFG motif in the "out" conformation. B). Showing the specific

interactions of corydaline (blue) within the ATP binding site, the blue lines represent hydrogen bonds while the dotted red lines represent hydrophobic interactions.



A. **B.**

Figure 7: A). Crystal structure of the EGFR kinase domain in complex with co-crystallized compound, 4-amino-6- {[1-(3-fluorobenzyl)-1-Hindazol-5-yl] amino} pyrimidine-5-carbaldehyde O- (2-methoxyethyl) oxime (pale yellow), it occupies the cleft between the N- and C-lobes of the kinase domain. The two residues of the DFG motif, asp-855 and phe-856 and the activation loop tyr-827 are shown. The co-crystallized binds to the active state of the kinase with the activation loop in an "open" conformation and the DFG motif in the "out" conformation. B). Showing the specific interactions of the co-crystallized (pale yellow) within the ATP binding site, the blue lines represent hydrogen bonds, the dotted lines represent hydrophobic interactions, the green line represents halogen interaction while the wheat yellow colour represents pication

Actinidine binds in between the N- and C- lobes (Figure 4A). Actinidine binds to the active state of the kinase with the activation loop in an "open" conformation and the DFG motif in the "out" conformation Figure 4A. Actinidine forms hydrogen bond interaction with thr-854, it also forms hydrophobic interactions with lys-745, met-766, leu-788 thr-790, phe-856 and leu-858, (Figure 4B). The formation of hydrophobic interaction with phe -856 of the DFG-out conformation is critical to the inhibitory potential of actinidine. This helps to prevent D and the F of the DFG motif from swapping positions (Vijayan *et al*., 2015). It is noteworthy that actinidine forms hydrophobic interaction with the gatekeeper residue thr-315. This hydrophobic bond not only strengthens binding potency, but also increases kinase specificity (Azam *et al*., 2008).

Berberine also binds in between the N- and C- lobes (Figure 5A). Berberine binds to the active state of the kinase with the DFG motif in the "out" conformation Figure 5A. It forms two hydrogen bonds with asp-855. This prevent the D and the F of the DFG motif from swapping positions and this in turn give credence to the inhibitory potential of berberine. It also forms hydrophobic interactions with val-726, ala-743, lys-745, leu-788 and with the gatekeeper residue thr-790. The hydrophobic interaction of berberine with the gatekeeper residue not only strengthens binding potency, but also increases kinase specificity (Azam *et al*., 2008).

Corydaline equally binds in between the N- and C- lobes (Figure 5A). It binds to the active state of the kinase with the DFG motif in the "out" conformation Figure 6A. The inhibitory potential of Corydaline on the EGFR kinase domain is enhanced by the formation of both hydrophobic and hydrogen bonds interactions with asp-855, thereby preventing the D and the F of the DFG motif from swapping positions (Figure 6B). It also forms hydrophobic interactions with leu-718, ala-743, lys-745, leu-792, leu-844, thr-854 and one additional hydrogen bond interaction with

lys-745 (Figure 6B). The extensive hydrophobic cum hydrogen bonds interactions of corydaline with important residues within the ATP pocket of the EGFR kinase domain depicts it as a good inhibitor. These lead compounds on the other hand share common interactions with the co-crystallized (val-726, lys-745, and thr-790) (Figure 7A and B).

Conclusion

Actinidine, berberine, and corydaline fit perfectly into the rule of five. They are potential EGFR kinase domain inhibitors. The QSAR model generated in the present study is statistically robust and thoroughly validated. The formation of hydrophobic interaction with phe -856 of the DFG-out conformation is critical to the inhibitory potential of actinidine. Berberine forms two hydrogen bonds with asp-855. Corydaline forms extensive hydrophobic and hydrogen bond interactions with important residues within the ATP pocket of the EGFR kinase domain. The QSAR model herein can reliably predict potential EGFR kinase domain inhibitors. Hence these procedures can help in the prediction of anti-tumour compounds.

Consent

Not applicable

Ethics approval

In this study, no method requiring the permission of the Ethics Committee was used

Availability of Data and Materials

All data generated or analysed during this study are included in this published article

Abbreviations:

EGFR: Epidermal Growth Factor Receptor

PTK: Protein Tyrosine Kinase

vHTS: Virtual High Throughput Screening

QSAR: Quantitative Structure-Activity Relationship

ATP: Adenosine Triphosphate

HER2: Human Epidermal Growth Factor Receptor 2

EGF: Epidermal Growth Factor

PDB: Protein Data Bank

SDF: Structure-Data File

MWT: Molecular Weight

CDK: Cyclin-Dependent Kinase

RMSD: Root-Mean-Square Deviation

ADME: Absorption, Distribution, Metabolism, Excretion

H.B.A: number of Hydrogen bond acceptors

H.B.D: number of Hydrogen bond donors

R.B: number of Rotatable bonds

XLogP: Octanol-water partition coefficient

M.W: Molecular weight

P.S.A: Polar surface area

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

References

Akinloye, O. A., Akinloye, D. I., Lawal, M. A., Shittu, M. T., & Metibemu, D. S. (2021). Terpenoids from Azadirachta indica are potent inhibitors of Akt: Validation of the anticancer potentials in hepatocellular carcinoma in male Wistar rats. *Journal of Food Biochemistry*, *45*(1), e13559. http://dx.doi.org/10.1111/jfbc.13559

Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J., & Daley, G. Q. (2008). Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nature structural & molecular biology*, *15*(10), 1109. https://doi.org/10.1038/nsmb.1486

Caminero Gomes Soares, A., Marques Sousa, G. H., Calil, R. L., & Goulart Trossini, G. H. (2023). Absorption matters: A closer look at popular oral bioavailability rules for drug approvals. *Molecular informatics*, *42*(11), e202300115. https://doi.org/10.1002/minf.202300115

Cheng, W. L., Feng, P. H., Lee, K. Y., Chen, K. Y., Sun, W. L., Van Hiep, N., ... & Wu, S. M. (2021). The role of EREG/EGFR pathway in tumor progression. *International journal of molecular sciences*, *22*(23), 12828. https://doi.org/10.3390%2Fijms222312828

Cohen, M. M. (2014). Tulsi-Ocimum sanctum: A herb for all reasons. *Journal of Ayurveda and integrative medicine*, *5*(4), 251. https://doi.org/10.4103/0975-9476.146554

Dai, Y., Hogan, S., Schmelz, E. M., Ju, Y. H., Canning, C., & Zhou, K. (2011). Selective growth inhibition of human breast cancer cells by graviola fruit extract in vitro and in vivo involving

downregulation of EGFR expression. *Nutrition and cancer*, *63*(5), 795-801. https://doi.org/10.1080/01635581.2011.563027

Da Rocha, M. N., Marinho, E. S., Marinho, M. M., & Dos Santos, H. S. (2022). Virtual screening in pharmacokinetics, bioactivity, and toxicity of the amburana cearensis secondary metabolites. *Biointerface Res Appl Chem*, *12*(6), 8471-8491. http://dx.doi.org/10.33263/BRIAC126.84718491

Gajalakshmi, S., Vijayalakshmi, S., & Devi, R. V. (2012). Phytochemical and pharmacological properties of Annona muricata: a review. *International Journal of Pharmacy and Pharmaceutical Sciences*, *4*(2), 3-6.

Halder, S., Basu, S., Lall, S. P., Ganti, A. K., Batra, S. K., & Seshacharyulu, P. (2023). Targeting the EGFR signaling pathway in cancer therapy: What's new in 2023?. *Expert Opinion on Therapeutic Targets*, *27*(4-5), 305-324. https://doi.org/10.1080/14728222.2023.2218613

Hossain, A. (2022). Genetic Algorithm for variable selection. figshare. Presentation. https://doi.org/10.6084/m9.figshare.21741248.v1

Ivanović, V., Rančić, M., Arsić, B., & Pavlović, A. (2020). Lipinski's rule of five, famous extensions and famous exceptions. *Popular Scientific Article*, *3*(1), 171-177. https://doi.org/10.46793/chemn3.1.171i

Kubo, I., Nitoda, T., Tocoli, F. E., & Green, I. R. (2010). *Multifunctional cytotoxic agents from Anacardium occidentale. Phytotherapy Research, 25(1), 38–45.* doi:10.1002/ptr.3109

Kulis, M., MacQueen, I., Li, Y., Guo, R., Zhong, X.-P. & Burks A. W. (2012). Pepsinized cashew proteins are hypoallergenic and immunogenic and provide effective immunotherapy in mice with cashew allergy. *Journal of Allergy and Clinical Immunology*; 130, 3: 716–723 https://doi.org/10.1016/j.jaci.2012.05.044

Le, V. T., Nguyen, T. H., & Do, P. C. (2024). Global Ligand-Protein Docking Tools: Comparation and Case Study. http://dx.doi.org/10.5772/intechopen.1005158

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, *23*(1-3), 3-25. https://doi.org/10.1016/s0169-409x(00)00129-0

Metibemu, D. S., Akinloye, O. A., Akamo, A. J., Ojo, D. A., Okeowo, O. T., & Omotuyi, I. O. (2019). Exploring receptor tyrosine kinases-inhibitors in Cancer treatments. *Egyptian Journal of Medical Human Genetics*, *20*(1), 1-16. http://dx.doi.org/10.1186/s43042-019-0035-0

Metibemu, D. S., Akinloye, O. A., Akamo, A. J., Okoye, J. O., Ojo, D. A., Morifi, E., & Omotuyi, I. O. (2020). Carotenoid isolates of Spondias mombin demonstrate anticancer effects in DMBA-induced breast cancer in Wistar rats through X-linked inhibitor of apoptosis protein (XIAP) antagonism and anti-inflammation. *Journal of Food Biochemistry*, e13523. https://doi.org/10.1111/jfbc.13523

Metibemu, D. S., Akinloye, O. A., Akamo, A. J., Okoye, J. O., Ojo, D. A., Morifi, E., & Omotuyi, I. O. (2021). VEGFR-2 kinase domain inhibition as a scaffold for anti-angiogenesis: Validation of the anti-angiogenic effects of carotenoids from Spondias mombin in DMBA model of breast carcinoma in Wistar rats. *Toxicology Reports*, *8*, 489-498. https://doi.org/10.1016%2Fj.toxrep.2021.02.011

Metibemu, D. S., Oyeneyin, O. E., Omotoyinbo, D. E., Adeniran, O. Y., Metibemu, A. O., Oyewale, M. B., ... & Omotuyi, I. O. (2020). Molecular Docking and Quantitative Structure Activity

Relationship for the Identification of Novel Phyto-inhibitors of Matrix Metalloproteinase-2. *Science Letters, 8*(2): 61-68. http://dx.doi.org/10.13187/ercr.2020.1.3

Metibemu, D. S. (2022). 3D-QSAR and Molecular Docking Approaches for the Identification of Phyto-Inhibitors of Hsp90. *LIANBS*, *11*, 3871-3886.

National Center for Biotechnology Information (2024). PubChem Compound Summary for CID 17747343. Retrieved June 13, 2024 from https://pubchem.ncbi.nlm.nih.gov/compound/17747343.

Olasupo, S. B., Uzairu, A., Shallangwa, G., & Uba, S. (2019). QSAR analysis and molecular docking simulation of norepinephrine transporter (NET) inhibitors as anti-psychotic therapeutic agents. *Heliyon*, *5*(10). https://doi.org/10.1016/j.heliyon.2019.e02640

Paul, M. D., & Hristova, K. (2019). The RTK Interactome: Overview and Perspective on RTK Heterointeractions. *Chemical reviews*, *119*(9), 5881–5921. https://doi.org/10.1021/acs.chemrev.8b00467

Rodriguez, S. M. B., Kamel, A., Ciubotaru, G. V., Onose, G., Sevastre, A. S., Sfredel, V., ... & Tataranu, L. G. (2023). An overview of EGFR mechanisms and their implications in targeted therapies for glioblastoma. *International Journal of Molecular Sciences*, *24*(13), 11110. https://doi.org/10.3390%2Fijms241311110

Roux, P. P., & Topisirovic, I. (2018). Signaling pathways involved in the regulation of mRNA translation. *Molecular and cellular biology*, *38*(12). https://doi.org/10.1128/mcb.00070-18

Rudrapal, M., Gogoi, N., Chetia, D., Khan, J., Banwas, S., Alshehri, B., ... & Walode, S. G. (2022). Repurposing of phytomedicine-derived bioactive compounds with promising anti-SARS-CoV-2 potential: Molecular docking, MD simulation and drug-likeness/ADMET

studies. *Saudi journal of biological sciences*, *29*(4), 2432-2446. https://doi.org/10.1016/j.sjbs.2021.12.018

Sousa, A. M. D., Moreira, P. O. L., & Monte Neto, R. L. D. (2024). AI is a viable alternative to high throughput screening: a 318-target study. https://doi.org/10.1038/s41598-024-54655-z

Taneja, S. C., & Qazi, G. N. (2007). Bioactive Molecues in Medicinal Plants: A perspective in their therapeutic action. *Drug discovery and development. Chorghade, MS., editor. John Wiley and Sons, Inc*, 1-50. http://dx.doi.org/10.1002/9780470085226.ch17

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2): 455-461. https://doi.org/10.1002%2Fjcc.21334

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, *45*(12): 2615-2623. https://doi.org/10.1021/jm020017n

Vijayan, R. S. K., He, P., Modi, V., Duong-Ly, K. C., Ma, H., Peterson, J. R., ... & Levy, R. M. (2015). Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *Journal of medicinal chemistry*, *58*(1): 466-479. https://doi.org/10.1021/jm501603h

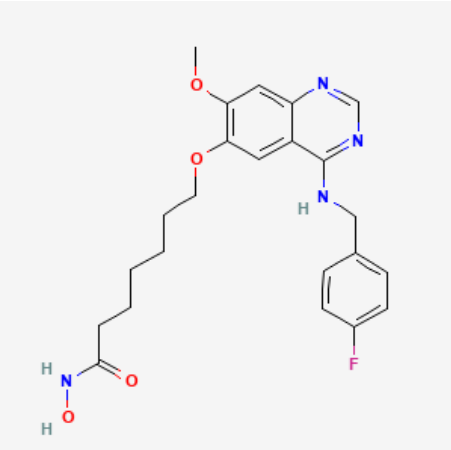
Willighagen, E. L., Mayfield, J. W., Alvarsson, J., Berg, A., Carlsson, L., Jeliazkova, N., ... & Steinbeck, C. (2017). The Chemistry Development Kit (CDK) v2. 0: atom typing, depiction, molecular formulas, and substructure searching. *Journal of cheminformatics*, *9*, 1-19. https://doi.org/10.1186/s13321-017-0220-4

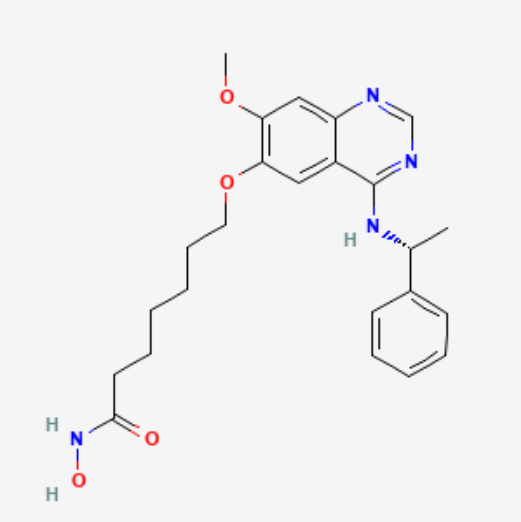
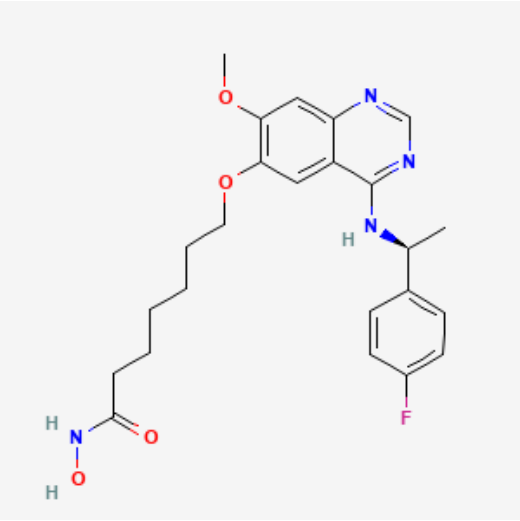
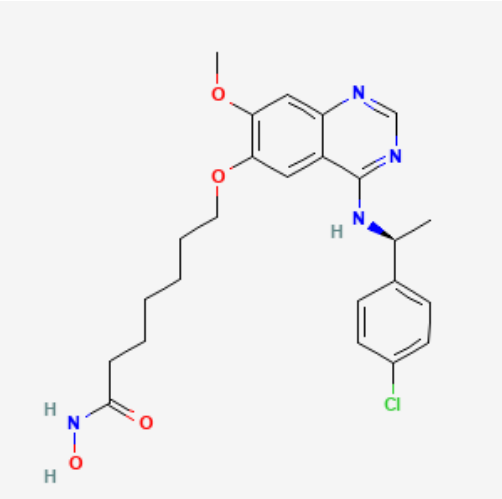
Table S1: The spearman rank correlation (R) coefficient of docking scores against the pIC_{50}.

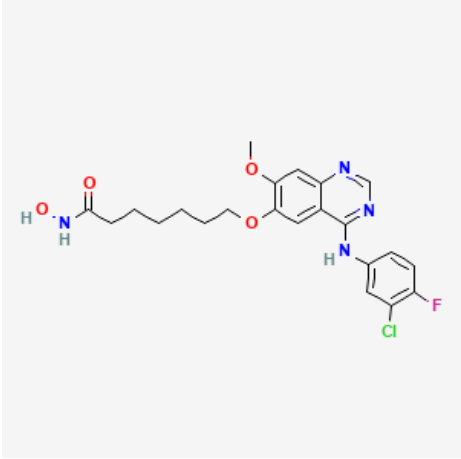
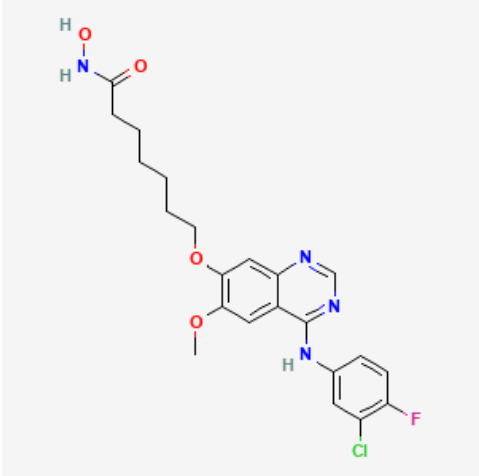
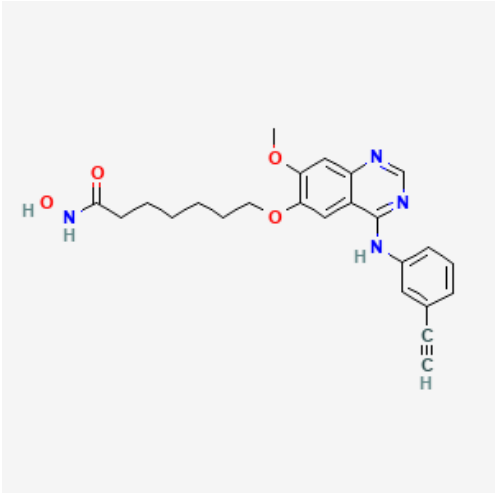
Correlation Coefficient			**.577****
Sig. (2-tailed)			**.000**
N			**85**
Bootstrap	**Bias**		**.006**
	Std. Error		**.076**
	95% Confidence interval	**Lower**	**.42**
		Upper	**.703**

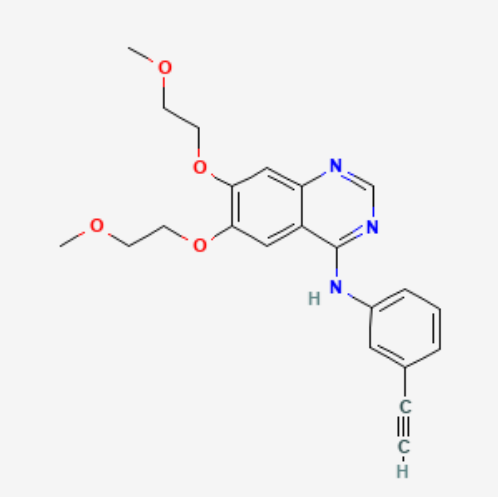
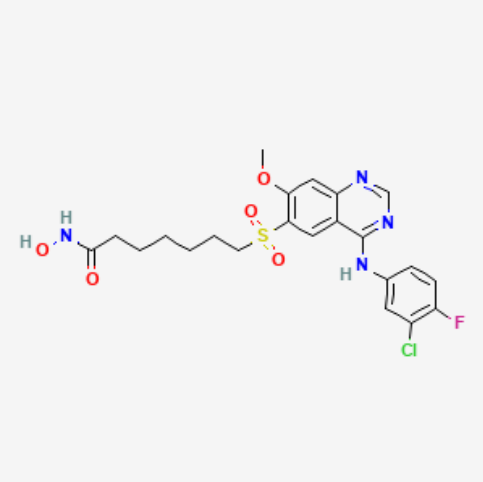
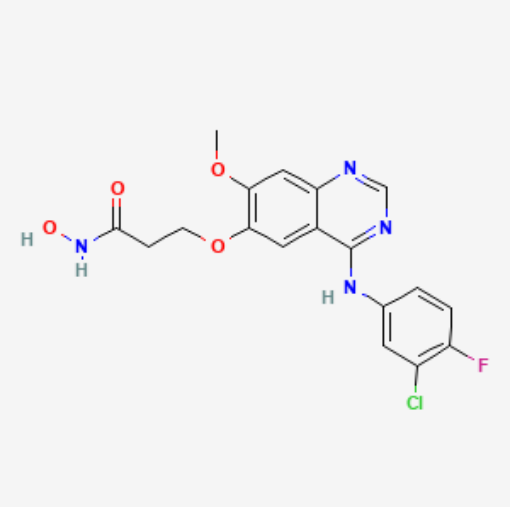
**** Correlation is significant at P<0.01 level**

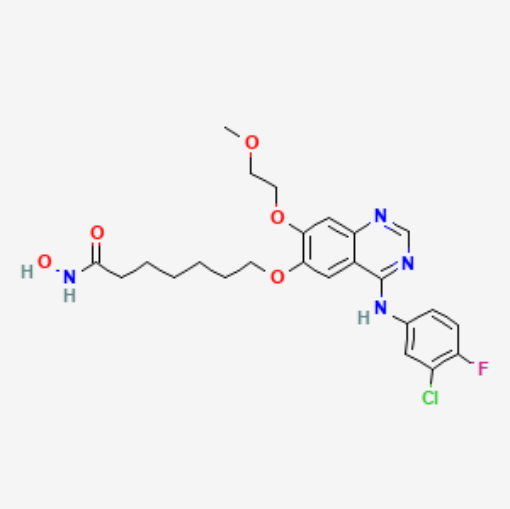
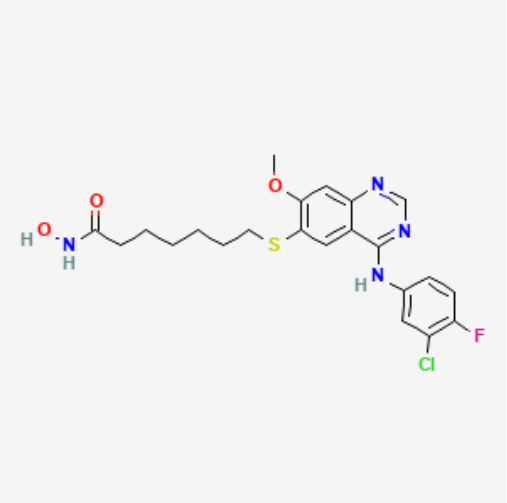
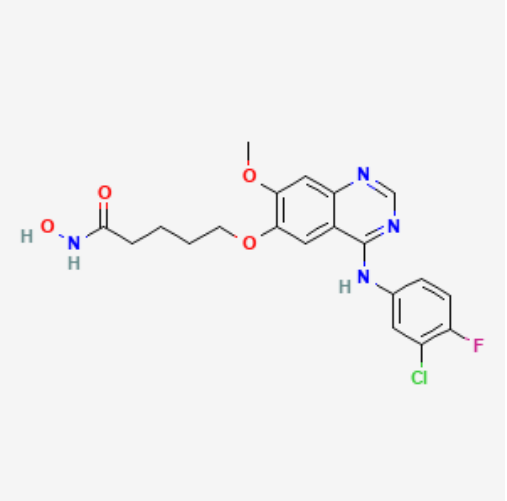
Table S2: Experimental and predicted pIC_{50} values of the of some of the Training set

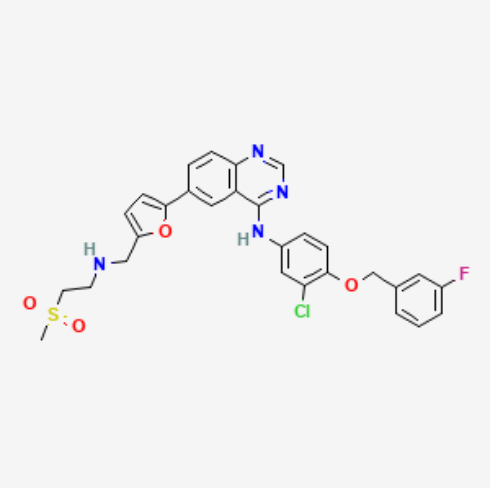
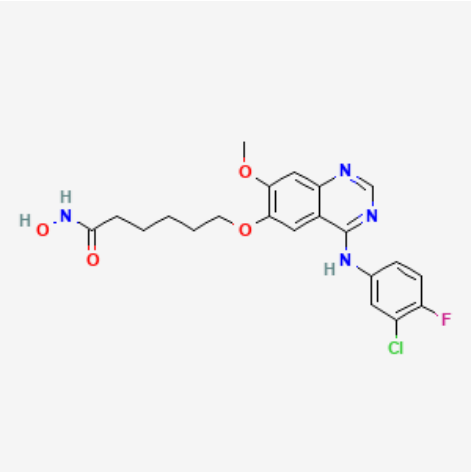
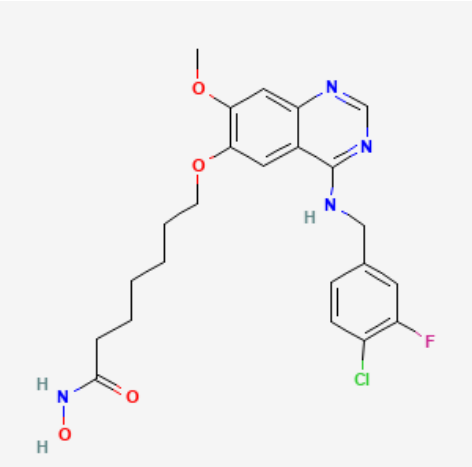
CHEMBL ID	Structures	Observed pIC_{50}	Predicted pIC_{50}	Residual
CHEMBL596964		6.57	6.56249	-0.10823

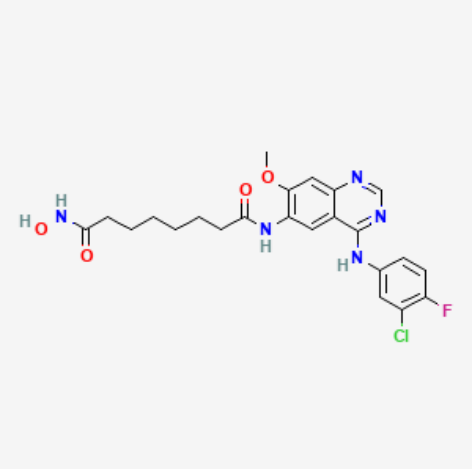
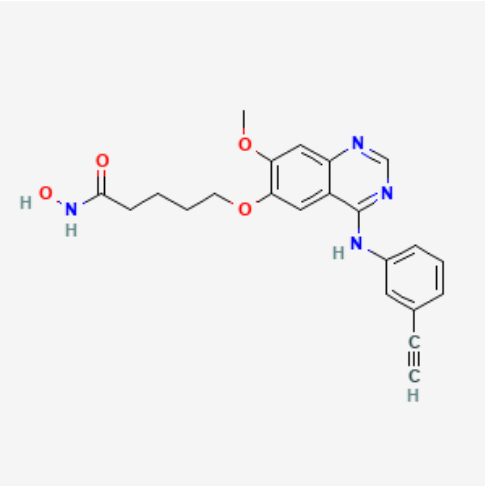
CHEMBL605976		7.89	7.90649	0.16206
CHEMBL598406		6.9	6.72146	0.04253
CHEMBL598407		6.21	6.37234	-0.2095

CHEMBL598377		8.51	8.65839	-0.21549
CHEMBL598610		8.16	8.20095	-0.20451
CHEMBL598797		8.62	8.56748	-0.1327

CHEMBL553		7.32	7.31971	0.12878
CHEMBL596755		7.81	7.86024	0.02803
CHEMBL597551		8.15	8.18002	-0.08444

CHEMBL597569		7.98	7.98699	0.1867
CHEMBL596754		8.7	8.75149	0.19166
CHEMBL596736		8.34	8.48269	0.05709

CHEMBL554		7.95	7.85316	0.04382
CHEMBL598163		8.47	8.17369	0.27881
CHEMBL596957		7.14	7.2049	-0.04972

CHEMBL597773		8.68	8.57047	-0.10916
CHEMBL599398		7.82	7.84704	-0.00573