Journal of Carcinogenesis
Covering Genesis to Therapeutics
Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 10s, https://carcinogenesis.com/

In Silico and Multi-Omics Exploration of EGFR-Targeting Hits Against Colon Cancer

Rana A. Alghamdi [12*]

[1]Department of Chemistry, Science and Arts College, King Abdulaziz University, Rabigh, Saudi Arabia; raalghamdi3@kau.edu.sa

https://orcid.org/0000-0003-3572-6947

[2]Regenerative Medicine Unit, King Fahd Medical Research Centre, King Abdulaziz University, Jeddah, Saudi Arabia

***Correspondence**: raalghamdi3@kau.edu.sa

ABSTRACT

Background: Colon cancer is a serious global health concern caused by genetic variability and abnormal signaling pathways, such as the epidermal growth factor receptor (EGFR) axis. Although EGFR is a well-known therapeutic target in colorectal cancer, its expression pattern and function in colon cancer are context dependent.

Methods: This study used an integrated in silico technique that included transcriptome analysis, pathway enrichment, molecular docking, pharmacophore modeling, virtual screening, ADMET profiling, and molecular dynamics (MD) simulation to find and assess new small-molecule EGFR inhibitors.

Results: TCGA data showed downregulated EGFR expression in colon cancer tissues, with no significant patient survival association. KEGG pathway analysis revealed EGFR's central role in oncogenic pathways like PI3K-Akt and MAPK. Molecular docking identified Encorafenib as a strong EGFR binder, serving as a template for pharmacophore generation. Virtual screening of ZINC database compounds identified ZINC103239230 as a top hit. ZINC103239230, a promising candidate for further development, demonstrated favorable ADMET properties and formed a stable complex with EGFR in a 100 ns MD simulation, indicating its potential as a therapeutic target in colon cancer. This strategy holds promises for accelerating drug development, reducing resistance, and promoting oncology customized treatment.

Keywords: *Colon cancer, Epidermal Growth Factor Receptor (EGFR), in silico technique, KEGG pathway, Encorafenib, ADMET properties*

How to Cite: Rana A. Alghamdi, (2025) In Silico and Multi-Omics Exploration of EGFR-Targeting Hits Against Colon Cancer, *Journal of Carcinogenesis, Vol.24, No.10s*, 471-484

1. INTRODUCTION

Colon cancer, sometimes called colorectal adenocarcinoma, is an aberrant proliferation of cells in the tissues of the colon, usually from glandular epithelial cells. Colon cancer is one of the most prevalent gastrointestinal tract cancers and is a significant worldwide health burden. When certain cells in the epithelial tissue experience epigenetic or genetic changes that permit cell proliferation, it arises. These genetic abnormalities can be caused by personality traits, lifestyle, or inheritance[1-3]. Colon cancer is caused by hundreds of mutations in various genes, however, the number of mutated genes that drive carcinogenesis remains limited. Due to lifestyle changes and environmental hazards, colon cancer is becoming more common in low- and middle-income nations, making it a significant worldwide cancer concern. Even with improvements in diagnosis and therapy, survival rates are still below ideal, especially for patients who are at an advanced stage. Effective treatment of colon cancer is significantly hampered by its complicated biological landscape, which calls for more specialized and focused therapeutic approaches[4-8].

The Epidermal Growth Factor Receptor (EGFR) is critical to the development and progression of colon cancer, particularly in metastatic cases. The EGFR gene, located on chromosome 7p12-13, encodes a 170 kDa transmembrane receptor with an extracellular ligand-binding domain and an intracellular tyrosine kinase domain. It is part of the ErbB receptor TK family[9]. EGFR triggers the PI3K–PTEN–Akt and RAS–RAF–MAP kinase pathways, which affect gene expression and transcription factors for cellular responses such as apoptosis, migration, proliferation, and differentiation [10-13]. Colon

cancer has a well-established history of aggressive tumor activity and poor clinical outcomes linked to aberrant activation of EGFR signaling, whether through ligand-dependent stimulation, overexpression, or interaction with other oncogenic pathways. Thus, in oncology, EGFR has become a crucial molecular target. Therapeutic strategies like monoclonal antibodies and TKIs, which inhibit EGFR signaling, have shown inconsistent clinical efficacy in colon cancer due to resistance mechanisms and compensatory signaling loops[14, 15].

To overcome these obstacles, the combination of multi-omics data and *in silico* drug discovery techniques is becoming more and more important to find and evaluate new EGFR-targeting hits. Virtual screening, molecular docking, and MD simulations are examples of *in silico* techniques that aid in assessing the stability and binding affinities of huge chemical libraries with target proteins. Rational design is made possible, and drug discovery is accelerated by these techniques. These techniques enhance the potency of *in silico* pipelines by offering a systems-level perspective of drug-target interactions and disease-specific molecular changes when paired with multi-omics analysis [16-22].

This project uses a combination of *in silico* and multi-omics approaches to develop novel EGFR-targeting drugs against colon cancer. This approach is promising for speeding up drug discovery, overcoming resistance, and advancing personalized medicine in oncology.

2. METHODOLOGY

2.1. Literature Search

A comprehensive literature search was conducted to gather relevant data on the role of epidermal growth factor receptor (EGFR) in colorectal cancer (CRC) and its potential as a therapeutic target. The search included peer-reviewed articles, reviews, and clinical guidelines from databases such as PubMed and PMC, focusing on studies published in the last decade that addressed EGFR expression, signaling pathways, and molecular biomarkers in CRC. Key terms included "EGFR," "colorectal cancer," "colon cancer," "molecular biomarkers," and "targeted therapy." The search strategy also incorporated reports on EGFR overexpression prevalence, mutational analyses, and therapeutic implications, ensuring inclusion of both experimental and clinical findings. This approach was aligned with previous comprehensive reviews and guideline recommendations that emphasize the importance of EGFR molecular testing to guide targeted therapies in CRC patients[23-25].

2.2. TCGA Analysis Using UALCAN

To investigate the expression profile and clinical relevance of EGFR in colon cancer, we utilized the UALCAN web portal, a comprehensive and user-friendly resource for analyzing cancer OMICS data derived from The Cancer Genome Atlas (TCGA) and other large-scale datasets. The UALCAN platform enables exploration of gene expression differences between tumors and normal tissues, analysis across pathological stages, and survival analysis based on gene expression levels. We accessed the colon adenocarcinoma (COAD) TCGA dataset within UALCAN and queried EGFR by its official gene symbol to obtain box plots depicting mRNA expression in tumor versus normal samples and across different tumor stages. Statistical significance was assessed using the platform's built-in Student's T-test with a p-value cutoff of 0.05. Furthermore, survival analysis was performed to evaluate the prognostic impact of EGFR expression in colon cancer patients. This approach allowed us to validate EGFR's overexpression and its association with clinical parameters in CRC, supporting its role as a potential therapeutic target[26].

2.3. KEGG Pathway Analysis

To analyze the role of EGFR within the molecular pathways implicated in colon cancer, the Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database (https://www.genome.jp/kegg/pathway.html) was utilized. KEGG provides manually curated pathway maps that represent molecular interaction and reaction networks, enabling a detailed understanding of biological processes relevant to disease mechanisms. The EGFR signaling pathway (hsa04012) was specifically examined to identify key components and interactions involved in colorectal cancer pathogenesis. Differentially expressed genes from the study were mapped onto the KEGG EGFR pathway to visualize their involvement and to elucidate potential regulatory mechanisms. Moreover, the known drugs against colorectal cancer were retrieved.

2.4. Molecular Docking

The list of drugs against colorectal cancer was obtained from KEGG and were docked against EGFR receptor. The crystal structure of EGFR was retrieved from PDB database (PDB ID: 4WKQ). The crystal structure was imported and processed by Protein Preparation Wizard in Maestro tool [27]. In preprocess step, bond orders were assigned, polar hydrogens were added, and zero bond orders were created to metals. Additionally, water molecules were removed beyond 5 Å, and het atom states were generated using Epik at pH 7.0. The missing residues were also added by using Prime. In the next step, the hydrogen bonds were optimized using PROPKA at pH 7.0 [28]. and the energy of structure was minimized by utilizing OPLS forcefield [29]. To perform the site-specific docking, a 3D grid was generated at the active site residues. The X, Y, Z coordinates were 1.46, 194.21, and 20.41 respectively. After protein preparation, the drugs were prepared for docking

using LigPrep and docked against prepared receptor by using SP mode of glide tool [30]. The docking results were analyzed, and drugs were selected based on glide scores.

2.5. Pharmacophore Modelling and Virtual Screening

The drug with highest binding affinity was selected to analyze the molecular interactions with EGFR receptor. The interacting groups of selected drugs were used to generate the Pharmacophore model using Pharmit server (https://pharmit.csb.pitt.edu/). Pharmit provides built in databases for virtual screening including PubChem, ZINC, ChEMBL, ChemDiv, etc. The pharmacophore model was generated, and virtual screening of PubChem database was conducted by applying screening filters. The hit screening filters were as follows: MW < 500, LogP < 5, PSA < 140, HBA < 10, HBD < 5. The screened hits were obtained and prepared by LigPrep tool and then docked to already prepared EGFR receptor.

2.6. Post Docking Analysis

The screened hits were docked to EGFR receptor to obtain the binding affinities. The top ten hits were selected based on glide scores and their molecular interactions were analyzed by using Discovery Studio [31].

2.7. Druglikeness and Toxicity Analysis

Drug erosion is attributed to toxicity and poor pharmacokinetics issues[32]. To overcome these issues, ADMET properties of the selected drugs were predicted using OSIRIS Property Explorer tool [33]. Various properties like, molecular weight, TPSA, solubility, and LogP values were predicted. Moreover, the potential toxicity risks of hits were measured.

2.8. MD Simulation

The stability of the protein-ligand complexes was evaluated by conducting 100 ns Simulation by Desmond [34]. The system preparation involved the solvation of complex in a periodic box containing TIP3P water molecules [35]. The physiological conditions were mimicked by adding Na^+ and Cl^- counter ions with the addition of 0.15 M NaCl. The temperature and pressure of the system was set to 300 K and 1 atm pressure by using NPT ensemble [36]. The system was subjected to a relaxation phase and the production run was started by storing MD snapshots at each 50 ps time interval. The simulation trajectory was analyzed by simulation interaction diagram tool.

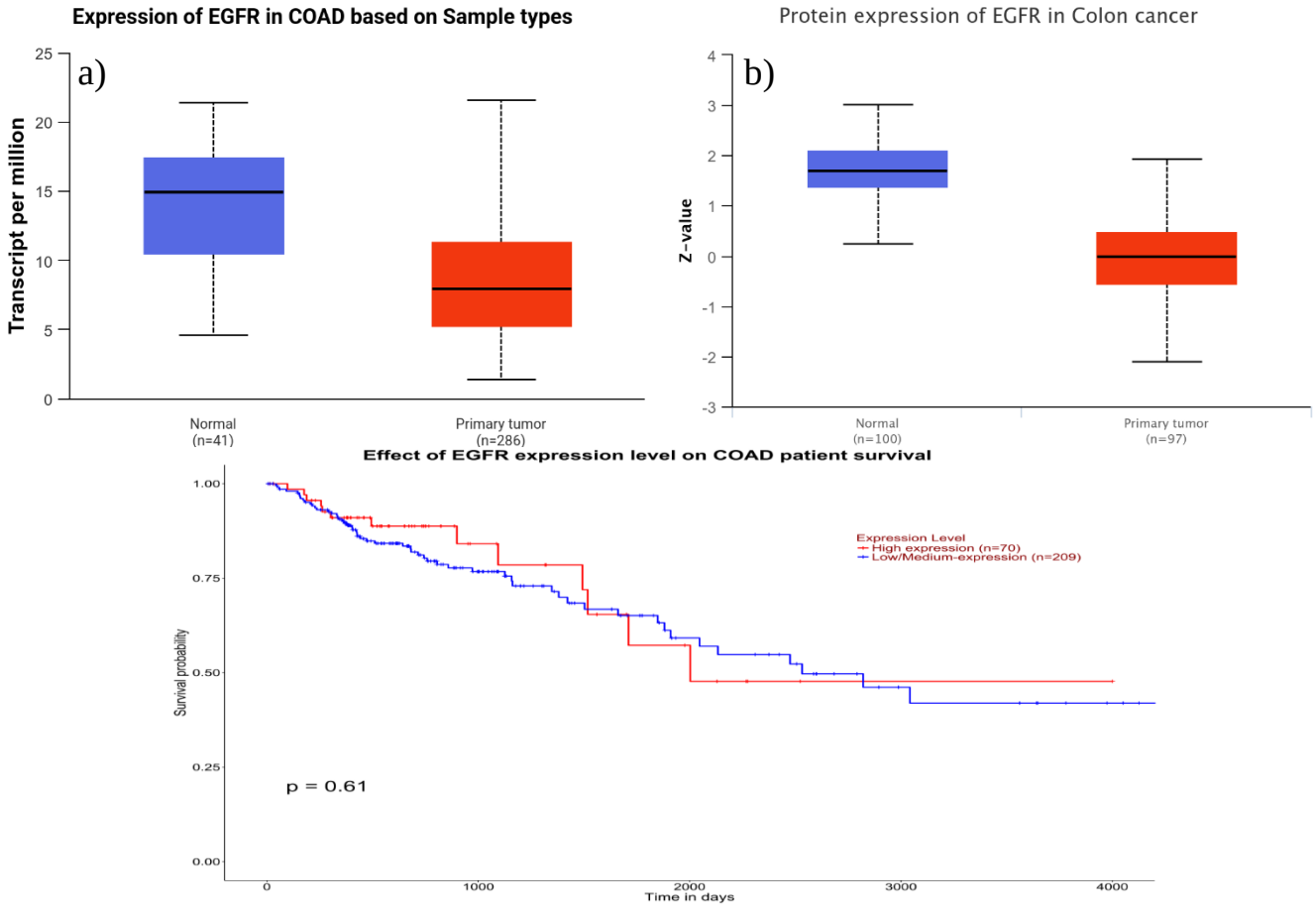
3. RESULTS

3.1. Literature Findings

The literature consistently indicates that EGFR is frequently overexpressed in colon cancer cells, with studies showing significant elevation of EGFR protein levels in the majority of colon cancer cell lines compared to normal colon epithelial cells[37]. EGFR overexpression correlates with tumor progression, metastasis, and poor prognosis, making it a critical biomarker and therapeutic target in CRC[38, 39]. Molecular testing for EGFR pathway mutations has become standard practice to predict response to anti-EGFR monoclonal antibody therapies, with evidence supporting its use in personalized treatment strategies[23, 24]. Moreover, recent studies highlight the complexity of EGFR signaling, including its regulation by ubiquitin ligases like FBXW7, which affects EGFR stability and therapeutic resistance[40]. Clinical data also reveal variability in anti-EGFR therapy efficacy depending on tumor location within the colon and rectum, underscoring the need for precise molecular characterization in treatment planning[41].

3.2. TCGA Analysis

The expression profile and clinical relevance of EGFR with colon cancer was analyzed by TCGA analysis using UALCAN server. Three analyses involving EGFR mRNA expression, EGFR protein expression in colon cancer and analysis of EGFR expression effect on COAD patients' survival was conducted. mRNA expression analysis showed the EGFR mRNA expression levels in normal colon tissues and primary tumor samples (Figure 1a). The EGFR expression was higher in normal tissues suggesting that EGFR is downregulated at transcriptional level in colon cancer. Similarly, EGFR protein expression analysis revealed that protein levels were decreased in tumor samples compared to normal tissues which also show the downregulation of EGFR at protein level, like EGFR mRNA expression in colon cancer (Figure 1b). The Kaplan-Meier survival analysis showed that patients with high EGFR expression and with low expression did not show any significant differences in survival probability ($p= 0.61$), indicating that EGFR expression did not impact patient prognosis (Figure 2). While EGFR expression did not significantly impact patient survival in this cohort, extensive research demonstrates that EGFR remains a key driver of tumor progression and a widely used therapeutic target in colorectal cancer[42].



ErbB signaling pathway and subsequently stimulates downstream effectors such as PI3K-Akt, MAPK, and mTOR, leading to enhanced cell proliferation, survival, and resistance to apoptosis. This activation integrates with other oncogenic events, including K-Ras mutations and the inactivation of tumor suppressors like p53 and APC, collectively driving tumor growth and genomic instability as shown in Figure 3. These findings underscore the critical importance of targeting EGFR in colon cancer, as inhibiting this pathway can disrupt multiple oncogenic signals and potentially impede tumor progression. Furthermore, the drugs reported against colorectal cancer were retrieved from KEGG database and their binding affinities were predicted against EGFR.

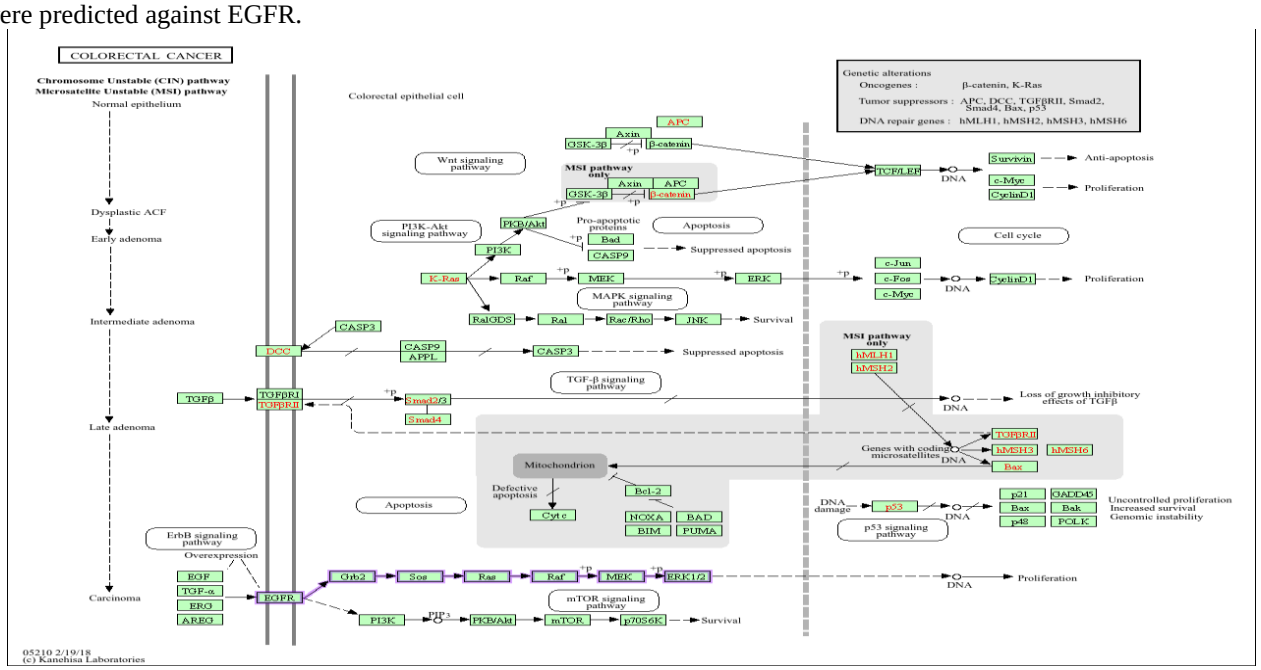
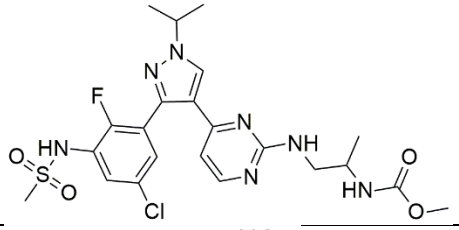
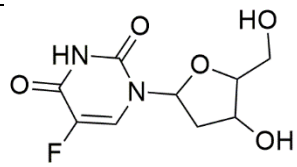
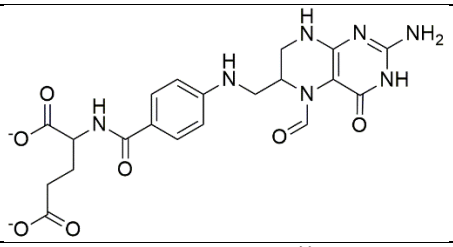
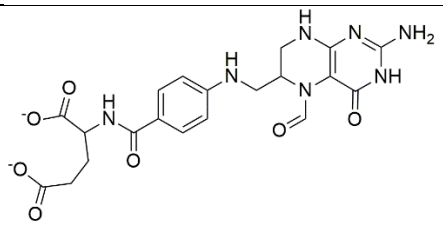
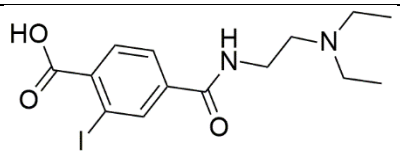
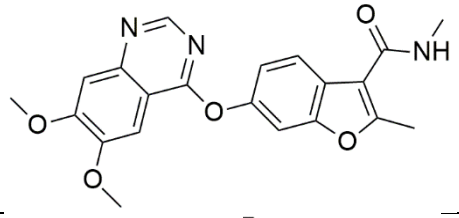
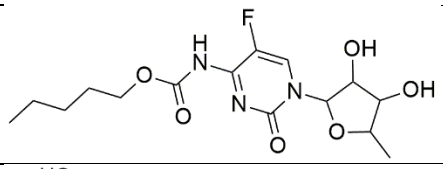
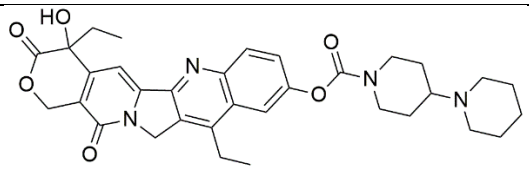


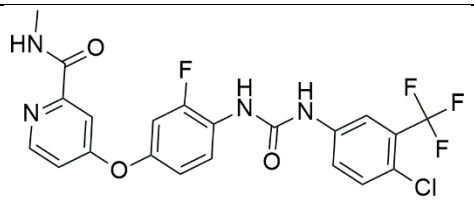
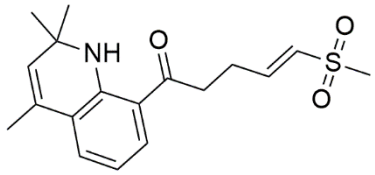
Figure 3: KEGG pathway of colorectal cancer highlighting the role of EGFR protein in the progression of cancer.

3.4. Molecular Docking

The known drugs for colorectal cancer obtained from KEGG database were docked against EGFR by glide tool to predict the binding affinities of all compounds. The binding affinities were analyzed based on the glide scores. The predicted binding affinities were in the range of -6.994 to -5.141 kcal/mol (Table 1). Based on the glide scores, Encorafenib was selected as templated for the development of pharmacophore hypothesis.

Table 1: The binding affinities of known colorectal drugs against EGFR receptor.

No.	Compound	Structure	Glide scores
1	Encorafenib		-6.994
2	Floxuridine		-6.873
3	Sodium levofolinate		-6.809
4	Levoleucovorin calcium		-6.809
5	Bevacizumab		-6.048
6	Fruquintinib		-5.97
7	Capecitabine		-5.847
8	Irinotecan hydrochloride		-5.715

9	Regorafenib hydrate		-5.32
10	Nivolumab		-5.141

3.5. Pharmacophore Modelling and Virtual Screening

Among the known drugs of colorectal cancer, Encorafenib showed highest binding affinity, so its molecular interactions were observed. The molecular interactions showed the interacting groups of Encorafenib (Figure 4a). Based on the interactions, the pharmacophore hypothesis was developed by using Pharmit server. A total of six pharmacophoric features were selected containing three hydrophobic groups (green sphere), one aromatic group (purple sphere), and two hydrogen bond acceptor groups (orange sphere) as shown in Figure 4b. After generating the model, the virtual screening of PubChem database was conducted by using screening filters. The threshold for the hits was set as: M.W < 500, HBD < 5, HBA < 10, TPSA < 140, and LogP < 5. A total of 742 hits were generated by virtual screening which were then subjected towards molecular docking against prepared EGFR receptor. Again, the binding affinities of hits were estimated based on glide scores and top ten hits were selected for further analysis. The binding affinities of selected hits were in range of -8.39 to -7.816 better than the template drug Encorafenib.

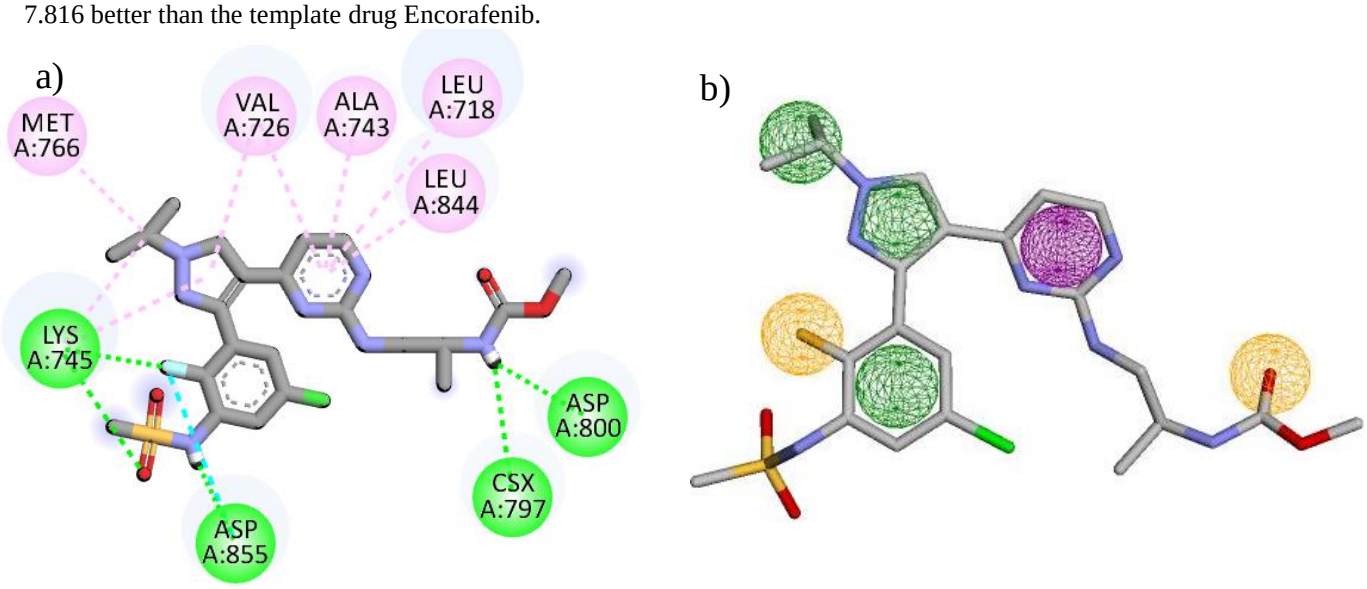
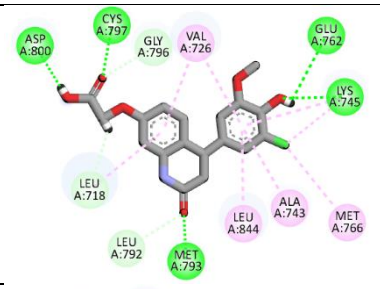
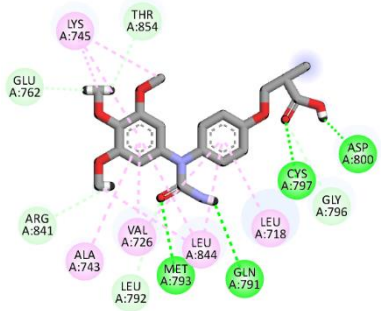
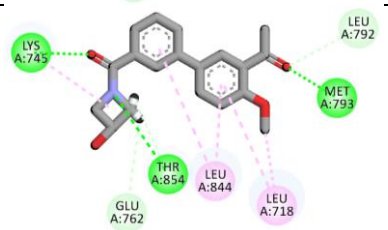
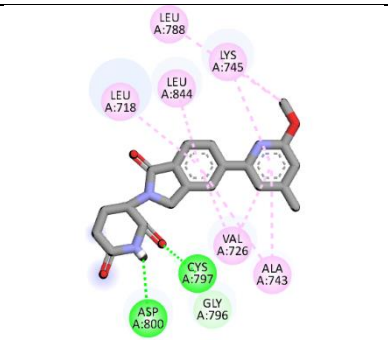
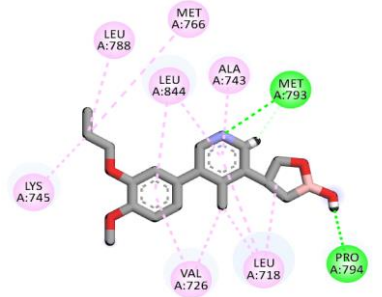
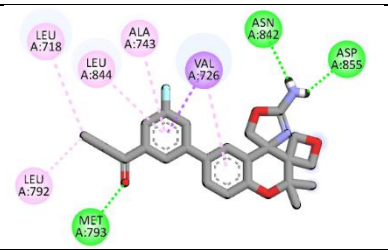


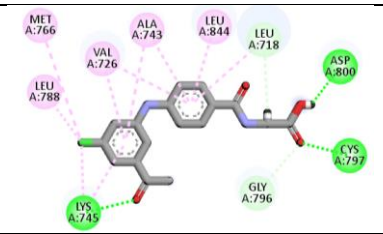
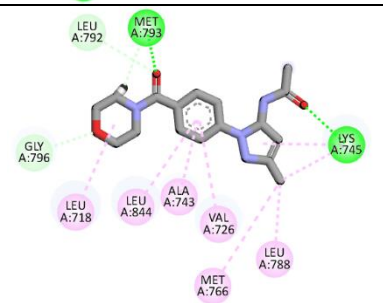
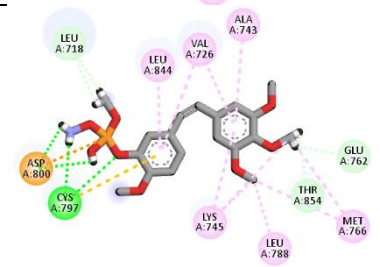
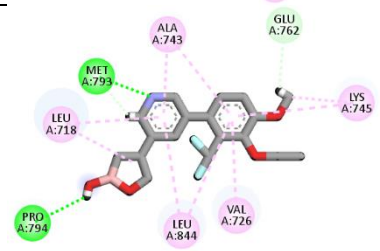
Figure 4: a) The molecular interactions between EGFR and Encorafenib showing the interacting groups of Encorafenib. b) The developed pharmacophore model on Encorafenib template indicating the pharmacophoric groups with different colors.

3.6. Post Docking Analysis

The molecular interactions of the top ten selected hits with EGFR binding sites were analyzed by using Discovery studio. The interactions mainly involve hydrogen bonds, Pi-Sigma interaction, Pi -sulfur, and alkyl interactions. These interactions play a key role in determining the binding affinities of candidate compounds. The overall stability of protein-ligand complex is dependent on these interactions [43]. The binding affinities and molecular interactions of selected hits are shown in Table. The summary of residues forming hydrogen bonds and hydrophobic interactions is also shown in Table 2. Among the interaction diagram, hydrogen bonds are shown with green sphere, Pi-Sigma interaction is shown with purple sphere, while alkyl/hydrophobic interactions are shown with pink spheres.

Table 2: The docking scores and molecular interactions of hit compounds against EGFR receptor.

Sr.	Compound ID	Glide score (kcal/mol)	Interactions	Hydrogen bonds	Hydrophobic interactions
1	95131112	-8.39		Asp800, Cys797, Glu62, Gly796, Lys745, Met793, Leu792, Leu718	Val726, Leu844, Ala743, Met766
2	24803478	-8.197		Glu762, Arg841, Leu792, Met793, Gln791, Gly796, Cys797, Asp800, Thr854	Lys745, Ala743, Val726, Leu844, Leu718
3	122566189	-8.113		Lys745, Glu762, Leu792, Met793, Thr854	Leu844, Leu718
4	156867440	-7.975		Gly796, Cys797, Asp800	Leu718, Leu788, Lys745, Leu844, Val726, Ala743
5	156778766	-7.961		Met793, Pro794	Lys745, Leu788, Met766, Leu844, Ala743, Val726, Leu718
6	149220690	-7.94		Met793, Asn842, Asp855	Leu718, Leu844, Ala743, Val726, Leu792

7	110453089	-7.901		Lys745, Gly796, Cys797, Asp800, Leu718	Leu788, Met766, Val726, Ala743, Leu844
8	35264260	-7.857		Leu792, Gly796, Met793, Lys745	Leu718, Leu844, Ala743, Val726, Met766, Leu788
9	101734289	-7.82		Leu718, Cys797, Thr854, Leu762, Asp800	Leu844, Val726, Ala743, Lys745, Leu788, Met766
10	150870843	-7.816		Met793, Pro794, Glu762	Leu718, Ala743, Lys745, Val726, Leu844

3.7. ADMET Analysis

The pharmacokinetic and toxicity analysis of the selected hits was conducted using OSIRIS Property Explorer. In pharmacokinetic properties, molecular weight, LogP, TPSA, and solubility (LogS) of the hits were estimated. Molecular weight helps in determining the easy distribution of drug un the cells so hit compound with lower molecular weight can easily dissolve in the body. Similarly, hydrophilicity indicates the absorption of compound which is determined by calculating LogP values. LogP values are more than 5 shows poor absorption of compounds. Another parameter is TPSA which is related to hydrogen bonding ability of a compound and a good predictor of bioavailability [44]. Compound having TPSA value less than 160 $Å^2$ have good oral bioavailability [45]. Lastly, solubility helps in measuring the ability of compound to dissolve in the solvent. The ADMET profile of the hits showed that the hits met the threshold values for M.W < 500, LogP < 5, TPSA < 160 $Å^2$, and LogS < -5 except for one compound 149220690 (Table 3). In addition to these parameters, the drug-likeness and drug scores of the hits were also predicted. Positive value for drug-likeness shows that the compound has some common structural features with known drugs while drug scores is the combination of drug-likeness, solubility, molecular weight, hydrophilicity, and toxicity risks. The higher the drug score for a compound, the higher the potential of that compound to be developed into a medication [46]. Moreover, the drug toxicity analysis was performed to evaluate the potential risks for hits to be a mutagenic, tumorigenic, irritant, and reproductive effect. The analysis revealed that all hits passed the toxicity risk test except for compound-156778966 which has mild risk for irritation and compound-101735289 which has mild mutagenic risk and high risk for reproductive effect (Table 3). Based on the ADMET profiles, the top four compounds were selected for further study.

Table 3: The physicochemical and toxicity profiles of the selected hits against EGFR.

	Pharmacokinetic Properties						Toxicity Profiles			
Compound codes	MW	cLogP	TPSA	LogS	Druglikeness	Drugscore	Mutagenic	Tumorigenic	Irritant	Reproductive effect
95131112	442	4.52	82.99	-5.55	2.34	0.44	Passed	Passed	Passed	Passed
24803478	400	4.21	58.64	-5	4.47	0.56	Passed	Passed	Passed	Passed
122566189	386	3.01	83.65	-5.16	0.27	0.39	Passed	Passed	Passed	Passed
156867440	396	3.97	63.21	-4.69	2.71	0.59	Passed	Passed	Passed	Passed
156778766	388	2.66	108.3	-4.96	2.8	0.63	Passed	Passed	Passed	Passed
149220690	416	3.4	58.08	-3.2	3.23	0.58	Medium risk	Passed	Passed	Passed
110453089	375	4.71	39.5	-5.38	-1.19	0.31	Passed	Passed	Passed	Passed
35264260	415	4.67	50.6	-4.85	2.99	0.51	Passed	Passed	Passed	Passed
101734289	400	2.9	74.86	-5.48	0.95	0.4	Passed	Passed	Passed	Medium risk
150870843	496	2.55	128.1	-5.05	3.34	0.32	Passed	Passed	Passed	High risk

3.8. Binding Pose Alignment

After selection of four compounds, their binding modes were aligned with the native co-crystal ligand to check their binding in receptor cavity. All four ligands were superimposed on co-crystal ligand which revealed that the hit compound occupied same space in the EGFR receptor like co-crystal ligand, indicating the accuracy of docking protocol (Figure 5).

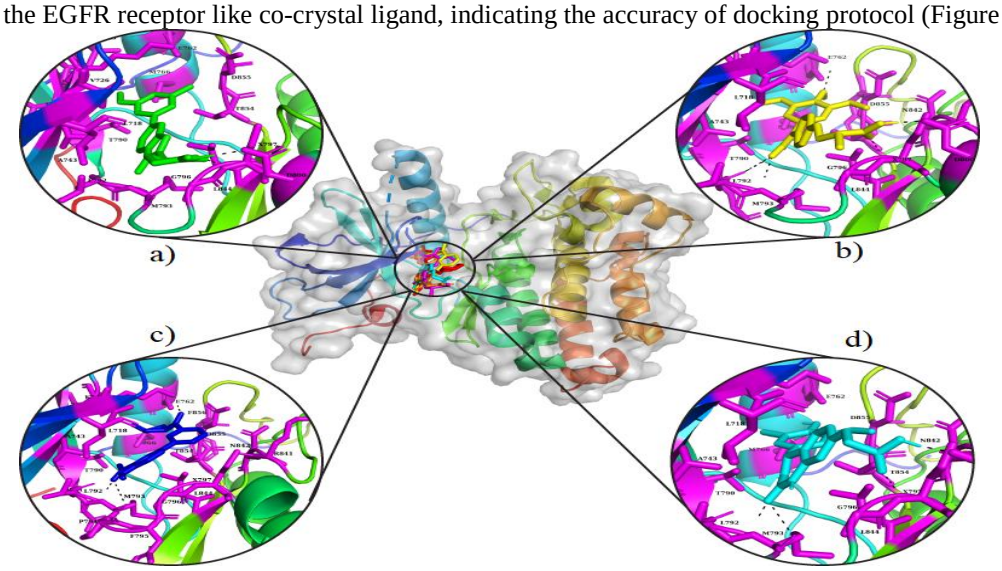


Figure 5: The alignment of binding modes of selected hits in EGFR binding pocket. a) 95131112-complex, b) 24803478-complex, c) 122566189-complex, d) 156867440-complex.

3.9. MD Simulation

The stability of selected hit compounds with EGFR receptor was estimated by conducting 100 simulation and the MD trajectory was analyzed to calculate RMSD, RMSF, and protein-ligand contacts during simulation.

3.9.1. RMSD

The RMSD of carbon alpha atoms of protein and ligand atoms was calculated to estimate the protein-ligand complex stability [47]. Protein RMSD shows the deviation of protein structure from its initial confirmation while the ligand RMSD values show how strongly or loosely ligand is bound to protein. In 95131112-complex, the protein RMSD fluctuates around 2.4–2.8 Å, indicating moderate structural stability with some flexibility. The ligand RMSD starts low (~0.5 Å) and gradually increases to around 2.5 Å, suggesting that the ligand remains relatively stable but with some movement or repositioning within the binding site over time (Figure 6a). Similarly, in complex of 24803478-compound, Protein RMSD fluctuates between approximately 2.5 and 3.7 Å, showing more flexibility or conformational changes than 95131112-complex. Ligand RMSD varies between 1.5 and 3.5 Å, indicating moderate ligand mobility but still generally stable binding (Figure 6b). In third complex, Protein RMSD increases progressively from about 1.5 Å to around 3.5–4.0 Å, indicating increasing structural changes or flexibility. Ligand RMSD shows a similar increasing trend, reaching up to about 9–10 Å, which suggests significant ligand movement or possible partial dissociation from the binding site (Figure 6c). While fourth complex analysis showed that the Protein RMSD stabilizes around 3.0–3.5 Å after an initial rise, indicating a relatively stable protein conformation after equilibration. Ligand RMSD remains mostly between 3 and 5 Å, showing moderate ligand mobility but no major dissociation (Figure 6d).

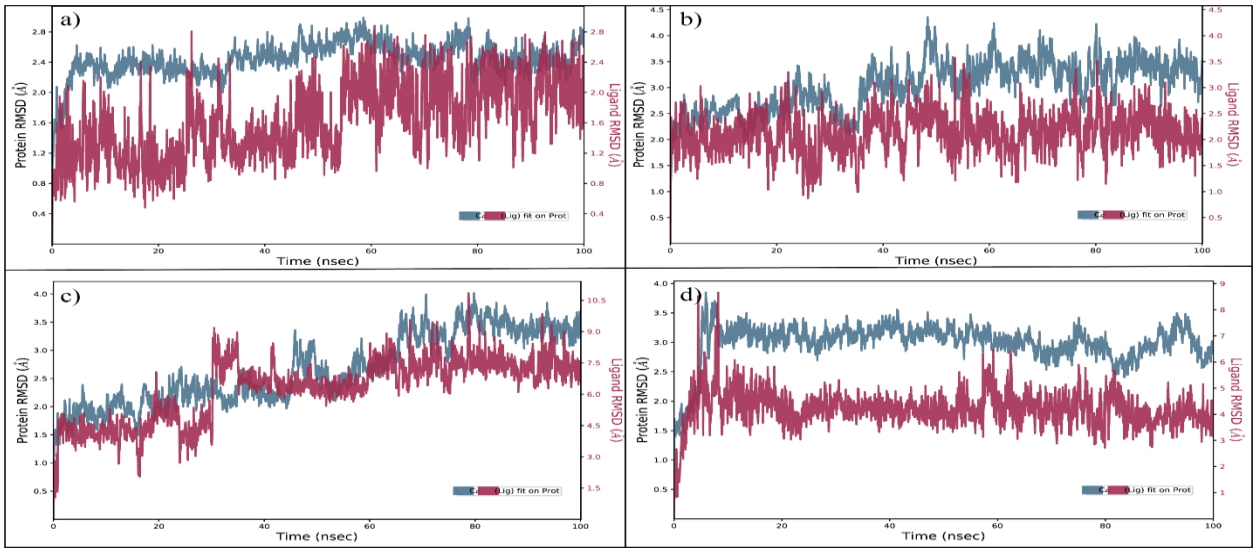


Figure 6: RMSD analysis of protein and ligand atoms during 100 ns molecular dynamics simulations for selected protein-ligand complexes. a) 95131112-complex, b) 24803478-complex, c) 122566189-complex, d) 156867440-complex.

3.9.2. RMSF

RMSF measures the average deviation of each residue from its mean position during the MD simulation, providing insight into the flexibility and mobility of different regions of the protein [48]. The RMSF analysis of four complexes was conducted as displayed in Figure 7. The Figure shows that all complexes display several peaks and troughs, indicating regions of high and low flexibility, respectively. The highest RMSF values (peaks) are typically observed at the N- and C-termini, as well as at specific internal segments, likely corresponding to loop or unstructured regions. Most of the protein residues exhibit RMSF values between 0.6 and 2.0 Å, suggesting these regions are relatively stable and less flexible. The RMSF profiles reveal that certain regions of the protein, particularly around residue 50 and the terminal residues, are consistently more flexible across all four systems.

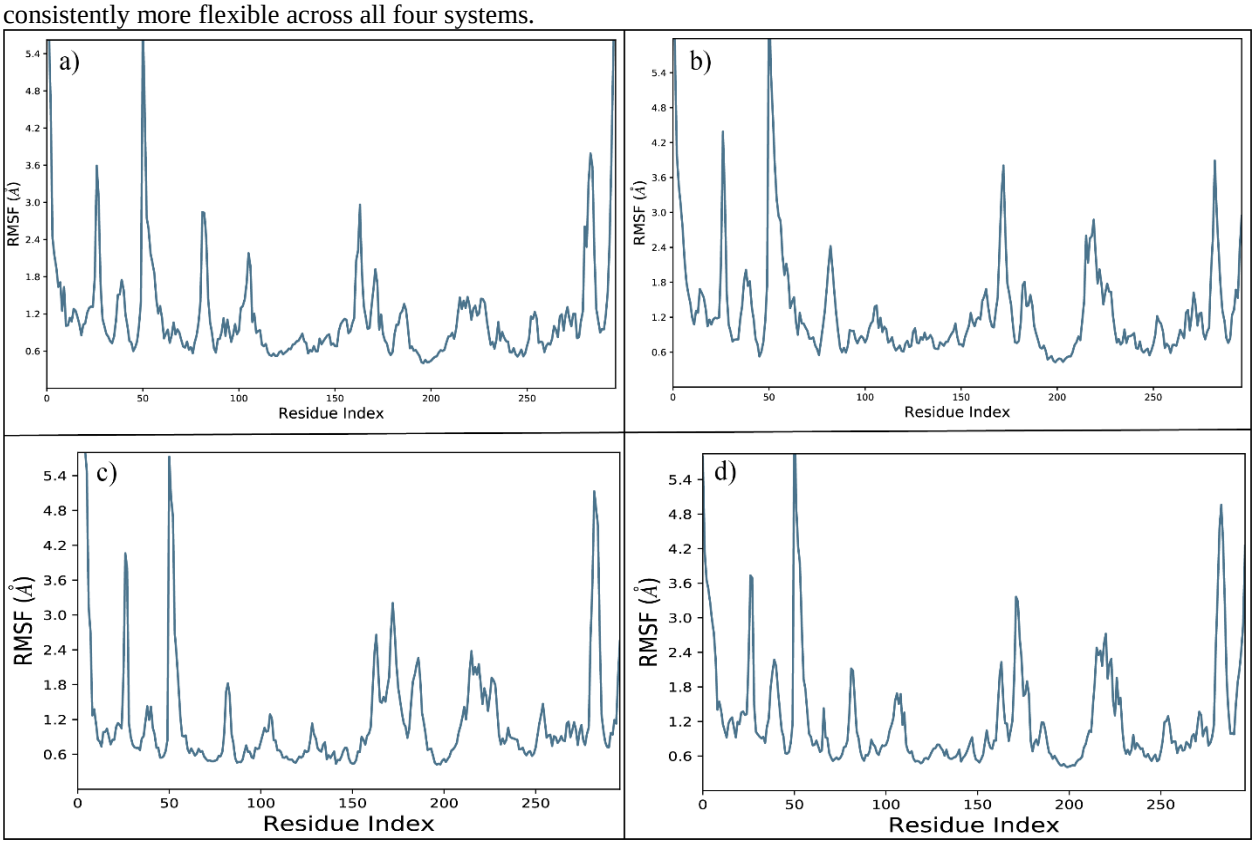


Figure 7. Root Mean Square Fluctuation (RMSF) profiles of Cα atoms for four different protein systems. a) 95131112-complex, b) 24803478-complex, c) 122566189-complex, d) 156867440-complex.

3.9.3. Protein-Ligand Interactions during Simulation

The protein-ligand interactions were observed during simulation and plotted as bar charts as shown in Figure. In the first complex, several residues (e.g., PHE-793, ASP-855, GLU-767) show high interaction fractions, indicating frequent and stable contacts with the ligand. PHE-793 stands out with the highest interaction fraction, suggesting it is a key residue for ligand binding in this complex. Multiple interaction types are observed for some residues, indicating diverse binding modes (Figure 8a). In the second complex, residues such as GLU-767 and PHE-793 again show prominent interactions, but the distribution is broader, with more residues participating at moderate levels. The interaction profile suggests this ligand engages the binding site with a different pattern, possibly involving more residues but with less dominance by a single residue (Figure 8b). The third complex showed that the interaction is highly focused on GLU-767, which dominates the profile, suggesting a strong, specific interaction with this residue. Most other residues show minimal interaction, indicating a more selective binding mode (Figure 8c). Lastly, the interaction pattern in the fourth complex is more distributed, with moderate interaction fractions for several residues i.e. VAL-726, PHE-793, ASP-855 (Figure 8d). In short, PHE-793 and GLU-767 consistently appear as key interacting residues across different ligands, highlighting their importance in ligand recognition and binding.

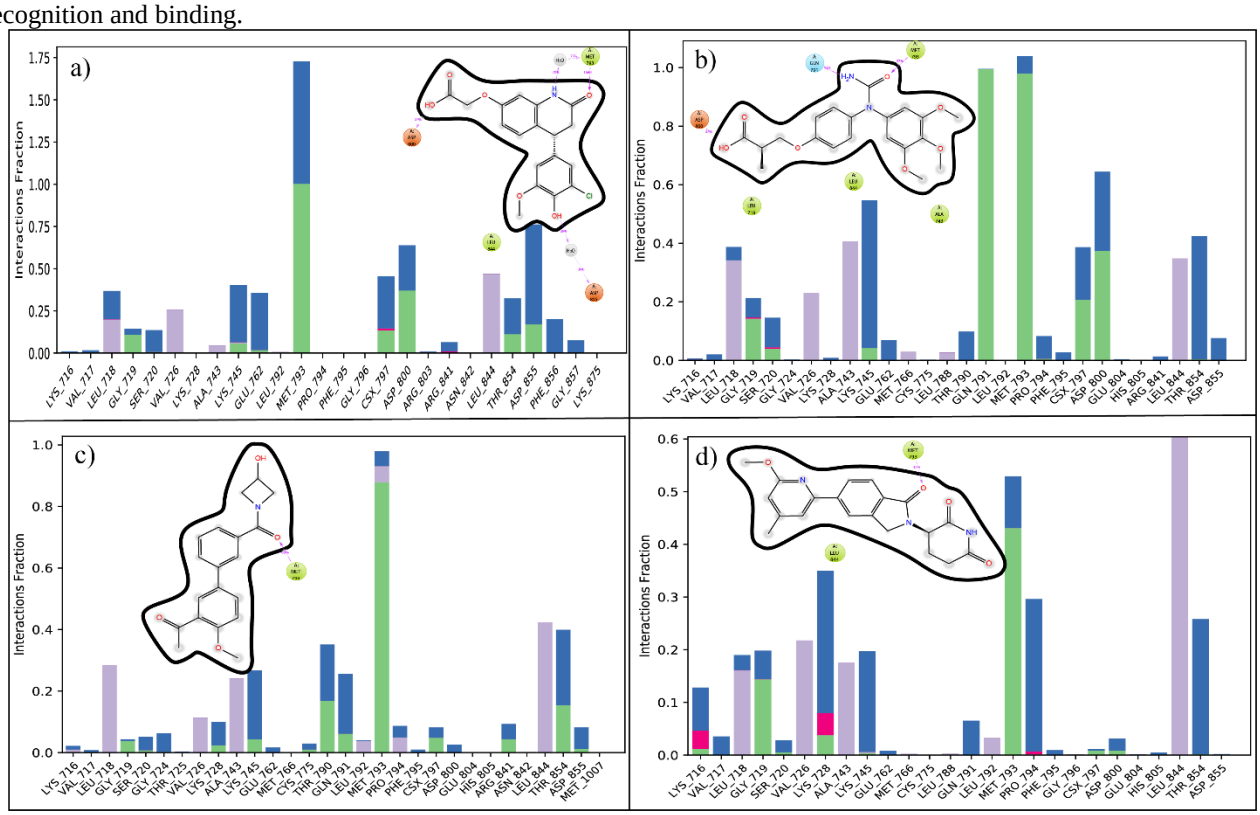


Figure 8. Interaction fraction profiles of four different ligands against EGFR key residues during molecular dynamics simulations. a) 95131112-complex, b) 24803478-complex, c) 122566189-complex, d) 156867440-complex.

4. DISCUSSION

The study investigates the possibility of employing an integrative *in silico* method to target the epidermal growth factor receptor (EGFR) in colon cancer. Although EGFR is implicated in several cancers, including colorectal cancer, its expression pattern and potential therapeutic applications are still unclear and occasionally conflicting[49-51]. The study used transcriptomic data from TCGA via the UALCAN platform and found that, in contrast to earlier reports of overexpression in colorectal cancers[52-54], EGFR mRNA and protein levels were significantly downregulated in colon tumor tissues. This finding suggests that the expression pattern may be subtype-specific or context-dependent. The lack of significant connection between EGFR expression and patient survival shows that EGFR's role is subtler than can be determined solely by expression levels.

Notwithstanding these discoveries, KEGG pathway analysis shows that EGFR continues to play a significant role in colorectal carcinogenesis by integrating with important signaling cascades such as PI3K-Akt and MAPK and promoting carcinogenic processes. EGFR interacts with mutant genes in colon cancer, including KRAS, APC, and TP53 [55, 56], potentially activating downstream pathways even when EGFR expression is low. This demonstrates the potential of EGFR-targeted treatments, particularly in molecular settings where EGFR signaling is impaired.

Molecular docking experiments used the Schrödinger suite to find small compounds capable of regulating EGFR activity. Encorafenib demonstrated the highest binding affinity among 11 FDA-approved colorectal cancer drugs, making it a feasible reference chemical for pharmacophore model generation. Encorafenib's pharmacophore hypothesis was applied to screen a vast chemical library from the ZINC database. The reference drug was surpassed by ZINC103239230, which formed stable contacts inside the EGFR binding pocket for inhibitory activity and had a promising docking score of -7.273 kcal/mol.

SwissADME and pkCSM were used in the study to assess ZINC103239230's drug-likeness and pharmacokinetic characteristics. Its promise as a drug candidate was indicated by its favorable absorption, distribution, metabolism, excretion, and toxicity qualities, as well as its adherence to Lipinski's rule of five. A 100 ns molecular dynamics simulation confirmed the stability and dynamic behaviour of a ligand-receptor complex within the EGFR binding site, revealing minimal conformational fluctuations and supporting the lead compound's docking results and potential efficacy under physiological conditions.

Thus, despite EGFR's downregulation in tumour tissues and its significance in important signalling pathways, the study underlines the use of computational approaches in the early stages of identifying novel EGFR inhibitors for colon cancer. The lead chemical, ZINC103239230, appears to be a viable candidate for further preclinical validation and development, with the potential to contribute to more effective and individualized therapy options in EGFR-driven colon cancer.

5. CONCLUSION

The study emphasizes the relevance of EGFR as a therapeutic target in colon cancer, despite its downregulation. It demonstrates that EGFR is an important node in oncogenic signaling networks, making it a promising target for small-molecule inhibitions. The discovery of ZINC103239230 as a lead molecule with high binding affinity, drug-likeness, and pharmacokinetic stability demonstrates the value of *in silico* pipelines in early-stage drug discovery. These findings lay the groundwork for future experimental validation and preclinical development of novel EGFR-targeted treatments specific to the molecular landscape of colon cancer.

REFERENCES

[1] Chen, W., et al., Cancer statistics in China, 2015. CA: a cancer journal for clinicians, 2016. 66(2): p. 115-132.
[2] Church, J. Molecular genetics of colorectal cancer. in Seminars in Colon and Rectal Surgery. 2016. Elsevier.
[3] Ridwan, S., et al., Identification and in silico analysis of inhibitor on the Wnt/β-catenin signaling pathway as potential drug for colon cancer. Int J Appl Pharm, 2023. 15: p. 111-20.
[4] Labianca, R., et al., Colon cancer. Critical reviews in oncology/hematology, 2010. 74(2): p. 106-133.
[5] Benson, A.B., et al., Colon cancer, version 3.2024, NCCN clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network, 2024. 22(2D).
[6] Dey, A., et al., Recent advancements, limitations, and future perspectives of the use of personalized medicine in treatment of colon cancer. Technology in Cancer Research & Treatment, 2023. 22: p. 15330338231178403.
[7] Burt, M., Randall W, M. DiSario, James A, and P.D. Cannon-Albright, Lisa, Genetics of colon cancer: impact of inheritance on colon cancer risk. Annual review of medicine, 1995. 46(1): p. 371-379.
[8] Rustgi, A.K., The genetics of hereditary colon cancer. Genes & development, 2007. 21(20): p. 2525-2538.
[9] Mitsudomi, T. and Y. Yatabe, Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. The FEBS journal, 2010. 277(2): p. 301-308.
[10] Arteaga, C.L., The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 2001. 19(18 Suppl): p. 32S-40S.
[11] Talapatra, S. and C.B. Thompson, Growth factor signaling in cell survival: implications for cancer treatment. The Journal of pharmacology and experimental therapeutics, 2001. 298(3): p. 873-878.
[12] Venook, A., Critical evaluation of current treatments in metastatic colorectal cancer. The Oncologist, 2005. 10(4): p. 250-261.
[13] Custodio, A. and J. Feliu, Prognostic and predictive biomarkers for epidermal growth factor receptor-targeted therapy in colorectal cancer: beyond KRAS mutations. Critical reviews in oncology/hematology, 2013. 85(1): p. 45-81.
[14] Saletti, P., et al., EGFR signaling in colorectal cancer: a clinical perspective. Gastrointestinal Cancer: Targets and Therapy, 2015: p. 21-38.
[15] Krasinskas, A.M., EGFR signaling in colorectal carcinoma. Pathology research international, 2011. 2011(1): p. 932932.
[16] Celebi, R., et al., In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. Scientific Reports, 2019. 9(1): p. 8949.

[17] Guo, S., et al., Computational and systematic analysis of multi-omics data for drug discovery and development. 2023, Frontiers Media SA. p. 1146896.
[18] Jiang, W., et al., Network-based multi-omics integrative analysis methods in drug discovery: a systematic review. BioData Mining, 2025. 18(1): p. 27.
[19] 19. Kontoyianni, M., Docking and virtual screening in drug discovery, in Proteomics for drug discovery: Methods and protocols. 2017, Springer. p. 255-266.
[20] Liu, X., et al., Molecular dynamics simulations and novel drug discovery. Expert opinion on drug discovery, 2018. 13(1): p. 23-37.
[21] Kitchen, D.B., et al., Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery, 2004. 3(11): p. 935-949.
[22] Hou, T. and X. Xu, Recent development and application of virtual screening in drug discovery: an overview. Current pharmaceutical design, 2004. 10(9): p. 1011-1033.
[23] Sepulveda, A.R., et al., Molecular biomarkers for the evaluation of colorectal cancer: guideline from the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology. 2017. 147(3): p. 221-260.
[24] Doleschal, B., A. Petzer, and H.J.F.i.O. Rumpold, Current concepts of anti-EGFR targeting in metastatic colorectal cancer. 2022. 12: p. 1048166.
[25] Shin, J.H., et al., USP21-EGFR signaling axis is functionally implicated in metastatic colorectal cancer. 2024. 10(1): p. 492.
[26] Chandrashekar, D.S., et al., UALCAN: An update to the integrated cancer data analysis platform. 2022. 25: p. 18-27.
[27] Schrödinger, L.J.S.S., Schrödinger, LLC; New York, NY: 2017. 2017. 2: p. 2017-1.
[28] Kim, M.O., et al., Effects of histidine protonation and rotameric states on virtual screening of M. tuberculosis RmlC. 2013. 27(3): p. 235-246.
[29] Jeong, S.W., et al., Prediction of enthalpy of vaporization for particulate matter through molecular dynamics using OPLS force field. 2025: p. 106595.
[30] Friesner, R.A., et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. 2004. 47(7): p. 1739-1749.
[31] Visualizer, D.J.A.s.i., Discovery Studio Visualizer. 2. 2005.
[32] Agamah, F.E., et al., Computational/in silico methods in drug target and lead prediction. 2020. 21(5): p. 1663-1675.
[33] 33. Ejeh, S., et al., In silico design and pharmacokinetics investigation of some novel hepatitis C virus NS5B inhibitors: pharmacoinformatics approach. 2022. 46(1): p. 1-11.
[34] Bowers, K.J., et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. in Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. 2006.
[35] Price, D.J. and C.L.J.T.J.o.c.p. Brooks III, A modified TIP3P water potential for simulation with Ewald summation. 2004. 121(20): p. 10096-10103.
[36] Kalibaeva, G., M. Ferrario, and G.J.M.P. Ciccotti, Constant pressure-constant temperature molecular dynamics: A correct constrained NPT ensemble using the molecular virial. 2003. 101(6): p. 765-778.
[37] SiShi, L., et al., EGFR and HER2 levels are frequently elevated in colon cancer cells. 2014. 1(1): p. e1.
[38] Yufeng, Z., Q. Ming, and W.J.F.i.G. Dandan, MiR-320d inhibits progression of EGFR-positive colorectal cancer by targeting TUSC3. 2021. 12: p. 738559.
[39] 39. Pabla, B., M. Bissonnette, and V.J.J.W.j.o.c.o. Konda, Colon cancer and the epidermal growth factor receptor: Current treatment paradigms, the importance of diet, and the role of chemoprevention. 2015. 6(5): p. 133.
[40] Boretto, M., et al., Epidermal growth factor receptor (EGFR) is a target of the tumor-suppressor E3 ligase FBXW7. 2024. 121(12): p. e2309902121.
[41] Lee, K.-H., et al., The efficacy of anti-EGFR therapy in treating metastatic colorectal cancer differs between the middle/low rectum and the left-sided colon. 2021. 125(6): p. 816-825.
[42] Del Carmen, S., et al., Prognostic implications of EGFR protein expression in sporadic colorectal tumors: Correlation with copy number status, mRNA levels and miRNA regulation. 2020. 10(1): p. 4662.
[43] Thillainayagam, M., et al., In-Silico molecular docking and simulation studies on novel chalcone and flavone hybrid derivatives with 1, 2, 3-triazole linkage as vital inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase. 2018. 36(15): p. 3993-4009.
[44] Husain, A., et al., Synthesis, molecular properties, toxicity and biological evaluation of some new substituted imidazolidine derivatives in search of potent anti-inflammatory agents. 2016. 24(1): p. 104-114.
[45] Gogoi, P., et al., In silico study, synthesis, and evaluation of the antimalarial activity of hybrid dimethoxy pyrazole 1, 3, 5-triazine derivatives. 2021. 35(3): p. e22682.
[46] Behrouz, S., et al., Design, synthesis, and in silico studies of novel eugenyloxy propanol azole derivatives having potent antinociceptive activity and evaluation of their β-adrenoceptor blocking property. 2019. 23: p. 147-164.

[47] Sargsyan, K., et al., How molecular size impacts RMSD applications in molecular dynamics simulations. 2017. 13(4): p. 1518-1524.
[48] Martínez, L.J.P.o., Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. 2015. 10(3): p. e0119264.
[49] Sigismund, S., D. Avanzato, and L. Lanzetti, Emerging functions of the EGFR in cancer. Molecular oncology, 2018. 12(1): p. 3-20.
[50] Nicholson, R.I., J.M.W. Gee, and M.E. Harper, EGFR and cancer prognosis. European journal of cancer, 2001. 37: p. 9-15.
[51] Fischer, O., et al., EGFR signal transactivation in cancer cells. Biochemical Society Transactions, 2003. 31(6): p. 1203-1208.
[52] Guo, G.F., et al., Overexpression of SGLT1 and EGFR in colorectal cancer showing a correlation with the prognosis. Medical oncology, 2011. 28: p. 197-203.
[53] SiShi, L., et al., EGFR and HER2 levels are frequently elevated in colon cancer cells. Discoveries Reports, 2014. 1(1): p. e1.
[54] Gu, X.-Y., et al., Over-expression of EGFR regulated by RARA contributes to 5-FU resistance in colon cancer. Aging (Albany NY), 2020. 12(1): p. 156.
[55] Conlin, A., et al., The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma. Gut, 2005. 54(9): p. 1283-1286.
[56] Palacio-Rúa, K., et al., Genetic analysis in APC, KRAS, and TP53 in patients with stomach and colon cancer. Revista de Gastroenterología de México (English Edition), 2014. 79(2): p. 79-89.