



# Identification of Novel EGFR Inhibitor for Hepatocellular Carcinoma: A Computational Study of Benzothiazole Derivatives

**Mahak, Dr. Neelam Vashisth, Aastha Sharma, Department of Pharmaceutical Sciences  
(Pharmaceutical Chemistry), Gurugram University, Gurugram**

## ABSTRACT

The most frequent form of primary liver cancer in adults is hepatocellular carcinoma (HCC), which also happens to be the leading cause of mortality for cirrhosis patients. The third most common cause of cancer-related mortality is hepatocellular carcinoma, which is also the sixth most common cancer overall. It originates from hepatocytes and is intimately associated with long-term liver illnesses, especially cirrhosis, which can be brought on by alcohol addiction, hepatitis B or hepatitis C virus (HCV) infections, and non-alcoholic fatty liver disease (NAFLD).

In order to find novel lead compounds—EGFR inhibitors—for cancer therapy, a sequential computational framework was used in the current study. Pharmacophoric hypotheses were produced by drawing on benzothiazole compounds that have been previously reported to inhibit EGFR. Upon validation, the AHRRR\_1 model showed strong performance and statistical significance ( $r^2 = 0.9877$ ,  $Q^2 = 0.7280$ ). It was used in a virtual screening workflow to produce innovative hits by examining the vital interactions between the ligands and important catalytic amino acid residues—achieved through induced fit docking—the process' specificity was further improved (PDB ID: 1M17). Upon evaluation, the drug-like properties of the filtered leads were optimal. A total of five leads with possible anti-cancer efficacy were suggested for EGFR inhibition.

**Keywords:** *EGFR-TK, HCC, Pharmacophore model, 3D-QSAR, Induce fit docking, ADMET.*

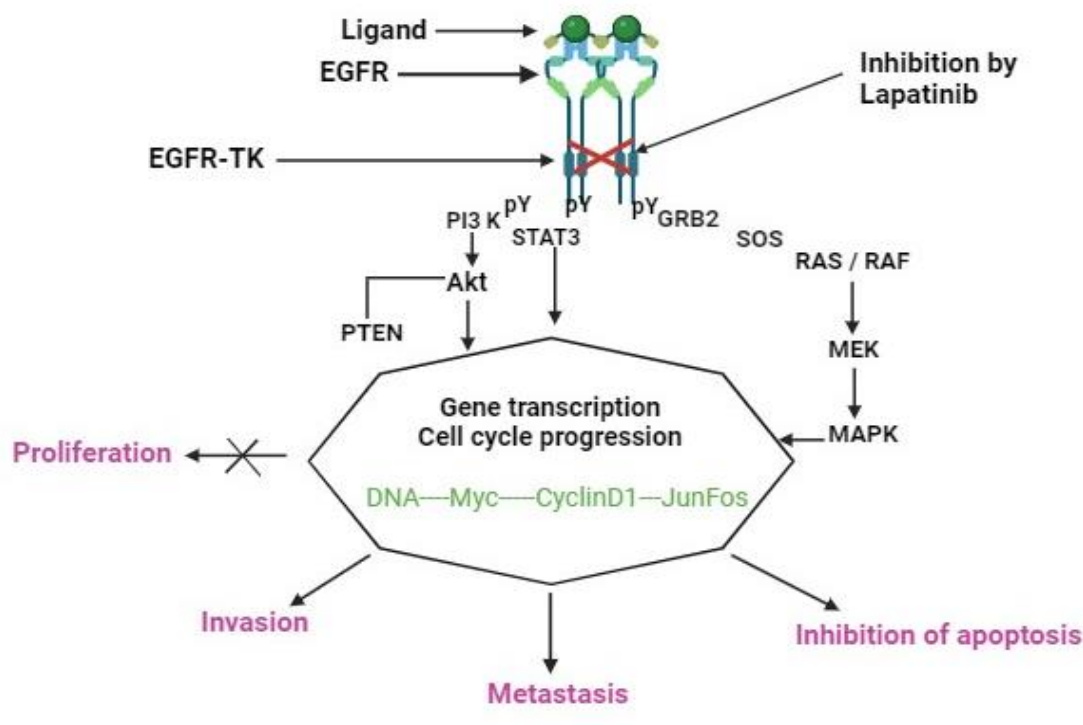
## HEPATOCELLULAR CARCINOMA

The most frequent form of primary liver cancer in adults is hepatocellular carcinoma (HCC), which also happens to be the leading cause of mortality for cirrhosis patients (1). According to USCS database incidence rates of HCC increased overall, with women's rates rising from 2.38 per 100,000 population in 2001 to 3.09 in 2020 and men's rates rising from 7.32 in 2001 to 9.82 in 2001(2). The third most common cause of cancer-related mortality is hepatocellular carcinoma, which is also the sixth most common cancer overall. It originates from hepatocytes and is intimately associated with long-term liver illnesses, especially cirrhosis, which can be brought on by alcohol addiction, hepatitis B or hepatitis C virus (HCV) infections, and non-alcoholic fatty liver disease (NAFLD). In the multistep process of hepatocarcinogenesis, integration of the hepatitis B viral DNA most likely serves as a genotoxic initiator; nevertheless, the precise mechanisms underlying the carcinogenesis remain unknown. Consumption of aflatoxin may also play an etiological role in high incidence areas, most likely acting as a genotoxic or epigenetic promoter to the carcinogenesis initiated by the hepatitis B virus. This setting encourages

the accumulation of genetic and epigenetic modifications, such as mutations in oncogenes (like CTNNB1) and tumor suppressor genes (like TP53), as well as changes in signaling pathways (including Wnt/ $\beta$ -catenin, PI3K/AKT/mTOR, and MAPK). Growing genetic instability and clonal expansion of altered hepatocytes are hallmarks of the transition from chronic liver disease to cirrhosis and ultimately to HCC (3). These pathways vary depending on the environment because different risk factors cause HCC in different ways (4). Epidermal growth factor (EGF) binding to its receptor initiates an enzymatic cascade that in turn involves major intracellular enzymatic pathways implicated in the process of hepatocarcinogenesis. Serum biomarkers and imaging investigations are the main sources of information used in HCC diagnosis. For early screening, ultrasound is frequently utilized, particularly in high-risk groups such those with cirrhosis.

### Mechanism of EGFR

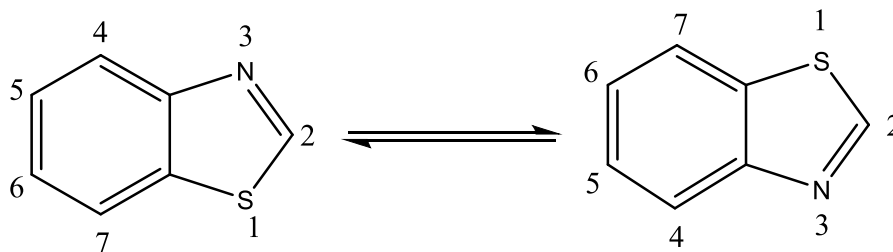
The type of enzyme known as epidermal growth factor receptor (EGFR) is essential for cell migration, survival, and proliferation. It belongs to the family of receptor tyrosine kinases known as ErbB (5). TGF- $\alpha$  and epidermal growth factor are examples of selective ligands for EGFR that bind to the receptor and activate it. Activation results in the conversion of EGFR-TK from an inactive monomeric form to an active homodimer. This, in turn, raises the intrinsic intracellular protein-tyrosine kinase activity of EGFR and triggers autophosphorylation of several tyrosine (Y) residues in the C-terminal domain of EGFR (6). Participates in the control of cell division, apoptosis, and proliferation after being activated by the binding of particular ligands. Cancer may arise from mutations that impact EGFR expression or activity. Tumors that express EGFR can't grow if EGFR is targeted (7).



**Fig1:** Diagrammatic view of mechanism of action of EGFR-TK

### BENZOTHAZOLE

Benzothiazole is a heterocyclic molecule with the chemical formula  $C_7H_5NS$ . Its core structure is made up of nitrogen and sulfur atoms, while the benzene ring is fused to the thiazole ring. Benzothiazole's molecular weight (135.1863 g/mol), melting point ( $20^\circ\text{C}$ ), and boiling point ( $227\text{--}228^\circ\text{C}$ ) are all known. Its liquid is colorless and slightly viscous. (8). The various positions on the Benzothiazole ring are numbered in the manner indicated as shown in Fig2 (9).



**Fig2:** Benzothiazole

Benzothiazole derivatives have garnered significant attention in the field of medicinal chemistry due to their diverse biological activities, including their potential as anticancer agent (10). These derivatives have demonstrated a broad spectrum of biological activities, such as anticancer (11), antioxidant (12), anti-inflammatory (13), anti-tubercular (14), antiviral (15), antibacterial (16), and anti-fungal properties (17). Benzothiazole is a crucial structural motif in medicinal chemistry because of the distinct methine center found in the thiazole ring. Its derivatives have demonstrated increased activity with less harmful effects (14). In order to better grasp the potential of novel benzothiazole derivatives as anticancer drugs, recent research has concentrated on their synthesis, assessment, and investigation of the structure-activity relationship (18).

### Dataset

82 benzothiazole compounds that have been identified as EGFR-TK inhibitors were selected from published research and are thought to have anticancer potential. Using the formula  $pIC_{50} = -6 \log_{10} IC_{50}$  for pharmacophore and 3D-QSAR, the  $IC_{50}$  values were transformed to  $pIC_{50}$ . From recent works, the minimum inhibitory concentration of all the chemicals proliferating within the range of 0.9 to 29  $\mu\text{g/ml}$  was chosen (19-25).

### Preparation of Protein

From the protein data library (<https://www.rcsb.org/search>), EGFR-TK (PDB ID- 1M17) with resolution 2.60 Å structures were obtained [25]. Protein was prepared in glide by removing the alternative conformations and water molecules in the incomplete residues. After pre-processing, relating Het states, adding missing residues, adding hydrogen's, and optimizing the chosen protein further grid generation, permitted (24).

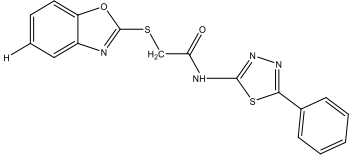
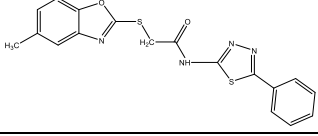
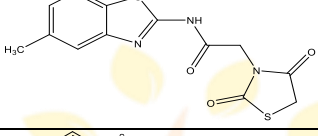
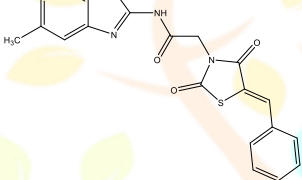
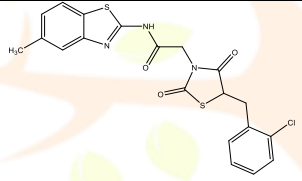
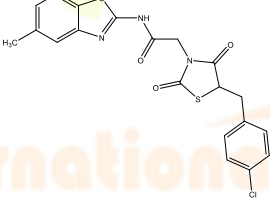
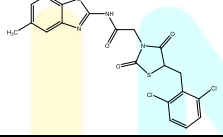
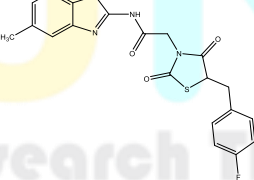
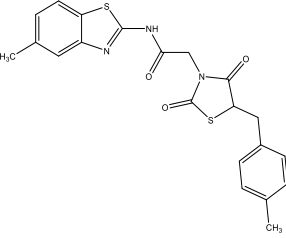
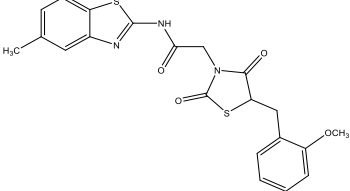
### Pharmacophore hypothesis and 3D-QSAR model

The generation of the pharmacophore hypothesis and subsequent 3D-QSAR model building for the prepared EGFR inhibitors were performed using Phase (26). The activity threshold was established when the synthesized compounds and matching  $pIC_{50}$  were imported into the module. In practice, the activity threshold is set using a difference. The chosen compounds, however, did not exhibit the indicated difference when subjected to the same procedure for enzyme inhibition analysis. As a result, molecules with an  $IC_{50}$  of less than 0.55  $\mu\text{M}$  ( $pIC_{50} = 4.003$ ) were considered to be better enzyme inhibitors (actives), whereas molecules having an  $IC_{50}$  of more than 5.5  $\mu\text{M}$  ( $pIC_{50} = 7.959$ ) were considered to be less active inhibitors (inactives). The remainder was classified as moderately engaged. Subsequently, the pharmacophoric sites were evaluated in light of the chemical characteristics included in the dataset. The six pharmacophoric features—hydrophobic group (H), negative ionisable group (N), positive ionisable group (P), hydrogen bond acceptor (A), hydrogen bond donor (D), and aromatic ring (R)—were automatically used in the development of each model. The parameters for creating the variation list of hypotheses were as follows: a minimum of five sites, a maximum of six sites, and a tolerance limit of 1 Å for pharmacophore matching. A list of variations with five points shared by all the molecules was the outcome of this. A scoring process was exercised to rank the generated hypotheses so as to rationalize selecting the hypothesis for developing the 3D-QSAR model.

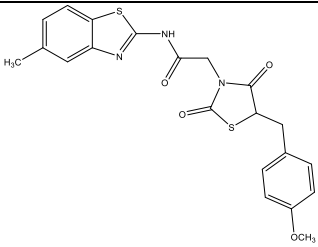
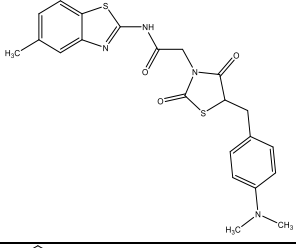
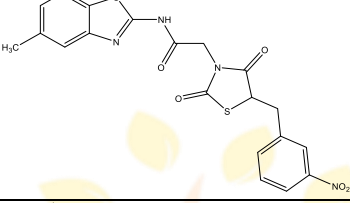
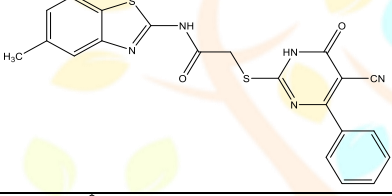
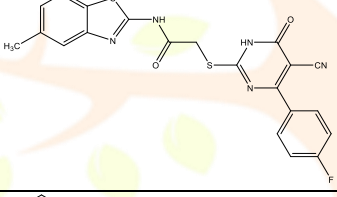
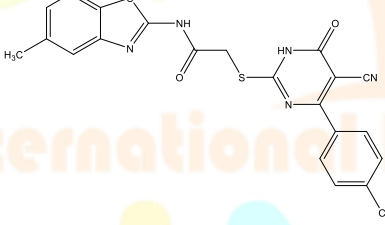
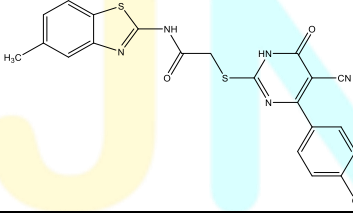
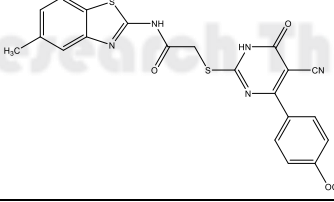
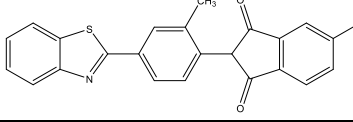
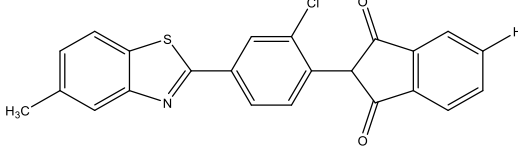
The three-dimensional partial least square partial least square (PLS) model based on atoms was built using a rectangular grid spaced one Å apart, and it may have up to four PLS components. Each of the 70 training and 30 test sets of molecules in the data set was randomly selected (Table 1). Finding the spatial locations that would

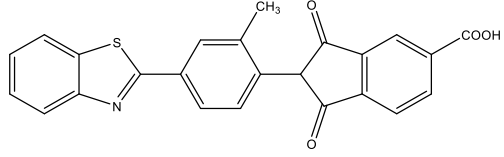
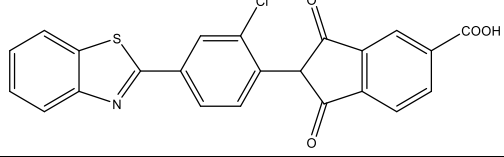
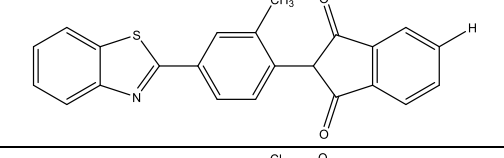
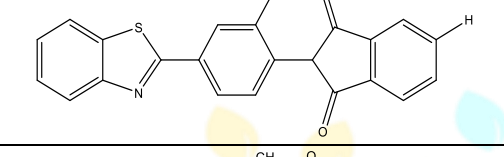
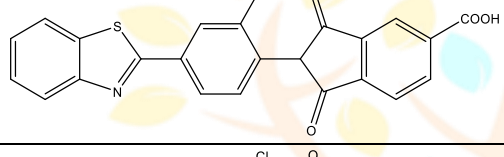
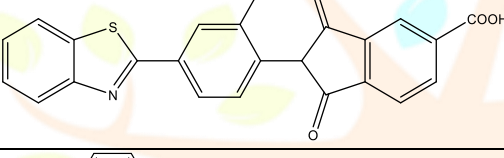
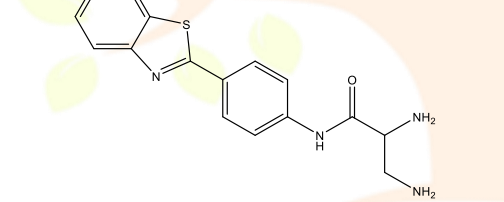
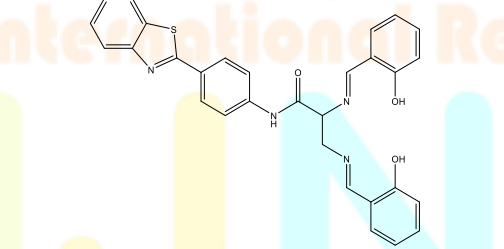
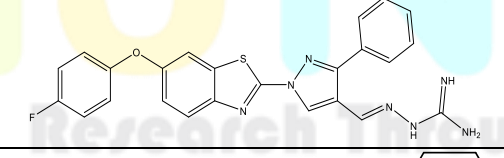
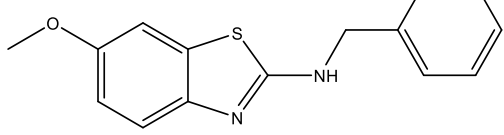
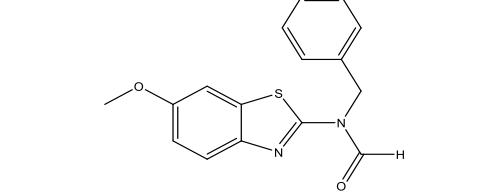
either directly or indirectly contribute to the biological activity was made easier by the model's inter-feature distances and the contour plots that followed.

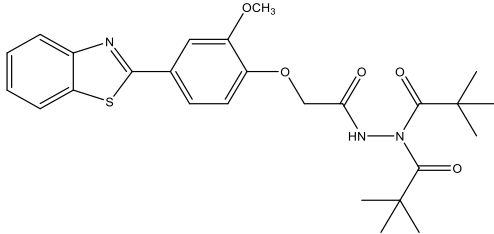
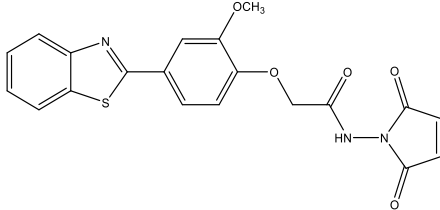
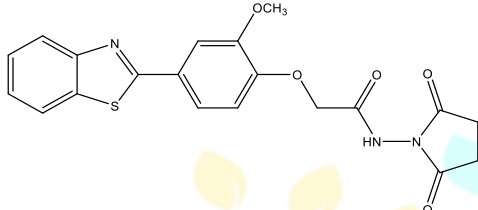
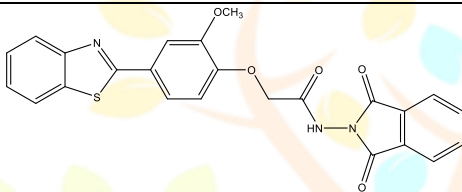
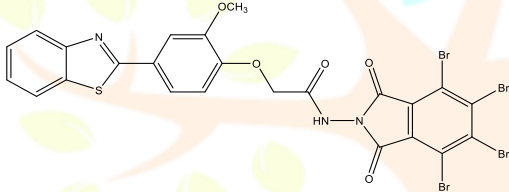
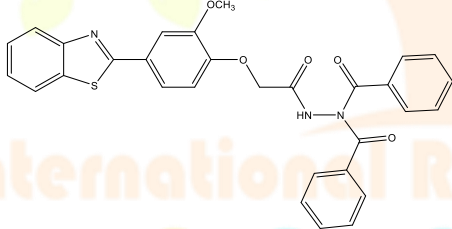
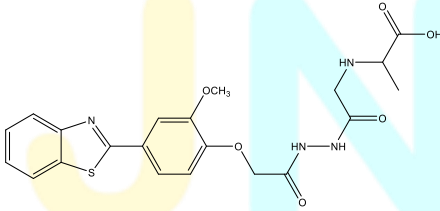
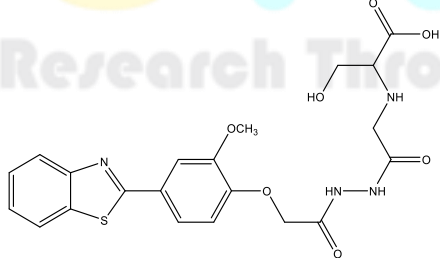
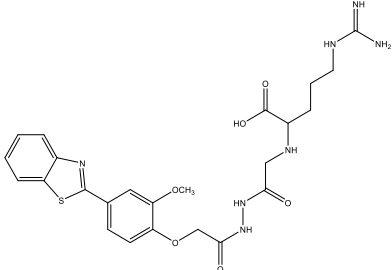
**Table1.** Library of benzothiazole derivatives with IC<sub>50</sub> values (μM).

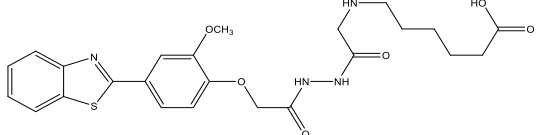
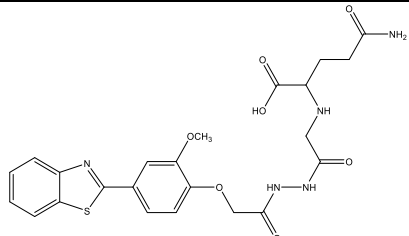
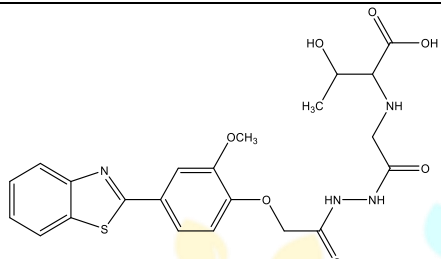
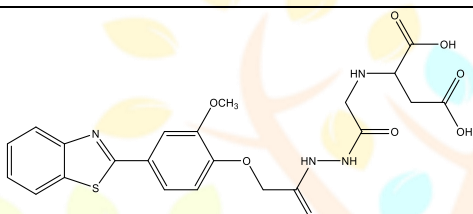
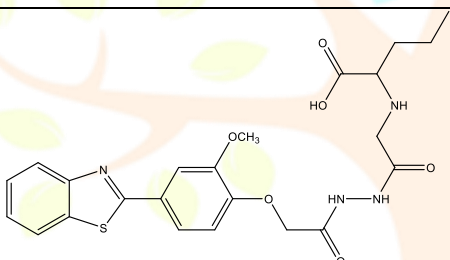
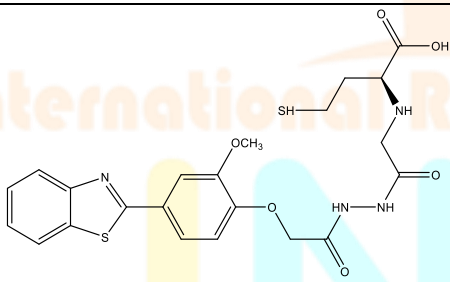
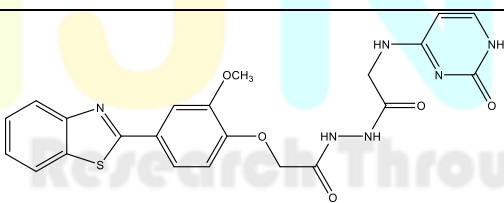
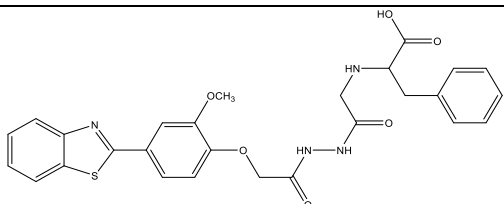
S. No	Derivatives	Observed (pIC <sub>50</sub> )	Predicted (pIC <sub>50</sub> )	Residual
J1		4.799	4.79436	-4.694
J2		5.001	4.98714	-4.496
J3		4.462	4.49172	-4.307
J4		4.591	4.44921	-6.097
J5		4.787	4.76329	-5.139
J6		4.958	4.9003	-4.81
J7		4.719	4.72647	-3.672
J8		4.933	4.89685	-5.781
J9		5.000	5.13902	-5.403
J10		4.607	5.12014	-4.578



J11		4.852	4.82953	-4.549
J12		4.634	4.95338	-4.922
J13		4.587	4.55575	-4.033
J14		4.599	4.56542	-6.271
J15		4.624	4.56111	-5.986
J16		4.916	4.86693	-6.451
J17		4.620	4.70058	-7.161
J18		4.745	4.77042	-5.894
J19		7.959	6.56399	-4.695
J20		6.140	6.52931	-4.781

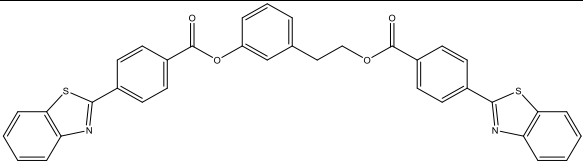
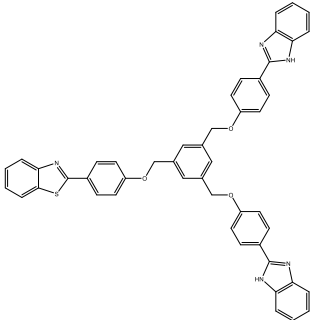
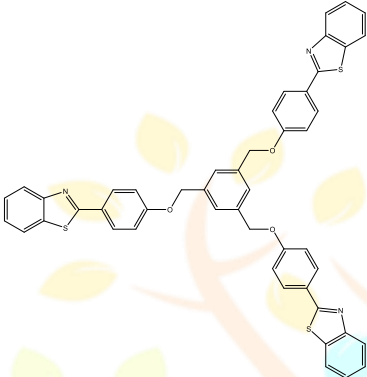
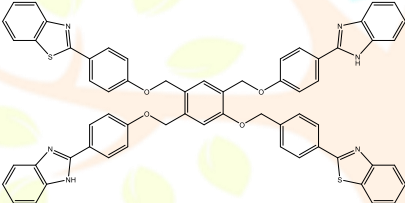
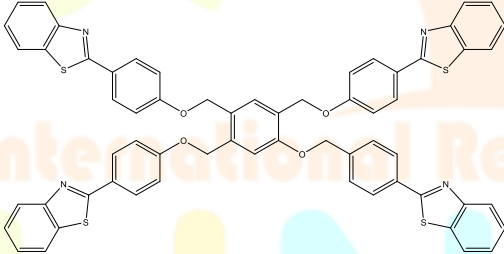
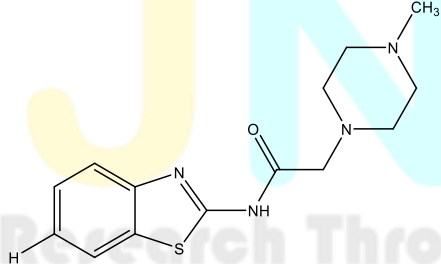
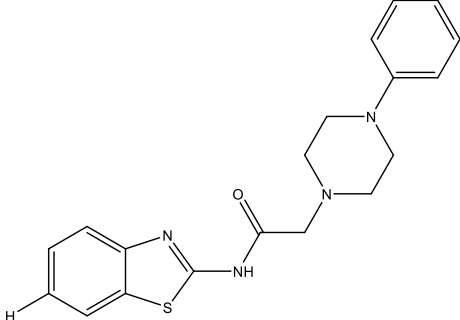
J21		7.252	7.19726	-4.859
J22		6.879	7.10977	-5.785
J23		7.009	6.94473	-4.719
J24		7.444	6.89524	-4.747
J25		7.721	7.48846	-3.56
J26		7.229	7.38233	-4.79
J27		5.229	5.16148	-5.996
J28		4.848	4.85923	-6.974
J29		5.292	4.74895	-5.443
J30		5.238	4.92761	-7.607
J31		5.256	5.32814	-6.349

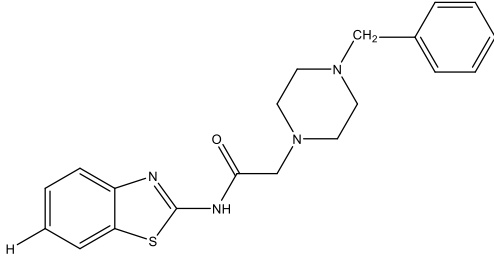
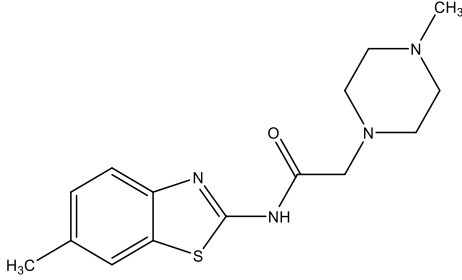
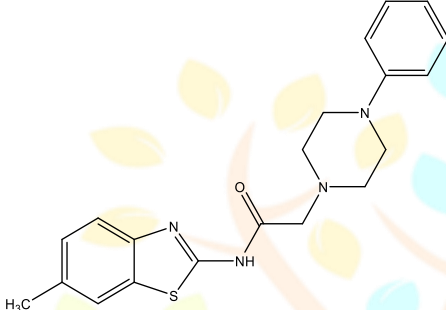
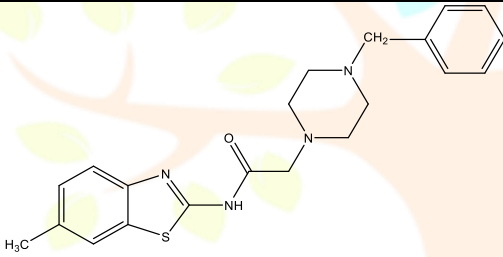
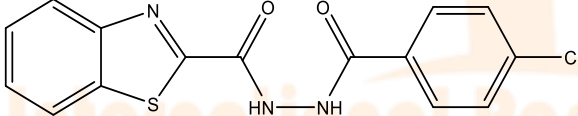
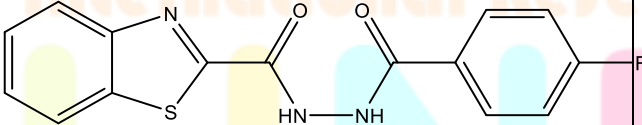
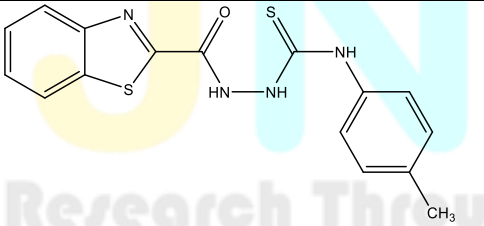
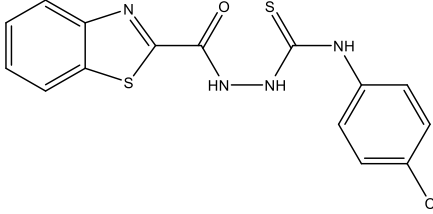
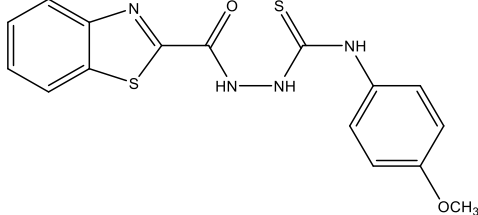
J32		5.597	5.56556	-6.198
J33		5.592	5.53975	-5.011
J34		5.588	5.55414	-6.852
J35		5.544	5.57239	-5.289
J36		5.575	5.57304	-1.632
J37		5.519	5.47335	-6.61
J38		5.526	4.93152	-6.664
J39		5.533	5.53278	-8.352
J40		5.590	5.77529	-9.413

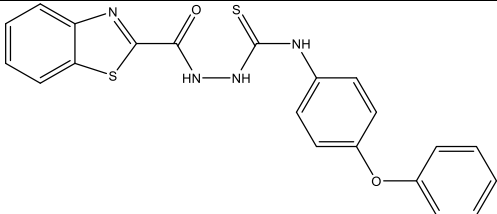
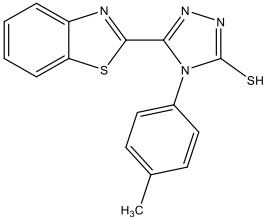
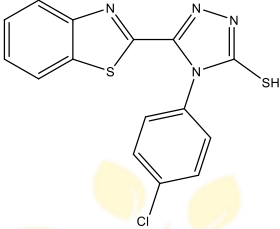
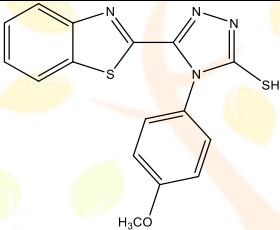
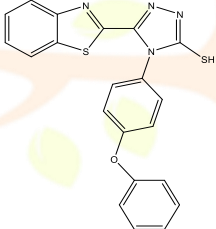
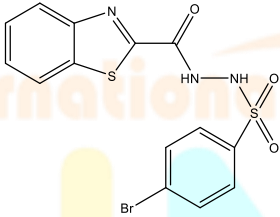
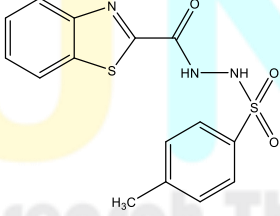
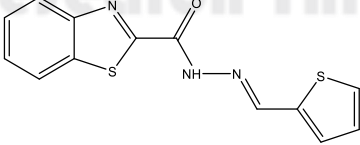
J41		5.526	5.53865	-7.711
J42		5.498	5.49298	-7.67
J43		5.587	5.58263	-8.736
J44		5.465	5.42335	-7.547
J45		5.460	5.4437	-7.236
J46		5.570	5.51989	-8.118
J47		5.498	5.50723	-6.6
J48		5.483	5.48624	-7.324



J49		4.345	4.31582	-4.04
J50		4.064	3.98673	-4.531
J51		4.016	4.02944	-4.529
J52		4.003	4.00357	-5.702
J53		4.366	4.2617	-4.874
J54		4.366	4.66725	-6.145
J55		4.425	4.34926	-5.203
J56		4.214	4.1898	-5.9
J57		4.178	4.27813	-5.873
J58		4.842	4.69663	-5.307

J59		4.131	4.74986	-1.269
J60		4.643	4.81962	-6.125
J61		4.509	4.88222	-3.026
J62		4.792	4.71487	-3.215
J63		4.689	4.87254	-
J64		4.799	4.42258	-4.699
J65		5.893	4.09156	-3.835

J66		4.147	4.40624	-4.02
J67		4.755	4.43066	-4.434
J68		4.096	4.03972	-4.416
J69		4.628	4.18679	-4.268
J70		4.028	4.17216	-5.988
J71		4.273	4.18519	-5.746
J72		4.273	5.10993	-5.006
J73		4.149	5.08689	-
J74		4.868	4.72316	-4.659

J75		5.135	4.49327	-4.986
J76		4.720	4.41655	-6.477
J77		4.490	4.4073	-5.919
J78		4.424	5.21143	-5.764
J79		4.361	4.50999	-5.433
J80		5.267	4.98059	-3.918
J81		4.530	4.78493	-4.361
J82		5.019	432766	-6.836

### Model validation

A strong 3D-QSAR model's dependability depends on its capacity to forecast biological activity, which can be verified statistically. PLS regression analysis was used to assess internal validity. Analysis of validation using enrichment descriptors was done to evaluate the model's quality. (27). For this purpose, a benchmark EGFR decoy set was retrieved (<https://dude.docking.org/targets/hdac8>) it showed characteristics similar to drugs, yet with differences in topology. Using the derived hypothesis, the entire list of 35524 compounds—35442 decoys and 82

active compounds—was screened. The rankings were evaluated using the following metrics: enrichment factor (EF), robust initial enhancement (RIE), receiver operating characteristic (ROC) curve, and Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC). External validation is also essential to measure the level of prediction of established models (28,29). For this reason, statistical metrics including root mean square error (RMSE), Pearson-r, variance ratio (F-value), variance ratio (R<sup>2</sup>), cross-validated (R<sup>2</sup>CV), regression coefficient (R<sup>2</sup>), and significance of variance ratio (P) were computed using the appropriate statistical equations.

### Screening databases and molecular docking workflow

Compounds with similar pharmacophoric characteristics were found by virtually screening the chemical databases with the Phase module. With 5712 compounds in several 2D compound libraries (PubChem), the proven 3D pharmacophore model was utilized to find novel drug-like compounds based on the pharmacophoric template. Prior to screening, the compounds underwent processing to improve their 2D structure and decrease energy use. A molecular weight between 150 and 500,  $x \log p \leq 5.0$ , number of H-bond donors  $\leq 5$ , and number of H-bond acceptors  $< 10$  were the search parameters that were established using Lipinski's Rule of Five. Several score factors, including site matching, RMSD, and vector and volume alignment, were used to categorize the hits that were received after screening. The hits obtained were taken up for molecular docking.

The high-resolution crystal structure of the EGFR enzyme (PDB ID: 1M17, resolution 2.60 Å) was obtained from the protein data bank prior to molecular docking. The protein preparation wizard was utilized for additional processing (30). Using OPLS\_2005 force field, the energy of the protein was minimized and optimized (31). The Side chains of the protein that have gaps and then undergo refining. Water molecules with three H bonds missing from crystallography were eliminated. When the pH was 7.4, many ligand ionization states and tautomeric forms were produced. In order to create ionization states, the essential residues were also chosen. An active site centroid was generated and a grid box was constructed around it using the co-crystallized ligand as the boundary within a 10-micron radius. Glide was utilized to enable the hits identified from database screening to dock into the catalytic site of the EGFR protein that was produced (32). To better analyze, high throughput virtual screening (HTVS) was first conducted, and then Glide—standard precision (Glide SP) and Glide—extra precision (Glide XP) docking. The docking method known as Glide SP is helpful for database screening since it can identify ligands with a fair inclination for binding to targets and eliminate false negatives. Robust scoring functions, including suitable penalties for optimal poses, are employed in the Glide XP module. Thus, it aids in lead optimization and eliminates the selection of false positives. Based on docking energy and scores, Glide XP docking's last hits were selected.

### Induced fit docking

The induced fit docking (IFD) module allows the option for both protein and ligand flexibility through the docking process (33). The three-step process began with the ligands (here, hits selected after Glide XP docking) being docked into the stiff protein. After that, the various ligand conformers that were discovered were refined using protein (side chain and backbone), including residues that had one atom involved within a radius of five Å. The energy-minimized complexes were then used in the Glide XP mode for the last docking and scoring step. Based on the IFD scores and the binding interactions of the hits with the receptor, the outcomes were assessed and contrasted with those of lapatinib, a conventional inhibitor.

### In Silico ADMET

The pharmacokinetic profile required for ADMET (absorption, distribution, metabolism, excretion and toxicity) prediction can be determined using the Lab3.0. Various pharmacokinetic parameters including human percent oral absorption, water coefficient, blood-brain barrier permeability, rule of five and Caco cell permeability were examined by Lipinski's rule of five. From docking, top 5 docked ligand was further selected for ADMET with the help of Lab3.0 (<http://admetlab3.scbdd.com>). From the PubChem database library, top hundred molecules were selected to perform the docking. By using Lab3.0 top 5 (Figure) molecules from PubChem database library was then performed for ADMET by using Lab2.0 [34].



## Results and discussion

### 3D-QSAR pharmacophore models and validation

The pharmacophore hypotheses were generated using the ligand-based approach, followed by atom-based 3D-QSAR model, to correlate the effect of different substituents on enzyme activity. Six common hypotheses were produced from the input data set by setting the five feature points and a suitable activity threshold. Following that, the resulting hypotheses underwent a scoring procedure using a number of modules, including survival, site, volume, vector, and bedroc (Table 2). The most optimal model was the hypothesis AHRRR\_1, which consists of one hydrophobic group (H), one hydrogen bond acceptor (A), and three aromatic rings (R). In the chosen data set, these site sites were discovered to be crucial for EGFR enzyme inhibition. The produced model's inter-feature distances (é) are depicted in Figure (Table 3).

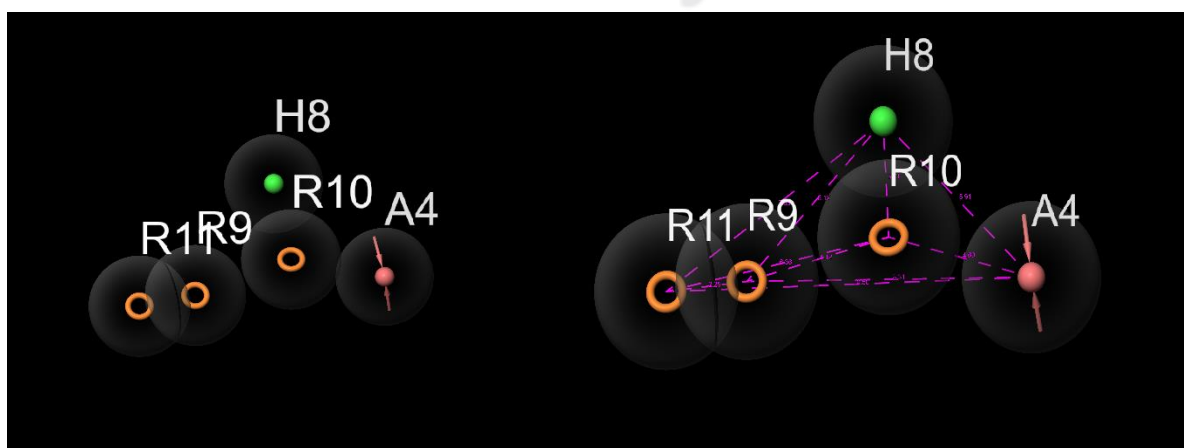
Table 3 summarizes the regression data of the optimal model, AHRRR\_1. The model's coefficient of determination ( $r^2$ ) for the training set is 0.93, which suggests that the constructed model is relevant, according to the data. The reliability of the model is supported by a strong F-value of (833.4) of PLS factor five, with a reduced P (41e-48) and RMSE (0.46). The model's credibility is additionally guaranteed by the Q2 (0.7820) and Pearson-r (0.8924) parameters. For both the training set (Fig.) and test set (Fig.) molecules, the residual or scatter plots of the observed V/s predicted activities (in pIC50) provide additional support for this.

The validation process using enrichment descriptors for the model, AHRRR\_1, was investigated for a decoy set of 1081 compounds (1055 decoys and 26 active compounds). Reliable parameters such as BEDROC (at  $\alpha = 20.0$ ), EF at 1% of the decoy set, RIE (19.92), and ROC (1.00) respectively. The area under the curve (AUC) of the ROC plot is another important metric to judge the performance of the model (Supplementary Fig). The plot shows an AUC of 1.00, and a ROC of 1.00.

External statistical validation is another factor that affects the model's predictive power. R2cv, or the computed cross-validated coefficient, is 0.7897. Between the predicted and observed actions, the correlation coefficient, or r (or  $r^2$ ), is 0.92 ( $r^2 = 0.9877$ ). The figures for stability (0.818) and R2 scramble (0.8937), respectively, lend additional credence to this. They are inside Table 4's statistical range. The model's statistical significance is confirmed as all analyzed external parameters fall inside the statistical bounds (Table 4).

**Table2:** Pharmacophore hypothesis generated by PHASE

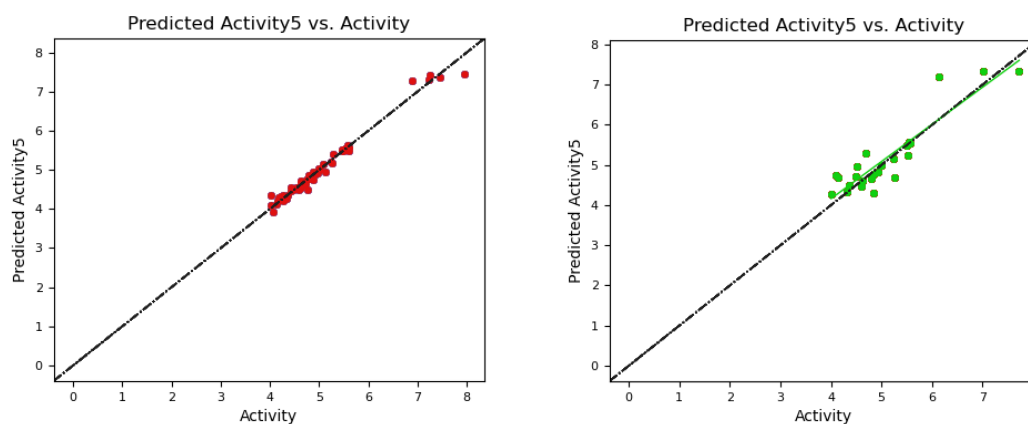
S. No	Hypothesis	Survival score	Site Score	Vector	Volume Score	BEDROC
1	AHRRR_1	5.7004	0.7468	0.961	0.5984	0.6393
2	AHRRR_2	5.6923	0.742	0.9605	0.5949	0.6361
3	AHRRR_3	5.5783	0.7096	0.9581	0.5302	0.6379
4	AAHRR_1	5.4868	0.7037	0.92	0.5846	0.6147
5	AAHRR_2	5.4476	0.6812	0.9175	0.5701	0.622
6	AAHRR_3	5.0205	0.4139	0.4139	0.4722	0.6381



**Fig3:** Five-point pharmacophore model (AHRRR\_1) by PHASE. The model show acceptor feature (A4 red-coloured arrows), hydrophobic feature (H8 green coloured).

**Table3:** Distance (in Å) between the sites.

Entry	Site 1	Site 2	Distance(Å)
AHRRR_1	A4	H8	5.91
AHRRR_1	A4	R9	8.31
AHRRR_1	A4	R10	4.60
AHRRR_1	A4	R11	10.50
AHRRR_1	H8	R9	6.12
AHRRR_1	H8	R10	3.71
AHRRR_1	H8	R11	7.85
AHRRR_1	R9	R10	4.12
AHRRR_1	R9	R11	2.25
AHRRR_1	R10	A4	4.60
AHRRR_1	R10	R11	6.36
AHRRR_1	R11	A4	10.50
AHRRR_1	R11	H8	7.85

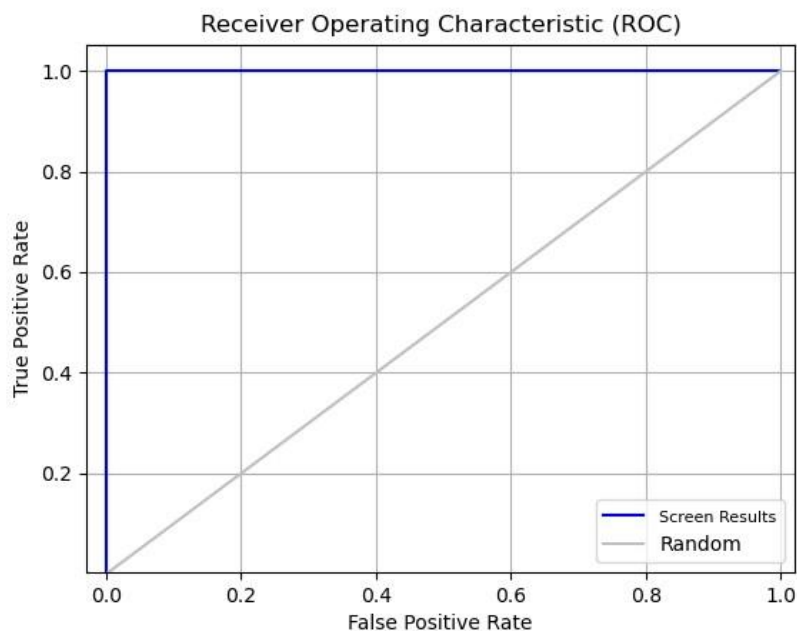


**Fig4:** Scatter plot for the 3D-QSAR model display the relationship between (a) training set's actual activity and predicted activity (b) Scatter plot for the test set with a cross-validation correlation coefficient with best fit line  $y = 0.93x + 0.46$  ( $R^2 = 0.83$ )

**Table4:** PLS regression analysis of the selected 3d-QSAR model AHRRR\_1

# Factors	SD	$R^2$	$R^2$ CV	$R^2$ Scramble	Stability	F	P	RMSE	$Q^2$	Pearson-r
1	0.4266	0.7272	0.5416	0.4759	0.93	149.3	1.97e-17	0.58	0.6416	0.8477
2	0.2037	0.9389	0.7836	0.6917	0.901	422.6	4.13e-34	0.43	0.8084	0.9068

3	0.1421	0.9708	0.8056	0.8087	0.878	598.7	2.13e-41	0.45	0.7893	0.8968
4	0.1092	0.9831	0.7978	0.8589	0.827	769.8	06e-46	0.46	0.7794	0.8919
5	0.0941	0.9877	0.7897	0.8937	0.818	833.4	41e-48	0.46	0.7820	0.8924



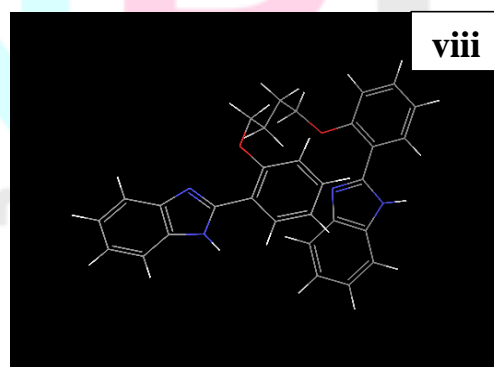
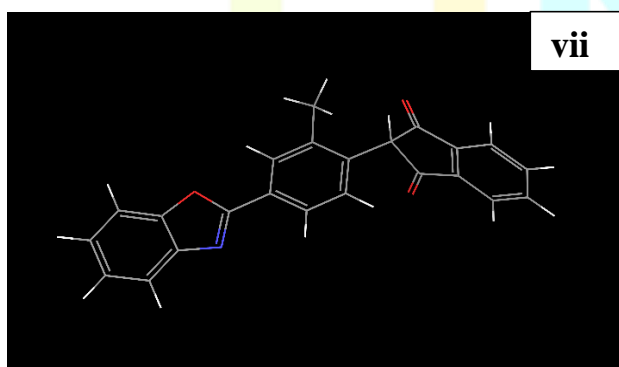
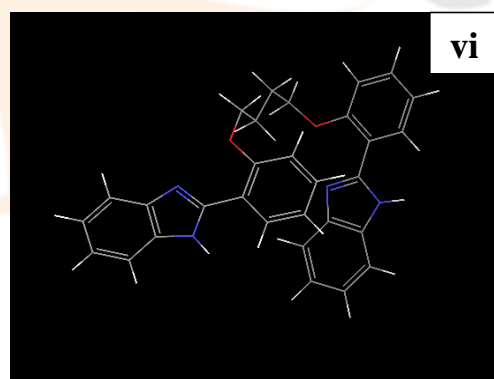
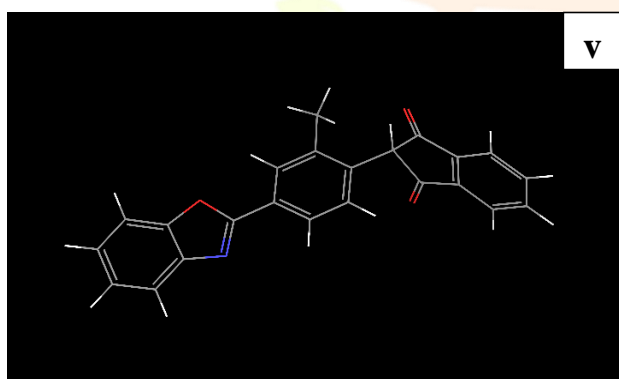
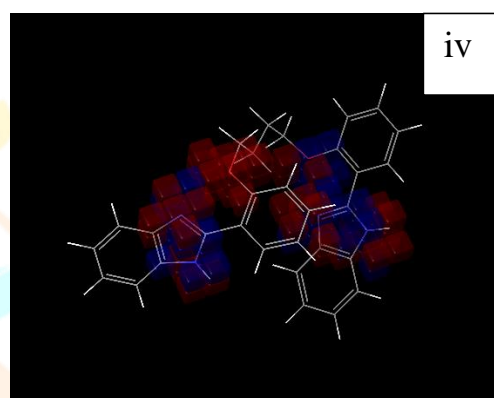
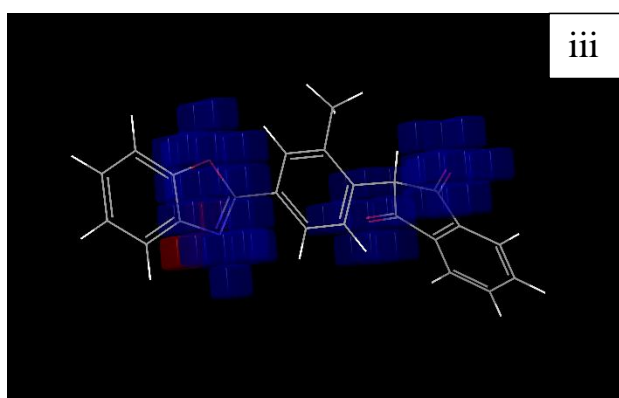
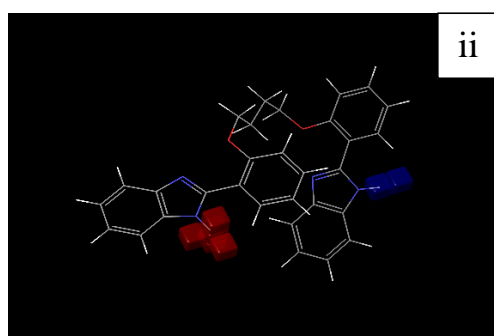
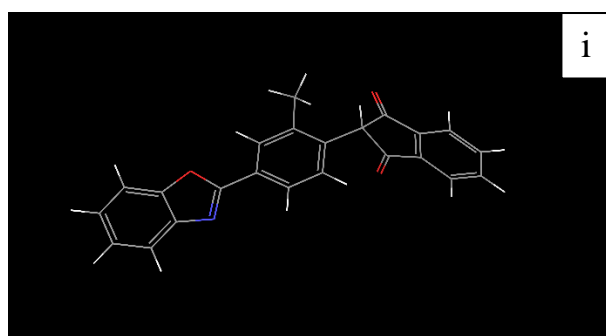
**Fig5:** ROC plot showing the protein-ligand interaction by PHASE pharmacophore

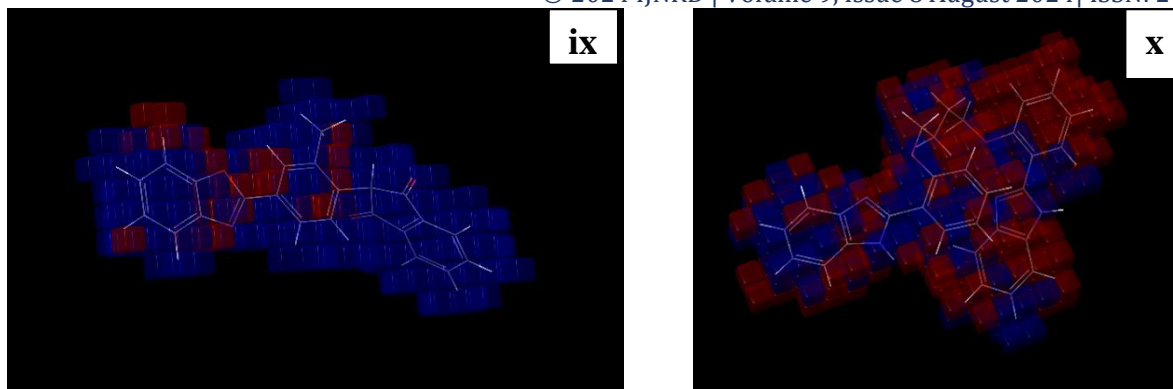
### Contour map analyses

The effects of the spatial arrangement of structural features hydrogen bonding, electron withdrawing, ionic, and hydrophobic were determined using contour plot analysis. The comparison of the most significant favorable and unfavorable interactions that result from applying the QSAR model (PHASE) to the most and least active ligands is displayed by blue cubes for each individual's positive contribution and red cubes for the negative contribution. Increased biological activity can be correlated to the blue-colored occlusion maps in the Schrödinger software tool for atom-based QSAR. Nevertheless, the decrease in biological activity has been correlated to the red-colored occlusion maps/cubes.

The counter maps are correlated with hydrogen bond donor surrounding the benzothiazole ring; they have no hydrogen bond properties in the most active compound J19 fig(i). The effectiveness of hydrogen bond donor was recognized (N-H) as blue in favorable cubes and red cubes in the benzothiazole of least compound 52. The counter maps correlated with electron with-drawing group attributes on the benzothiazole ring positively correspond with biological activity. In the presence of electron withdrawing group surrounding the benzothiazole of the least active compound 52, characterized in red cubes as an unfavorable region in fig (iv). The counter maps correlated negative ionic surrounding the benzothiazole ring they have no negative ionic properties in most active compound J19 fig (v) and the least active compound fig (vi). The counter maps correlated positive ionic surrounding the benzothiazole ring they have no positive ionic properties in most active compound J19 fig (vii) and the least active compound fig (viii). The maps correlated with hydrophobic properties on benzothiazole ring demonstrate highly blue occlusion maps and improve biological activity. The maps consistent with other attributes on benzothiazole and 2-(2-(4-phenoxybutoxy) phenyl) benzo[d]thiazole indicate an increased in activity of the most active

compound19. In the contour map, the blue cubes indicate the interaction in the region connected with nitrogen (N) at 3 position in benzothiazole, which is favorable in activity, and in the surrounding benzothiazole ring, they have red cubes to decrease the biological activity.





**Fig6:** Visual representation of 3D-QSAR model with (a) Hydrogen bond donor of best active compound J19, (b) Hydrogen bond donor of least active compound J52, (c) electron withdrawing compound J19 most active, (d) electron withdrawing compound J52 least active, (e) negative ionic compound J19 most active, (f) negative ionic compound J52 least active, (g) positive ionic compound J19 most active, (h) positive ionic compound J52 least active, (i) hydrophobic/non-polar compound J19 most active, (j) ) hydrophobic/non-polar compound J52 least active.

### Virtual screening and molecular docking

The primary goal of the research was to find new leads, and this was achieved by virtual screening of chemical databases (Figure 4). Compounds with a maximum root mean square deviation (RMSD) of 0.46 Å were found using the Phase module's sophisticated pharmacophore screening, matching the five sites of the approved AHRRR\_1 model.

After obtaining the hits, they underwent additional filtering according to scoring factors such as vector and volume alignment, RMSD, and site matching, in preparation for molecular docking research. Three steps of docking were completed, starting with HTVS of 3459 compounds. The molecules were filtered down to 328 using docking energy and score analysis of the data, after which Glide XP docking was applied. Table 5 lists the top hits with good docking score and energy from each database, derived from a selection of about 100 molecules derived from the Glide XP docking findings.

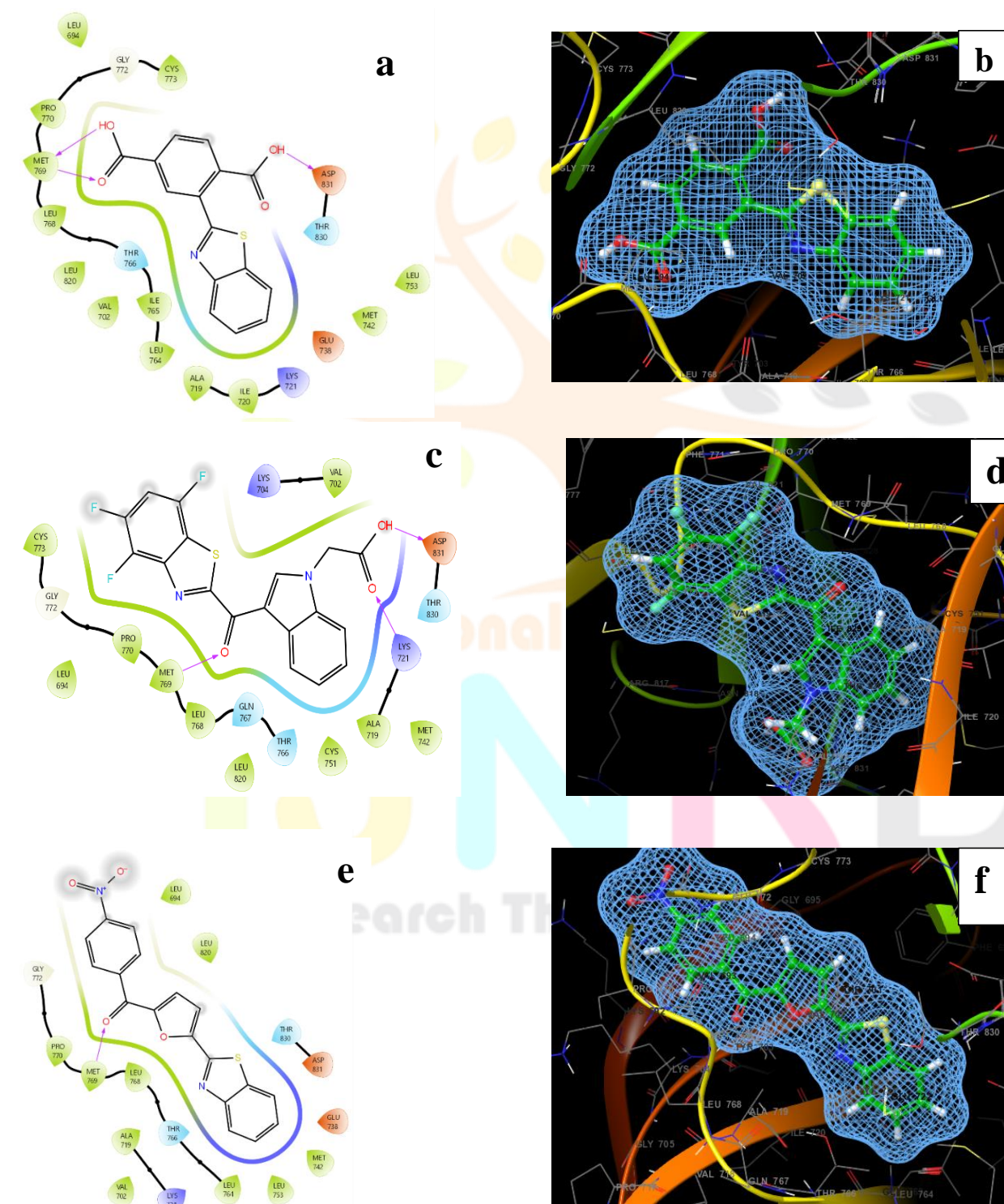
### Induced fit docking, site interaction analysis

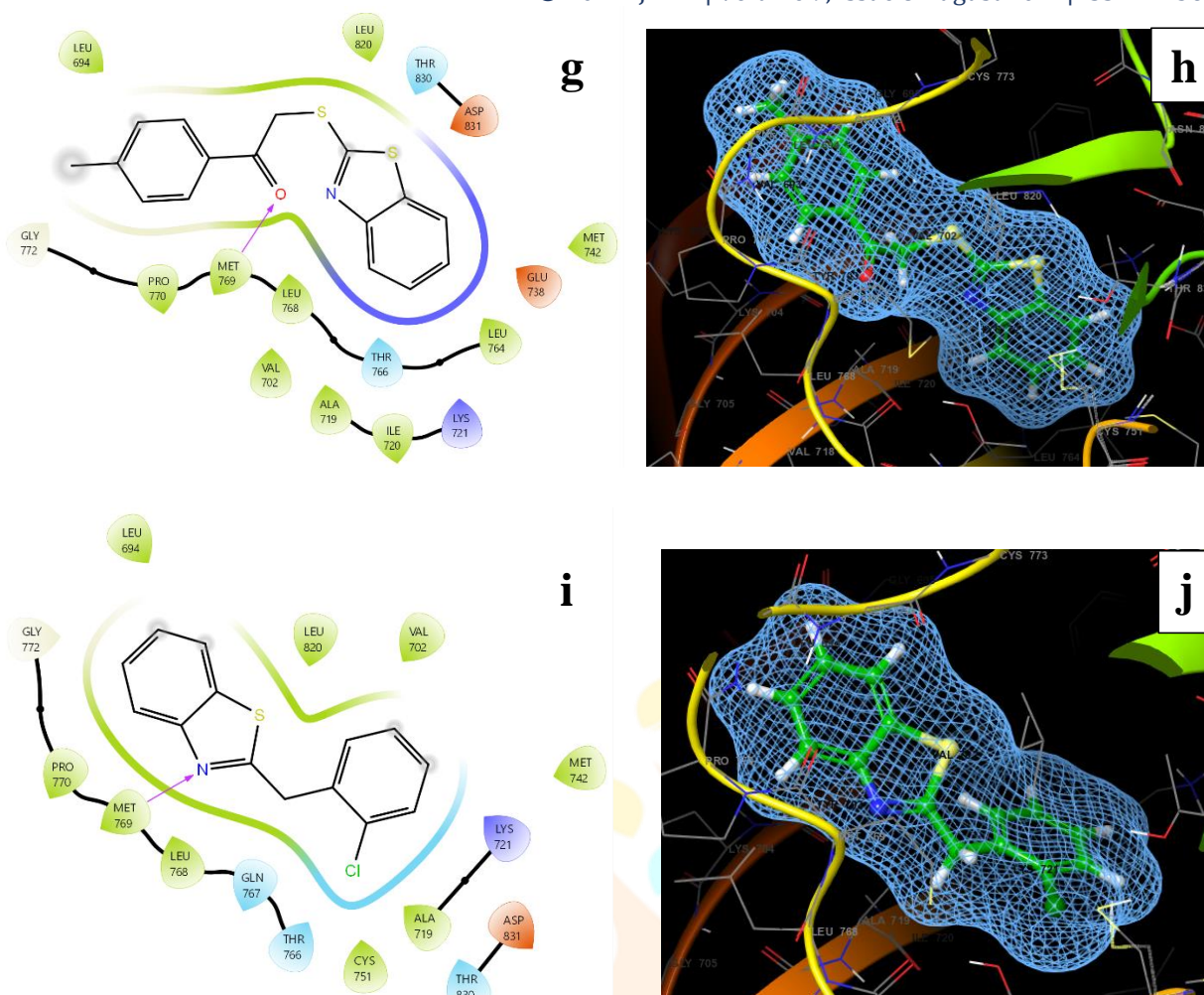
From the hits obtained, a short list of five molecules, 4302236, 11292212, 129799968, 576793, 151967363 that displayed the highest Glide XP docking score, -8.367, -7.651, -7.284, -7.191, -7.153 kcal/mol, respectively, were chosen for induced fit docking (IFD). In this protocol, both the ligands as well as the receptor undergo flexibility throughout the docking process. As observed from the interaction images, the ligand 4302236 (Fig: a) display hydroxyl (OH) bond interaction with amino acid residues ASP831, MET769 and another one interaction oxy (O) with MET769. As for the ligand 11292212 (Fig: b) they display oxy (O) bond interaction with amino acid residues LYS721, MET769 another one is hydroxyl bond (OH) with ASP831. For this ligand 129799968 (Fig: c) display oxy (O) bond interaction with amino acid residue MET769. A similar network of interaction for the ligand 576793 (Fig: d) shows the oxy (O) bond interaction with amino acid residue MET769. As per the ligand 151967363 (Fig:e) display nitrogen(N) bond interaction with amino acid residues MET769. These interactions indicating the binding of the ligands to theses amino acid residues of EGFR is beneficial in enzyme inhibition.



**Table6:** Docking scores of 5 compounds using XP mode

S. No	Compounds	Docking Score
1	4302236	-8.367
2	11292212	-7.651
3	129799968	-7.284
4	576793	-7.191
5	151967363	-7.153
Lapatinib	208908	-6.435





**Fig7:** 2D and 3D site interaction images of (a) 4302236, (b) 11292212, (c)129799968, (d)576793, (e)151967363

### In Silico ADMET Prediction

We estimated ADMET predictions for each compound in the dataset as well as encountered PubChem database hits using the Lab3.0. Results showed that compounds 1-5 possess crucial druggable characteristics. Additionally, compounds had significant caco-2 cell permeability (-5.469 to -4.663) with the acceptable range (>5.15), Human intestinal absorption shows as HIA (0.056 to 0.003) with the acceptable range (0 to 1.0), BBB (0.076 to 0.804) with the acceptable range (0 to 1.0), clearance (0.886 to 6.213),  $T^{1/2}$  (0.773 to 0.101) with the acceptable range (0 to 1.0) values indicating that they are in consistent, respectively. In Silico toxicity estimation is an essential procedure instead of drug candidate undertaken clinical trials. Computational based emanate from In Silico toxicity measurements have been broadly used due to their precision, rapidity, approachability, which can provide details about natural compounds. To determine the toxicity and adverse effect of specify 5 compounds, we used Lab3.0 online tool. This tool was used to estimate different toxicological parameters like hERG liability, human hepatotoxicity, Drug-induced liver injury (DILI), carcinogenicity for compounds, 4302236, 11292212, 129799968, 576793, 151967363 (Table).

**Table7:** ADME Profile

CI-D	4302236	11292212	129799968	576793	151967363	Acceptable range
MW	299.030	390.030	35.040	299.040	259.020	100-600
Log P	3.447	3.896	4.195	4.201	4.648	0-3
CaCo2	-5.469	-4.7	-4.719	-4.629	-4.663	>-5.15
HIA	0.056	0.004	0.008	0.004	0.003	Excellent: 0-0.3

						Medium: 0.3-0.7 Poor: 0.7-1.0
BBB	0.076	0.062	0.036	0.076	0.804	Excellent:0-0.3 Medium:0.3-0.7 Poor:0.7-1.0
CyP1A2 (inhibitor)	0.325	0.674	0.926	0.325	0.978	0-1
CyP1A2 (substrate)	0.047	0.15	0.122	0.047	0.556	0-1
CyP2C19 (inhibitor)	0.048	0.61	0.709	0.048	0.953	0-1
CyP2C19 (substrate)	0.035	0.063	0.059	0.035	0.074	0-1
Clearance	0.886	2.809	1.658	0.886	6.213	Excellent: $\geq 5$ Poor: $< 5$
$T^{1/2}$	0.773	0.143	0.039	0.773	0.101	Excellent:0-0.3 Medium:0.3-0.7 Poor:0.7-1.0

**Table8:** Toxicity Profile

CID	4302236	11292212	129799968	576793	151967363
hERG	0.114	0.013	0.443	0.14	0.016
H-HT	0.686	0.895	0.317	0.015	0.106
DILI	0.989	0.981	0.964	0.965	0.937
Carcinogenicity	0.013	0.817	0.935	0.74	0.727
Acceptable range	Excellent:0-0.3, Medium:0.3-0.7, Poor:0.7-1.0				

## Conclusion

The enormous research has been done in an effort to find more innovative medications for cancer treatment. In the current work, lead compounds for selective EGFR inhibition as a cancer treatment were found using an integrated multi-computational strategy. Structurally diverse molecules that have been reported to inhibit EGFR were used as a template to create ligand-based pharmacophore models. Three aromatic rings, one hydrophobic group, and one hydrogen bond acceptor make up the best-generated hypothesis. Using these, a 3D-QSAR model was constructed. Through the use of validation techniques including test sets, training sets, and enrichment factors, the model's resilience was guaranteed. The locations that are advantageous or detrimental in enzyme inhibition are indicated by the relevant structural properties (hydrogen bond donors, electron donating or withdrawing groups, hydrophobic regions) that are detected as contour maps. Using a combination of molecular docking and virtual screening, the model was further configured in three dimensions to map leads for EGFR inhibition. After additional examination by ligand-protein interaction insights, the procedure produced three lead compounds, of which only two were found. The compounds' drug-like characteristics were also discovered by



the in silico ADMET studies. Thus, new leads—4302236, 11292212, 129799968, 576793, and 151967363—are suggested by this work as prototypical leads for EGFR inhibition. Through the process of lead optimization and biological examinations for possible inhibition of EGFR and, thus, their efficiency in cancer therapy, the likelihood of this proposition could be further strengthened.

## Acknowledgement

One of the authors, Aastha Sharma, would like to thank the Department of Pharmaceutical Science, MDU University, Rohtak, Haryana, for their helpful guidance and providing the software assistance required for the completion the work.

## References

1. Lafaro KJ, Demirjian AN, Pawlik TM. Epidemiology of hepatocellular carcinoma. *Surgical Oncology Clinics*. 2015 Jan 1;24(1):1-7.
2. Abboud, Y., Ismail, M., Khan, H., Medina-Morales, E., Alsakarneh, S., Jaber, F., & Pyrsopoulos, N. T. (2024). Hepatocellular Carcinoma Incidence and Mortality in the USA by Sex, Age, and Race: A Nationwide Analysis of Two Decades. *Journal of Clinical and Translational Hepatology*, 12(2), 172.
3. Toh, M. R., Wong, E. Y. T., Wong, S. H., Ng, A. W. T., Loo, L. H., Chow, P. K. H., & Ngeow, J. (2023). Global epidemiology and genetics of hepatocellular carcinoma. *Gastroenterology*, 164(5), 766-782.
4. Yoon, S. K. (2018). Molecular mechanism of hepatocellular carcinoma. *Hepatoma Res*, 4(8), 42.
5. Zubair, T., & Bandyopadhyay, D. (2023). Small molecule EGFR inhibitors as anti-cancer agents: discovery, mechanisms of action, and opportunities. *International Journal of Molecular Sciences*, 24(3), 2651.
6. Nie, W., Tang, L., Zhang, H., Shao, J., Wang, Y., Chen, L., ... & Guan, X. (2012). Structural analysis of the EGFR TK domain and potential implications for EGFR targeted therapy. *International journal of oncology*, 40(6), 1763-1769.
7. Chang, Y. S., Choi, C. M., & Lee, J. C. (2016). Mechanisms of epidermal growth factor receptor tyrosine kinase inhibitor resistance and strategies to overcome resistance in lung adenocarcinoma. *Tuberculosis and respiratory diseases*, 79(4), 248.
8. Tariq, S., Kamboj, P., & Amir, M. (2019). Therapeutic advancement of benzothiazole derivatives in the last decennial period. *Archiv der Pharmazie*, 352(1), 1800170.
9. Gao, X., Liu, J., Zuo, X., Feng, X., & Gao, Y. (2020). Recent advances in synthesis of benzothiazole compounds related to green chemistry. *Molecules*, 25(7), 1675.
10. Irfan, A., Batool, F., Zahra Naqvi, S. A., Islam, A., Osman, S. M., Nocentini, A., ... & Supuran, C. T. (2020). Benzothiazole derivatives as anticancer agents. *Journal of enzyme inhibition and medicinal chemistry*, 35(1), 265-279.
11. Islam, M. K., Baek, A. R., Sung, B., Yang, B. W., Choi, G., Park, H. J., ... & Chang, Y. (2021). Synthesis, Characterization, and Anticancer Activity of Benzothiazole Aniline Derivatives and Their Platinum (II) Complexes as New Chemotherapy Agents. *Pharmaceuticals*, 14(8), 832.
12. Djuidje, E. N., Barbari, R., Baldisserotto, A., Durini, E., Sciabica, S., Balzarini, J., ... & Manfredini, S. (2022). benzothiazole derivatives as multifunctional antioxidant agents for skin damage: Structure–activity relationship of a scaffold bearing a five-membered ring system. *Antioxidants*, 11(2), 407.
13. Kumar, G., & Singh, N. P. (2021). Synthesis, anti-inflammatory and analgesic evaluation of thiazole/oxazole substituted benzothiazole derivatives. *Bioorganic Chemistry*, 107, 104608.
14. Yadav, K. P., Rahman, M. A., Nishad, S., Maurya, S. K., Anas, M., & Mujahid, M. (2023). Synthesis and biological activities of benzothiazole derivatives: A review. *Intelligent Pharmacy*.
15. Asiri, Y. I., Alsayari, A., Muhsinah, A. B., Mabkhot, Y. N., & Hassan, M. Z. (2020). Benzothiazoles as potential antiviral agents. *Journal of Pharmacy and Pharmacology*, 72(11), 1459-1480.
16. Ikpa, C. B., Onoja, S. O., & Okwaraji, A. O. (2020). Synthesis and antibacterial activities of benzothiazole derivatives of sulphonamides. *Acta Chemica Malaysia*, 4(2), 55-57.

17. Morsy, M. A., Ali, E. M., Kandeel, M., Venugopala, K. N., Nair, A. B., Greish, K., & El-Daly, M. (2020). Screening and molecular docking of novel benzothiazole derivatives as potential antimicrobial agents. *Antibiotics*, 9(5), 221.
18. Sharma, P. C., Sharma, D., Sharma, A., Bansal, K. K., Rajak, H., Sharma, S., & Thakur, V. K. (2020). New horizons in benzothiazole scaffold for cancer therapy: Advances in bioactivity, functionality, and chemistry. *Applied Materials Today*, 20, 100783.
19. El-Helby, A. G. A., Sakr, H., Eissa, I. H., Abulkhair, H., Al-Karmalawy, A. A., & El-Adl, K. (2019). Design, synthesis, molecular docking, and anticancer activity of benzoxazole derivatives as VEGFR-2 inhibitors. *Archiv der Pharmazie*, 352(10), 1900113.
20. Philoppes, J. N., & Lamie, P. F. (2019). Design and synthesis of new benzoxazole/benzothiazole-phthalimide hybrids as antitumor-apoptotic agents. *Bioorganic chemistry*, 89, 102978.
21. Islam, M. K., Baek, A. R., Sung, B., Yang, B. W., Choi, G., Park, H. J., ... & Chang, Y. (2021). Synthesis, Characterization, and Anticancer Activity of Benzothiazole Aniline Derivatives and Their Platinum (II) Complexes as New Chemotherapy Agents. *Pharmaceuticals*, 14(8), 832.
22. Liu, D. C., Gao, M. J., Huo, Q., Ma, T., Wang, Y., & Wu, C. Z. (2019). Design, synthesis, and apoptosis-promoting effect evaluation of novel pyrazole with benzo [d] thiazole derivatives containing aminoguanidine units. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 34(1), 829-837.
23. Gallego-Yerga, L., Ceña, V., & Peláez, R. (2023). Potent and Selective Benzothiazole-Based Antimitotics with Improved Water Solubility: Design, Synthesis, and Evaluation as Novel Anticancer Agents. *Pharmaceutics*, 15(6), 1698.
24. Abd El-Meguid, E. A., Moustafa, G. O., Awad, H. M., Zaki, E. R., & Nossier, E. S. (2021). Novel benzothiazole hybrids targeting EGFR: Design, synthesis, biological evaluation and molecular docking studies. *Journal of Molecular Structure*, 1240, 130595.
25. Rezeki, N., Almeahmadi, M. A., Ihmaid, S., Shehata, A. M., Omar, A. M., Ahmed, H. E., & Aouad, M. R. (2020). Novel scaffold hopping of potent benzothiazole and isatin analogues linked to 1, 2, 3-triazole fragment that mimic quinazoline epidermal growth factor receptor inhibitors: Synthesis, antitumor and mechanistic analyses. *Bioorganic chemistry*, 103, 104133.
26. Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., & Friesner, R. A. (2006). PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *Journal of computer-aided molecular design*, 20, 647-671.
27. Kirchmair, J., Markt, P., Distinto, S., Wolber, G., & Langer, T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes. *Journal of computer-aided molecular design*, 22, 213-228. P.P. Roy and K. Roy, On some aspects of variable selection for partial least squares regression models, *QSAR Comb. Sci.* 27 (2008), pp. 302–313.
28. Roy, P. P., & Roy, K. (2008). On some aspects of variable selection for partial least squares regression models. *QSAR & Combinatorial Science*, 27(3), 302-313. Golbraikh, A., & Tropsha, A. (2002). Beware of q<sup>2</sup>! *Journal of molecular graphics and modelling*, 20(4), 269-276.
29. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design*, 27, 221-234.
30. Jiao, J., Fang, H., Wang, X., Guan, P., Yuan, Y., & Xu, W. (2009). Design, synthesis and preliminary biological evaluation of N-hydroxy-4-(3-phenylpropanamido) benzamide (HPPB) derivatives as novel histone deacetylase inhibitors. *European journal of medicinal chemistry*, 44(11), 4470-4476.
31. R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, and P.S. Shenkin, Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.* 47 (2004), pp. 1739–1749.
32. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., & Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. *Journal of medicinal chemistry*, 49(2), 534-553.



33. Gupta, S., Baweja, G. S., Singh, S., Irani, M., Singh, R., & Asati, V. (2023). Integrated fragment-based drug design and virtual screening techniques for exploring the antidiabetic potential of thiazolidine-2, 4-diones: Design, synthesis and in vivo studies. *European Journal of Medicinal Chemistry*, 261, 115826.

