

Identification of oxazolo[4,5-g]quinazolin-2(1H)-one Derivatives as EGFR Inhibitors for Cancer Prevention

Renganthan Senthil¹, Kishore Kumar Meenakshi Sundaram², Giridharan Bupesh³, Singaravelu Usha^{4*}, Konda Mani Saravanan^{2*}

Abstract

Objective: Abnormal expression of EGFR (epidermal growth factor receptor) results in different types of human tumors. Quinazoline-containing derivative signify an attractive platform for EGFR inhibitors. The present study aims to discover the potential binders of a group of compounds belonging to oxazolo[4,5-g]quinazolin-2(1H)-one derivative as EGFR inhibitors. **Methods:** We apply multiple computational procedures, including pharmacophore-based virtual screening methods, to perform a preliminary screening against EGFR over compounds belonging to oxazolo[4,5-g]quinazolin-2(1H)-one derivative. Then, we carried a fine screening by molecular dynamics simulations, followed by free energy calculations. **Results:** The best pharmacophore model created has five characteristics. Three hydrogen bonds acceptors (A) and two aromatic rings (R) make up AAARR (a sequential representation of chemical features of ligands). We have performed pharmacophore-based screening with different databases like Asinex, Chembridge, Lifechemicals, Maybridge, Specs, and Zinc. Top-scoring 30 molecules were considered for performing induced fit docking. Molecular Dynamics Simulations executed for the top five ligands confirmed that throughout the simulation, the protein-ligand complexes remained stable. **Conclusion:** Thus, the results obtained from this research provide insights for the development of oxazolo[4,5-g]quinazolin-2(1H)-one derivative as potent EGFR inhibitors.

Keywords: EGFR- tyrosine kinase- pharmacophore-based virtual screening- oxzolo[4,5-g]quinazolin-2(1H)-one

Asian Pac J Cancer Prev, 23 (5), 1687-1697

Introduction

The ability to differentiate between a malignant and a noncancerous cell is essential to intervene in treating cancer cell progression so that the normal cells remain undisturbed. Hence, current research efforts mainly focus on developing targeted therapies on key molecules elaborate in tumor proliferation and survival (Citri and Yarden, 2006; Du et al., 2021; Lee et al., 2021). Epidermal growth factor receptors (EGFR) is one such target that has been successful in curing solid tumors (Viswanathan et al., 2019; Abeer et al., 2021; Zhang et al., 2021). The role of EGFR is to initiate the signalling events and tumors of epithelial cell origin. It is a transmembrane glycoprotein that's one of four tyrosine kinase receptors in the erbB family. The external receptor domain, transmembrane region, and intracellular domain with tyrosine kinase activity make up the EGFR glycoprotein (170kDa) (Saravanan et al., 2018; Nguyen et al., 2021). The association of EGFR to its cognate molecules causes autophosphorylation of receptor tyrosine kinase and subsequent activation of signal transduction pathways

important in controlling cell growth, viability, and differentiation (Herbst, 2004).

It is significantly expressed in several tumour cells lines and has been linked to a bad prognosis and shorter survival time. The efforts made in the past two decades leading the development of anticancer agents that can affect EGFR activity. The most widely investigated pharmacologic strategy for affecting EGFR function aims to inhibit cellular phosphorylation of the EGFR tyrosine kinase with low-molecular weight compounds (Izadian et al., 2017). Inhibitors operate on the intracellular part of the receptor to block tyrosine kinase phosphorylation and consequent stimulation of signal transduction pathways. The four classifications of EGFR inhibitors are 4-anilinoquinazolines, 4-(ar (alk) yl amino) pyridopyrimidines, 4-phenylaminopyrrolo-pyrimidines (Elrayess et al., 2021), and oxime inhibitors (Xu et al., 2008a, 2008b).

The quinazoline derivatives are a popular family of synthetic compounds that might be used as a scaffold for EGFR inhibitors. Gefitinib (Ranson et al., 2002), Erlotinib (Luo et al., 2011), and Lapatinib (Collins et al.,

¹Marudupandiyar College (Affiliated to Bharathidasan University), Thanjavur – 613403, Tamilnadu, India. ²Department of Biotechnology, Bharath Institute of Higher Education and Research, Chennai – 600073, Tamilnadu, India. ³Department of Forest Science, Nagaland University, Zunheboto, Nagaland -798627, India. ⁴Department of Bioinformatics, Bharathiar University, Coimbatore - 641 046, Tamilnadu, India. *For Correspondence: usha@buc.edu.in, saravananbioinform@gmail.com

2009) are examples of anti-tumor medicines developed and marketed using the 4-anilinoquinazoline scaffold. Gefitinib was used as a key molecule in the development of oxazolo[4,5-g]quinazolin-2(1H)-one derivatives (Yin et al., 2015). We have attempted to identify novel potent EGFR inhibitors through pharmacophore-based virtual screening in the present work.

Materials and Methods

Data Set

Twenty four compounds from oxazolo[4, 5-g]quinazolin-2(1H)-one derivative's series data were retrieved to study EGFR inhibitors with available IC_{50} value. All the compounds were refined by Gefitinib as a lead compound (Sim et al., 2018). ChemAxon's Marvin Sketch chemical software constructed the 2D structure of these derivatives. All the in silico approaches of these compounds was done by Schrodinger software.

Ligand Preparation

The preparation process of these derivatives was performed by using the ligprep module in the Schrodinger package. This module confirms an ionization state tautomers and ring conformation to filter the molecules based on the various criteria. The structure with the best chirality is imported to the workspace. This module confirms the 3D conformations and possible combinations. The hydrogen bonds were added, and the bond length was evaluated using OPLS_2005 (Banks et al., 2005).

Pharmacophore Generation

The pharmacophore hypothesis generation of these molecules was done by the PHASE module (Dixon et al., 2006). Rapid torsion angle search approach used to generate conformers and OPLS_2005 force field with an implicit generalised Born and surface area solvation (GB/SA) model used to minimize each generated structure. Following minimization, each conformer was filtered via a 10kcal/mol relative energy window with a minimal atom deviation of 1.0Å. The six pharmacophoric features which are defining the chemical features of all ligands are, hydrogen bond donor (D), hydrogen bond acceptor (A), hydrophobic group (H), negatively charged group (N), positively charged group (P), and aromatic ring (R). A method for identifying common pharmacophore hypotheses (CPHs) using active analogues. Using a tree-based partitioning technique, the common pharmacophores are selected from the conformations of the set of active ligands. The tree-based partitioning approach clusters comparable pharmacophores collectively based on their intersite distances, the distances between the pairs of sites in the pharmacophore.

The pharmacophores that resulted were then assessed and ranked. The scoring goal was to find the best prospective hypothesis that would offer an overall ranking of all the hypotheses. The scoring method additionally considers site point and vector alignment, volume overlap, selectivity, the number of ligands matched, relative conformational energy, and activity. Based on the score analysis and the matching of active compounds to the

developed hypothesis, the best highest representative pharmacophore hypothesis was chosen (Figure 1a). The chosen 3D pharmacophore hypothesis (Figure 1b) has the following characteristics: with a survival value of 3.823, there are three hydrogen bond acceptors (A) and two aromatic rings (R).

Pharmacophore Validation

The validation of the pharmacophore was done by using a decoy set. This process validates the discriminative ability of the pharmacophore hypothesis to differentiate the activated compounds. To develop the database, the 24 compounds are combined with the 1000 compound decoy set database. The screening of the databases is done by using "find matches to hypotheses in the PHASE module" (Dixon et al., 2006). The enrichment factor (EF) and goodness of hit score (GH) were computed to estimate the hypothesis capacity for all the active compounds.

Pharmacophore based Virtual Screening

In the previous step, the PHASE module generated different pharmacophore hypotheses. The motif of chemical features of ligands with three hydrogen bond acceptors and two aromatic rings (AAARR) ranked top. The generated hypothesis was used to evaluate the chemical features of compounds. We explored the database and recovered the molecules with the desired chemical features from the Asinex (<http://www.asinex.com>), Chembridge (<http://www.hit2lead.com>), Enamine (<http://www.enamine.net>), Lifechemicals (<http://www.lifechemicals.com>), Maybridge (<http://www.maybridge.com>), Specs (<http://www.specs.net>), Zinc (<http://zinc15.docking.org>), and from each database 1000 best compounds were retrieved.

Protein Preparation

The atomic coordinates of EGFR kinase protein with the resolution of 1.71Å was obtained from the protein databank (PDB) (Berman et al., 2000). The preparation of the crystal structure was done by the protein preparation wizard. The protein molecule is imported into the workspace and missing side chains and loops are built with primes. The hydrogen bond in the protein structure was optimized. Water molecules under 3Å were deleted. Optimized crystal structure subjected for minimization to confirm the energy of the structure. All the processes were done by the OPLS_2005 force field. Following the preparation of the protein receptor, the grid is generated by the receptor grid generation module.

Structure - Based Virtual Screening (SBVS) Using Glide

The compounds retrieved over pharmacophore-based screening were integrated into Maestro. All the compounds were allowed to the Ligprep module for the addition of hydrogen, salt removal, ionization, and ring conformations (Schrödinger, 2018). The chirality of the original compounds was saved. 3D conformation structures of all the compounds were generated. Pharmacodynamic and pharmacokinetic parameters analyzed drug likeness of all the compounds. For the docking and scoring functions, Virtual Screening Workflow was used. Glide

was originally performed in a high-throughput virtual screen module (Friesner et al., 2004). 50% of the leading score compounds were docked in standard precision mode and 25% of leading score ligands were docked in extra precision mode (Friesner et al., 2006). All the docking processes were done by default parameters. Scoring functions were used to select the best conformations of each ligand.

Induced Fit Docking Procedure

For initial molecular docking, standard precision mode was used. 20 poses for each ligand were retrieved for structural refinement. The ligand docking process was done by the induced-fit docking procedure in the Schrodinger suite. Induced fit docking process was performed for the high scoring 30 compounds retrieved by using glide docking. First, the ligands were docked with a rigid receptor model. The energy minimization of protein structure was done by using the OPLS-2005 force field with the default parameters (Banks et al., 2005). Three phases are involved in this procedure: a) Glide-based docking of the ligands and receptors in standard precision (SP) mode, with a van der Waals scaling factor of 0.5 provided to the non-polar atoms of the receptor and compounds (Friesner et al., 2004). For the structural refinement, 20 poses of all the ligands were retrieved. b) To generate the protein and ligand complexes, the docked conformers are allowed to side-chain and backbone refinements. c) In the extra precision (XP) mode of Glide docking, the ligands are optimized at each of the binding pockets. Induced-fit docking scoring functions calculates both protein interaction and total energies of the system for ranking of Induced fit docking poses. Thirty compounds that contain the best glide score were selected for further analysis.

Rescoring Using Prime/MM-GBSA

The induced fit docking pose with the best Glide score value was calculated using the Prime/MM-GBSA approach (Genheden and Ryde, 2015). This process has been used to calculate the free binding energy of all the ligands to the receptor. The energy of all complexes was calculated using the OPLS-AA force field and the Generalized-Born Surface Area (GBSA) continuum solvent model. The binding free energy (G_{bind}) is calculated by $G_{bind} = ER:L - (ER + EL) + G_{solv} + GSA$, where ER:L - energy of the protein-ligand complex, ER + EL - sum of energies of the ligand and unliganded receptor, G_{solv} - difference between GBSA solvation energy (surface area energy) of the complex and GSA is the sum of corresponding energies of the ligand and un-liganded protein.

Molecular Electrostatic Potential (MESP) Calculation

The five molecules tabbed from the induced fit docking study were subjected to the density functional theory (DFT) calculations (Becke, 1993). The DFT calculations of all the compounds were done by the Jaguar module (Bochevarov et al., 2013). The geometry optimization of all the compounds was done by the hybrid density functional theory. The energy calculations in

the aqueous environment were done with the Poisson-Boltzmann solver MESP, HOMO, LUMO, and aqueous solvation energy were determined during this operation (Lee et al., 1988). The red and blue colours correspond to the high negative and positive electrostatic potentials, respectively. Other colours (orange, yellow, green) represent intermediary reactivity ranges.

Molecular Dynamics Simulations

The five compounds with the highest glide score and high binding energy were subjected to the molecular dynamic simulations. The molecular dynamic simulations were performed by the Desmond module in the Schrodinger package (Chow et al., 2008). The best-docked structure complexes were allowed to 20ns simulations. The dynamic simulations help to understand the structural stability and structural fluctuation of the protein complexes. The protein-ligand complexes were fixed in the orthorhombic water box that contains TIP3P as water molecules, and neutralized by adding 0.15M Na⁺ Cl⁻ ions (Mark and Nilsson, 2001). The system equilibrium was maintained at 300K and 1 bar pressure for the whole process. The built-in function of Desmond trajectory analysis was used to determine RMSD, RMSF, backbone comparison in Initial and Final structures, and hydrogen bond interactions (Chow et al., 2008).

Results

Pharmacophore based Virtual Screening

The set of 55 compounds is the precise compound to inhibit EGFR activity. From the set of 55 compounds, 24 compounds were chosen for their better inhibiting activity. Pharmacophore was formed to these compounds respective to the features named as Hydrogen bond

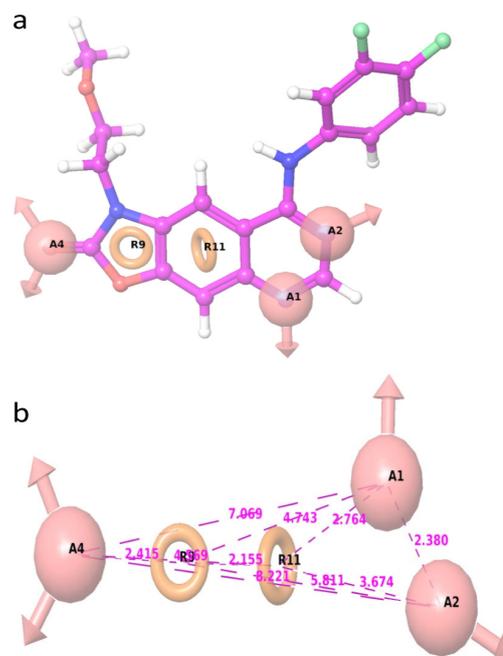


Figure 1. (a). Generation of a pharmacophore hypothesis based on properties of potent 24 oxazolo [4, 5-g] quinazolin-2(1H)-one derivatives as EGFR kinase inhibitors. (b). Intersite distance of the pharmacophore model.

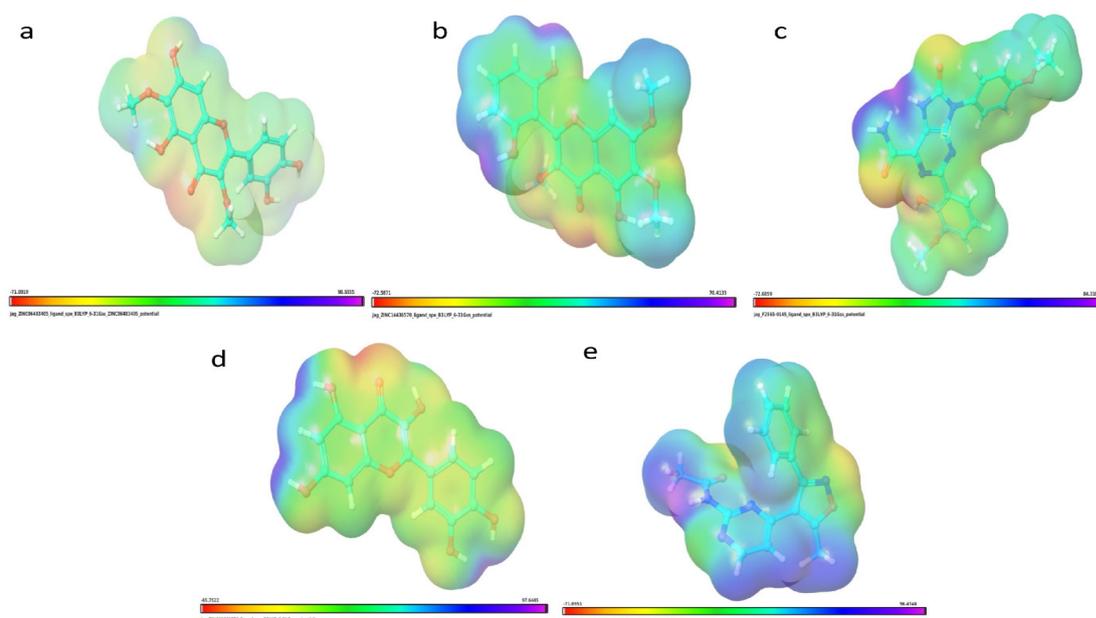


Figure 2. Molecular Electrostatic Profiles of Compounds. (a), ZINC06483405; (b), ZINC14436570; (c), F2565-0145; (d), ZINC03861279; (e), ZINC01031908.

acceptor and donor, hydrophobic group, positive and negative ions, and aromatic ring. Pharmacophore features of these compounds were developed by default parameter in PHASE. Common features using a 1Å terminal box size and tree-based algorithms that developed together according to their intersite distances, applying scoring functions for five featured pharmacophore hypothesis was ranked considering site point alignment and vector alignment, volume overlap, selectivity, number of leads.

The general pharmacophore alignment of all these compounds was retrieved based on five characteristics, the AAARR hypothesis with the survival score of

3.283 shown in Figure 1A and distance in 1B. This hypothesis includes five pharmacophore features with two acceptors, two donors, and two aromatic rings. These values show that all compounds have the same structural framework and binding orientation. Using these screening pharmacophore models helps extract the best pharmacophore features and active compounds for EGFR protein. Small molecules are screened using seven different databases and the fitness criteria given as >2.0 to test new compounds for a model. We obtained 1,000 compounds from each database and docked the target with EGFR by virtual glide workflow screening.

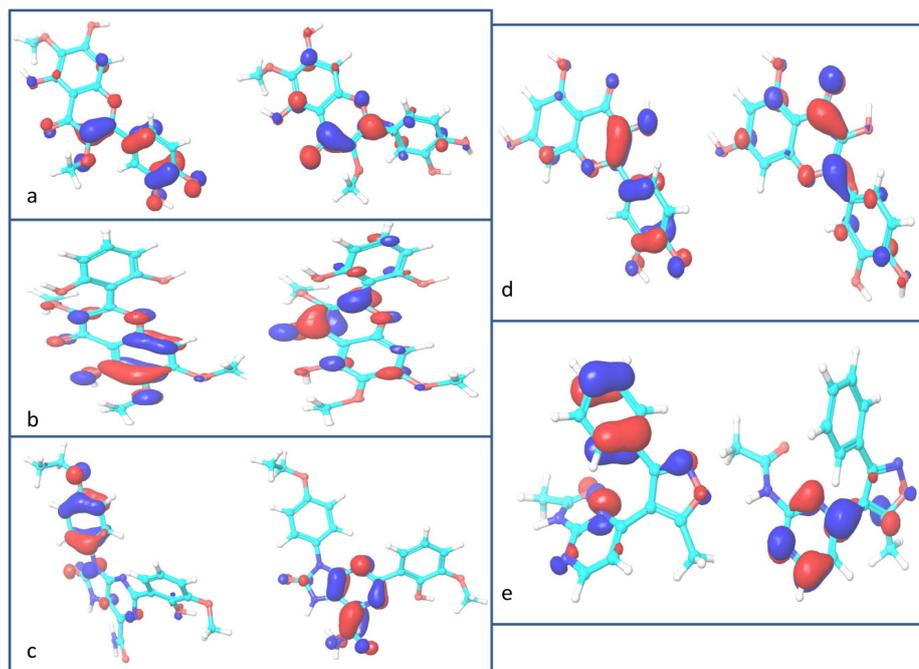


Figure 3. HOMO and LUMO of Compounds (a). ZINC06483405, (b). ZINC14436570, (c). F2565-0145, (d). ZINC03861279 and (e). ZINC01031908.

Table 1. Glide Docking Results of the Screened Compounds with EGFR

| S.No. | Title | Docking score | Glide gscore | Glide evdw | Glide ecoul | Glide energy | Glide einternal | Glide emodel |
|-------|---------------------------|---------------|--------------|------------|-------------|--------------|-----------------|--------------|
| 1 | SYN 22986007 | -10.59 | -11.00 | -42.72 | -4.18 | -46.90 | 7.77 | -77.41 |
| 2 | ZINC06483405 | -10.95 | -10.95 | -36.87 | -12.00 | -48.87 | 3.14 | -68.45 |
| 3 | SYN 22985919 | -10.14 | -10.56 | -42.54 | -3.65 | -46.20 | 5.45 | -77.36 |
| 4 | ZINC05999141 | -10.53 | -10.53 | -35.34 | -16.95 | -52.29 | 3.35 | -80.39 |
| 5 | ZINC01324210 | -10.47 | -10.47 | -37.67 | -6.14 | -43.81 | 8.43 | -61.81 |
| 6 | ZINC14644232 | -10.36 | -10.36 | -39.14 | -7.57 | -46.71 | 1.95 | -61.40 |
| 7 | ZINC01324204 | -10.34 | -10.34 | -41.13 | -4.70 | -45.83 | 7.46 | -71.09 |
| 8 | Chembridge29121658 | -9.30 | -10.32 | -36.84 | -6.97 | -43.82 | 1.09 | -52.59 |
| 9 | Chembridge47369784 | -10.32 | -10.32 | -31.81 | -6.53 | -38.33 | 2.75 | -58.79 |
| 10 | ZINC01324205 | -10.31 | -10.31 | -39.98 | -6.41 | -46.39 | 14.30 | -62.79 |
| 11 | ZINC14436570 | -10.28 | -10.28 | -40.35 | -8.25 | -48.60 | 6.15 | -58.48 |
| 12 | ZINC03870887 | -10.15 | -10.15 | -35.73 | -11.66 | -47.39 | 5.48 | -67.88 |
| 13 | ZINC04394877:ZINC12370003 | -10.14 | -10.14 | -39.10 | -4.77 | -43.87 | 3.69 | -69.05 |
| 14 | Chembridge28686225 | -9.71 | -10.12 | -36.90 | -7.25 | -44.15 | 1.99 | -70.54 |
| 15 | Chembridge16689284 | -9.87 | -9.95 | -50.02 | -7.98 | -58.00 | 4.27 | -71.06 |
| 16 | AG-205/08399053 | -9.53 | -9.94 | -37.42 | -5.79 | -43.21 | 0.34 | -61.31 |
| 17 | ZINC14644235 | -9.94 | -9.94 | -40.15 | -7.58 | -47.73 | 2.67 | -63.16 |
| 18 | ZINC13382721 | -9.89 | -9.89 | -35.41 | -6.02 | -41.42 | 3.39 | -61.57 |
| 19 | Chembridge45202431 | -8.55 | -9.88 | -46.66 | -6.38 | -53.03 | 9.13 | -71.28 |
| 20 | F2565-0145 | -9.86 | -9.87 | -45.62 | -10.46 | -56.08 | 7.94 | -81.87 |
| 21 | Chembridge34691117 | -9.84 | -9.86 | -44.55 | -5.89 | -50.43 | 8.76 | -64.49 |
| 22 | Chembridge92217067 | -9.74 | -9.74 | -36.22 | -4.55 | -40.76 | 1.36 | -58.99 |
| 23 | SYN 17738772 | -9.73 | -9.73 | -41.37 | -10.87 | -52.24 | 16.10 | -73.29 |
| 24 | AO-022/43452440 | -9.71 | -9.71 | -46.67 | -5.58 | -52.25 | 8.95 | -75.07 |
| 25 | ZINC01031908 | -9.67 | -9.67 | -34.20 | -6.46 | -40.66 | 2.44 | -59.21 |
| 26 | ZINC03861279 | -9.66 | -9.66 | -34.63 | -13.38 | -48.01 | 1.18 | -67.77 |
| 27 | Chembridge41553241 | -8.64 | -9.57 | -45.72 | -5.14 | -50.86 | 4.43 | -68.88 |
| 28 | LEG 07404742 | -8.90 | -9.57 | -40.57 | -6.42 | -46.99 | 4.79 | -60.07 |
| 29 | ZINC14757287 | -9.57 | -9.57 | -36.86 | -7.00 | -43.86 | 0.00 | -57.08 |
| 30 | ZINC08917876 | -9.56 | -9.56 | -21.29 | -3.59 | -24.87 | 0.00 | -52.88 |

Pharmacophore Validation (Decoy Set Method)

The AAARR hypothesis has been validated and used as an external decoy and active database to control its predictive ability. The data set contains 1,024 compounds that include all 24 compounds from the database and 1,000 decoy compounds from Schrodinger. The 24 active compounds have been obtained as hits through the screening of the database using AAARR, which indicate it is a successful model.

Molecular Docking

All the oxazolo[4, 5-g]quinazolin-2(1H)-one derivative were developed using Gefitinib, the documented inhibitor of EGFR [9]. Gefitinib mimics ATP, and the ATP binding site is sandwiched between the lobes where ATP forms essential interactions with hydrogen bonding and the hinge zone. The target protein 4I22 is bound to Gefitinib thus; the interactions formed were taken as a reference point. A total of 1,000 compounds from each database were obtained through pharmacophore-based virtual screening

of seven databases. Via virtual screening workflow using Glide, these compounds were docked with EGFR (PDB: 4I22). The glide score value was used to choose the best-docked compounds out of the 1,000 compounds from each set. We got the 30 best-docked conformations with glide values greater than 9.5kcal/mol (Table 1). The identified inhibitor Gefitinib forms close contact of hydrogen bond with the protein residue Met793. Hydrogen bond interaction with the backbone atoms of Met793 was found to form the selected 34 compounds. Even in several of the docked compounds, the residue Gln791 is found to form interactions. It has also been found that the residues Asp855, Asp800, Lys745, Cys797, Thr854, Arg841, Pro794 form interactions with the target protein. It was also found that these residues form interactions with the reference compound Gefitinib, which indicates that the selected 30 compounds also form many interactions of the receptor. These compounds were subjected to additional analysis.

Table 2. Induced Fit Docking and Free Energy Binding Results for the Screened Compounds with Interacting Amino Acids in the Active of EGFR

| S.No. | Ligands | Glide GScore | Glide Evidw | Glide Ecol | Glide energy | Glide Emodel | IFD score | Amino acid interaction | $\Delta G_{bind}^{(kcal\ mol^{-1})}$ |
|-------|--------------------|--------------|-------------|------------|--------------|--------------|-----------|---|--------------------------------------|
| 1 | ZINC06483405 | -13.31 | -38.69 | -19.27 | -57.96 | -82.85 | -662.70 | Met793(2), Cys797, Gln791, Thr854(SC), Asp855(SC), Asp800(2)(SC) | -89.87 |
| 2 | ZINC14436570 | -13.28 | -39.35 | -13.40 | -52.75 | -89.13 | -661.37 | Met793(3), Thr854(SC), Asp855(SC) | -94.84 |
| 3 | F2565-0145 | -13.25 | -46.77 | -16.74 | -63.51 | -109.30 | -666.01 | Met793(3), Pro794, Asp800(SC) | -90.25 |
| 4 | ZINC03861279 | -12.70 | -28.36 | -24.47 | -52.84 | -77.69 | -660.64 | Met793, Cys797, Gln791(2), Thr854(SC), Asp855(SC), Asp800(2)(SC), Lys745(2)(SC) | -77.26 |
| 5 | ZINC01031908 | -12.48 | -42.49 | -12.19 | -54.68 | -79.37 | -666.16 | Met793, Thr854(SC), Gln791, Lys745(SC) | -86.41 |
| 6 | ZINC14644232 | -12.25 | -41.92 | -11.19 | -53.11 | -77.80 | -660.25 | Met793(3), Asp855(SC) | -91.91 |
| 7 | ZINC01324210 | -12.03 | -49.80 | -7.18 | -56.99 | -91.27 | -660.17 | Met793, Gln791 | -93.46 |
| 8 | F0683-0320 | -11.94 | -55.70 | -15.73 | -71.43 | -104.32 | -665.12 | Thr854(SC), Asn842(SC), Lys745(SC), Arg841 | -111.62 |
| 9 | ZINC04394877 | -11.75 | -51.83 | -3.37 | -55.19 | -75.23 | -660.27 | Cys797, Gln791 | -95.12 |
| 10 | ZINC01324204 | -11.74 | -46.20 | -4.84 | -51.04 | -78.35 | -659.84 | Met793 Gln791 | -85.69 |
| 11 | ChemBridge29121658 | -11.42 | -46.10 | -5.37 | -51.47 | -73.03 | -659.33 | Met793(2) | -94.36 |
| 12 | ChemBridge16689284 | -11.39 | -48.98 | -8.70 | -57.68 | -80.38 | -660.23 | Met793, Thr854, Lys745 | -91.40 |
| 13 | ZINC14644235 | -11.39 | -39.78 | -11.43 | -51.21 | -71.88 | -659.56 | Met793, Thr854(SC), Asp855(SC) | -95.35 |
| 14 | ZINC08917876 | -11.17 | -37.73 | -7.98 | -45.71 | -68.62 | -661.25 | Met793, Thr854, Asp855 | -76.03 |
| 15 | ZINC05999141 | -11.12 | -46.72 | -11.33 | -58.05 | -94.16 | -659.72 | Met793(3), Asp855(SC) | -97.83 |
| 16 | AG-205/08399053 | -11.11 | -40.42 | -7.00 | -47.41 | -69.49 | -660.03 | Met793, Gln791 | -79.15 |
| 17 | AO-022/43452440 | -11.11 | -54.52 | -11.45 | -65.97 | -110.87 | -663.18 | Met793(2), Lys728 | -125.25 |
| 18 | ZINC13382721 | -11.10 | -32.51 | -6.42 | -38.93 | -61.85 | -661.41 | Gln791(2) | -96.31 |
| 19 | ZINC01324205 | -11.07 | -43.94 | -3.14 | -47.08 | -74.72 | -658.90 | Met793, Gln791 | -80.99 |
| 20 | LEG 07404742 | -10.84 | -42.87 | -6.49 | -49.37 | -74.64 | -659.01 | Met793(2), Asp855 | -77.80 |
| 21 | SYN 22986007 | -10.69 | -48.37 | -5.73 | -54.10 | -77.97 | -660.30 | Gln791 | -87.38 |
| 22 | ZINC15222500 | -10.67 | -40.73 | -4.94 | -45.67 | -66.63 | -661.40 | Met793, Thr854(SC) | -103.68 |
| 23 | ChemBridge41553241 | -10.65 | -52.19 | -8.87 | -61.06 | -88.97 | -659.12 | Met793 | -105.57 |
| 24 | SYN 22985919 | -10.64 | -45.19 | -5.16 | -50.35 | -88.87 | -659.56 | Gln791 | -81.18 |
| 25 | ZINC03870887 | -10.62 | -34.43 | -17.73 | -52.17 | -73.69 | -659.01 | Met793, Thr854(SC), Gln791, Asp855(SC), Asp800(SC) | -81.41 |
| 26 | ChemBridge34691117 | -10.60 | -48.81 | -5.63 | -54.44 | -80.81 | -661.00 | Met793, Lys745(SC) | -97.91 |
| 27 | ChemBridge45202431 | -10.65 | -50.10 | -8.60 | -58.70 | -91.55 | -660.59 | Met793(2), Lys745 | -89.33 |
| 28 | ChemBridge92217067 | -10.49 | -42.13 | -6.17 | -48.29 | -72.75 | -656.88 | Met793, Gln791 | -74.48 |
| 29 | ChemBridge47369784 | -10.46 | -42.24 | -5.20 | -47.44 | -67.39 | -657.02 | Met793, Gln791 | -78.14 |
| 30 | ZINC14757287 | -10.29 | -38.21 | -5.70 | -43.92 | -64.94 | -657.55 | Met 793 (3) | -80.03 |

Table 3. Solvation Energy, HOMO, LUMO and HOMO-LUMO Gap of the Top Five Compounds

| S.No. | Title | Solvation Energy (kcal/mol) | HOMO | LUMO | HLG (eV) |
|-------|--------------|-----------------------------|-------|-------|----------|
| 1 | ZINC06483405 | -29.61 | -0.22 | -0.06 | 0.16 |
| 2 | ZINC14436570 | -25.24 | -0.22 | -0.06 | 0.16 |
| 3 | F2565-0145 | -29.02 | -0.22 | -0.08 | 0.14 |
| 4 | ZINC03861279 | -33.38 | -0.21 | -0.06 | 0.15 |
| 5 | ZINC01031908 | -17.22 | -0.24 | -0.06 | 0.18 |

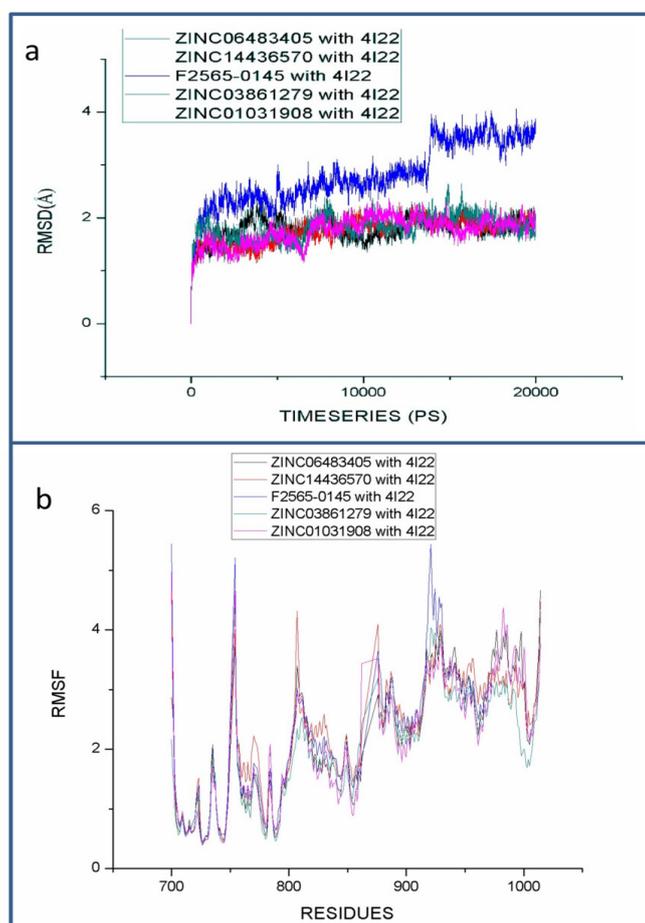


Figure 4. (a). RMSD and (b). RMSF graph of the top five compounds with EGFR.

Induced Fit Docking (IFD)

In the Glide docking, the protein was rigidly placed when binding the ligand was made flexible. In the Induced fit docking, both the protein and the ligand were treated flexibly. For each compound, we built induced fit models and the significance of ligand fitting in the binding pocket was taken into account. IFD score had used to rate IFD poses. Thirty (30) compounds were selected out of thirty-four (34) glide docked compounds for IFD

analysis. The findings show that Met793 residues formed more numbers of backbone hydrogen bond contacts compared to glide docking. The compounds tested were also found to interact strongly with Gln791 residues. The residues Thr854, Asp855, Asp800, and Lys745 form good contacts through their side-chain interactions. All of these interactions were found to be similar to those of the interactions between Gefitinib and the EGFR target, which shows the reliability of the selected compounds (Table 2).

Table 4. ADME Properties of the Top-Scoring Five Compounds

| S.No. | Title | mol MW | SASA | QPlogPo/w | QPlogHERG | QPPCaco | QPlogBB | QPP MDCK | %Human Oral Absorption |
|-------|--------------|--------|--------|-----------|-----------|---------|---------|----------|------------------------|
| 1 | ZINC06483405 | 346.29 | 547.19 | 1.30 | -4.68 | 98.39 | -1.71 | 40.35 | 70.24 |
| 2 | ZINC14436570 | 360.32 | 578.41 | 2.32 | -4.89 | 402.60 | -1.15 | 185.04 | 87.16 |
| 3 | F2565-0145 | 421.41 | 706.59 | 2.17 | -6.14 | 56.90 | -2.34 | 22.32 | 71.07 |
| 4 | ZINC03861279 | 302.24 | 510.65 | 0.39 | -4.95 | 21.66 | -2.29 | 7.86 | 53.13 |
| 5 | ZINC01031908 | 294.31 | 554.11 | 2.32 | -5.21 | 728.33 | -0.60 | 351.19 | 91.73 |

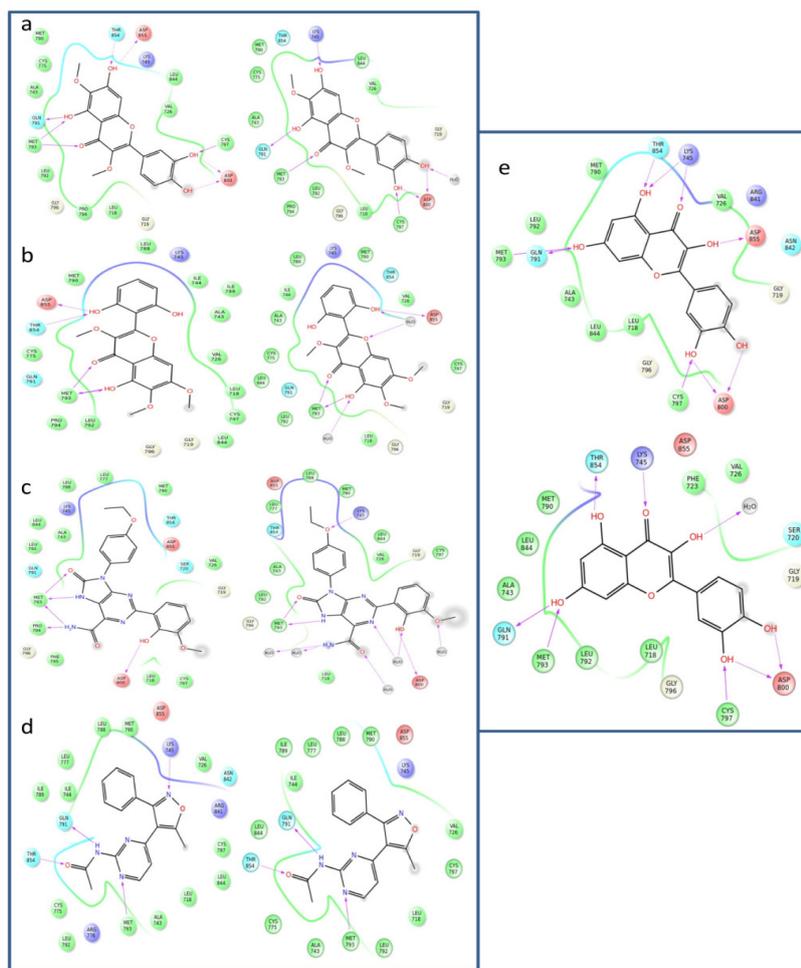


Figure 5. Protein-Ligand Interactions of the Compounds (a), ZINC06483405; (b), ZINC14436570; (c), F2565-0145; (d), ZINC03861279; (e), ZINC01031908with EGFR (PDB: 4I22) before and after dynamics.

Binding Energy Calculations

The binding free energy analysis was performed using the Prime/MM-GBSA method for the docked compounds induced to match. The calculated G_{bind} values ranged from -80.03 kcal/mol to -125.25 kcal/mol. The more negative the beliefs, the more successful they are in binding. The primary residue that forms the most important contact with EGFR is Met793, followed by Gln791, Thr854, and Asp855. These are the residues that help effectively bind the compounds to the protein (Table 2).

Molecular Electrostatic Potential (MESP) Profiles

For electronic structure analysis, the best five compounds from induced fit docking and five highly potent compounds (possessing high IC₅₀ values) were taken from the dataset sequence. The results show that all substances share electronic properties identical to each other. The MESP maps displayed the MESP profiles for the five induced fit docked compounds in Figures 2A, 2B, 2C, 2D, and 2E respectively, in which blue colour reflects the potentially electropositive region and the potentially electronegative region reflected by red colour. The negative potential values ranged from -65.75 kcal/mol to -72.69 kcal/mol and the positive potential ranged

from 56.47 kcal/mol to 98.56 kcal/mol for the induced fit docked compounds, whereas the negative potential ranged from -43.98 kcal/mol to -47.77 kcal/mol and the positive potential ranged from 54.09 kcal/mol to 82.95 kcal/mol for the compounds in the dataset. The positive ranges of electrostatic potentials for the dataset and the screened compounds were almost identical. The green and the light yellow colour represent the molecules' neutral zone, and the reactivity of the selected molecules can be understood as a whole.

Lowest Unoccupied and Highest Occupied Molecular Orbital

The frontier orbitals' energies contribute to the electron donor and acceptor's molecular properties, thus affecting the reactions of chemical compounds. The LUMO energy is related to the affinity of electrons and the measure of possible nucleophilic attack site of molecules. The data set and screened compounds of HOMO and LUMO sites are shown in Figure 3. From this, we found that HOMO and LUMO energies are weak, respect to the range between -0.208 and 0.242 eV and between -0.060 and -0.075 eV, ranging between -0.202 and -0.223 eV (HOMO) for screened compounds and for dataset compounds between -0.074 and -0.084 eV that indicates the fragile existence of the bound electrons. The energy difference between

the HOMO and LUMO energies (HLG) of screened compounds, ranges between 0.144 and 0.182eV, but for dataset compounds 0.123 and 0.149eV. It indicates that less stability making these compounds very reactive. It allows both rapid electron transfer and exchange equally possible. It was also discovered that the compounds screened were more reactive than the compounds from the dataset. The compounds ZINC01031908, F2565-0145, and ZINC14436570 have high energy of HOMO and different energies of LUMO. The highly active ZINC01031908 has a lower LUMO energy of -0.060459 and has a larger band difference. Due to its low band gap, ZINC03861279 is a more reactive compound among the three compounds. All the screened compounds have nearly the same HLG values that suggest that they are similarly reactive (Figure 3A, 3B, 3C, 3D, 3E, 3F, 3G, 3H, 3I & 3J and Table 3). Consequently, the effect of HOMO and LUMO energies may directly affect the inhibitory action against EGFR.

Aqueous Stabilization

Aqueous solubility acts as a major property leading to the drug's oral bioavailability. For the 30 induced fit docked compounds, the solvation-free energies ranged from -4.86kcal/mol to -60.2kcal/mol. The compound LEG 07404742 has -60.2kcal/mol of high-solvation capacity. Next the compounds ZINC03861279, ZINC06483405, and F2565-0145 possess high solvent capacity -33.38 kcal/mol, -29.61 kcal/mol, and -29.02 kcal/mol (Table 4) respectively. Solvation energy indicates the compound's solubility. The more negative the values are, the more is the aqueous solubility of the compound. The solvation energies calculated could provide knowledge about the pharmacokinetic optimization of these compounds.

ADME

The compounds analysed were subjected for property analysis by ADME, as the compounds have to be pharmacologically suitable to serve as potential leads against EGFR objectives. QPlogK_{sh} is -6.5 and -0.5, QPlogK_{oa} is -1.5 and 1.5, QPPCaco is < 25 and > 500, QPlogBB is -3.0 and -1.2, QPPMDCK is < 25 and > 500 and total human oral absorption is between 25 and 80 percent, according to the Qikprop software. For the top five compounds, the obtained values of all parameters are coming under the acceptable range and are further subject to molecular dynamics simulations.

Molecular Dynamics Simulation

For the top-scoring five substances, molecular dynamics (MD) simulations were performed using the Schrödinger Desmond module. The stability of the protein-ligand complexes examined after 20ns of simulations indicates that all ligands have time-period deviations of up to 5ns, the initial fluctuations were considered as a time taken for equilibrium. Throughout the simulation time depicted in the RMSD graph (Figure 4A), all protein-ligand complexes were later found to be stable. The RMSF graph in Figure 4B shows fluctuations among the target protein residues and we observe that smaller fluctuations in the residues with interaction. The Met793, Asp800 active site residues form stable interactions in the

simulations and additional interactions are also formed with Lys745, Glu791, Cys797, Thr854, and Asp855 with water-mediated interactions as shown in Figures 5A, 5B, 5C, 5D, 5E, 5H, 5I, 5J. Water-mediated protein-ligand interactions play a crucial role in the binding affinity, potency, and selectivity of the small molecules.

Discussion

Discovering drug leads can be greatly aided by computer-aided drug screening. Finding effective inhibitors with a limited number of experiments would be difficult. Molecular modelling and molecular dynamics simulation, according to several publications in the literature, considerably assist in narrowing down the candidate pool for EGFR inhibitors at a reliable computational cost (Li et al., 2021). Further, molecular dynamics simulations give more precise atomic details of the protein-ligand interaction, making the prediction much more understandable and aiding in discovering ligand binding mechanisms (Zhang et al., 2021). Considering the above, twenty-four highly active compounds with the five features of the pharmacophore hypothesis were predicted against EGFR. The predicted pharmacophore hypothesis was further carried out by screening seven separate databases for pharmacophore-based virtual screening to carry new and the most active compounds against epidermal growth factor receptors (EGFR). The obtained 7,000 screened compounds were further implemented in induced fit docking protocol with EGFR structure. To assess the inhibitory properties of these compounds, the five best compounds, ZINC06483405, ZINC14436570, F2565-0145, ZINC03861279, and ZINC01031908 were docked with EGFR using the standard protocols. Induced fit docking of the selected five compounds had strong interactions with EGFR. Binding-free energy analysis confirms the findings of IFD as the binding energy of protein-ligand complexes appears to have low values which suggest more stable and feasible binding. Furthermore, molecular dynamics research shows that the interaction and structure of protein-ligand complexes remained constant during simulation. The electronic structure of the top five compounds is analysed to determine their essence, which can aid in the understanding of their chemical and pharmacokinetic features. The ADME prediction of these compounds implies that they have drug-like features, and more investigation into their potential as EGFR inhibitors should have been done. Our findings can greatly aid in selecting prospective compounds from a much larger chemical space, which will greatly aid in discovering true innovative drugs against EGFR.

In summary, we have screened seven small molecular chemical databases to discover molecules with desired pharmacokinetic properties that inhibit EGFR. The hits obtained were subjected to multiple computational methods, including molecular dynamics simulations. Our results suggest oxazolo[4,5-g]quinazolin-2(1H)-one derivative that have the potential to facilitate drug development for EGFR. Thus, this computational research gives the first glimpses of a possible binding between oxazolo[4,5-g]quinazolin-2(1H)-one derivative and

EGFR, and additional biochemical tests will shed light on inhibitory modalities.

Author Contribution Statement

SU and KMS conceived the study. All authors participated in the development of the protocol. RS and KKMS collected and cleaned the data, and obtained ethics approval and consent. RS and GB analyzed the data. SU and KMS wrote the first draft of the paper. GB edited the manuscript and provided expert advice based on their medical specialist knowledge. All authors prepared and revised the manuscript, including relevant scientific content. All authors approved the final version of the manuscript.

Acknowledgements

Nil.

Ethics Approval and Consent to Participate
Not applicable.

Consent for Publication
Not applicable.

Availability of Data and Materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Abeer I, Khan S, Hasan MJ, Hussain M (2021). EGFR and HER2neu Expression in Gall Bladder Carcinoma and Its Association with Clinicopathological Parameters - An Institutional Experience from North India. *Asian Pac J Cancer Biol*, **6**, 57–65.

Banks JL, Beard HS, Cao Y, et al (2005). Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem*, **26**, 1752–80.

Becke AD (1993). Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys*, **98**, 5648.

Berman HM, Westbrook J, Feng Z, et al (2000). The protein data bank. *Nucleic Acids Res*, **28**, 235–42.

Bochevarov AD, Harder E, Hughes TF, et al (2013). Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int J Quantum Chem*, **113**, 2110–42.

Chow E, Rendleman CA, Bowers KJ, et al (2008). Desmond Performance on a Cluster of Multicore Processors. *Simulation*, **1**, 1–14.

Citri A, Yarden Y (2006). EGF–ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol*, **7**, 505–16.

Collins D, Hill ADK, Young L (2009). Lapatinib: A competitor or companion to trastuzumab?. *Cancer Treat Rev*, **35**, 574–81.

Dixon SL, Smondyrev AM, Knoll EH, et al (2006). PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des*, **20**, 647–71.

Du X, Yang B, An Q, et al (2021). Acquired resistance to third-generation EGFR-TKIs and emerging next-generation EGFR inhibitors. *Innovation (China)*, **2**, 100103.

Elrayess R, Abdel Aziz YM, Elgawish MS, et al (2021). Discovery of potent dual egfr/her2 inhibitors based on thiophene scaffold targeting h1299 lung cancer cell line. *Pharmaceuticals*, **14**, 1–21.

Friesner RA, Banks JL, Murphy RB, et al (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem*, **47**, 1739–49.

Friesner RA, Murphy RB, Repasky MP, et al (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem*, **49**, 6177–96.

Genheden S, Ryde U (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*, **10**, 449–61.

Herbst RS (2004). Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys*, **59**, 21–6.

Izadian Z, Bahrami N, Naji T, Mohamadnia A (2017). Surveying the Expression of CDH1 and EGFR Genes in Patients with Anthracosis and Its Relationship with Lung Carcinoma. *Asian Pac J Cancer Care*, **2**, 65.

Lee C, Yang W, Parr RG (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B*, **37**, 785–9.

Lee JC, Hung JY, Kim YC, et al (2021). Real-world Treatment Patterns in Patients with EGFR Mutation-positive NSCLC Receiving a First-Line, First- or Second-generation EGFR Tyrosine Kinase Inhibitor in South Korea and Taiwan. *Asian Pac J Cancer Biol*, **6**, 123–32.

Li M, Xue N, Liu X, et al (2021). Discovery of Potent EGFR Inhibitors With 6-Arylureido-4-anilinoquinazoline Derivatives. *Front Pharmacol*, **12**, 647591.

Luo Q, Gu Y, Zheng W, et al (2011). Erlotinib inhibits T-cell-mediated immune response via down-regulation of the c-Raf/ERK cascade and Akt signaling pathway. *Toxicol Appl Pharmacol*, **251**, 130–6.

Mark P, Nilsson L (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A*, **105**, 9954–60.

Nguyen P, Doan P, Rimpilainen T, et al (2021). Synthesis and Preclinical Validation of Novel Indole Derivatives as a GPR17 Agonist for Glioblastoma Treatment. *J Med Chem*, **64**, 10908–18.

Ranson M, Hammond LA, Ferry D, et al (2002). ZD1839, a selective oral epidermal growth factor receptor-tyrosine kinase inhibitor, is well tolerated and active in patients with solid, malignant tumors: Results of a phase I trial. *J Clin Oncol*, **20**, 2240–50.

Saravanan KM, Palanivel S, Yli-Harja O, Kandhavelu M (2018). Identification of novel GPR17-agonists by structural bioinformatics and signaling activation. *Int J Biol Macromol*, **106**, 901–7.

Schrödinger (2018). LigPrep | Schrödinger. Schrödinger Release 2018-1.

Sim EH, Yang IA, Wood-Baker R, Bowman RV, Fong KM (2018). Gefitinib for advanced non-small cell lung cancer. *Cochrane database Syst Rev*, **1**, CD006847.

Viswanathan A, Kute D, Musa A, et al (2019). 2-(2-(2,4-dioxopentan-3-ylidene)hydrazineyl)benzotrile as novel inhibitor of receptor tyrosine kinase and PI3K/AKT/mTOR signaling pathway in glioblastoma. *Eur J Med Chem*, **166**, 291–303.

Xu G, Abad MC, Connolly PJ, et al (2008a). 4-Amino-6-arylaminopyrimidine-5-carbaldehyde hydrazones as potent ErbB-2/EGFR dual kinase inhibitors. *Bioorg Med Chem Lett*, **18**, 4615–9.

Xu G, Searle LL, Hughes TV, et al (2008b). Discovery of novel

4-amino-6-arylaminopyrimidine-5-carbaldehyde oximes as dual inhibitors of EGFR and ErbB-2 protein tyrosine kinases. *Bioorganic Med Chem Lett*, **18**, 3495-9.

Yin S, Zhou L, Lin J, Xue L, Zhang C (2015). Design, synthesis and biological activities of novel oxazolo[4,5-g]quinazolin-2(1H)-one derivatives as EGFR inhibitors. *Eur J Med Chem*, **101**, 462-75.

Zhang H, Li J, Saravanan KM, et al (2021). An Integrated Deep Learning and Molecular Dynamics Simulation-Based Screening Pipeline Identifies Inhibitors of a New Cancer Drug Target TIPE2. *Front Pharmacol*, **12**, 772296.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.