

Design of Novel Quinazoline-Based Allosteric EGFR Inhibitors against C797S-Mediated Resistance: An Integrated QSAR, Molecular Docking, Dynamics and ADMET Study

Abdurrahman hudu

Huduabdu1rahman45@gmail.com

Nigerian Army University Biu

Stephen Eyije Abechi

Ahmadu Bello University

Somdutt Mujwar

Chitkara University

Adamu Uzairu

Ahmadu Bello University

Mohamed El Fadili El Fadili

Sidi Mohamed Ben Abdellah University

Gideon Adamu Shallangwa Shallangwa

Ahmadu Bello University

Zakari Ya'u Ibrahim Ibrahim

Ahmadu Bello University

Sagir Hamza Abdullahi Abdullahi

Ahmadu Bello University

Research Article

Keywords: EGFR, NSCLC, QSAR, molecular docking, molecular dynamics, drug design

Posted Date: September 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10963870/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Design of Novel Quinazoline-Based Allosteric EGFR Inhibitors against C797S-Mediated Resistance: An Integrated QSAR, Molecular Docking, Dynamics and ADMET Study

Abdurrahman Hudu^{a,b,*}, Stephen Eyije Abechi^b, Somdutt Mujwar^c, Adamu Uzairu^b, Mohamed El Fadili^d, Gideon Adamu Shallangwa^b, Zakari Ya'u Ibrahim^b, Sagir Hamza Abdullahi^b

^a Department of chemistry, Nigerian Army university Biu, Borno, Nigeria

^b Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria

^c Chitkara College of Pharmacy, Chitkara University, Rajpura, 140401, Punjab, India

^d Laboratory of Engineering, Modeling, and Systems Analysis (LIMAS), chemistry department, Faculty of science Dhar El Mahraz, Sidi Mohamed Ben Abdellah University

Corresponding Author

Email address: Huduabdurrahman45@gmail.com (A.Hudu), stephenabechi@gmail.com (S.E. Abechi)

ABSTRACT

Resistance mutations, particularly C797S, have emerged and dramatically shorten the clinical durability of third generation EGFR inhibitors such as osimertinib in non-small cell lung cancer (NSCLC). For this purpose, a series of computational approaches such as 2D-QSAR, molecular docking, molecular dynamics simulation, MM/GBSA binding free energy calculation and ADMET profiling was used to design novel quinazoline based EGFR inhibitors. A predictive QSAR model was developed with 52 reported inhibitors with good internal and external validation metrics ($R^2 = 0.7811$, $Q^2 = 0.7088$, $R^2_{\text{ext}} = 0.7032$). Next, four analogues (D1–D4) with increased predicted potency to the template compound were designed. Docking analysis showed that these compounds have higher binding affinity to allosteric pocket of EGFR than osimertinib and more importantly, do not interact with C797 residue, which may help to maintain the activity against C797S-mediated resistance. Molecular dynamics simulations of 100 ns confirmed the sustained stability of the complex and MM/GBSA analysis ranked D4 as the most energetically favourable binder. All synthesized compounds were drug-like with good pharmacokinetics and toxicity profiles. Thus, compound D4 (MM/GBSA $\Delta G_{\text{bind}} = -66.5365 \text{ kcal/mol}$) can be considered as a promising allosteric lead for NSCLC therapy that needs to be validated by experimental in vitro and in vivo studies.

EGFR, NSCLC, QSAR, molecular docking, molecular dynamics, drug design

INTRODUCTION

Lung cancer remains the leading cause of cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) accounting for approximately 85% of all cases (Luke et al., 2026a), and comprising principally adenocarcinoma, squamous cell carcinoma and large cell carcinoma (Lemjabbar-Alaoui et al., 2015). Prognosis remains poor, with a five-year survival rate of approximately 25% in the United States, largely reflecting diagnosis at advanced or metastatic stages (Sharma & Khubchandani, 2024). Established risk factors include tobacco smoking, occupational exposure to asbestos, arsenic, nickel and radon, ionizing radiation, and ambient air pollution. Although surgery, radiotherapy and cytotoxic chemotherapy remain therapeutic mainstays, molecularly targeted agents have delivered superior outcomes by selectively disrupting the signalling pathways required for tumour growth, thereby limiting the off-target toxicity associated with non-selective chemotherapy (Thirunavukkarasu et al., 2021).

Receptor tyrosine kinases (RTKs) catalyze the transfer of phosphate from ATP to tyrosine residues on substrate proteins and govern cellular growth, differentiation, metabolism and motility (Sever et al., 2019). Among these, the epidermal growth factor receptor (EGFR); a member of the ErbB family comprising ErbB1/HER1, ErbB2/HER2, ErbB3/HER3 and ErbB4/HER4; is a principal regulator of proliferation and survival (Wang et al., 2025a). Under physiological conditions EGFR signalling is tightly regulated, but activating mutations lead to constitutive downstream signalling and oncogenic transformation (Kang et al., 2024). These changes are particularly common in NSCLC, where exon 19 deletions and the L858R substitution are the most frequent oncogenic drivers and validated biomarkers for targeted therapy. About 14% of NSCLC cases show EGFR mutations, which are more common in never-smokers and patients of East Asian descent (Nakamura et al., 2018).

Consequently, EGFR has become a central therapeutic target in NSCLC. First-generation reversible inhibitors (gefitinib, erlotinib) and second-generation irreversible agents (afatinib) produce substantial initial responses in EGFR-mutant disease (Tan et al., 2018), but their durability is limited by acquired resistance. The gatekeeper mutation T790M is the most extensively characterised mechanism, increasing ATP affinity and impairing inhibitor binding (Hayes et al., 2024). Third-generation inhibitors, notably osimertinib, were developed specifically to overcome T790M-mediated resistance while sparing wild-type EGFR, and act by forming a covalent bond

with Cys797 within the ATP-binding site. However, substitution of this residue by serine, the C797S mutation abolishes covalent adduct formation and represents the predominant mechanism of acquired resistance to osimertinib (Deshmukh et al., 2025; Li et al., 2023). Disease progression generally occurs after approximately 19 months of first-line and 11 months of second-line osimertinib therapy (Cheng et al., 2021; Schmid et al., 2020), highlighting the need for agents active against C797S-mutant EGFR.

To tackle this problem, several approaches are being pursued, including fourth-generation inhibitors, proteolysis-targeting chimeras (PROTACs) and allosteric modulators (Lin et al., 2026; Yu et al., 2022).

Allosteric inhibitors are of particular interest because they bind outside the ATP pocket and therefore do not depend on Cys797; EAI045, for example, exhibits selectivity for T790M/C797S-mutant EGFR while sparing the wild-type receptor (Patel et al., 2017). In parallel, the quinazoline scaffold remains a privileged pharmacophore in EGFR inhibitor development, with recently reported derivatives demonstrating inhibitory potency exceeding that of approved agents in preclinical evaluation (Nematpour et al., 2017).

Computational approaches have become integral to early-stage drug discovery, enabling prediction of biological activity and target engagement prior to synthesis and thereby reducing cost and attrition (Daoui et al., 2023; Gao et al., 2024; Talukder et al., 2025). In this study, a combination of quantitative structure-activity relationship (QSAR) modeling, molecular docking, molecular dynamics simulation, MM-GBSA binding free energy analysis and ADMET prediction has been used for the design of novel quinazoline-based inhibitors of mutant EGFR. Since loss of the Cys797 thiol for covalent binding is the major cause of resistance to osimertinib, we chose the allosteric pocket as the target site, instead of the orthosteric ATP-binding cleft. Compounds that target this regulatory site are not dependent on Cys797 for binding and therefore may retain activity in C797S mutant disease. Thus, the crystal structure of mutant EGFR in complex with a selective allosteric inhibitor (PDB ID: 5D41) was utilized for the entire docking and simulation studies.

Materials and methods

Data collection

A dataset of 52 EGFR inhibitor was retrieved from BindingDB (Ahmad, 2025) . For 2D-QSAR modeling, the IC₅₀ values were converted to molar pIC₅₀ values [$-\log_{10}$ (IC₅₀ in M)] and subsequently divided into a training set of 47 compounds for model development and a test set of 5 compounds for external validation. The structure and pIC₅₀ of the data set are presented in Table S1 (supplementary file).

Molecular structure preparation and optimization

2D structures of the compounds were drawn with ChemDraw Ultra 23.1.1 and converted into 3D models, which were pre-minimized. The molecular geometries were optimized at the B3LYP/6-31G level of theory using Spartan 14 software (D. B, 2025; Nyijime et al., 2025) so as to obtain accurate low-energy conformations for subsequent computational studies. The optimized structures were saved in SDF format for further calculation of QSAR descriptors and molecular docking.

QSAR model Development

The initial set of 797 descriptors (2D, 3D and fingerprints) was calculated using PaDEL-Descriptor v2.21 (Ahmadi et al., 2023; Ramalakshmi et al., 2021). In order to reduce the statistical noise in the descriptor matrix, 785 descriptors were removed because of missing values, near constant variance, or high intercollinearity ($r > 0.90$). The improved dataset was used to correlate descriptor features with biological activity by Multiple Linear Regression (MLR) using QSARINS software (Subramani et al., 2022). The best QSAR model was selected from twenty models with up to seven descriptors on the basis of the lowest Lack-of-Fit (LOF) and the highest predicted accuracy on the external test set (R^2_{ext}).

QSAR-model Validation and Applicability domain

The QSAR models were thoroughly validated in terms of their statistical validity, robustness and prediction ability to avoid over fitting (Shayanfar & Shayanfar, 2022). Internal validation was achieved by leave-one-out cross-validation (R^2_{cv} (Q^2)) (Er-raiy et al., 2022) and external validation was achieved by an independent test set (R^2_{ext}) (Abdullahi et al., 2025). Moreover, the cR^2_p ($cR^2_p > 0$.) measure was used to confirm the robustness of the model and Y-randomization was performed to exclude chance correlations (Kashyap et al., 2025). To determine the acceptable chemical space described by the training set and to mark possible structural outliers to guarantee the accuracy of

future predictions. The leverage cut-off value, h^* , was used to define the model's application domain (AD) (Abdullahi et al., 2021; Toropov et al., 2018).

$$R^2_{cv} = Q^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - Y_{train})^2} \quad (1)$$

$$R^2_{ext} = 1 - \frac{\sum (Y_{pred(test)} - Y_{test})^2}{\sum (Y_{test} - Y_{train})^2} \quad (2)$$

$$cR^2_p = R^2 * \sqrt{R^2 - R^2_r} \quad (3)$$

$$h^* = \frac{3(p+1)}{n} \quad (4)$$

Where Y_{obs} , Y_{pred} , and Y_{train} represents the experimental, the predicted and mean values of experimental activity of training set compounds respectively. y is the vector descriptor of the sample and Y indicates the matrix of the descriptor which are obtained from the training set descriptor values. n represents the number of training set molecules, and P , the number of descriptors (independent variables) used to develop the model.

Docking studies

Molecular docking was carried out using Schrödinger maestro (v11.1). The best docked compounds with the lowest binding energies were chosen for visualization in Discovery Studio and further analysis (Panigrahi & Sahu, 2025).

Ligand preparation

Ligand preparation was performed by using the LigPrep module of Schrödinger Maestro (Hasan et al., 2022). The 2D structures of the quinazoline derivatives were converted to 3D and energy minimized using the OPLS4 force field. Probable tautomers and protonation states at biological

pH (pH 7.4 ± 2.0) were generated using Epik. For each molecule, up to 32 stereoisomers were generated while maintaining all original chiralities.

Protein preparation

The X-ray crystal structure of mutant EGFR enzyme (PDB ID: 5D41) in complex with selective allosteric inhibitor was retrieved from Protein Data Bank and prepared utilizing the Protein Preparation Wizard. The structure was prepared by assigning bond orders, adding hydrogens, modeling missing residues and loops and removing water molecules beyond 5.0 Å from the ligand. Protonation states (Epik at pH 7.4) were assigned. The protein was optimized and minimized with the OPLS4 force field (Wang et al., 2025b) to a heavy-atom RMSD of 0.30 Å after optimization of the hydrogen bond network with PROPKA (pH 7.4).

Molecular Docking and Grid Generation

Molecular docking of the quinazoline derivatives with the active site of EGFR in GLIDE module of Schrödinger Maestro was performed. First, receptor grid was generated around the native co-crystallized ligand to map the dimensions and physicochemical properties of the active site for optimal pose scoring (M. B et al., 2021a). GLIDE Extra Precision (XP) docking protocol was used to predict and evaluate all protein-ligand interactions.

In silico drug likeness and pharmacokinetic (ADME-T) analysis

In silico toxicity and pharmacokinetic profiling of the designed quinazoline derivatives were performed to speed up the early phase drug evaluation (Panigrahi et al., 2022). pkCSM web server was used to predict the ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) parameters. Moreover, the platform SwissADME (Luke et al., 2026b) was used to establish the lipophilicity, physicochemical properties and drug-likeness, to evaluate the compounds' oral bioavailability according to Lipinski's Rule of Five.

Molecular dynamic simulation and MM-GBSA

Molecular dynamics simulations of the D1, D4 and osimertinib complexes were performed using the Desmond module in schrodinger mastro. The D1 (most favourable docking score) and D4 (highest predicted activity) were chosen for MD simulation based on docking and QSAR results, and the osimertinib complex was simulated under the same condition as a clinical reference. Systems were solvated in orthorhombic boxes with the SPC water model, with a 10 Å buffer, neutralised with counterions and adjusted to 0.15 M NaCl. The system was minimized and equilibrated using the default Desmond relaxation protocol and then simulated for 100 ns in the NPT ensemble (300 K, 1.01325 bar) using the OPLS4 force field. Trajectory stability was evaluated by RMSD, RMSF, solvent accessible surface area, radius of gyration, hydrogen bond occupancy and interaction fraction analysis, and binding free energies were computed. MM-GBSA calculations were performed on frames extracted at regular intervals over the last 50 ns of each trajectory and values are reported as mean binding free energy (M. B et al., 2021b)

RESULTS AND DISCUSSION

MODEL

$$\text{PIC50} = -9.65627 + 11.07295 (\text{RotBFrac}) + 0.02004 (\text{TDB6i}) - 0.10467 (\text{RDF120u}) \\ - 0.15469 (\text{RDF15s}) + 0.02367 (\text{RDF50s}) + 0.54524 (\text{RDF140s}) \quad (5)$$

$R^2 = 0.7811$, $R^2_{\text{adj}} = 0.7483$, $R^2 - R^2_{\text{adj}} = 0.0328$, $\text{LOF} = 0.2847$, $\text{RMSE}_{\text{tr}} = 0.3973$, $F = 23.7900$, $Q^2_{\text{loo}} = 0.7088$, $\text{RMSE}_{\text{cv}} = 0.4583$, $\text{PRESS}_{\text{cv}} = 9.8724$, $\text{CCC}_{\text{cv}} = 0.8386$, $R^2_{\text{ext}} = 0.7032$, $\text{PRESS}_{\text{ext}} = 0.8928$, $\text{RMSE}_{\text{ext}} = 0.4226$, $\text{CCC}_{\text{ext}} = 0.7806$, $Q^2 - F1 = 0.8053$, $Q^2 - F^2 = 0.6776$, $Q^2 - F3 = 0.7524$ $N_{\text{train}} = 47$, $N_{\text{test}} = 5$

The final QSAR model was built using six molecular descriptors reflecting the structural and physicochemical properties affecting pIC50. RotBFrac (fraction of rotatable bonds) describes the molecular flexibility influencing the conformational adaptability of the compound in the binding site of the biological target (Papa et al., 2018). TDB6i encodes the topological complexity and atomic connectivity. Moreover, four Radial Distribution Function (RDF) descriptors are employed to characterize the 3D spatial arrangement of atoms at different interatomic distances: RDF15s (short-range), RDF50s (medium-range), and RDF140s and RDF120u (long-range, the latter being unweighted). The importance of the RDF descriptors used in the model suggests that the 3D atomic arrangement and the molecular shape are the main determinants of biological activity with secondary effects of the molecular flexibility and topological branching.

Table 1. Predicted activities of training test and test set by the developed model

Name	RotBF rac	TDB6i	RDF12 0u	RDF1 5s	RDF5 0s	RDF14 0s	PIC50	pred. PIC50	resid ue
Training									
set	0.1578	888.24	0.0049	7.9314	55.105	1.96E-	9.6989		0.267
1	95	71	95	06	14	130	7	9.9667	7
									-
	0.1666	884.35	1.6332	10.931	33.808	6.51E-	9.3010		0.453
2	67	59	07	79	74	156	3	8.8471	9
									-
	0.1621	856.12	1.93E-	11.467	54.372	3.45E-	9.3979		0.591
3	62	82	16	19	04	295	4	8.8066	4
									-
	0.0909	928.72	0.6156	9.2244	28.796	0.0029	9.1549		0.004
4	09	58	75	94	75	44	02	9.1508	1
	0.1578	894.10	2.6111	14.412	41.953	1.5790	9.0457		0.312
5	95	26	36	74	42	12	57	9.3582	4
									-
	0.2121	875.47	2.2102	13.643	39.074	1.73E-	9.0000		0.182
6	21	57	74	17	61	11	0	8.8173	7
	0.1621	868.36	1.4719	11.281	52.027	0.3353	8.9208		0.132
7	62	24	5	68	8	04	19	9.0537	8
		776.30	1.40E-	16.126	51.905	6.96E-	6.9172		-
8	0.175	95	07	85	42	250	15	6.5702	0.347
		886.06	8.42E-	22.394	53.354		6.8996		0.657
9	0.15	04	42	7	82	0	29	7.5571	5
	0.1707	869.52	2.1113	24.257	53.286	3.12E-	6.8507		
10	32	73	27	07	91	23	81	6.9447	0.094
		831.19	1.93E-	20.102	57.365	3.63E-	6.6840		0.223
12	0.15	21	16	2	02	295	3	6.9073	3
	0.1686	853.26	0.7324	17.467	37.566	4.37E-	6.6252		0.503
13	67	22	92	72	99	56	52	7.1287	5
	0.1442	871.60	3.8209	16.276	47.303	4.85E-	6.4571		0.535
15	6	11	71	7	15	16	75	6.993	8
	0.1538	825.89	1.93E-	20.773	59.243	3.70E-	6.0814		0.702
16	46	18	16	59	26	295	45	6.7843	9
	0.2195	834.12		14.194	62.940	0.3260	7.9586		0.071
18	12	01	8.8807	48	5	24	07	8.0298	2
	0.1463	872.05	1.93E-	17.293	54.111	3.67E-	7.8538		0.189
19	41	54	16	2	87	295	72	8.0431	2
	0.1590	843.41	0.7695	17.389	72.501	5.79E-	7.8181		0.132
20	91	69	93	86	9	184	56	7.9503	1

21	0.1944	838.01	2.8871	14.078	48.512	8.31E-	7.7695		0.186
	44	72	58	73	2	76	51	7.9562	7
									-
22	0.1333	867.58	6.3811	20.423	38.318	1.2662	7.7423		0.768
	33	15	76	32	87	58	21	6.9739	5
23	0.1538	878.87	5.3541	18.459	48.827	0.7977	7.6777		0.154
	46	67	09	83	89	06	81	7.8319	1
									-
24	0.1578	863.57	0.7692	20.079	47.895	7.04E-	7.6038		0.261
	95	34	81	32	95	156	01	7.3424	4
									-
25	0.1578	843.28	1.20E-	18.796	44.583	7.25E-	7.5391		0.402
	95	7	12	33	97	278	02	7.1365	6
26	0.1860	888.92	1.5664	23.047	49.016	1.41E-	7.4089		
	47	32	42	89	96	116	35	7.646	0.237
27	0.1842	854.10	2.3002	18.504	53.573	4.09E-	7.3635		0.298
	11	55	87	25	8	26	12	7.6619	4
28		849.41	1.40E-	18.682	56.884	6.96E-	7.3124		
	0.175	11	07	57	89	250	71	7.7574	0.445
29	0.1860	891.77	1.5664	26.211	51.050	1.41E-	7.2218		0.040
	47	54	42	07	42	116	49	7.2619	1
30	0.0909	881.65	4.8390	10.035	25.799	0.0031	7.1518		0.417
	09	75	39	13	14	1	11	7.5694	6
31	0.1666	849.26	0.7549	17.133	27.566	3.37E-	7.0861		0.042
	67	22	29	72	99	67	86	7.1287	5
33	0.1388	865.88	1.6542	14.924	45.266	5.96E-	6.9625		0.858
	89	01	93	58	54	123	74	7.8207	1
34	0.1282	867.75	7.2329	15.060	33.436	3.6275	8.7958		
	05	54	22	53	27	85	8	8.8329	0.037
35	0.1818	857.78	1.93E-	12.586	56.600	3.49E-	8.7447		0.192
	18	79	16	55	63	295	27	8.9371	3
									-
36	0.1714	886.31	0.7693	15.401	38.744	7.09E-	8.7212		0.266
	29	39	2	91	53	156	46	8.4548	4
									-
37	0.1842	850.69	3.6369	15.528	54.693	1.85E-	8.6989		0.758
	11	39	73	69	92	14	7	7.9405	5
									-
38	0.1621	839.94	1.93E-	14.694	64.251	3.66E-	8.6197		0.402
	62	38	16	49	15	295	89	8.2169	9
									-
39	0.2093	849.88	3.4897	17.594	72.418	1.43E-	8.5686		0.251
	02	18	44	39	95	27	36	8.3175	1

									-
	0.1515	864.20	5.49E-	13.856	39.783	8.05E-	8.6020		0.466
40	15	19	15	65	79	289	6	8.1355	6
	0.1707	861.67	5.2379	14.270	49.800	1.3626	8.5528		0.112
41	32	75	63	64	8	34	42	8.6655	6
									-
		897.47	2.0126	15.722	44.584	2.77E-	8.4948		0.371
43	0.125	81	03	37	32	23	5	8.123	9
	0.1842	845.00	1.93E-	17.924	88.050	3.77E-	8.4559		0.170
44	11	2	16	16	04	295	32	8.6262	2
	0.1379	875.57	1.93E-	9.8572	34.639	3.47E-	8.4089		0.300
45	31	63	16	53	25	295	35	8.7098	9
	0.1860	899.07	1.2597	23.023	46.160	2.46E-	8.3467		-
46	47	94	53	19	3	62	87	7.8177	0.529
		852.34	1.93E-	13.728	58.578	3.45E-	8.3010		0.598
47	0.2	9	16	53	36	295	3	8.8996	5
									-
	0.1578	864.57	2.55E-	16.567	46.797		8.2596		0.299
48	95	59	35	3	3	0	37	7.9603	3
									-
	0.1538	841.78	1.93E-	18.016	64.227	3.63E-	8.2218		0.574
49	46	79	16	42	6	295	49	7.6473	5
									-
		847.58	1.93E-	17.360	70.039	3.61E-	8.2076		0.247
50	0.15	89	16	49	33	295	08	7.96	6
									-
	0.1388	861.51	1.6543	14.762	42.818	5.94E-	8.1549		0.454
51	89	44	16	84	35	123	02	7.7003	6
									-
	0.1538	837.23	8.76E-	16.329	62.694	7.03E-	8.0604		0.279
52	46	59	07	99	27	246	81	7.7807	8
									-
Test set	0.1463	827.63	1.93E-	24.273	79.015	3.84E-	6.8386		0.175
11	41	98	16	98	53	295	32	6.6628	8
	0.1842	821.60	5.8209	15.276	47.303	1.85E-	6.5575		0.435
14	11	11	71	7	15	14	2	6.993	5
									-
	0.1621	867.02	0.7692	15.623	36.822	6.59E-	8.0177		0.131
17	62	6	81	74	81	156	29	7.886	8
		859.30	2.3291	16.746	40.413	4.04E-	7.0343		
32	0.175	05	88	55	34	26	28	7.6214	0.587
	0.1621	861.88	0.7692	14.374	35.784	7.26E-	8.5086		-
42	62	32	81	31	02	156	38	7.9516	0.557

Table 2. Criteria of Golbraikh and Tropsha for external validation

Name	Symbol	Threshold value	Model score	
Internal validation (Training set)				
Coefficient of Determination	R^2	> 0.6	0.7811	passed
Adjusted R-squared	R^2_{adj}	> 0.5	0.7483	passed
validated Q-squared (Q^2_{LOO})	(Q^2_{LOO})	> 0.5	0.7088	passed
External validation (Test set)				
Predicted R^2 for test set	R^2_{Ext}	> 0.6	0.7032	passed
Number of tests set compounds	N_{test} set	≥ 5	5.000	passed
$\left r^2 - \frac{r_0'^2}{r^2} \right $		< 0.1	0.03506	passed
Slope of the plot of observed activity against predicted activity values at zero intercept	K'	$0.85 < k < 1.15$	0.99995	passed
Slope of the plot of predicted against observed activity at zero intercept	K	$0.85 < k < 1.15$	0.99682	passed

The statistical robustness and predictive capacity of the QSAR model (Eq. 5) were evaluated by standard validation metrics, including the coefficient of determination (R^2), leave-one-out cross-validation coefficient (Q^2_{LOO}), and root-mean-square error (RMSE). The pIC_{50} values (experimental) were compared to the model predicted pIC_{50} values (Table 1), which show a high degree of consistency and low residual errors. The structure-activity relationship of the dataset was successfully modeled with a good goodness-of-fit for the training set ($R^2 = 0.7811$) and good predictive ability for the external test set ($R^2_{ext} = 0.7032$). It can be observed from Table 2 that all the metrics for internal and external validation are within the widely accepted threshold limits (Tropsha et al., 2003), which confirms the reliability of the developed model and its applicability in predicting the EGFR inhibitory activity of novel quinazoline derivatives.

Variance Inflation Factor (VIF)

The Variance Inflation Factor (VIF) was calculated to estimate the degree of multicollinearity among the selected descriptors. Typically, VIF values of less than 5 indicate an acceptable level of intercorrelation, while values above 10 indicate severe collinearity (Adeniji et al., 2020). Table 3 reveals that all the selected descriptors had VIF values < 5 , indicating the absence of harmful multicollinearity and the statistical reliability of the model. Furthermore, the contribution of each of the descriptor was evaluated individually using a Student's t-test. All the descriptors were

statistically significant ($p < 0.01$), indicating their meaningful contributions in the predictive model. TDB6i and RDF15s were by far the most highly significant predictors ($p < 0.0001$).

Mean Effect (ME)

The relative contribution and the direction of influence of each descriptor were quantified by the Mean Effect (ME) analysis (El fadili et al., 2022) (Table 3). Largest positive mean effect was observed in RDF50s followed by TDB6i, RotBFrac and RDF140s. On the opposite side, RDF15s and RDF120u had negative mean effects. RDF15s had the strongest inverse effect on activity. We link the ME signs to the physical binding mechanism, which enables us to infer crucial structural requirements for EGFR inhibition. For example, the strong negative contribution of the RDF15s suggests that the high compactness of the short-range atomic packing could lead to steric clashes in the binding pocket. However, large positive contributions from RDF50s, RDF140s, and RotBFrac suggest that an adaptable, flexible structure (RotBFrac) capable of projecting atoms into favorable medium and long-range spatial arrangements (RDF50s, RDF140s) is required for target affinity maximization.

Table 3. Correlation matrix for descriptors, P-value, VIF and ME of the developed model

	RotBFrac	TDB6i	RDF120u	RDF15s	RDF50s	RDF140s	P-value	VIF	ME
RotBFrac	1						0.00049	1.84	0.214
TDB6i	-0.346	1					2.1×10^{-9}	2.66	0.462
RDF120u	0.1259	-0.0093	1				0.0072	1.92	-0.031
RDF15s	0.1996	-0.129	-0.076	1			1.0×10^{-11}	3.11	-0.298
RDF50s	0.3547	-0.4411	-0.2005	0.3309	1		0.00035	2.87	0.587
RDF140s	-0.1987	0.1359	0.5499	-0.0549	-0.2039	1	0.00011	1.76	0.066

Domain of applicability

Applicability domain analysis (Williams plot, Figure 1.) shows the cut-off leverage (h^*) to be 0.447. The Williams plot identified compound 34 as the sole training set molecule with leverage exceeding the warning threshold. This arises from its distinctive position in descriptor space exhibiting the highest RDF140s (3.628) and second highest RDF120u (7.233) of all 52 compounds, together with one of the lowest rotatable bond fractions (0.128), consistent with a rigid, extended molecular architecture that is under-represented in the remainder of the series. Rather than compromising the model, this compound provides experimental validation of the trends identified by the mean effect analysis. It couples a high value of the positively contributing RDF140s descriptor with a below median value of the strongly negative RDF15s descriptor, and is correspondingly one of the most potent inhibitors in the dataset ($pIC_{50} = 8.796$). Its standardized residual lies well within $\pm 3\sigma$ (Abdullahi et al., 2025), with an absolute residual of only 0.037, confirming it as an influential but accurately predicted observation rather than an outlier. All other compounds in both sets fell within the applicability domain.

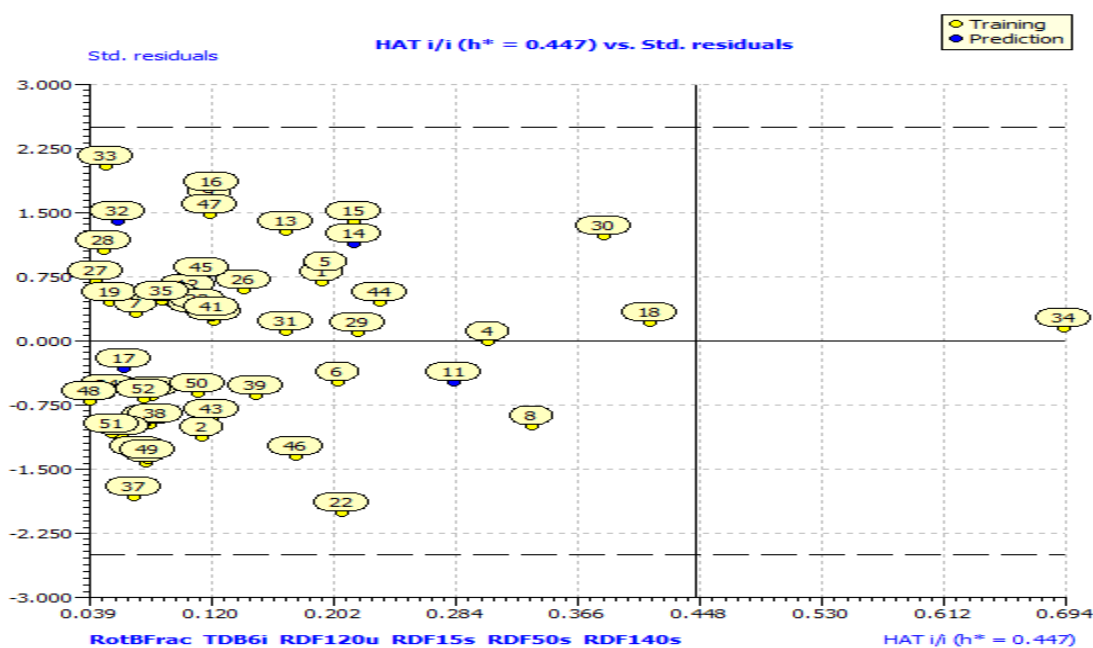


Fig. 1. William's plot of the build model

Scrambling or Y-randomization

To confirm that the developed QSAR model was not a result of chance correlation, a Y-randomization test was performed. The dependent variable (pIC_{50}) was randomly scrambled 50 times against the unperturbed descriptor matrix, generating 50 randomized models (Table S2 in supplementary file). The average correlation coefficient of these randomized models ($R_{\text{rand}} = 0.3618$) was drastically lower than that of the original model ($R_{\text{original}} = 0.8838$). The calculated cR^2_{p} value of 0.7167 (Table 4) comfortably exceeded the standard acceptability threshold of 0.5. These observed results conclusively demonstrate that the robust structure-activity relationship established by the six descriptors is genuine and not a product of accidental statistical correlation. Furthermore, the average cross-validation coefficient of the randomized models ($Q^2_{\text{rand}} = -0.20981$) which is within the acceptable limit ($Q^2_{\text{rand}} < 0.05$) was notably poor, further confirming the absence of chance correlation.

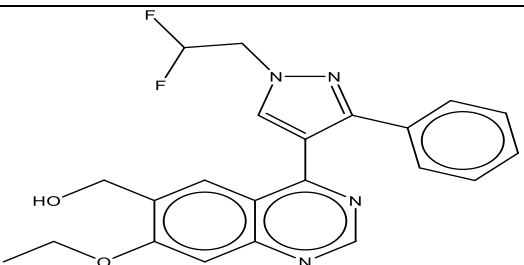
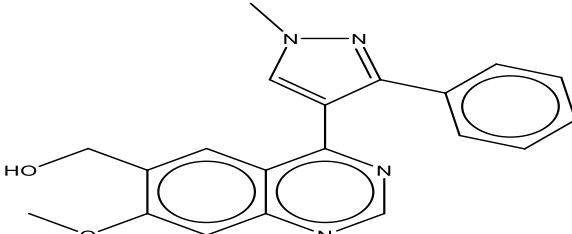
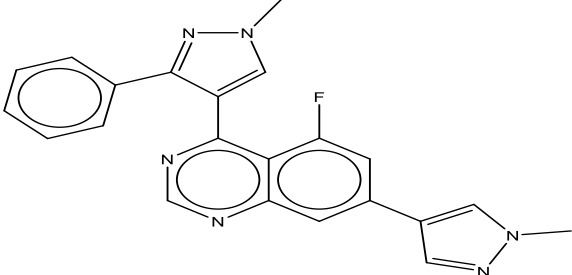
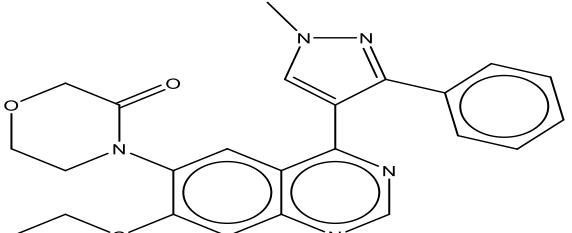
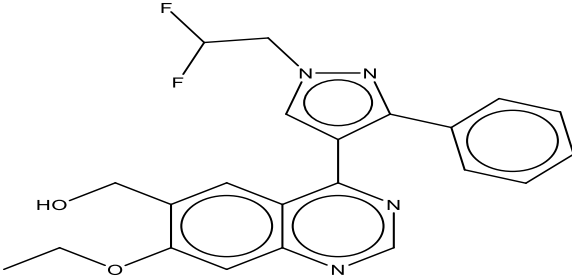
Table 4. MLR Y-randomization model parameter

Random Models	Parameters
Average R	0.361806
Average R^2	0.144809
Average $Q^2(\text{LOO})$	-0.20981
cR^2_{p}	0.716727

Design of new compounds

Virtual screening led to the docking of the entire data set into the binding pocket of EGFR (PDB ID: 5D41), with the respective binding affinities. The 5 top ranked compounds are presented in Table 4. Compound 6 was selected as the design template (Figure 2) due to its good docking score (-10.41 kcal/mol), high experimental potency ($\text{pIC}_{50} = 9.00$) and a small residual value (-0.1827) indicating high accuracy of the QSAR model in predicting this scaffold. Based on this template, four new quinazoline analogues (D1–D4) were designed (Figures 3), and their molecular descriptors and predicted activities were calculated (Tables 5). All four of our designed compounds showed predicted pIC_{50} values higher than both template and parent dataset indicating enhanced inhibitory potential. The activity gradually increased from D1 to D4 and D4 was the most predicted potent, showing that the structural modifications introduced were effective and that D3 and D4 were the most promising designed candidates for further investigation.

Table 4. Structure, PIC50 and docking scores of the top docked dataset

S/N	STRUCTURE	IC50(nm)	PIC50	Docking score
1		1	9.000	-10.41
2		3.9	8.408935	-10.012
3		0.7	9.154902	-9.857
4		109	6.962574	-9.63
5		0.9		-9.462

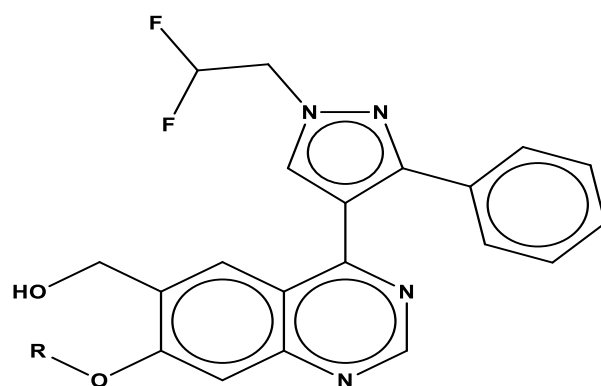


Fig 2. Structure of Template

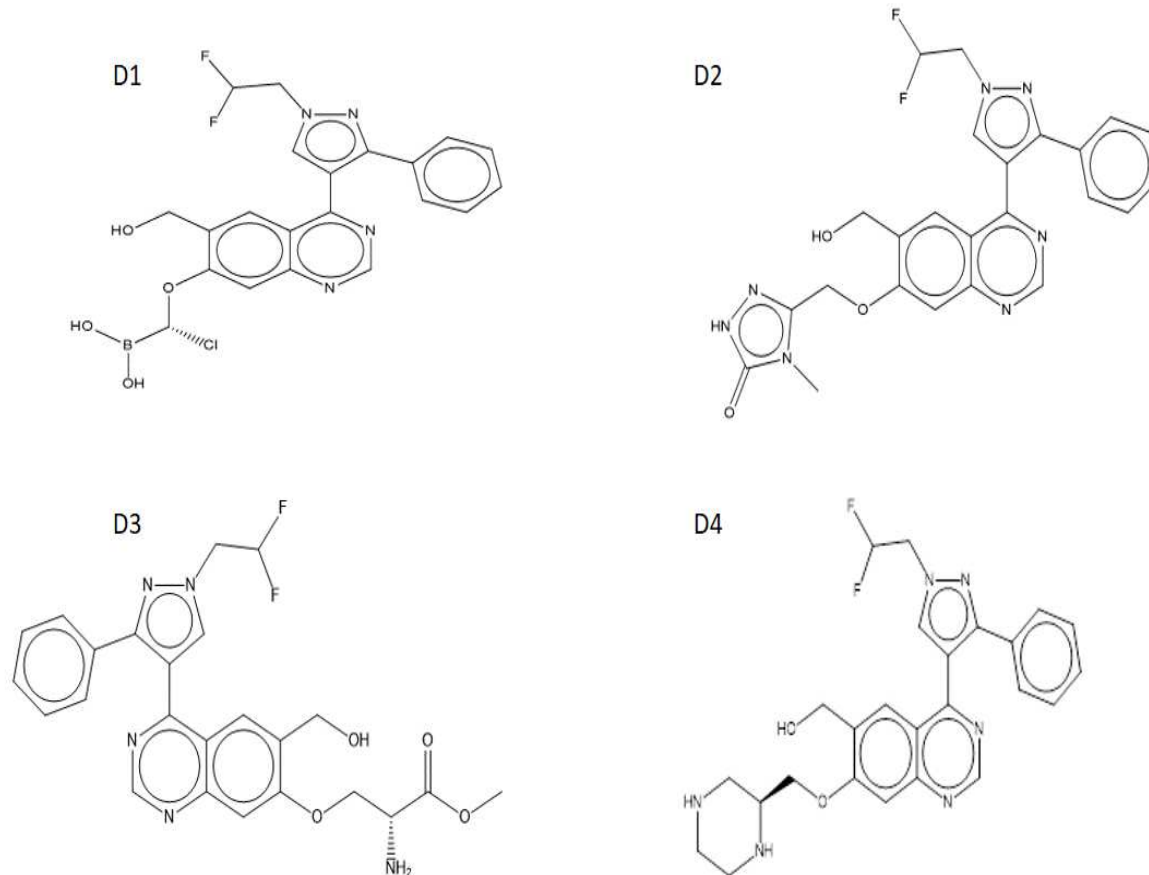


Fig.3. The structure of design compound and their predicted activities

Table 5. Generated descriptors of the design compounds

Comp.ID	RotBFrac	TDB6i	RDF120u	RDF15s	RDF50s	RDF140s	PIC50
D1	0.222222	883.096	3.976174	8.002017	38.68082	1.317172	10.4785
D2	0.2	905.5262	6.459623	9.665942	34.90699	2.962393	10.9723
D3	0.263158	880.4913	12.56028	11.39942	31.17086	4.528506	11.0288
D4	0.205128	883.0701	13.13224	13.66634	42.62691	6.43413	11.3375

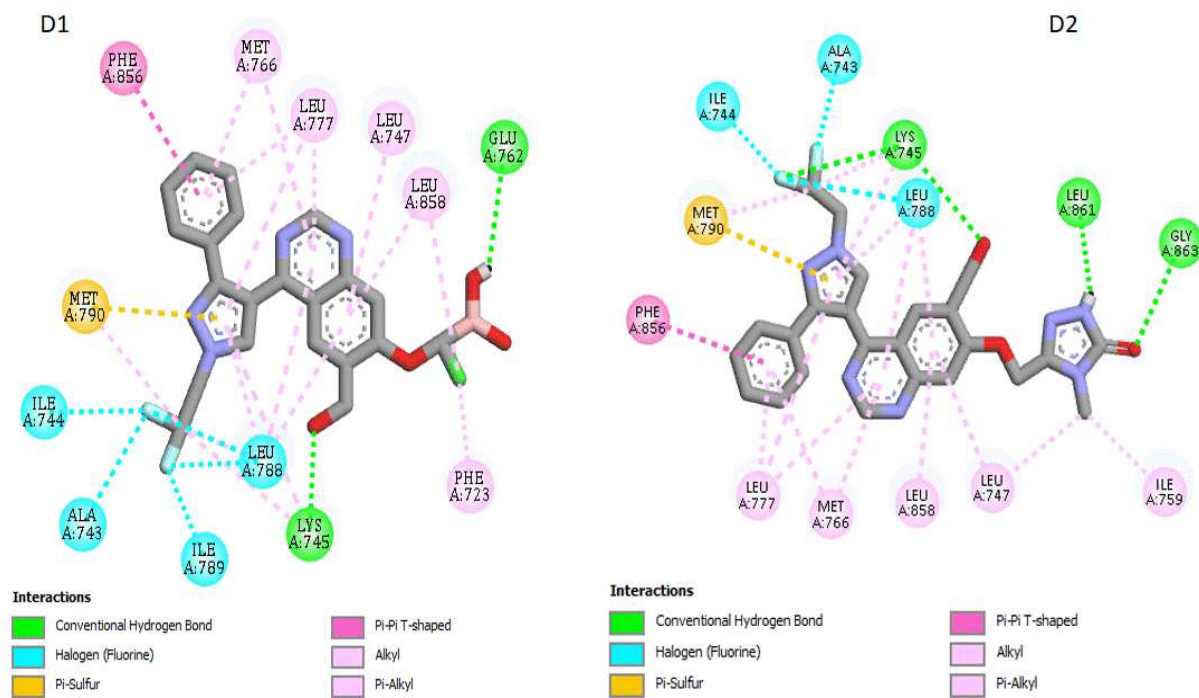
Docking studies

The binding modes of the designed compounds with the mutant EGFR active site of (PDB ID: 5D41) were using molecular docking in comparison with Osimertinib (reference ligands), the

docking interaction scores and interacting residues are given in Table 6. The binding affinities of all four designed compounds (D1–D4) were found to be strong, with D1 showing the best affinity score (−11.428 kcal/mol) followed by D2 (−11.134), D3 (−11.032) and D4 (−10.594). These values are comparable to the co-crystallised ligand (−10.428 kcal/mol) and the reference drug Osimertinib (−7.556 kcal/mol), suggesting better predicted affinity for the designed analogues. The 2D interaction profiles (Figures 4) indicate that all the designed compounds have consistent interaction with a key residue of the active site such as ALA-743, LYS-745, MET-766, LEU-777, LEU-788, MET-790 and PHE-856, similar to the interaction pattern of the co-crystallised ligand (Figure 5) and confirming occupancy of the same binding region. D1 made two hydrogen bonds (LYS-745 and GLU-762), D2 made three (LYS-745, LEU861 and GLY-863), D3 made two (LYS-745 and GLY-863) while D4 interacted with ARG-748, CYS-775 and ASP-855. These complexes were further stabilized by hydrogen bonding. Importantly, LYS-745 was occupied by co-crystallized ligand, Osimertinib (Figure 5) and all the designed compounds except D4. None of the designed compounds interacted with the C797 residue consistent with their allosteric binding mode, suggestive of potential activity against C797S mediated resistance. The docking analysis reveals D1-D3 as the leading candidates due to their combined binding affinities and interaction profiles. Although QSAR ranks D4 highest, docking favors D1, a divergence that reflects the complementary nature of the two approaches, as QSAR captures whole-molecule physicochemical trends whereas docking evaluates static pose-specific complementarity. Therefore, the MD simulations were employed to discriminate between these predictions.

Table 6. Docking interaction and total scoring of designed compound and control with EGFR receptor

Compound	Residue	scoring
D1	PHE;856,LEU;777,MET;766,ILE;789,GLU;762,PHE;723,LEU;858,LEU;788,LYS;745,MET;790,ILE:744,ALA:743,LEU:747	-11.428
D2	LEU;858,LEU;788,ALA,743,LYS;745,MET;766,MET;790,LEU;777, PHE;856,ILE;759,ILE:744, ALA:743, LEU:861, GLY:863,	-11.134
D3	MET;790, LEU;777, MET;766, PHE;856, ASP;855, ILE;759, LEU;858, LEU,788, ILE;789, LYS;745, ALA:859, GLY:863, ALA,743, ILE:744, LEU;862, LEU:747	-11.032
D4	MET;766,PHE;723,ASP;855,CYS;775,ILE;759,GLU;866,LEU;858, LEU;788,ARG:748,LEU:747	-10.594
Co-crystalize	MET;766,LEU;777,LEU;788,VAL;726,MET;790,ALA;743,LYS:745 ,PHE;856,LEU;858, ASP:855	-10.428
Osimertinib	GLU;762,LEU;747,LEU;858,LEU;777,LEU;788,VAL;726,MET;766 ,ALA;743,ILE;759,LYS;745,MET:790, LEU:844,ALA:763	-7.556



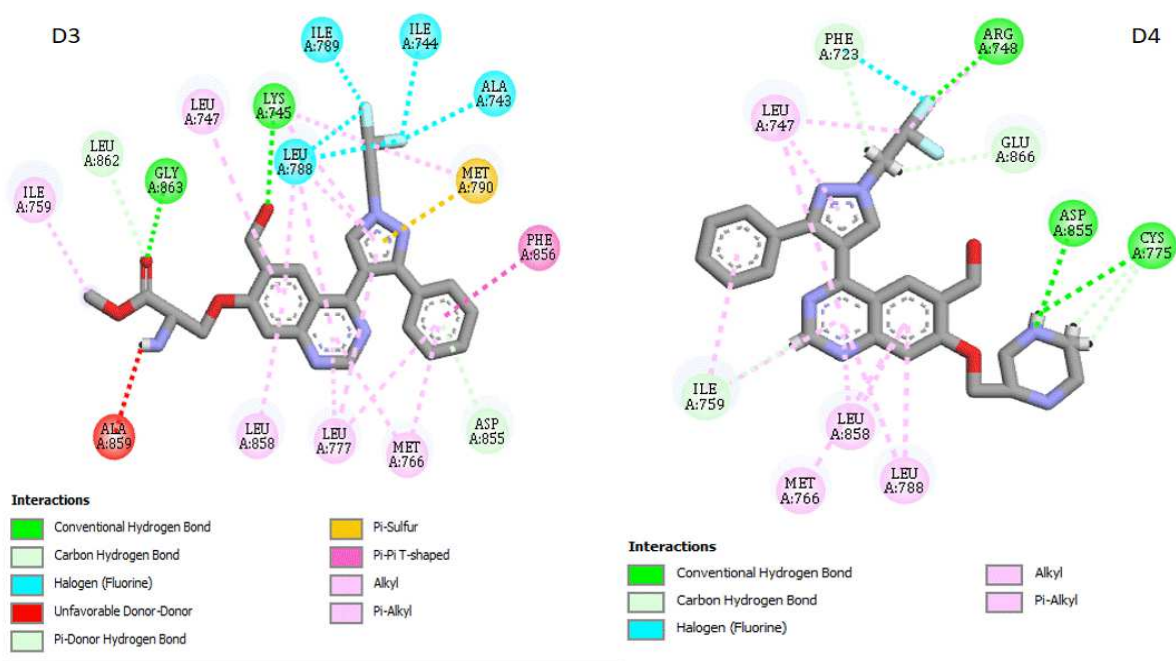


Fig. 4. 2D interaction of designed compounds with the EGFR receptor

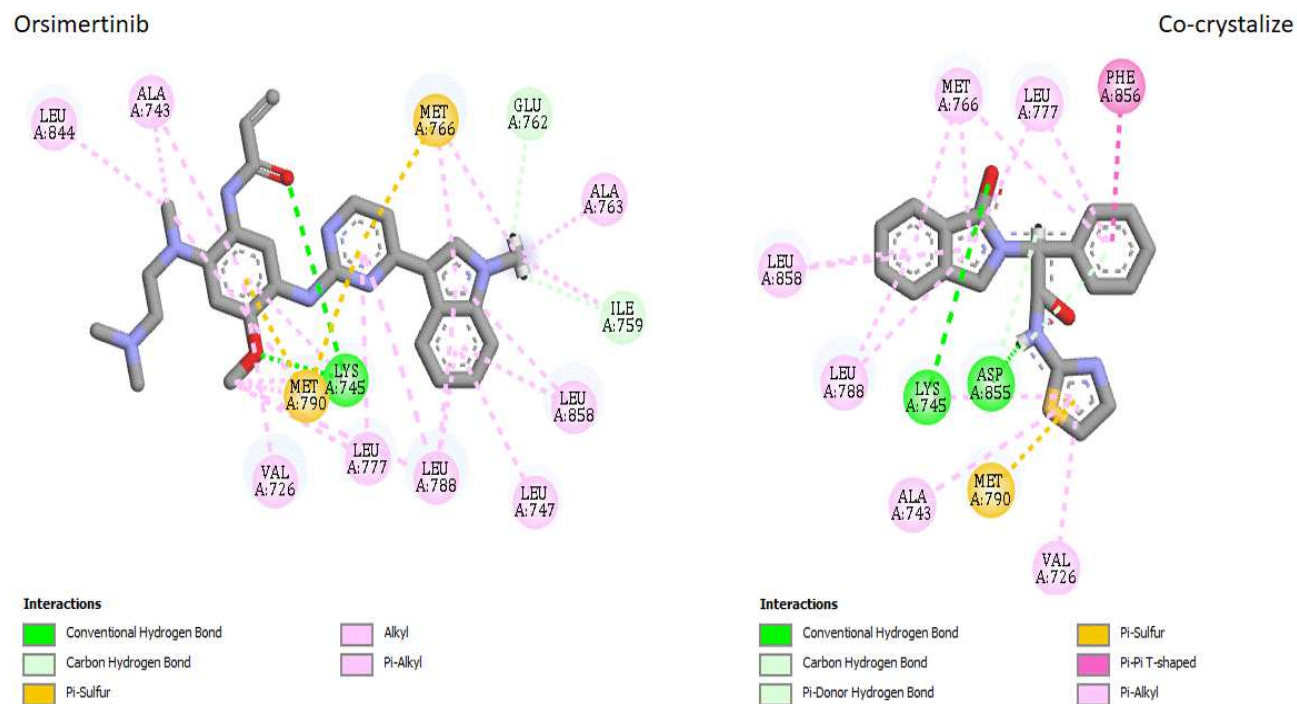


Fig. 5. 2D interactions of osimertinib and co-crystallize compound with the EGFR receptor

Docking validation protocol

To check docking method reliability in this study, the co-crystallized was re-docked into the EGFR (PDB ID: 5D41). As shown in Fig. 6, the re-docked ligand aligns very closely with the original ligand in the binding pocket. The RMSD value obtained is 0.3866 Å, which is well below the acceptable limit of 2 Å. This indicates that schrodinger can accurately predict the binding mode and is suitable for this type of study.

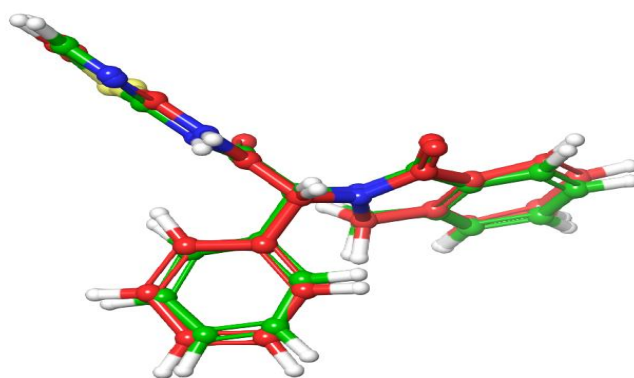


Fig.6. Re-docking poses of with RMSD value of 0.39 Å (Orange, original; Green, docked)

Lipinski's rule and ADMET prediction

The drug-likeness of the designed compounds was predicted with respect to the Lipinski rule of five (Table 7). The LogP values for all the four analogues were found to be less than 5, with acceptable molecular weights and the number of hydrogen bond donors and acceptors, indicating a balanced hydrophilic–lipophilic character. Additional filters were also used to evaluate the compliance (Table S3 in supplementary file). All compounds passed Lipinski, Veber and Egan criteria. D2, D3 and D4 were out of Ghose criteria and so was osimertinib. The bioavailability score of 0.55 is the same for all and the synthetic accessibility values are in the range of 3.10-4.03 indicating a moderate potential for oral absorption and synthetic feasibility.

ADMET properties of the designed compounds of the co-crystallized ligand and osimertinib are summarized in Table 8. The four designed compounds were all predicted to act as substrate and

inhibitor of P-glycoprotein. While P-gp substrate recognition can in principle restrict oral absorption and facilitate efflux-mediated resistance, several observations pointed that this is unlikely to compromise the developability of the present series. The first was the prediction of a similar dual substrate/inhibitor profile for osimertinib and the co-crystallised ligand, which confirms that this feature is compatible with clinically successful EGFR inhibitors. Second, all analogues showed high predicted human intestinal absorption (80.5–92.2%) with D1 (92.24%) being very similar to the co-crystallised ligand (92.99%) and better than osimertinib (90.38%), suggesting that efflux transport is not a major limitation to absorption. Third, simultaneous P-gp inhibition may reduce self-efflux at therapeutic concentrations and, because P-gp overexpression is a well-established contributor to multidrug resistance in NSCLC, may provide further benefit. Total clearance values (0.561–0.856 mL/min/kg) are also consistent with that of osimertinib (0.663) and indicate sufficient systemic exposure. However, dual P-gp and CYP3A4 inhibition requires careful assessment of the drug-drug interaction potential in subsequent experimental studies. The designed compounds had lower Caco-2 permeability values than osimertinib and the co-crystallised ligand, though these were uniformly high for predicted human intestinal absorption (>80%) suggesting that in vivo absorption is unlikely to be limiting. All the designed compounds showed negative LogBB values (–1.806 to –1.105) indicating limited exposure in the central nervous system and hence lower risk of CNS-related toxicity. However, it is worth noting that the osimertinib was also predicted to have poor BBB-penetration (LogBB= –1.062) despite its documented clinical CNS activity, which indicates that these insilico estimates could underestimate brain distribution for this chemo type and should be experimentally confirmed. These results (taken together) indicate that the designed compounds have acceptable pharmacokinetic and safety profiles, and D1 and D4 are the most promising candidates for further development.

Table 7. Lipinski's properties, TPSA and NROTB of the newly designed compound

Inhibitors	property					
	Mw	LogP	HBA	HBA	TPSA	NROTB
D1	474.65	2.8734	3	9	113.52	8
D2	493.47	2.9183	2	9	123.74	8
D3	483.47	2.7968	2	10	125.38	10
D4	480.51	2.8579	3	9	97.12	8
Co-crys	349.41	3.4789	1	3	90.54	5
Osimert.	499.61	4.5098	2	5	87.55	11

Table 8. Insilico ADMET prediction of the newly designed inhibitors

ADMET properties	Compounds					
	D1	D2	D3	D4	Co-cryst	Osimertinib
Water solubility	-4.137	-4.275	-4.603	-3.99	-4.567	-5.12
CaCO ₂ permeability	0.178	-0.077	0.091	0.38	1.062	0.998
Human Intestinal Absorption (HIA)	92.24	80.523	83.613	84.917	92.997	90.38
BBB permeability (LogBB)	-1.806	-1.523	-1.503	-1.105	-0.021	-1.062
Skin permeability (Logkp)	-2.735	-2.857	-2.926	-2.997	-3.281	-2.938
p-glycoprotein substrate	YES	YES	YES	YES	YES	YES
p-glycoprotein inhibitor	YES	YES	YES	YES	YES	YES
CYP2D6 substrate	NO	NO	NO	NO	NO	NO
CYP3A4 substrate	YES	YES	YES	YES	YES	YES
CYP2D6 inhibitor	NO	NO	NO	NO	NO	NO
CYP3A4 inhibitor	YES	YES	YES	YES	NO	YES
Total clearance (mL/min/kg)	0.625	0.856	0.561	0.716	-0.01	0.663
Renal OCT2 inhibitor	NO	NO	NO	NO	NO	NO
hERG inhibitor	NO	NO	NO	NO	NO	NO
AMES toxicity	NO	NO	NO	NO	NO	NO
Skin sensitization	NO	NO	NO	NO	NO	NO
synthetic Accessibility	3.92	3.64	4.03	4.12	3.10	4.01
Bioavailability score	0.55	0.55	0.55	0.55	0.55	0.55

MOLECULAR DYNAMICS SIMULATION

To evaluate the dynamic behaviour and conformational stability of the designed inhibitors within the mutant EGFR binding site, 100 ns molecular dynamics simulations were performed for the D1 and D4 complexes, with the osimertinib complex simulated in parallel as a reference.

Stability Analysis of the Protein–Ligand Complexes

Backbone and ligand RMSD.

The C α RMSD profiles (Figure 7) indicate that both designed complexes attained equilibrium within the simulation period. The D4 complex converged most rapidly, plateauing after approximately 20 ns and maintaining a stable backbone deviation of 2.4–2.8 Å throughout the

remainder of the trajectory. The D1 complex exhibited an initial adjustment phase before stabilising after approximately 40 ns at 2.4–2.8 Å. By contrast, the osimertinib reference complex displayed a progressive upward drift from 1.2 Å to approximately 3.2 Å, reaching a plateau only after 80 ns, indicating slower conformational relaxation. The RMSD analysis of the ligand showed that the designed compounds remained securely accommodated within the binding site, whereas osimertinib fluctuated between 2.4 and 3.3 Å, reflecting greater positional mobility within the pocket.

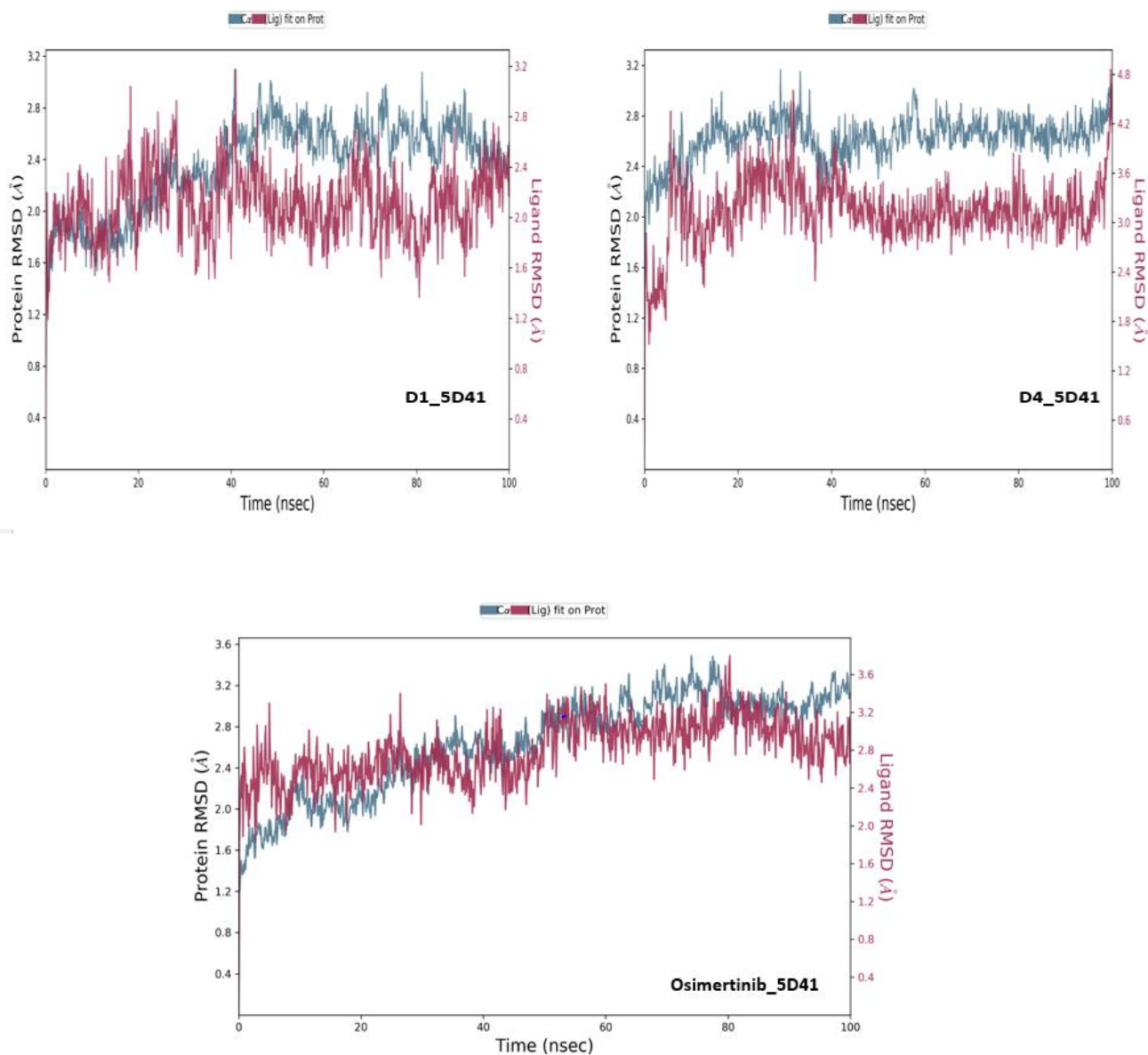


Fig. 7. Protein-ligand RMSD trajectories of complexes D1_5D41, D4_5D41 and Osimertinib_5D41

Residue Flexibility Analysis (RMSF)

RMSF analysis of the C α atoms (Figure 8) revealed that ligand binding differentially rigidified the receptor. The D4 complex exhibited the lowest overall flexibility, with the majority of residues fluctuating below 1.5 Å and a baseline of 0.5–1.0 Å; elevated mobility was confined to the terminal region beyond residue 280, which is solvent-exposed and not involved in binding. In contrast, both the D1 and osimertinib complexes displayed pronounced flexibility in the loop region near residue 170, reaching maxima of approximately 4.7 Å, together with additional mobility around residues 40–50. Critically, the residues comprising the allosteric binding site remained rigid (< 1.2 Å) in all systems, confirming that ligand engagement stabilises the pocket. The markedly reduced global fluctuation observed for D4 indicates that this compound confers the greatest conformational restraint on the receptor.

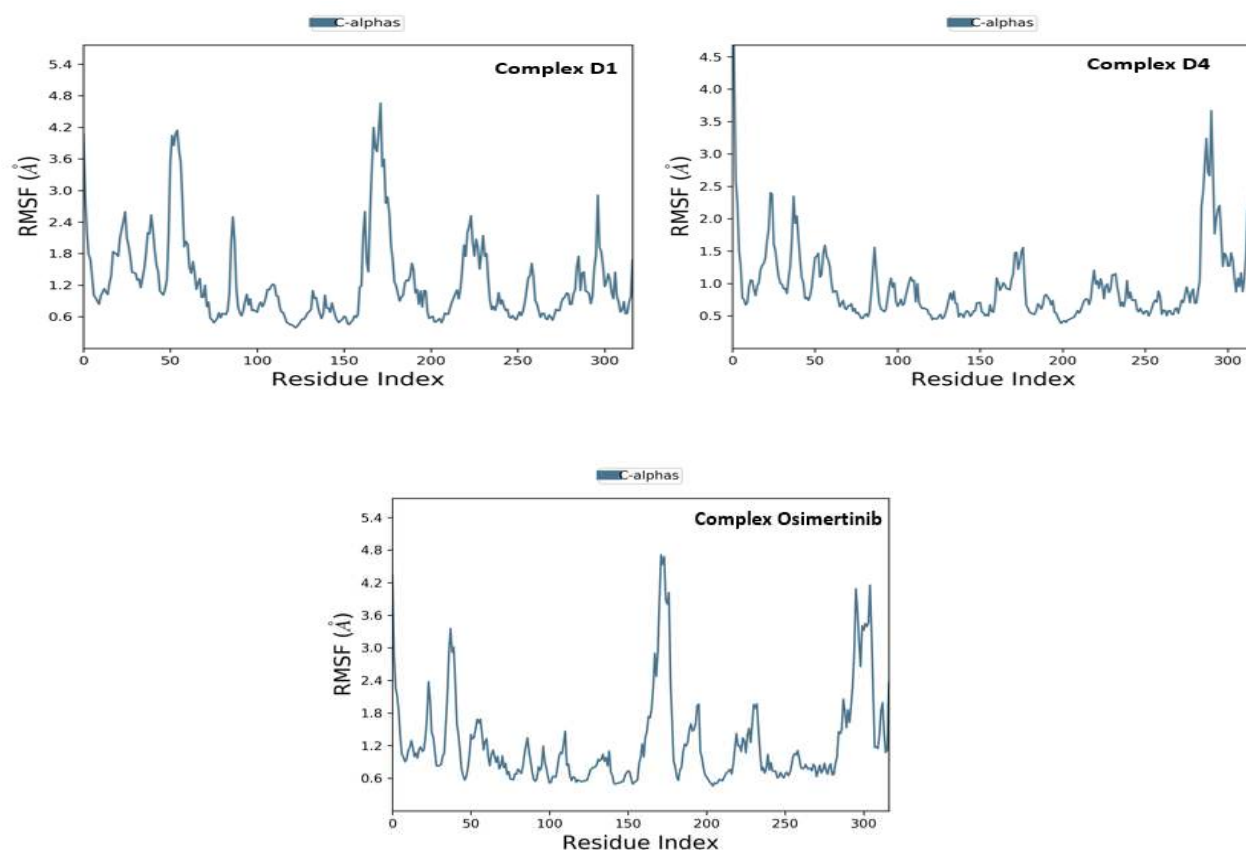


Fig.8. Root means square fluctuation (RMSF) graph of the complexes over the 100 ns MD simulation; D1_5D41, D4_5D41 and Osimertinib_5D41

Ligand conformational properties

The radius of gyration (Figure 9) remained essentially invariant for both designed compounds (D1: 4.6–4.9 Å; D4: 4.8–5.2 Å), indicating retention of a compact bound conformation, whereas osimertinib adopted a more extended geometry (5.3–5.7 Å). Solvent accessible surface area was consistently lowest for D4 (25–50 Å²), intermediate for D1 (25–75 Å²), and substantially higher for osimertinib (60–120 Å²), demonstrating that D4 remains the most deeply buried within the hydrophobic allosteric pocket and is least exposed to solvent. Intramolecular hydrogen bonding was sparse and transient in all systems, indicating that the ligands do not adopt self-stabilised folded conformations that would compete with receptor engagement.

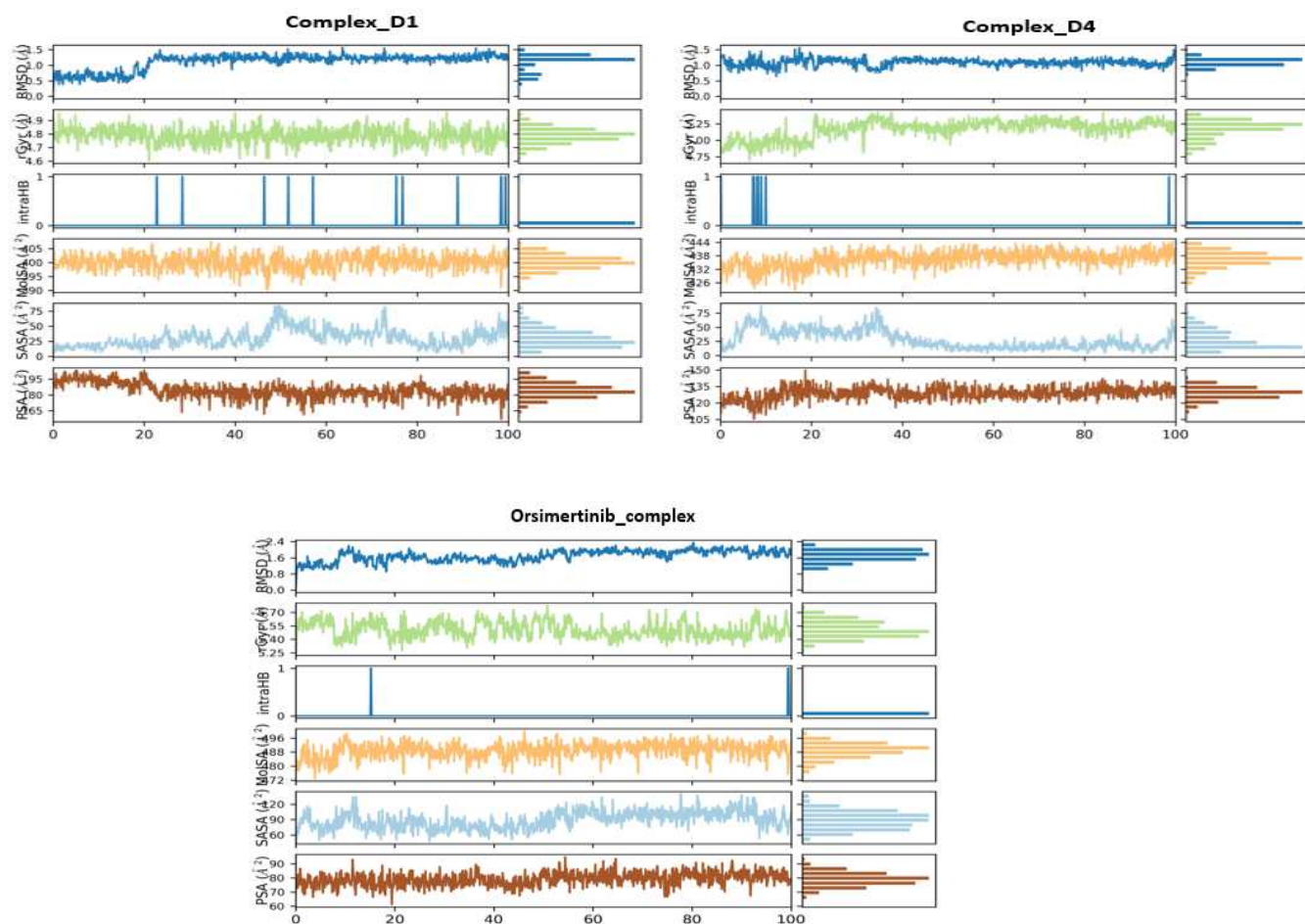


Fig.9. Ligand property trajectory of the selected complexes; D1_5D41, D4_5D41 and Orsimertinib_5D41

Protein-Ligand Interaction Analysis

The protein-ligand interaction profiles derived from the molecular dynamics simulations revealed the key amino acid residues engaged in maintaining ligand binding within the EGFR active site throughout the 100 ns simulation period. Interaction fraction indicates the proportion of simulation time during which a particular residue remained in contact with the ligand through hydrophobic interactions, hydrogen bonds, ionic contacts, or water-bridges. These interactions are presented in figure 10. The interaction histograms showed that D1, D4 and osimertinib stable interactions with key residues in the EGFR binding pocket throughout the 100 ns simulation. The major interactions for D1 were LYS-745, GLU-746, ARG-748, MET-766, LEU-777, LEU-788, PHE-856, LEU-858, LEU-861, GLU-863 and GLU-866 via hydrogen bond, hydrophobic contacts, water bridges and ionic interactions. Among these, LYS-745 showed the highest frequency of interaction indicating its significant role in stabilization of ligand. D4 formed a stable interaction network with LYS-745, LEU-747, MET-766, CYS-775, LEU-788, ASP-854, PHE-856, LEU-862 and GLU-863. PHE-856 contributed maximum fraction of interaction mainly through hydrophobic and water-mediated interactions while ASP-854 further enhanced the binding by hydrogen-bonding and ionic interactions. The persistent contacts indicate that D4 maintained a well-oriented and stable binding pose throughout the simulation. Osimertinib mainly interacted with LYS-745, MET-766, LEU-777, LEU-788, MET-790, ARG-841, ASN-842, LEU-844, ASP-855, and LEU-861. Among these residues, ASP-855 was the most frequently interacting by hydrogen-bonding, ionic and water-bridge interactions, while the other residues were mainly involved in hydrophobic contacts. The interaction profiles overall confirmed that all three ligands established stable complexes with EGFR, with D4 displaying the most persistent and favorable interaction network during the simulation.

(−39.31 kcal/mol). Decomposition of the energy terms indicates that binding is driven predominantly by van der Waals (D1: − 36.56; D4: − 31.60 kcal/mol) and lipophilic contributions (D1: −14.79; D4: −20.24 kcal/mol), consistent with the hydrophobic character of the allosteric pocket and with the low solvent exposure observed for these ligands. The superior affinity of D4 arises principally from its enhanced lipophilic complementarity and a markedly favourable generalised born solvation term (−15.75 kcal/mol), whereas osimertinib incurred a substantial electrostatic desolvation penalty that offset its otherwise favourable van der Waals contribution (−59.62 kcal/mol). Taken together, the RMSD, RMSF, ligand property, interaction and binding free energy analyses consistently identify D4 as the most stable and energetically favourable complex. Although molecular docking ranked D1 highest on the basis of static pose scoring, the dynamic analyses which account for conformational sampling, solvation and entropic accommodation establish D4 as the superior candidate, and this is in agreement with the QSAR model, which also predicted D4 to be the most potent analogue.

Table 9. Binding energies of the top hit complexes (MM/GBSA).

COMPLEX ID	MMGBSA dG Bind (kcal/mo)	MMGBSA dG Bind Coulomb (kcal/mo)	MMGBSA dG Bind Covalent (kcal/mo)	MMGBSAdG Bind Hbond (kcal/mo)	MMGBSA dG Bind Lipo (kcal/mo)	MMGBSA dG Bind Packing (kcal/mol)	MMGBSA dG Bind Solv GB (kcal/mol)	MMGBSA dG Bind vdW (kcal/mol)
D1	-49.8598	1.96154	-2.93145	-1.06552	-14.7918	-1.87657	5.407883	-36.5598
D4	-66.5365	18.46865	-19.7982	1.694413	-20.2436	0.751466	-15.7478	-31.5996
osimertinib	-39.3117	71.99904	-6.19477	1.515235	-21.072	0.181462	-25.936	-59.6203

CONCLUSION

An integrated computational workflow encompassing 2D-QSAR modelling, molecular docking, molecular dynamics simulation (MD), MM-GBSA analysis, and ADMET prediction was applied to the discovery of novel quinazoline-based EGFR inhibitors for resistant non-small cell lung cancer. A statistically robust QSAR model guided the rational design of four derivatives with improved predicted activities. Although D1 achieved the most favourable docking score, D4 emerged as the superior candidate following dynamic evaluation, forming the most conformationally stable complex and yielding the strongest binding free energy (−66.54 kcal/mol) relative to osimertinib (−39.31 kcal/mol). Critically, the designed compounds avoided contact with CYS797 throughout, consistent with an allosteric binding mode with potential utility against

C797S resistance. All analogues displayed acceptable drug-likeness, favourable pharmacokinetics, and low predicted toxicity. D4 is therefore proposed as a priority lead for synthesis and experimental validation. It should be emphasised that the present findings are derived entirely from computational prediction; synthesis followed by in vitro kinase inhibition and cellular assays against C797S-mutant EGFR is required to confirm the proposed activity.

REFERENCES

- Abdullahi, S. H., Patil, R. B., Uzairu, A., Ibrahim, M. T., Abdullahi, M., & Saje, B. A. (2025). Unveiling novel VEGFR-2 inhibitors: QSAR modeling, molecular docking, MD and MM-GBSA calculations, pharmacokinetics profiling and DFT studies. *Scientific African*, 28, e02692. <https://doi.org/10.1016/j.sciaf.2025.e02692>
- Abdullahi, S. H., Uzairu, A., Ibrahim, M. T., & Umar, A. B. (2021). Chemo-informatics activity prediction, ligand based drug design, Molecular docking and pharmacokinetics studies of some series of 4, 6-diaryl-2-pyrimidinamine derivatives as anti-cancer agents. *Bulletin of the National Research Centre*, 45(1), 167. <https://doi.org/10.1186/s42269-021-00631-w>
- Adeniji, S. E., Arthur, D. E., Abdullahi, M., & Haruna, A. (2020). Quantitative Structure–Activity Relationship Model, Molecular Docking Simulation and Computational Design of Some Novel Compounds Against DNA Gyrase Receptor. *Chemistry Africa*, 3(2), 391–408. <https://doi.org/10.1007/s42250-020-00132-9>
- Ahmadi, S., Ketabi, S., & Jebeli Javan, M. (2023). *Molecular Descriptors in QSPR/QSAR Modeling* (pp. 25–56). https://doi.org/10.1007/978-3-031-28401-4_2
- Ahmad, O. B. K. C. J. D. T. E. M. F. D. P. E. R. da S. V. P. Q. (2025). *EGFR INHIBITORS* (Patent US20250236608).
- B, D. (2025). Predictive QSAR modeling and ligand-based design of novel EGFR kinase inhibitors. *International Journal of Zoology and Applied Biosciences*, 10(6S), 505–508. <https://doi.org/10.55126/ijzab.2025.v10.i06.SP102>
- B, M., Manandhar, S., Hari, G., Priya, K., Kumar B, H., & Pai, K. S. R. (2021a). Virtual structure-based docking, WaterMap, and molecular dynamics guided identification of the potential natural compounds as inhibitors of protein-tyrosine phosphatase 1B. *Journal of Molecular Structure*, 1226, 129396. <https://doi.org/10.1016/j.molstruc.2020.129396>
- B, M., Manandhar, S., Hari, G., Priya, K., Kumar B, H., & Pai, K. S. R. (2021b). Virtual structure-based docking, WaterMap, and molecular dynamics guided identification of the potential natural compounds as inhibitors of protein-tyrosine phosphatase 1B. *Journal of Molecular Structure*, 1226, 129396. <https://doi.org/10.1016/j.molstruc.2020.129396>

- Cheng, Y., He, Y., Li, W., Zhang, H., Zhou, Q., Wang, B., Liu, C., Walding, A., Saggese, M., Huang, X., Fan, M., Wang, J., & Ramalingam, S. S. (2021). Osimertinib Versus Comparator EGFR TKI as First-Line Treatment for EGFR-Mutated Advanced NSCLC: FLAURA China, A Randomized Study. *Targeted Oncology*, 16(2), 165–176. <https://doi.org/10.1007/s11523-021-00794-6>
- Daoui, O., Elkhatabi, S., & Chtita, S. (2023). Design and Prediction of ADME/Tox Properties of Novel Magnolol Derivatives as Anticancer Agents for NSCLC Using 3D-QSAR, Molecular Docking, MOLCAD and MM-GBSA Studies. *Letters in Drug Design & Discovery*, 20(5), 545–569. <https://doi.org/10.2174/1570180819666220510141710>
- Deshmukh, V. G., Sapkal, S. B., Gadekar, S. S., & Deshmukh, V. (2025). EGFR inhibitors across generations: Progress, challenges, and future directions. *Journal of Molecular Structure*, 1339, 142326. <https://doi.org/10.1016/j.molstruc.2025.142326>
- El fadili, M., Er-Rajy, M., Kara, M., Assouguem, A., Belhassan, A., Alotaibi, A., Mrabti, N. N., Fidan, H., Ullah, R., Ercisli, S., Zarougui, S., & Elhallaoui, M. (2022). QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia. *Pharmaceuticals*, 15(6), 670. <https://doi.org/10.3390/ph15060670>
- Er-rajy, M., El Fadili, M., Hadni, H., Mrabti, N. N., Zarougui, S., & Elhallaoui, M. (2022). 2D-QSAR modeling, drug-likeness studies, ADMET prediction, and molecular docking for anti-lung cancer activity of 3-substituted-5-(phenylamino) indolone derivatives. *Structural Chemistry*, 33(3), 973–986. <https://doi.org/10.1007/s11224-022-01913-3>
- Gao, H., Liu, H., Hou, J., Gu, Q., Shi, M., Wu, Q., & Zheng, L. (2024). 3D-QSAR and Molecular Docking Studies of Pyrimidine-based EGFR Inhibitors. *Letters in Drug Design & Discovery*, 21(14), 3049–3059. <https://doi.org/10.2174/0115701808267565231012095231>
- Hasan, M. M., Khan, Z., Chowdhury, M. S., Khan, M. A., Moni, M. A., & Rahman, M. H. (2022). In silico molecular docking and ADME/T analysis of Quercetin compound with its evaluation of broad-spectrum therapeutic potential against particular diseases. *Informatics in Medicine Unlocked*, 29, 100894. <https://doi.org/10.1016/j.imu.2022.100894>
- Hayes, T. K., Aquilanti, E., Persky, N. S., Yang, X., Kim, E. E., Brenan, L., Goodale, A. B., Alan, D., Sharpe, T., Shue, R. E., Westlake, L., Golomb, L., Silverman, B. R., Morris, M. D., Fisher, T. R., Beyene, E., Li, Y. Y., Cherniack, A. D., Piccioni, F., ... Meyerson, M. (2024). Comprehensive mutational scanning of EGFR reveals TKI sensitivities of extracellular domain mutants. *Nature Communications*, 15(1), 2742. <https://doi.org/10.1038/s41467-024-45594-4>

- Kang, X., Li, R., Li, X., & Xu, X. (2024). EGFR mutations and abnormal trafficking in cancers. *Molecular Biology Reports*, 51(1), 924. <https://doi.org/10.1007/s11033-024-09865-z>
- Kashyap, M., Gupta, S., Bansal, Y., & Bansal, G. (2025). 2D-QSAR driven design, molecular docking, molecular dynamics simulation and MM/GBSA studies on quinazoline derivatives for development of VEGFR-2 inhibitors. *Discover Chemistry*, 2(1), 111. <https://doi.org/10.1007/s44371-025-00173-4>
- Lemjabbar-Alaoui, H., Hassan, O. U., Yang, Y.-W., & Buchanan, P. (2015). Lung cancer: Biology and treatment options. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1856(2), 189–210. <https://doi.org/10.1016/j.bbcan.2015.08.002>
- Lin, L., Zheng, F., Wei, H., Wu, X., Yan, H., & Ding, Y. (2026). Design of conditional and tumor-activated PROTACs for selective degradation in lung cancer. *European Journal of Medicinal Chemistry*, 303, 118432. <https://doi.org/10.1016/j.ejmech.2025.118432>
- Luke, U., Emeng, I. E., Gber, N. I., & Iyam, S. O. (2026a). In silico evaluation of (benzo)chromeno[3,4-c]pyridine derivative as dual EGFR/HER2 inhibitors for non-small cell lung cancer. *Scientific Reports*, 16(1), 4449. <https://doi.org/10.1038/s41598-025-34392-7>
- Luke, U., Emeng, I. E., Gber, N. I., & Iyam, S. O. (2026b). In silico evaluation of (benzo)chromeno[3,4-c]pyridine derivative as dual EGFR/HER2 inhibitors for non-small cell lung cancer. *Scientific Reports*, 16(1), 4449. <https://doi.org/10.1038/s41598-025-34392-7>
- Nakamura, T., Nakashima, C., Komiya, K., Kitera, K., Hirai, M., Kimura, S., & Aragane, N. (2018). Mechanisms of acquired resistance to afatinib clarified with liquid biopsy. *PLOS ONE*, 13(12), e0209384. <https://doi.org/10.1371/journal.pone.0209384>
- Nematpour, M., Rezaee, E., Tabatabai, S., & Jahani, M. (2017). A Copper-Catalyzed Synthesis of Functionalized Quinazolines from Isocyanides and Aniline Tri- and Dichloroacetonitrile Adducts through Intramolecular C–H Activation. *Synlett*, 28(12), 1441–1444. <https://doi.org/10.1055/s-0036-15588166>
- Nyijime, T. A., Shallangwa, G. A., Uzairu, A., Umar, A. B., Ibrahim, M. T., Madkhali, H. A., Hamadi, A., Barnawi, J., & Abdalla, M. (2025). Computational design, pharmacokinetics, molecular docking and molecular dynamic simulation of novel anti-tubercular inhibitors. *In Silico Research in Biomedicine*, 1, 100012. <https://doi.org/10.1016/j.insr.2025.100012>
- Panigrahi, D., Behera, B. K., & Sahu, S. K. (2022). Docking Based Identification of Bioactive Diosmin as Potential Multi-Targeted Anti SARS-Cov-2 Agent. *Journal of the Mexican Chemical Society*, 66(3). <https://doi.org/10.29356/jmcs.v66i3.1683>

- Panigrahi, D., & Sahu, S. K. (2025). Computational approaches: atom-based 3D-QSAR, molecular docking, ADME-Tox, MD simulation and DFT to find novel multi-targeted anti-tubercular agents. *BMC Chemistry*, 19(1), 39. <https://doi.org/10.1186/s13065-024-01357-2>
- Papa, E., Sangion, A., Arnot, J. A., & Gramatica, P. (2018). Development of human biotransformation QSARs and application for PBT assessment refinement. *Food and Chemical Toxicology*, 112, 535–543. <https://doi.org/10.1016/j.fct.2017.04.016>
- Patel, H., Pawara, R., Ansari, A., & Surana, S. (2017). Recent updates on third generation EGFR inhibitors and emergence of fourth generation EGFR inhibitors to combat C797S resistance. *European Journal of Medicinal Chemistry*, 142, 32–47. <https://doi.org/10.1016/j.ejmech.2017.05.027>
- Ramalakshmi, N., Manimegalai, P., Bhandare, R. R., Arun Kumar, S., & Shaik, A. B. (2021). 2D-Quantitative structure activity relationship (QSAR) modeling, docking studies, synthesis and in-vitro evaluation of 1,3,4-thiadiazole tethered coumarin derivatives as antiproliferative agents. *Journal of Saudi Chemical Society*, 25(7), 101279. <https://doi.org/10.1016/j.jscs.2021.101279>
- Schmid, S., Li, J. J. N., & Leighl, N. B. (2020). Mechanisms of osimertinib resistance and emerging treatment options. *Lung Cancer*, 147, 123–129. <https://doi.org/10.1016/j.lungcan.2020.07.014>
- Sever, B., Altıntop, M. D., Radwan, M. O., Özdemir, A., Otsuka, M., Fujita, M., & Ciftci, H. I. (2019). Design, synthesis and biological evaluation of a new series of thiazolyl-pyrazolines as dual EGFR and HER2 inhibitors. *European Journal of Medicinal Chemistry*, 182, 111648. <https://doi.org/10.1016/j.ejmech.2019.111648>
- Sharma, R., & Khubchandani, J. (2024). Global, Regional, and National Burden of Tracheal, Bronchus, and Lung Cancer in 2022: Evidence from the GLOBOCAN Study. *Epidemiologia*, 5(4), 785–795. <https://doi.org/10.3390/epidemiologia5040053>
- Shayanfar, S., & Shayanfar, A. (2022). Comparison of various methods for validity evaluation of QSAR models. *BMC Chemistry*, 16(1), 63. <https://doi.org/10.1186/s13065-022-00856-4>
- Subramani, A. K., Sivaperuman, A., Natarajan, R., Bhandare, R. R., & Shaik, A. B. (2022). QSAR and Molecular Docking Studies of Pyrimidine-Coumarin-Triazole Conjugates as Prospective Anti-Breast Cancer Agents. *Molecules*, 27(6), 1845. <https://doi.org/10.3390/molecules27061845>
- Talukder, Md. E. K., Atif, Md. F., Siddiquee, N. H., Rahman, S., Rafi, N. I., Israt, S., Shahir, N. F., Islam, Md. T., Samad, A., Wani, T. A., Rahman, Md. M., & Ahammad, F. (2025). Molecular docking, QSAR, and simulation analyses of EGFR-targeting phytochemicals in non-small cell lung cancer. *Journal of Molecular Structure*, 1321, 139924. <https://doi.org/10.1016/j.molstruc.2024.139924>

- Tan, C.-S., Kumarakulasinghe, N. B., Huang, Y.-Q., Ang, Y. L. E., Choo, J. R.-E., Goh, B.-C., & Soo, R. A. (2018). Third generation EGFR TKIs: current data and future directions. *Molecular Cancer*, 17(1), 29. <https://doi.org/10.1186/s12943-018-0778-0>
- Thirunavukkarasu, M. K., Suriya, U., Rungrotmongkol, T., & Karuppasamy, R. (2021). In Silico Screening of Available Drugs Targeting Non-Small Cell Lung Cancer Targets: A Drug Repurposing Approach. *Pharmaceutics*, 14(1), 59. <https://doi.org/10.3390/pharmaceutics14010059>
- Toropov, A. A., Toropova, A. P., Cappellini, L., Benfenati, E., & Davoli, E. (2018). QSPR analysis of threshold of odor for the large number of heterogenic chemicals. *Molecular Diversity*, 22(2), 397–403. <https://doi.org/10.1007/s11030-017-9800-5>
- Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science*, 22(1), 69–77. <https://doi.org/10.1002/qsar.200390007>
- Ugbe, F. A., Edache, E. I., Adeniji, S. E., Arthur, D. E., Ajala, A., Adawara, S. N., Ejeh, S., & Ibrahim, Z. Y. (2024). Computational evaluation of the inhibitory potential of some urea, thiourea, and selenourea derivatives of diselenides against leishmaniasis: 2D-QSAR, pharmacokinetics, molecular docking, and molecular dynamics simulation. *Journal of Molecular Structure*, 1302, 137473. <https://doi.org/10.1016/j.molstruc.2023.137473>
- Wang, J., Yuan, F., Kendre, M., He, Z., Dong, S., Patil, A., & Padvī, K. (2025a). Rational design of allosteric inhibitors targeting C797S mutant EGFR in NSCLC: an integrative in silico and in-vitro study. *Frontiers in Oncology*, 15. <https://doi.org/10.3389/fonc.2025.1590779>
- Wang, J., Yuan, F., Kendre, M., He, Z., Dong, S., Patil, A., & Padvī, K. (2025b). Rational design of allosteric inhibitors targeting C797S mutant EGFR in NSCLC: an integrative in silico and in-vitro study. *Frontiers in Oncology*, 15. <https://doi.org/10.3389/fonc.2025.1590779>
- Yu, X., Xu, J., Cahuzac, K. M., Xie, L., Shen, Y., Chen, X., Liu, J., Parsons, R. E., & Jin, J. (2022). Novel Allosteric Inhibitor-Derived AKT Proteolysis Targeting Chimeras (PROTACs) Enable Potent and Selective AKT Degradation in KRAS/BRAF Mutant Cells. *Journal of Medicinal Chemistry*, 65(20), 14237–14260. <https://doi.org/10.1021/acs.jmedchem.2c01454>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryfile.docx](#)