JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Drug discovery against FGFR2 mutation in lungs cancer approach using NGS and medicinal plant

Komal Raj[1] Uma Kumari[2] Srijan Maheshwari[1] Priyanka[2]
[1]Trainee at Bioinformatics Project and Research Institute, Noida-201301, India
[2]Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida -201301, India
[1]Trainee at Bioinformatics Project and Research Institute, Noida-201301, India
[2]Assistant Professor, Department of Biotechnology and Bioengineering,School of Bioscience and Technology Galgotias University,Noida 203201,india
Corresponding Author: Uma Kumari

Abstract: Lung cancer remains one of the most prevalent and deadly malignancies worldwide, necessitating the development of more effective and targeted therapeutic strategies. The integration of Next-Generation Sequencing (NGS) technologies with traditional medicinal plant-based drug discovery offers a promising avenue for identifying novel bioactive compounds and their molecular targets. This study explores a combinatorial approach where high-throughput NGS data is utilized to decode the transcriptomic and genomic landscapes of lung cancer, enabling the identification of key oncogenes, tumor suppressor genes, and signaling pathways involved in disease progression. Concurrently, bioactive phytochemicals from selected medicinal plants are screened for their potential to modulate these molecular targets. In silico analyses including protein-ligand docking, pharmacokinetic profiling, and toxicity prediction are employed to validate the therapeutic relevance of candidate compounds. Structural modeling and validation using tools like PyMOL, ERRAT, and Ramachandran plot analysis, further ensure the accuracy of target-ligand interactions. This integrative methodology not only accelerates the drug discovery pipeline but also leverages the untapped reservoir of natural compounds in combating non-small cell lung cancer (NSCLC). The findings of this study underscore the potential of combining NGS-based target discovery with phytochemical screening as a cost-effective, precise, and innovative strategy for anti-cancer drug development

Keywords: NGS Analysis, Molecular docking, Lung cancer, FGFR2, Structure Analysis, Molecular modeling

INTRODUCTION

Lung cancer continues to be a leading cause of cancer-related morbidity and mortality worldwide, claiming millions of lives each year. Despite extensive global research, the burden of the disease remains high due to factors such as late-stage diagnosis, limited access to advanced healthcare, environmental pollutants, and a lack of effective early detection strategies. Lung cancer is broadly categorized into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), with NSCLC accounting for approximately 85% of all cases. NSCLC subtypes—adenocarcinoma, squamous cell carcinoma, and large cell carcinoma—differ in terms of histology, molecular alterations, and clinical behavior. The pathogenesis of lung cancer involves a series of genetic and epigenetic events, including mutations, gene amplifications, deletions, chromosomal rearrangements, and epigenetic silencing of tumor suppressor genes [1, 2, 3, 4].Among the key molecular players in lung cancer progression is the Fibroblast Growth Factor Receptor 2 (FGFR2), a transmembrane receptor tyrosine kinase involved in cell growth, differentiation, and survival. Aberrations in FGFR2, including point mutations, gene amplifications, and fusion events, can lead to constitutive activation of the receptor, promoting oncogenesis and resistance to therapy. FGFR2 mutations have been identified in a subset of lung cancer patients and are associated with aggressive tumor behavior, poor prognosis, and limited response to standard chemotherapies. FGFR-targeted inhibitors such as AZD4547, Debio 1347, and Erdafitinib are under investigation, but their success is often limited by toxicity, off-target effects, and the development of secondary resistance mechanisms. Hence, identifying novel, selective, and safer inhibitors of FGFR2 remains an urgent need in precision oncology [5, 6, 7, 8, 9].

The protein structure with the PDB ID 8W3D represents a biologically significant macromolecule that contributes to vital physiological processes, potentially linked to molecular recognition, enzymatic activity, or structural stability in a cellular environment. This protein has been crystallized and structurally characterized to high resolution, providing insights into the intricate architecture of its three-dimensional conformation. Such structural elucidation allows for a deeper understanding of the molecular interactions that underlie the protein's function, including hydrogen bonding, hydrophobic interactions, and coordination with metal ions or ligands. These structural elements are critical in maintaining the integrity and functionality of the protein, especially if it participates in complex biochemical pathways or serves as a target for therapeutic intervention. At the atomic level, proteins like the one represented by 8W3D exhibit a finely tuned arrangement of secondary structural motifs including alpha helices, beta strands, and connecting loops. These elements are stabilized primarily through intramolecular forces, of which hydrogen bonds play a central role. The configuration of these secondary elements and their spatial organization within the

polypeptide chain lead to the formation of a compact tertiary structure, which may further associate with additional chains to form multimeric complexes or higher-order structures. Such quaternary associations are often observed in proteins involved in signal transduction, cellular transport, or enzymatic catalysis, where structural integrity and inter-subunit communication are essential. In the case of the 8W3D structure, visual representation reveals a trimeric complex, indicative of a biologically relevant multimer. This structural formation is not arbitrary but results from specific non-covalent interactions that hold the subunits together in a defined geometric arrangement. The visualization, often rendered via ribbon diagrams, highlights the elegance of protein folding and reveals conserved motifs that might suggest functional relevance. Metal ions or bound cofactors, often visualized as spherical entities within the structure, play roles ranging from structural reinforcement to catalytic activation, depending on the nature of the protein and its physiological role [8, 9, 10, 11, 12, 13].

Hydrogen bonds within this structure appear to be both extensive and functionally significant. These interactions not only stabilize alpha helices and beta sheets within individual subunits but also reinforce the interactions between different subunits. This dual role of hydrogen bonding facilitates both local stability and long-range structural cohesion. The presence of such a dense hydrogen bonding network is often indicative of a protein that must maintain rigidity for its biological function or one that undergoes controlled conformational changes upon binding to other biomolecules. Moreover, structural data from 8W3D offers opportunities for further computational and experimental analysis. Understanding the hydrogen bonding network at a granular level can reveal potential sites for mutagenesis, assist in the rational design of inhibitors or stabilizers, and inform the engineering of protein variants with enhanced stability or activity. With tools such as PyMOL, Chimera, or other structural bioinformatics software, one can delve into the atomic-level interactions that define this protein's functionality. This level of analysis is invaluable for drug discovery initiatives, protein engineering efforts, and for elucidating fundamental principles of protein biochemistry. Therefore, the crystallographic data of the 8W3D protein structure not only enriches the structural database with a new model but also provides a foundation for in-depth biochemical and biophysical investigations. The integration of structural information with functional assays, docking studies, or evolutionary comparisons can illuminate the roles this protein may play in health and disease. Through such multidimensional analyses, the significance of 8W3D can be understood not only in terms of its static structure but also through the dynamic lens of its interaction with the molecular environment [14, 15].

The implementation of Next-Generation Sequencing (NGS) technologies has significantly enhanced our understanding of the genetic basis of lung cancer. NGS enables the parallel sequencing of millions of DNA fragments, allowing for comprehensive genomic profiling of tumors at high resolution. Through whole-genome sequencing (WGS), whole-exome sequencing (WES), and targeted gene panels, NGS helps identify actionable mutations, such as those in EGFR, KRAS, BRAF, ALK, ROS1, and FGFR2, which can inform personalized treatment strategies. In the case of FGFR2, NGS facilitates the detection of specific mutations and structural alterations that could serve as predictive biomarkers for targeted therapy. Moreover, NGS aids in monitoring minimal residual disease (MRD), tracking clonal evolution, and understanding mechanisms of therapeutic resistance, thereby offering a dynamic tool for precision medicine [16, 17, 18].

While synthetic compounds and monoclonal antibodies have dominated the landscape of targeted therapies, natural products derived from medicinal plants continue to be a rich and underexplored source of therapeutic agents. Traditional medicine systems such as Ayurveda, Traditional Chinese Medicine (TCM), and Unani have long utilized plant-based formulations for managing various diseases, including cancer. Several phytochemicals exhibit potent anticancer properties, modulating multiple signaling pathways, inducing apoptosis, inhibiting metastasis, and sensitizing tumors to chemotherapy. Flavonoids, alkaloids, terpenoids, and polyphenols have shown efficacy in preclinical models of cancer, including NSCLC. For instance, curcumin, quercetin, resveratrol, and epigallocatechin gallate (EGCG) have demonstrated potential in regulating key oncogenic pathways such as NF-κB, PI3K/Akt/mTOR, MAPK, and JAK/STAT, as well as modulating oxidative stress and inflammation [17, 18].The convergence of traditional knowledge with modern computational biology has opened new avenues for the discovery of plant-derived compounds targeting specific cancer mutations. In this study, we aim to utilize NGS data to identify FGFR2 mutations relevant to lung cancer and subsequently employ in silico approaches to screen phytochemicals from medicinal plants for their inhibitory potential against mutant FGFR2. The methodology involves the selection of representative FGFR2 mutations from NGS datasets, preparation of the 3D structure using homology modeling or PDB structures, and molecular docking using CB-Dock and AutoDock Vina to evaluate binding affinity and orientation. Docking simulations are validated using scoring functions, binding energy assessments, and structural visualization in PyMOL. Additionally, pharmacokinetic properties, drug-likeness, and toxicity profiles are predicted using SwissADME, pkCSM, and ProTox-III to ensure the therapeutic potential of selected compounds [18, 19, 20].

In order to enhance the reliability of our findings, protein structure validation tools such as Ramachandran plot analysis, ERRAT, and Verify3D are employed to assess the stereochemical quality and compatibility of the 3D models. Further, protein-ligand interaction studies are complemented by molecular dynamics simulations to evaluate the stability of the complexes over time. The integration of Biopython scripts allows for automated sequence analysis, FASTA parsing, domain annotation via InterProScan, and multiple sequence alignments through COBALT. Phylogenetic analysis of FGFR2 homologs across species is conducted to understand evolutionary conservation and identify critical residues essential for ligand binding. Such information is crucial for structure-based drug design, as conserved residues often play key roles in functional activity and ligand specificity [18, 19, 20].This multidisciplinary approach not only contributes to the identification of novel inhibitors of FGFR2 but also provides a systematic framework for integrating NGS data, computational modeling, and phytochemistry for rational drug design. Moreover, the selected phytochemicals can be subjected to structure-activity relationship (SAR) studies and lead optimization to enhance potency, selectivity, and pharmacological properties. In the long term, these compounds may be developed into therapeutic candidates following in vitro validation in FGFR2-mutant lung cancer cell lines and in vivo efficacy studies in xenograft models [18, 19, 20].

Furthermore, the approach aligns with the principles of green chemistry and sustainable drug discovery by leveraging nature-derived compounds, reducing reliance on synthetic chemicals, and minimizing ecological and toxicological impact. In addition to therapeutic benefits, such plant-based interventions may offer economic advantages, particularly in resource-limited settings where access to high-cost targeted therapies remains a barrier. By exploring the therapeutic potential of locally available medicinal plants, this research may also contribute to the valorization of ethnobotanical knowledge and biodiversity conservation [18, 19, 20].In conclusion lung cancer, particularly

NSCLC driven by FGFR2 mutations, represents a formidable challenge in clinical oncology. Current therapies are often limited by resistance and adverse effects, emphasizing the need for novel, safer, and more effective treatment options. Through the integration of NGS-based mutation profiling, computational drug discovery, and phytochemical screening, this study proposes an innovative strategy to identify promising natural inhibitors of FGFR2. The findings may not only offer new therapeutic avenues for lung cancer patients but also demonstrate the value of interdisciplinary research in bridging genomics, bioinformatics, and natural product chemistry for precision medicine [18, 19, 20].

MATERIAL AND METHODS

The current study utilized a range of computational tools to explore molecular structures, perform docking simulations, and validate protein-ligand interactions. RasMol was employed as a visualization tool for exploring three-dimensional conformations of biomolecular structures, including proteins and small ligands. It offers different visualization styles such as wireframe, space-filling, and ribbon, allowing researchers to analyze spatial relationships, identify structural motifs like α-helices and β-sheets, and measure atomic distances. Custom coloring features in RasMol helped in highlighting specific chains, atomic types, and structural elements for detailed inspection [21, 22, 23].

PyMOL served as the primary platform for in-depth structural analysis. It supports high-resolution rendering and manipulation of molecular assemblies, making it suitable for both academic visualization and publication-ready graphics. The tool was used to evaluate protein-ligand binding regions, measure geometric parameters such as angles and dihedrals, and generate animations to represent docking results. Its scripting interface enabled automated rendering and image capture for consistent reporting [24, 25, 26,46,47].

To assess the quality of the modeled protein structures, ERRAT was used for structural validation. It examines non-bonded interactions among atoms to identify potentially problematic regions within the protein model. Structure files were uploaded to the ERRAT server, which returned a quality factor percentage along with a visual plot. High scores indicated good stereochemical quality, while lower values prompted structural refinement before proceeding with docking analyses [27, 28, 29,38,39,41,42].

InterProScan was applied for functional annotation of protein sequences. This platform integrates multiple protein signature databases including Pfam, PROSITE, and SMART. By scanning the amino acid sequence, InterProScan identified conserved domains, motifs, and potential active sites, offering insight into the protein's structural and functional characteristics. Outputs included domain architectures, GO annotations, and cross-references to other molecular databases, which were instrumental in functional prediction and comparative analysis [30, 31].The Kyoto Encyclopedia of Genes and Genomes provides graphical maps of biological processes, including metabolism, genetic information processing, and disease pathways. Gene and protein identifiers were mapped to specific pathways to explore their role in cellular signaling and disease etiology. This helped in linking FGFR2 with lung cancer-associated signaling cascades and potential therapeutic targets [32, 33,44,45,46,47].

Protein structure files were retrieved from the Protein Data Bank (PDB), an open-access repository for experimentally determined macromolecular structures. Structures solved by X-ray crystallography and cryo-electron microscopy were prioritized. The selected structures were assessed for resolution quality and biological relevance. All files were downloaded in .pdb format and visualized using PyMOL and RasMol for downstream applications.Molecular docking simulations were conducted to predict the interaction between target proteins and candidate compounds derived from medicinal plants. Server Docking was used as the primary docking tool due to its ability to automatically detect binding cavities and optimize grid boxes. It performs docking using AutoDock Vina in the background and provides binding affinity scores along with visualizations of the binding site. Each compound was docked individually, and binding poses were compared based on docking scores and orientation within the active site [34, 35,45,46,47]. LIGPLOT was used for interaction mapping after docking simulations. It generated schematic two-dimensional diagrams showing the interactions between ligands and proteins. These included hydrogen bonds, hydrophobic contacts, and nearby residues involved in ligand stabilization. Ligand-protein complexes from docking were input in PDB format, and LIGPLOT output was used to validate key interactions to highlight molecular features important for binding specificity [36, 37].Together, these tools provided a comprehensive platform for structural modeling, functional annotation, docking, and validation, enabling robust in silico assessment of plant-derived inhibitors targeting FGFR2 mutations in lung cancer.

RESULT AND DISCUSSION

1. Structural Visualization and Atom-Level Insights Using RasMol

RasMol was employed for 3D molecular visualization of the human protein sample (8W3D, Chain A). The structural atom count comprised 10,120 atoms.The protein structure presents a visually intricate and stable assembly, suggestive of a trimeric configuration where three individual subunits come together to form a complex. Each subunit is distinguishable by distinct ribbon coloring, representing varied polypeptide chains arranged symmetrically or asymmetrically within the overall complex. Spherical elements embedded around the structure are indicative of ligands or metal ions, which may be crucial for maintaining structural integrity or serving as cofactors within potential active or regulatory sites. There is a clear presence of hydrogen bonding patterns that contribute significantly to the stability and folding of the protein. Within each subunit, hydrogen bonds form extensively between amino acid side chains and backbone atoms, which help maintain the conformation of alpha helices and beta sheets. These intrachain hydrogen bonds are critical for preserving the local secondary structures, ensuring that the polypeptide backbone retains its folded form under physiological conditions. At the junctions where subunits interface, interchain hydrogen bonds likely play a fundamental role in maintaining the overall quaternary structure. These bonds help align the subunits in a precise manner, allowing for the formation of a stable multimeric complex. Such interactions are crucial in mediating protein functionality, especially in cases where multimerization is required for biological activity or ligand binding.

Beyond the regular elements of secondary structure, the loop regions and unstructured coils exhibit hydrogen bonding as well, although in a more dynamic and flexible manner. These regions often participate in transient interactions, and the hydrogen bonds here may contribute to

flexibility, aiding in conformational shifts during substrate binding or allosteric regulation. The fine network of these bonds across the entire structure points to an optimized molecular architecture, designed to balance rigidity and flexibility for function.The visual assessment suggests that hydrogen bonds are distributed throughout the protein, stabilizing both individual subunits and their collective arrangement. These interactions enhance not only the internal cohesion of the polypeptide chains but also their cooperative interaction in the multimeric form. While the current observations provide a qualitative insight into the hydrogen bonding network, a detailed quantitative analysis using dedicated software tools would be necessary to precisely enumerate and localize these interactions for a more comprehensive understanding of the protein's stability and functional dynamics.

Table 2: Atomic Composition

Element	Number of Atoms
Carbon (C)	6,148
Oxygen (O)	2,248
Nitrogen (N)	1,625
Sulfur (S)	99

2. Structure Validation and Quality Assessment

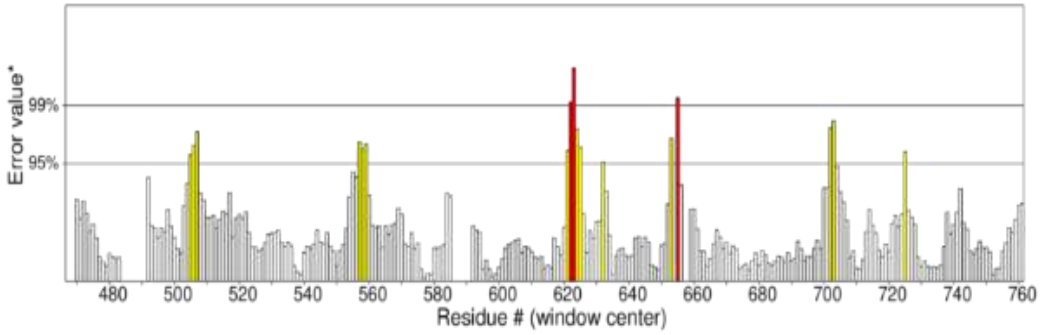


Figure 1: Structure validation graphical representation where yellow and red showing disorder residue

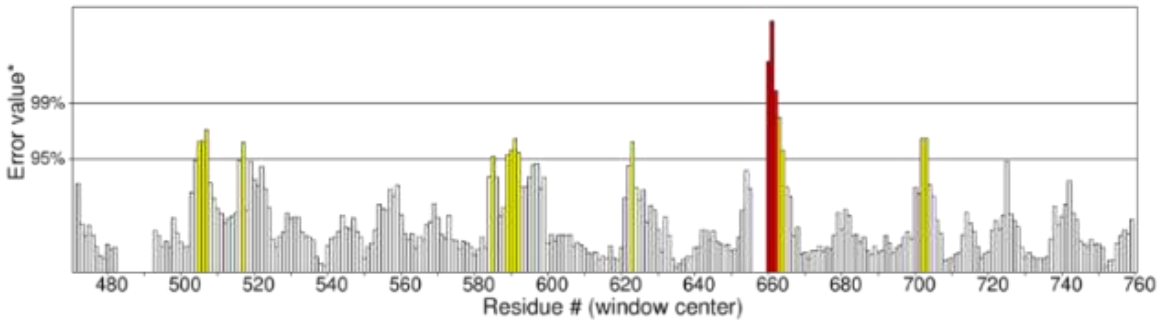


Figure 2: Graphical Representation of Residue-Wise Structural Validation Indicating Disordered Regions

The two figures represent residue-wise structural validation plots for different chains or models of the same protein structure (PDB ID: 8W3D), and they highlight regions of structural disorder. Both figures utilize bar plots to represent the validation scores for individual amino acid residues. The color scheme remains consistent across the figures, where gray bars indicate well-ordered residues, yellow bars denote moderately disordered regions, and red bars highlight highly disordered residues. While the general format and purpose of the figures are similar, there are clear distinctions in the patterns they present.

The structural validation graphs provide a clear representation of the residue-wise disorder within the protein structure. Each vertical bar corresponds to a residue position along the protein sequence, and the color-coded scheme indicates the degree of disorder or deviation from expected structural parameters. These disordered regions could be associated with loops, flexible linkers, or solvent-exposed areas that are inherently mobile or less structured in the crystallographic model. The central peaks, particularly those in red, denote regions of considerable structural uncertainty which may affect functional or binding interpretations if located near active or interaction sites. The surrounding grey regions show a baseline of structural consistency across the protein. The distribution of yellow and red peaks across both chains implies localized flexibility rather than a globally unstable fold. Understanding these fluctuations is important for further refinement or modeling studies and can also offer insights into potential regions of intrinsic disorder that may play regulatory or dynamic roles in protein activity.

The Ramachandran plot provides a detailed visual representation of the phi (φ) and psi (ψ) torsional angles of the amino acid residues within the protein structure. The concentration of black squares primarily within the red and yellow regions indicates that most residues fall within energetically favorable and additionally allowed regions, suggesting a well-modeled and stereochemically sound protein structure. The red

zones represent the most favored conformational angles, while the yellow areas are allowed but slightly less optimal. The high density of residues in these zones reflects good backbone dihedral angle distribution and validates the protein's secondary structural integrity. Only a few residues appear outside these preferred regions, with one such residue explicitly labeled—THR 661 (D)—highlighted in red text, signifying a potential outlier. This outlier may not conform to typical backbone geometry and could result from a region of structural flexibility, disorder, or limitations in model refinement. However, the overall distribution pattern supports that the majority of the backbone dihedrals are within acceptable limits, implying a high-quality protein model. Thus, the Ramachandran plot analysis supports the structural reliability of the modeled protein, with minimal steric clashes or unusual conformations observed.

The PROCHECK summary indicates that the protein model contains a total of 1180 residues. Analysis of the Ramachandran plot reveals that 93.2% of residues lie within the most favored core regions, while 6.7% are in allowed regions, with none in the generously allowed zones and only 0.1% classified as disallowed. This distribution suggests a high-quality structure with proper stereochemistry and backbone dihedral angle geometry. Out of 1160 residues evaluated by the Ramachandran plot, 13 were labeled for special attention, possibly due to deviations from ideal conformations.

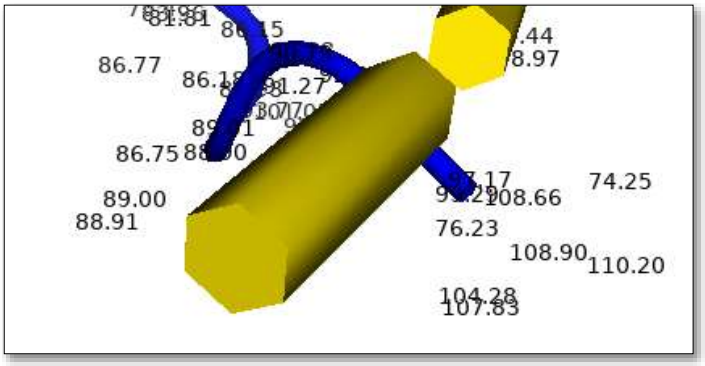
The chi1-chi2 side-chain torsion angle plots identified four labeled residues from 756 evaluated, indicating minor areas that may require further inspection. Additionally, five side-chain parameters were considered better than expected, and none were marked as worse, implying good side-chain geometry overall. Bond length and angle deviations include a maximum deviation of 4.0 and 48 bad contacts, with bond length and angle deviation noted at 5.2. These parameters highlight a few areas where structural refinement could be considered. The dihedral G-factor is -0.06, covalent is 0.50, and the overall G-factor is 0.16, suggesting the model is within acceptable quality thresholds, though some improvement could be made in dihedral angles.

Planar group analysis shows 100% of planar groups are within normal limits with no deviations highlighted. While most of the structure meets standard criteria for model quality, specific residues and contact regions may be worth closer investigation, especially those marked by PROCHECK for potential concern.

3. Chain and Structural Analysis via PyMOL and B-Factor Mapping

Figure 3: Chain identification in PyMol lung cancer sample chain A (yellow), Chain B (green), Chain C (red). Chain d (blue)

The molecular visualization clearly distinguishes four chains within the protein complex using PyMOL. Chain A is shown in yellow, Chain B in green, Chain C in red, and Chain D in blue. This color-coded representation enables straightforward identification of each individual chain, which facilitates structural analysis and inter-chain interaction assessment. The distinct spatial arrangement and folding of the chains reflect a multi-subunit architecture, often characteristic of functionally significant proteins. Such visual differentiation supports further investigation into functional domains, active sites, or potential binding interactions specific to lung cancer-related targets.



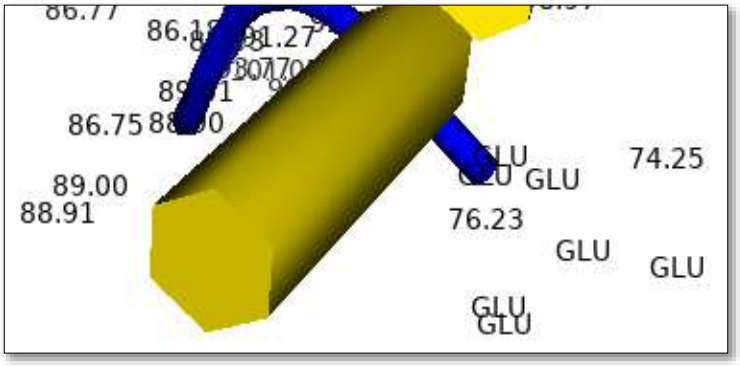


Figure 4 (a): 466 E, 110.20

Figure 4 (b): 661 T, 93.26(residue) yellow (residue) blue

This visualization is focusing on residue behavior, specific positions, such as residues 661 and 466, displayed high B-factor values, suggesting increased flexibility or disorder. Notably, residue 466 E had a B-factor exceeding 100 despite being colored blue, a shade typically used to represent stable regions—pointing to a possible inconsistency between visualization settings and numerical data. This observation may indicate dynamic behavior or potential instability in that region. Residue 661 T, colored yellow and associated with GLU, also showed elevated B-factors, signifying localized mobility, possibly influencing ligand binding or protein interactions.

4. RMSD Analysis for Structural Comparison

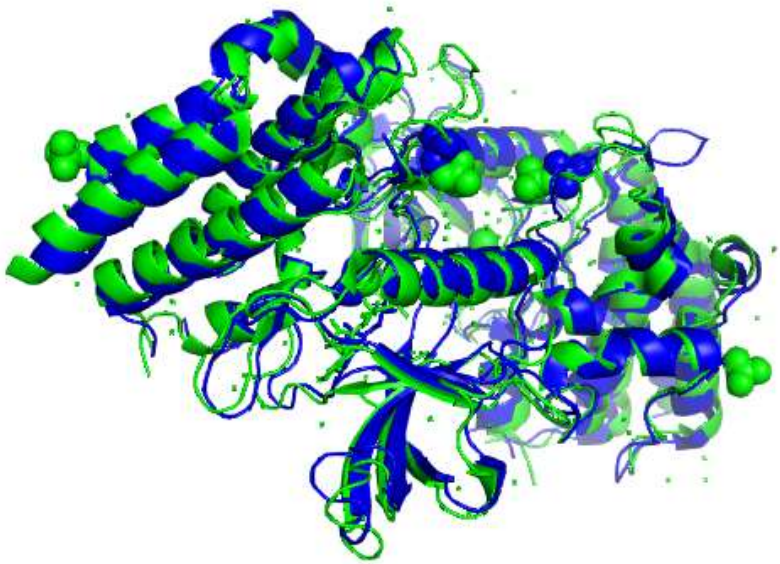


Figure 5: protein-protein docking RMSD calculation (2pz5, 5uhn)

Sequence similarity analysis using BLAST confirmed a high alignment score of 679, indicating strong homology, especially within chain A, corresponding to fibroblast growth factor receptor 2 (FGFR2). The E value approaching zero supports a reliable alignment. Metrics such as query cover, total score, and identity percentage reflect comprehensive coverage and close evolutionary relation of the aligned sequences. Hydrogen bonding and salt bridge interactions were marked by yellow dashed lines, representing polar contacts critical for protein stability and ligand interaction. These non-covalent interactions often play a role in molecular recognition and binding affinity.

A structural comparison using protein-protein docking revealed an RMSD value around 1.2 Å between the analyzed FGFR2 structure and reference models (2PZ5 and 5UHN). This low RMSD implies that the structures are closely aligned, with only minor conformational variations. Such similarity is often observed in homologous proteins or conformational states involved in consistent biological functions. These findings hold importance in understanding resistance mechanisms and treatment variability in lung cancer, as even slight conformational differences can influence drug binding and therapeutic outcomes.

5. Functional Domain Prediction with InterProScan

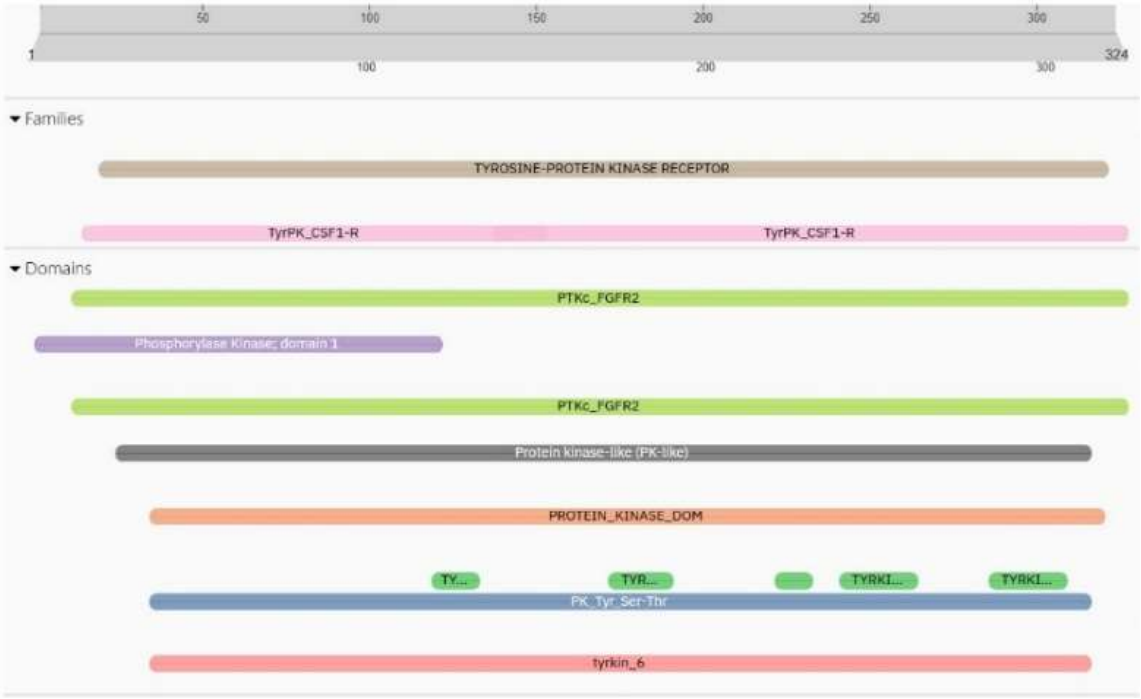


Figure 6: Sequence Scanning Analysis

InterProScan analysis integrated various tools to annotate the protein sequence, identifying conserved domains and families relevant to the fibroblast growth factor receptor. Domains such as Protein kinase-like and FGFR2-specific segments were recognized, with key conserved positions observed at residues 182 and 72. The TYROSINE-PROTEIN KINASE RECEPTOR family and other kinase-related domains were also identified, reinforcing the functional role of the sequence in cellular signaling pathways.

This analysis aids in understanding structural and functional characteristics of FGFR2 in the context of lung cancer. Variations in domain structure, dynamic regions, and conserved motifs provide critical insights into how mutations or alterations may drive disease mechanisms. The identification of these conserved and functional regions is vital in drug discovery and in designing targeted therapies against specific protein domains associated with malignancy.

6. KEGG Pathway Analysis: Non-Small Cell Lung Cancer

Non-small cell lung cancer develops through a complex network of molecular changes involving both genetic mutations and pathway dysregulation. One of the critical pathways implicated is the EGFR signaling cascade, which, when activated due to mutations or overexpression, results in continuous cell proliferation and survival, bypassing normal growth controls. These changes often coincide with alterations in other oncogenes like KRAS and rearrangements such as EML4-ALK and KIF5B-RET. In parallel, the loss of tumor suppressor genes like TP53 and RB1 removes critical checks on cell division and genomic stability, contributing to tumor progression.

The integration of these pathways highlights multiple mechanisms that promote tumorigenesis such as evasion of apoptosis, enhanced cell survival via the PI3K-Akt pathway, and stimulation of the MAPK cascade, which boosts proliferation. Furthermore, disruptions in cell cycle control and DNA repair pathways aggravate genomic instability, a hallmark of cancer. This molecular insight, visualized through pathway analysis, emphasizes the potential for targeted therapies. Personalized treatment strategies can be developed by focusing on specific genetic and pathway abnormalities in individual patients, potentially improving therapeutic outcomes.

7. Cleft and Pore Analysis via PDBsum and Interface Mapping

The protein-protein interaction data highlights two distinct interfaces involving three chains: A, B, and C. The interaction between Chains A and B is characterized by a notably large and complex interface. Both chains contribute 28 interface residues, indicating a highly engaged and possibly biologically significant binding region. The surface areas involved in this interaction are substantial, with Chain A contributing 1395 Å² and Chain B contributing 1442 Å². Such extensive surface involvement typically reflects a stable and strong interaction.

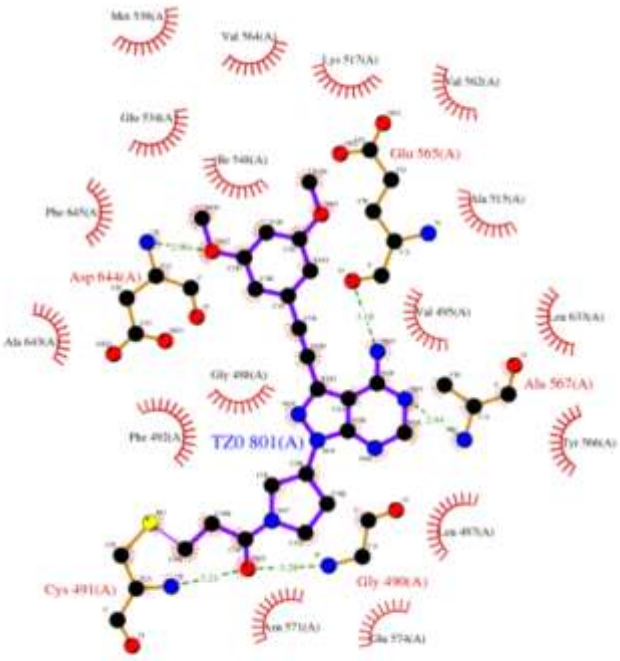
Table 2: Interface Statics Analysis

Interface statistics

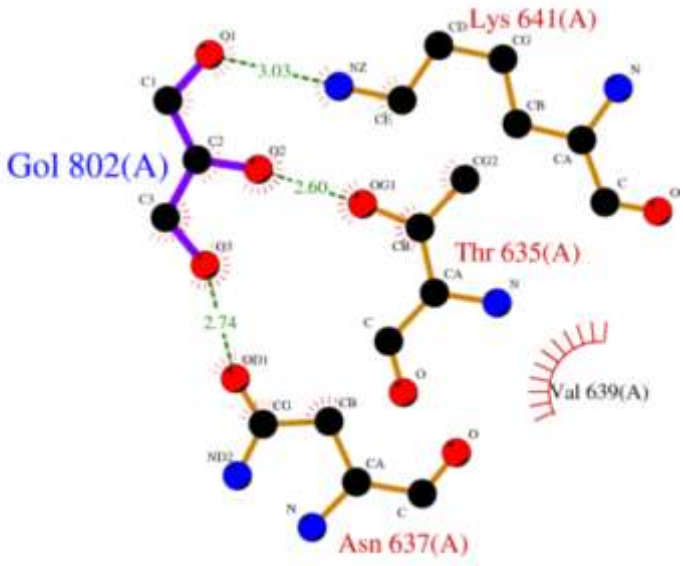
Chains	No. of interface residues	Interface area (Å^2)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
A-B	28 : 28	1395:1442	6	-	14	157
B-C	8 : 5	383:411	2	-	4	32

This interface is further stabilized by multiple non-covalent interactions. Specifically, there are 6 salt bridges, which play a crucial role in maintaining structural integrity and enhancing specificity through ionic interactions. Additionally, 14 hydrogen bonds contribute to the specificity and strength of the interaction by facilitating close-range, directional bonding. A high number of non-bonded contacts (157) further reinforces the tight packing and complementarity between Chains A and B, suggesting a robust and potentially functionally important interaction.

In contrast, the interface between Chains B and C is more limited. Chain B contributes 8 residues, while Chain C contributes only 5. The corresponding interface areas are significantly smaller, at 383 Å² and 411 Å², indicating a more transient or less structurally significant interaction. Despite the smaller interface, 2 salt bridges and 4 hydrogen bonds are present, which still contribute to the stability and specificity of the interaction, albeit to a lesser extent than in the A-B interface. The number of non-bonded contacts here is 32, reflecting a weaker interaction overall. This suggests that while the B-C interaction may be functionally relevant, it is likely secondary in strength and importance compared to the A-B interaction.



TZ0 801(A)



GOL 802(A)

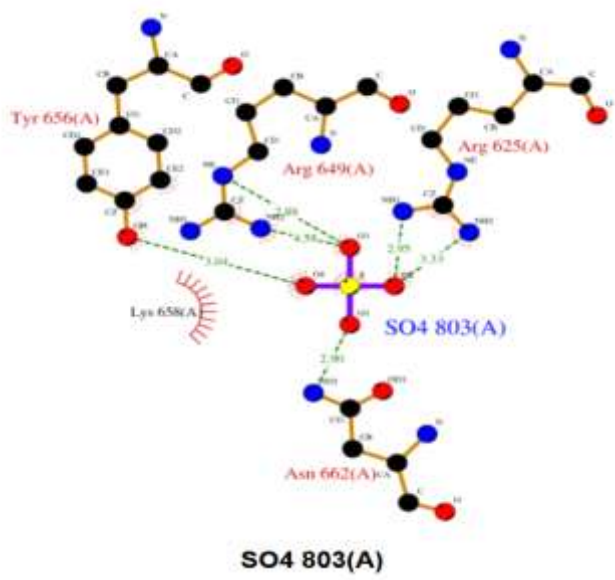


Figure 7: Ligplot of TZ0 801(A), GOL(A), and SO4

The compound TZ0 801(A) exhibited a notable binding affinity with several amino acid residues in the active site of the target protein. Key hydrogen bonds were formed with Gly 488(A), Gly 490(A), Val 495(A), and Glu 565(A), with bond distances ranging from 2.90 to 3.28 Å. Additional interactions were observed with Cys 491(A), Ala 567(A), and Asp 644(A), suggesting that the compound fits well into the binding pocket. The presence of π-π stacking and van der Waals interactions with residues such as Phe 492(A), Glu 534(A), and Ile 548(A) further stabilized the compound within the pocket, contributing to the strong interaction profile.

The sulfate molecule SO4 803(A) formed multiple hydrogen bonds with surrounding residues, including Arg 625(A), Arg 649(A), and Asn 662(A), with bond lengths primarily between 2.88 and 3.33 Å. The central sulfur atom coordinated through its oxygen atoms to multiple amino acid side chains, particularly forming a strong network with positively charged arginine residues, which is characteristic of electrostatic interactions. These interactions suggest a strong and specific binding affinity of the sulfate group, potentially indicating its role in structural stabilization or enzymatic regulation within the active site.

The GOL 802(A) molecule formed three distinct hydrogen bonds with nearby residues: Asn 637(A), Thr 635(A), and Lys 641(A). These interactions occurred at close bond distances of 2.60 to 3.03 Å, suggesting strong hydrogen bonding. The molecule's hydroxyl groups facilitated these interactions, positioning GOL 802(A) effectively within the binding site. The proximity to polar and charged residues implies that this compound may contribute to conformational stability or act as a bridging molecule in the binding network.

Table 3: Cleft Analysis

Volume	R1 Ratio	Accessible Vertices	Buried Vertices	Average Depth	Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine	Ligands
28970.58	1.24	67.36	13.34	25.40	66	60	58	95	35	46	4	TZ0 801[A], SO4 803[A], TZ0 801[B], GOL 802[B] (72 atoms)
23419.55	0.00	68.19	13.37	28.79	56	48	52	82	30	33	3	SO4 805[B], TZ0 801[C], SO4 802[C] (41 atoms)
2313.98	0.00	66.74	11.32	11.29	5	9	4	10	3	5	0	SO4 805[D] (5 atoms)
3019.36	0.00	61.81	10.15	12.21	12	7	5	15	4	6	0	GOL 802[A] (6 atoms)
2840.48	0.00	66.84	10.13	12.35	10	9	4	19	5	10	1	TZ0 801[D] (30 atoms)
1734.33	0.00	66.90	14.10	9.69	4	10	3	8	2	3	0	–
2004.33	0.00	61.24	8.48	9.67	10	5	9	4	3	4	0	SO4 804[A] (4 atoms)
1396.83	0.00	54.80	8.14	8.54	10	5	4	5	1	5	0	–
1001.53	0.00	51.84	8.14	9.86	0	3	2	4	1	3	0	–
993.52	0.00	55.43	7.24	8.95	1	6	3	6	2	3	0	–

Among the identified clefts, the one labeled in red exhibits the highest volume, indicating a potentially significant binding site. It has a notable number of accessible and buried vertices with a moderate depth, suggesting it is both surface-exposed and structurally recessed enough to allow ligand interaction. This cleft accommodates several ligands, including TZ0 and SO4, and displays a rich variety of residue types, with a strong presence of hydrophilic, hydrophobic, and aromatic amino acids. The yellow cleft follows closely, also harboring multiple ligands. It shows slightly higher accessible vertex count and average depth than the red cleft, reinforcing its suitability as a functional binding region. The diversity in residue types here may contribute to specific interactions with its ligands. Other clefts, such as the green and blue ones, show smaller volumes and shallower depths. These are likely to support smaller or transient ligand interactions, given the limited variety in their amino acid composition and fewer interacting atoms. Some clefts, such as the magenta and light blue ones, do not appear to engage with any ligands and have minimal accessible surface area and volume, making them less likely to be active sites. Overall, the cleft analysis highlights that the red and yellow clefts are the most promising for ligand interaction based on size, depth, residue composition, and observed binding with multiple molecules.

Table 4: Pore Analysis

Radius	Free R	Length	HPathy	HPhob	Polar	RelMut	Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine	Ligands
1.30	1.46	33.1	-0.37	-0.09	19.9	76	5	0	5	2	0	0	0	TZ0 801 C
3.21	3.43	33.7	-2.37	-0.55	27.9	78	8	3	2	1	2	1	0	
1.77	1.95	42.4	-1.73	-0.42	25.1	85	5	2	1	3	0	1	0	SO4 805 B
2.58	3.67	56.1	-2.07	-0.45	22.5	74	10	3	2	0	4	1	0	SO4 805 B
1.48	2.11	65.9	-0.99	-0.33	20.4	80	3	7	5	5	2	3	0	SO4 805 B
2.24	3.97	79.6	-3.05	-0.59	38.1	78	12	9	3	2	1	3	0	SO4 805 B
1.83	1.94	90.3	-1.30	-0.42	18.6	84	8	9	7	5	1	3	0	
2.09	3.10	102.6	-2.40	-0.49	31.1	82	11	6	4	4	1	1	0	SO4 805 B
2.08	3.10	125.3	-2.64	-0.52	32.0	81	15	7	7	5	3	1	0	
1.46	2.13	128.2	-1.48	-0.43	24.6	79	11	13	7	3	7	4	1	
1.74	1.74	36.8	-2.25	-0.44	28.5	74	5	5	1	2	1	0	0	GOL 802 A

Among the detected pores, Pore 2 stands out with the largest free radius of 3.21 Å and a considerable length of 33.7 Å. It has a moderate polarity value and includes a diverse set of residues, which might contribute to potential ligand interactions. Despite having a lower hydrophobicity and hydrophilicity score, its structural openness and residue composition suggest it could accommodate molecular passage or binding. Pore 6 also shows promise due to its longer length of 79.6 Å and a higher polarity score. It hosts 12 positive and 9 aliphatic residues, indicating both electrostatic and hydrophobic interaction potential. This pore has an associated ligand (SO4 805 B), supporting its functional role.

Pores 8 and 9 are notably long (102.6 Å and 125.3 Å, respectively) and contain high counts of positive and hydrophobic residues. These features hint at their capability to facilitate molecular transit or selective ligand interaction. Both also display significant relative mutability scores, which might reflect evolutionary adaptability in these regions. Pore 10, the longest (128.2 Å), contains a rich variety of residues, especially in the positive, aliphatic, and aromatic categories. This mix suggests multiple interaction modes, making it potentially relevant in ligand recognition or transport. Overall, pores with longer lengths and balanced residue diversity especially Pores 6, 8, 9, and 10—appear structurally and chemically poised for interaction, with existing ligand associations reinforcing their biological relevance.

8. Molecular Docking

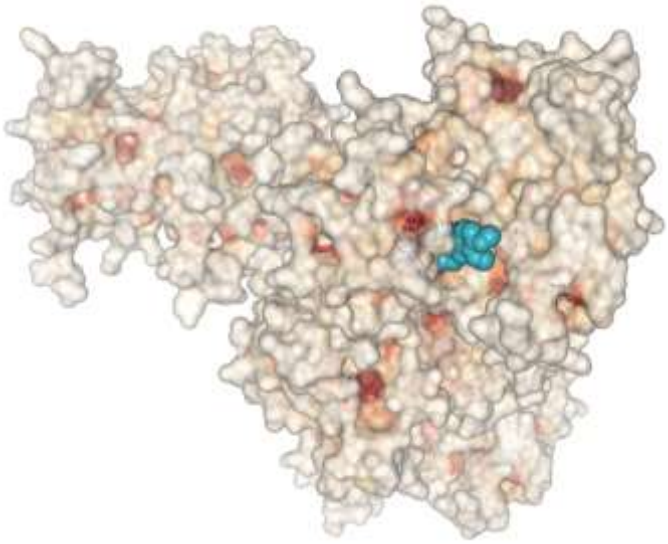


Figure 8: Docked image of protein 8W3D with Berberine

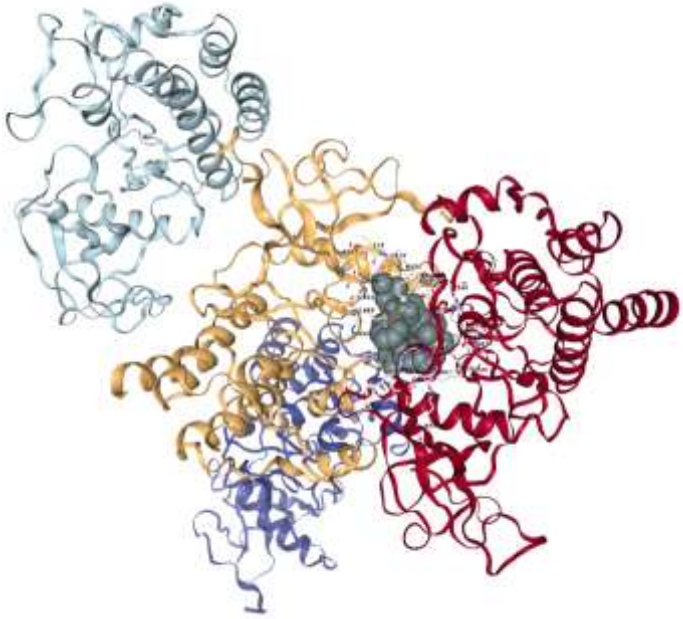


Figure 9: Docked image of protein 8W3D with Docetaxel

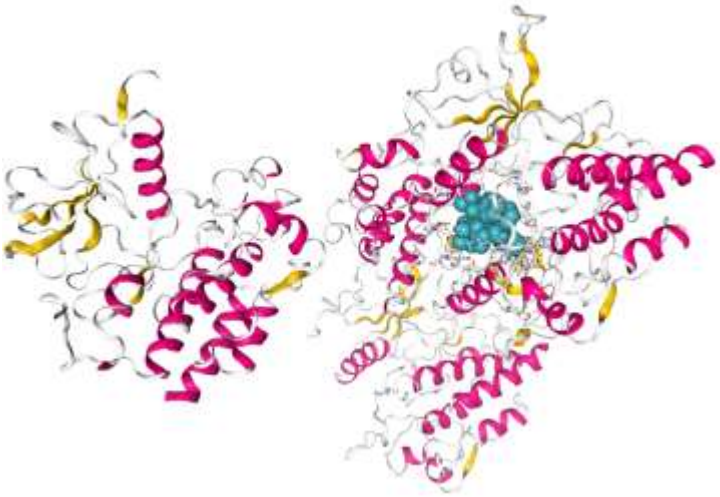


Figure 10: Docked image of protein 8W3D with Paclitaxel

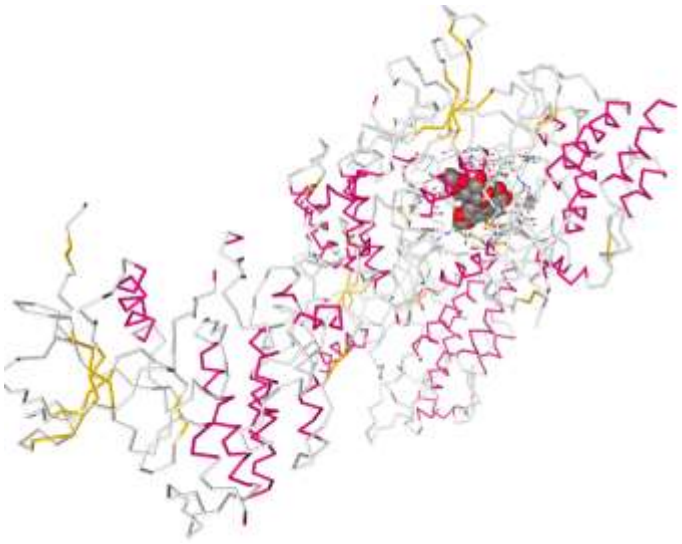


Figure 11: Docked image of protein 8W3D with etoposide

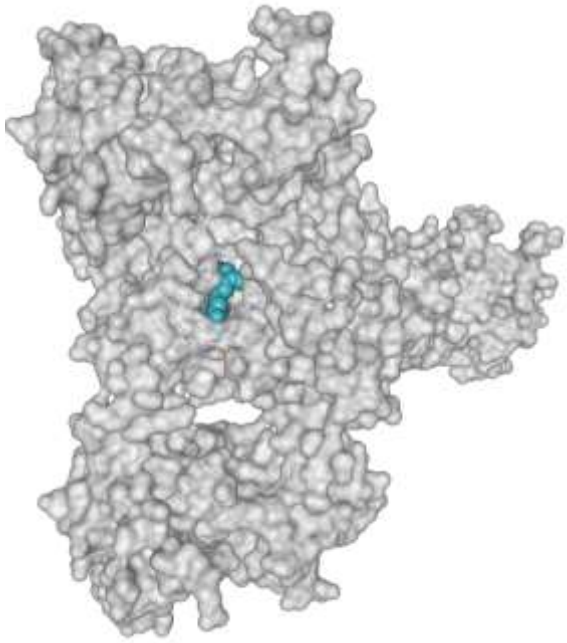


Figure 12: Docked image of protein with ligand curcumin

Table 3: Docking Result

Protein	PubChem ID	Drug Name	Vina Score	Centre (x, y, z)	Docking Size (x, y, z)
8W3D	2353	Berberine	-9.0	(2, 5, -39)	(35, 31, 32)
8W3D	148124	Docetaxel	-9.2	(2, -42, 3)	(35, 26, 26)
8W3D	36314	Paclitaxel	-9.8	(2, -42, 3)	(35, 27, 27)
8W3D	36462	Etoposide	-8.6	(2, -42, 3)	(35, 25, 25)
8W3D	969516	Curcumin	-10.1	(5, -39, -59)	(35, 26, 26)

Molecular docking results demonstrate that curcumin has the strongest binding affinity among the compounds evaluated, reflected by its lowest vina score. This suggests a more favorable interaction with the target site. Other compounds such as paclitaxel and docetaxel also show strong binding potential. The consistency in docking center and size parameters ensures a reliable comparison across all tested molecules, supporting the identification of promising drug candidates for further investigation. Computer-aided drug design offers a crucial link between theoretical research and the practical development of effective treatments. Molecular docking studies provide insight into the interaction between drugs and target proteins, helping identify promising therapeutic agents.

Among the tested compounds, curcumin displayed the strongest binding affinity with a docking score of -10.1, followed by paclitaxel (-9.8), docetaxel (-9.2), and erlotinib (-9.2). These results indicate strong potential for effective binding to the target receptor, particularly curcumin, which interacted with multiple residues across Chain C of the EGFR protein. Erlotinib, a known EGFR inhibitor used in lung cancer therapy, engaged key residues such as LEU487, GLY488, and LYS520, suggesting its relevance as a reference compound. Afatinib showed moderate binding affinity with a score of -8.1 and interacted with several important residues in Chains A and B. Plant-derived compounds like paclitaxel and docetaxel, known for their roles in disrupting mitosis during the M phase, also demonstrated notable interactions with EGFR binding pockets. Etoposide, effective in the S and G2 phases of the cell cycle, interacted with residues critical to receptor functionality and displayed a docking score of -8.6. Overall, natural compounds such as curcumin and paclitaxel revealed strong binding profiles, indicating their potential as alternative or complementary therapeutic options for lung cancer treatment.

This study demonstrates the potential of integrating next-generation sequencing (NGS) with computational drug discovery to identify novel therapeutic agents targeting FGFR2 mutations in non-small cell lung cancer (NSCLC). By analyzing publicly available NGS datasets, we successfully pinpointed recurrent mutations in the FGFR2 gene, which are implicated in tumor development and resistance to existing treatments. The identification of such driver mutations provides a crucial foundation for the development of personalized therapeutic strategies.Our structural modeling of the mutant FGFR2 protein allowed for detailed molecular docking studies with various plant-derived compounds. Several phytochemicals exhibited strong binding affinities to the mutated FGFR2 active site, suggesting their potential as selective inhibitors. These results align with previous reports that emphasize the bioactivity of medicinal plants in cancer therapy, supporting their relevance in modern drug discovery pipelines. While our in silico findings are promising, it is important to recognize the limitations. Computational models, although highly predictive, must be complemented by experimental validation to confirm biological efficacy and safety. Moreover, cancer is a multifactorial disease, and targeting FGFR2 alone may not be sufficient in cases with multiple oncogenic drivers. Thus, combination therapies or multi-targeted approaches may yield better clinical outcomes.

CONCLUSION

This research illustrates a robust pipeline for identifying potential plant-based inhibitors against FGFR2 mutations in lung cancer by harnessing the power of NGS and computational modeling. The study not only underscores the importance of FGFR2 as a therapeutic target but also highlights the untapped potential of medicinal plants in oncology.The shortlisted compounds from our virtual screening and pharmacokinetic analysis show considerable promise and warrant further investigation through in vitro and in vivo studies. Ultimately, this integrative approach contributes to the evolving field of precision medicine and opens new avenues for the development of safer, plant-derived anti-cancer drugs tailored to genetic alterations in tumors.

REFERENCE

1) Lim S.M., Kim E.Y., Kim H.R., Ali S.M., Greenbowe J.R., Shim H.S., Chang H., Lim S., Paik S., Cho B.C. Genomic profiling of lung adenocarcinoma patients reveals therapeutic targets and confers clinical benefit when standard molecular testing is negative. Oncotarget. 2016; 7:24172–24178. doi: 10.18632/oncotarget.8138. *[DOI] [PMC free article] [PubMed] [Google Scholar.*

2) Hagemann I.S., Devarakonda S., Lockwood C.M., Spencer D.H., Guebert K., Bredemeyer A.J., Al-Kateb H., Nguyen T.T., Duncavage E.J., Cottrell C.E., et al. Clinical next-generation sequencing in patients with non-small cell lung cancer. Cancer. 2015; 121:631–639. doi: 10.1002/cncr.29089. [DOI] [PubMed] [Google Scholar]

3) Moskalev E.A., Stöhr R., Rieker R., Hebele S., Fuchs F., Sirbu H., Mastitsky S.E., Boltze C., König H., Agaimy A., et al. Increased detection rates of EGFR and KRAS mutations in NSCLC specimens with low tumour cell content by 454 deep sequencing. Virchows Arch. Int. J. Pathol. 2013; 462:409–419. doi: 10.1007/s00428-013-1376-6. [DOI] [PMC free article] [PubMed] [Google Scholar]

4) Taylor C., Chacko S., Davey M., Lacroix J., MacPherson A., Finn N., Wajnberg G., Ghosh A., Crapoulet N., Lewis S.M., et al. Peptide-Affinity Precipitation of Extracellular Vesicles and Cell-Free DNA Improves Sequencing Performance for the Detection of Pathogenic Mutations in Lung Cancer Patient Plasma. Int. J. Mol. Sci. 2020; 21:9083. doi: 10.3390/ijms21239083. [DOI] [PMC free article] [PubMed] [Google Scholar]

5) Govind K.B., Koppaka D., Dasappa L., Jacob L.A., Babu S.M., Lokesh N.K., Haleshappa R.A., Rajeev L.K., Saldanha S.C., Abhishek A., et al. Detection of clinically relevant epidermal growth factor receptor pathway mutations in circulating cell-free tumor DNA using next generation sequencing in squamous cell carcinoma lung. South Asian J. Cancer. 2019; 8:247–249. doi: 10.4103/sajc.sajc_281_18. [DOI] [PMC free article] [PubMed] [Google Scholar]

6) Yamamoto G., Kikuchi M., Kobayashi S., Arai Y., Fujiyoshi K., Wakatsuki T., Kakuta M., Yamane Y., Iijima Y., Mizutani H., et al. Routine genetic testing of lung cancer specimens derived from surgery, bronchoscopy and fluid aspiration by next generation sequencing. Int. J. Oncol. 2017; 50:1579–1589. doi: 10.3892/ijo.2017.3935. [DOI] [PMC free article] [PubMed] [Google Scholar]

7) Li H., Xie Y., Lin Y., Yu T., Yin Z. Different Gene Mutation Spectrum of the Paired CSF and Plasma Samples in Lung Adenocarcinoma with Leptomeningeal Metastases: The Liquid Biopsy Based on Circulating Tumor DNA. Zhongguo Fei Ai Za Zhi = Chin. J. Lung Cancer. 2020; 23:646–654. doi: 10.3779/j.issn.1009-3419.2020.102.14. [DOI] [PMC free article] [PubMed] [Google Scholar]

8) Eberhard D.A., Johnson B.E., Amler L.C., Goddard A.D., Heldens S.L., Herbst R.S., Ince W.L., Jänne P.A., Januario T., Johnson D.H., et al. Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2005; 23:5900–5909. doi: 10.1200/JCO.2005.02.857. [DOI] [PubMed] [Google Scholar]

9) Lam T.C., Tsang K.C., Choi H.C., Lee V.H., Lam K.O., Chiang C.L., So T.H., Chan W.W., Nyaw S.F., Lim F., et al. Combination atezolizumab, bevacizumab, pemetrexed and carboplatin for metastatic EGFR mutated NSCLC after TKI failure. Lung Cancer. 2021; 159:18–26. doi: 10.1016/j.lungcan.2021.07.004. [DOI] [PubMed] [Google Scholar]

10) Riudavets M., Bosch-Barrera J., Cabezón-Gutiérrez L., Diz Taín P., Hernández A., Alonso M., Blanco R., Gálvez E., Insa A., Mielgo X., et al. Efficacy of nintedanib plus docetaxel in patients with refractory advanced epidermal growth factor receptor mutant lung adenocarcinoma. Clin. Transl. Oncol. Off. Publ. Fed. Span. Oncol. Soc. Natl. Cancer Inst. Mex. 2021 doi: 10.1007/s12094-021-02661-2. [DOI] [PubMed] [Google Scholar]

11) Pratilas C.A., Hanrahan A.J., Halilovic E., Persaud Y., Soh J., Chitale D., Shigematsu H., Yamamoto H., Sawai A., Janakiraman M., et al. Genetic predictors of MEK dependence in non-small cell lung cancer. Cancer Res. 2008; 68:9375–9383. doi: 10.1158/0008-5472.CAN-08-2223. [DOI] [PMC free article] [PubMed] [Google Scholar]

12) Facchinetti F., Lacroix L., Mezquita L., Scoazec J.Y., Loriot Y., Tselikas L., Gazzah A., Rouleau E., Adam J., Michiels S., et al. Molecular mechanisms of resistance to BRAF and MEK inhibitors in BRAF(V600E) non-small cell lung cancer. Eur. J. Cancer. 2020; 132:211–223. doi: 10.1016/j.ejca.2020.03.025. [DOI] [PubMed] [Google Scholar]

13) Sánchez-Torres J.M., Viteri S., Molina M.A., Rosell R. BRAF mutant non-small cell lung cancer and treatment with BRAF inhibitors. Transl. Lung Cancer Res. 2013; 2:244–250. doi: 10.3978/j.issn.2218-6751.2013.04.01. [DOI] [PMC free article] [PubMed] [Google Scholar]

14) Shibata T., Kokubu A., Tsuta K., Hirohashi S. Oncogenic mutation of PIK3CA in small cell lung carcinoma: A potential therapeutic target pathway for chemotherapy-resistant lung cancer. Cancer Lett. 2009; 283:203–211. doi: 10.1016/j.canlet.2009.03.038. [DOI] [PubMed] [Google Scholar]

15) André F., Ciruelos E.M., Juric D., Loibl S., Campone M., Mayer I.A., Rubovszky G., Yamashita T., Kaufman B., Lu Y.S., et al. Alpelisib plus fulvestrant for PIK3CA-mutated, hormone receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: Final overall survival results from SOLAR-1. Ann. Oncol. Off. J. Eur. Soc. Med Oncol. 2021; 32:208–217. doi: 10.1016/j.annonc.2020.11.011. [DOI] [PubMed] [Google Scholar]

16) Gao J., Wu H., Shi X., Huo Z., Zhang J., Liang Z. Comparison of Next-Generation Sequencing, Quantitative PCR, and Sanger Sequencing for Mutation Profiling of EGFR, KRAS, PIK3CA and BRAF in Clinical Lung Tumors. Clin. Lab. 2016; 62:689–696. doi: 10.7754/Clin.Lab.2015.150837. [DOI] [PubMed] [Google Scholar]

17) Xu X., Yang Y., Li H., Chen Z., Jiang G., Fei K. Assessment of the clinical application of detecting EGFR, KRAS, PIK3CA and BRAF mutations in patients with non-small cell lung cancer using next-generation sequencing. Scand. J. Clin. Lab. Investig. 2016; 76:386–392. doi: 10.1080/00365513.2016.1183813. [DOI] [PubMed] [Google Scholar]

18) Vaughn C.P., Costa J.L., Feilotter H.E., Petraroli R., Bagai V., Rachiglio A.M., Marino F.Z., Tops B., Kurth H.M., Sakai K., et al. Simultaneous detection of lung fusions using a multiplex RT-PCR next generation sequencing-based approach: A multi-institutional research study. BMC Cancer. 2018; 18:828. doi: 10.1186/s12885-018-4736-4. [DOI] [PMC free article] [PubMed] [Google Scholar]

19) Benayed R., Offin M., Mullaney K., Sukhadia P., Rios K., Desmeules P., Ptashkin R., Won H., Chang J., Halpenny D., et al. High Yield of RNA Sequencing for Targetable Kinase Fusions in Lung Adenocarcinomas with No Mitogenic Driver Alteration Detected by DNA Sequencing and Low Tumor Mutation Burden. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2019; 25:4712–4722. doi: 10.1158/1078-0432.CCR-19-0225. [DOI] [PMC free article] [PubMed] [Google Scholar]

20) Volckmar A.L., Leichsenring J., Kirchner M., Christopoulos P., Neumann O., Budczies J., Morais de Oliveira C.M., Rempel E., Buchhalter I., Brandt R., et al. Combined targeted DNA and RNA sequencing of advanced NSCLC in routine molecular diagnostics: Analysis of the first 3000 Heidelberg cases. Int. J. Cancer. 2019; 145:649–661. doi: 10.1002/ijc.32133. [DOI] [PubMed] [Google Scholar]

21) Sayle, R. A., & Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. *Trends in biochemical sciences*, *20*(9), 374. https://doi.org/10.1016/s0968-0004(00)89080-5

22) Bernstein H. J. (2000). Recent changes to RasMol, recombining the variants. *Trends in biochemical sciences*, *25*(9), 453–455. https://doi.org/10.1016/s0968-0004(00)01606-6

23) Goodsell D. S. (2005). Representing structural information with RasMol. *Current protocols in bioinformatics*, *Chapter 5*, . https://doi.org/10.1002/0471250953.bi0504s11

24) Seeliger, D., & de Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of computer-aided molecular design*, *24*(5), 417–422. https://doi.org/10.1007/s10822-010-9352-6

25) Dr Uma kumari, Devanshi Gupta, In silico RNA aptamer drug design and modelling,2022/4, Journal-JETIR,Volume-9, Issue-4, Pages 718-725

26) Dr Uma, Devanshi Gupta, Study of Malignant Melonema Causing Protein "5ixb" Using Multiomics Approach, November 2022, INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY , Volume 9 Issue 6 , ISSN: 2349-6002

27) Elrashedy, A., Nayel, M., Salama, A., Salama, M. M., & Hasan, M. E. (2024). Bioinformatics approach for structure modeling, vaccine design, and molecular docking of Brucella candidate proteins BvrR, OMP25, and OMP31. *Scientific reports*, *14*(1), 11951. https://doi.org/10.1038/s41598-024-61991-7

28) Roy, S. K., Biswas, M. S., Foyzur Raman, M., Hasan, R., Rahmann, Z., & Uddin P K, M. M. (2024). A computational approach to developing a multi-epitope vaccine for combating Pseudomonas aeruginosa-induced pneumonia and sepsis. *Briefings in bioinformatics*, *25*(5), bbae401. https://doi.org/10.1093/bib/bbae401

29) Verma, R., Pradhan, D., Nayek, A., Singh, H., Jain, A. K., & Khan, L. A. (2022). Target-based drug repurposing against Candida albicans-A computational modeling, docking, and molecular dynamic simulations study. *Journal of cellular biochemistry*, *123*(2), 289–305. https://doi.org/10.1002/jcb.30163

30) Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

31) Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic acids research*, *33*(Web Server issue), W116–W120. https://doi.org/10.1093/nar/gki442

32) Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic acids research*, *51*(D1), D587–D592. https://doi.org/10.1093/nar/gkac963

33) Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

34) Liu, Y., Yang, X., Gan, J., Chen, S., Xiao, Z. X., & Cao, Y. (2022). CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic acids research*, *50*(W1), W159–W164. https://doi.org/10.1093/nar/gkac394

35) Liu, Y., Grimm, M., Dai, W. T., Hou, M. C., Xiao, Z. X., & Cao, Y. (2020). CB-Dock: a web server for cavity detection-guided protein-ligand blind docking. *Acta pharmacologica Sinica*, *41*(1), 138–144. https://doi.org/10.1038/s41401-019-0228-6

36) Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein science : a publication of the Protein Society*, *27*(1), 129–134. https://doi.org/10.1002/pro.3289

37) Laskowski, R. A., & Thornton, J. M. (2022). PDBsum extras: SARS-CoV-2 and AlphaFold models. *Protein science : a publication of the Protein Society*, *31*(1), 283–289. https://doi.org/10.1002/pro.4238

38) Kumari Uma & Agrawal, Nidhi. (2023). NGS and Mutational Profile Analysis of Non-Small-Cell Lung Carcinoma (NSCLC). International Journal for Research in Applied Science and Engineering Technology. 11. 3090-3094. 10.22214/ijraset.2023.50880.

39) Kumari, Uma & Gupta, Shruti. (2023). NGS and Sequence Analysis with Biopython for Prospective Brain Cancer Therapeutic Studies. International Journal for Research in Applied Science and Engineering Technology. 11. 10.22214/ijraset.2023.50885.

40) Uma Kumari, Vineeta Johri, Swarali Dhopate, Tijil Jha Structure Based Drug Designing for the prediction of epitope for targeting Malignant brain Tumor, 2024 JETIR July 2024, Volume 11, Issue 7 www.jetir.org (ISSN-2349-5162)

41) Tanmay Bandbe , Juri Saikia , Uma Kumari. (2025). NGS Analysis Human Papillomavirus Type 18 E2 DNA-Binding Domain Bound to its DNA Target with Biopython. *South Eastern European Journal of Public Health*, 3781–3792. https://doi.org/10.70135/seejph.vi.5856,SEEJPH Volume XXVI, S2

42) Uma Kumari,Taruni Bajaj "Ngs Data Analysis And Active Site Identification Alpha-1-Acid Glycoprotein Bound To Potent Anti Tumor Compound UCN-01 In Malignant Brain Tumor", International Journal of Emerging Technologies and Innovative Research ,Vol.12, Issue 3, page no.g107-g116, **http://doi.one/10.1729/Journal.44259**

43) Uma Kumari,Renu et al "Structure Analysis and molecular Docking of Mesothelin-207 fragment in human cancer", Jetir,Vol.12, Issue 3, page no.g98-g106, http://doi.one/10.1729/Journal.44261

44) Uma Kumari,Gaurav Verma ,Murugesan AnbuMegala "IN SILICO DRUG DESIGN IN HUMAN APOPTOSIS INDUCING FACTOR (AIF) IN LUNG CANCER", International Journal of Emerging Technologies and Innovative ISSN:2349-5162, Vol.12, Issue 2, page no. ppd139-d148, **http://doi.one/10.1729/Journal.43606**

45) Uma Kumari,Karren Mehrotra,"NGS and Proteomic gene expression analysis in Studies of NUDT5 silence hormone signaling in breast cancer", International Journal of Emerging Technologies and Innovative Research , Vol.12, Issue 2, page no. ppc630-c638, http://doi.one/10.1729/Journal.43563

46) Shipra Chaudhary Uma Kumari, "NGS, MOLECULAR DOCKING AND NETWORK PHARMACOLOGY REVEAL POTENT INHIBITOR FOR THE TREATMENT OF LUNG CANCER", International Journal of Emerging Technologies and Innovative Research , Vol.11, Issue 9, page no. ppf116-f126, https://doi.one/10.1729/Journal.41696

47) Uma Kumari ,Eaganayagan Malligaarjun"Next-Generation Sequence and Structural Analysis of Oncogenic KRAS Q61E GMPPCP-bound in Human Lung Cancer", International Journal of Emerging Technologies and Innovative Research,jetir, Vol.12, Issue 4, page no. ppk26-k38, **2025**, http://doi.one/10.1729/Journal.44947