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

Integrating NGS and CADD for the Identification of Medicinal Plant Compounds Targeting EGFR Mutants in Non-Small Cell Lung Cancer

Uma Kumari[1] Vineeta Johri[2]

[1]Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida-201301, India
[2]Project Trainee at Bioinformatics Project and Research Institute, Noida-201301, India
Corresponding Author: Uma Kumari (uma27910@gmail.com

Abstract: Non-Small Cell Lung Cancer remains a significant global contributor to cancer deaths, largely due to Epidermal Growth Factor Receptor (EGFR) mutation that drive disease progression and treatment resistance. This research seeks to uncover new therapeutic avenues for EGFR-mutant NSCLC by combining Next-Generation Sequencing (NGS) with Computer-Aided Drug Design (CADD). We'll leverage NGS to thoroughly examine genetic changes in tumor samples from patients, specifically looking at the LRTM mutation within the 8D76 EGFR structure. This analysis will help us identify biomarkers and new drug targets associated with these EGFR mutations. Building on this genetic data, we will apply CADD methods including molecular docking, virtual screening, and molecular dynamics simulations to screen a library of natural compounds derived from medicinal plants. The goal is to identify plant-based compounds with strong binding affinity and inhibitory activity against the mutant EGFR protein. This perspective accentuates the therapeutic capacity of natural products in oncology and the imperative of amalgamating indigenous knowledge with sophisticated biotechnological methodologies. Through the identification of efficacious lead compounds, this inquiry is poised to facilitate the evolution of more durable and precise interventions for individuals afflicted with non-small cell lung carcinoma (NSCLC) exhibiting refractory EGFR mutations. In doing so, it showcases a modern, holistic approach to drug discovery, harnessing both cutting-edge genomic analysis and the therapeutic potential of medicinal plants.

Keywords: **Next-Generation Sequencing (NGS); Computational Aided Drug Design (CADD); Medicinal Plant Compounds; EGFR Mutants; Non-Small Cell Lung Cancer (NSCLC); 8D76; Molecular Docking; Phytochemicals; Targeted Therapy; Precision Medicine**

INTRODUCTION

Lung cancer continues to be a formidable global health crisis, impacting diverse populations and socioeconomic groups. Non-small cell lung cancer (NSCLC) accounts for the vast majority of cases, nearly 85%, and encompasses various histological subtypes like adenocarcinoma, squamous cell carcinoma, and large cell carcinoma, each with distinct biological and clinical characteristics. Despite advancements in early detection, surgical techniques, and therapeutic interventions, the prognosis for NSCLC patients, especially those diagnosed at advanced stages, remains poor. Lung cancer's persistent standing as the leading cause of cancer-related deaths worldwide underscores an urgent need for innovative treatment strategies [1, 2, 3, 4, 5].

Recent years have seen a transformative shift in NSCLC diagnosis and treatment, driven by a deeper understanding of its molecular foundations. A pivotal discovery has been the role of the Epidermal Growth Factor Receptor (EGFR) and its genetic alterations in promoting tumor growth. EGFR, a transmembrane receptor tyrosine kinase, is crucial for regulating essential cellular processes such as proliferation, differentiation, and survival. Mutations in the EGFR gene can lead to the constant activation of its kinase domain, resulting in uncontrolled cell growth and resistance to apoptosis—a hallmark of cancer. This insight has spurred the development of targeted therapies designed to inhibit the abnormal signaling driven by mutant EGFR proteins [1, 2, 3, 4, 5]. Initially, first-generation EGFR tyrosine kinase inhibitors (TKIs) like gefitinib and erlotinib showed significant promise by specifically targeting these mutations, yielding substantial clinical benefits. However, their effectiveness is often short-lived. Patients typically respond to these treatments only to develop resistance within about a year of starting therapy. The emergence of resistance mutations, notably T790M, C797S, and other alterations, has been identified as a primary mechanism of acquired resistance. These mutations

either modify the drug-binding site or activate alternative signaling pathways, rendering existing therapies ineffective and necessitating a continuous quest for new, more durable treatments [1, 2, 3, 4, 5]. A unique aspect of our methodology lies in the source of potential therapeutic agents. While synthetic molecules have historically dominated the pharmaceutical landscape, natural products derived from medicinal plants have consistently demonstrated remarkable bioactivity and structural diversity. Historically, nature has served as a rich source of medicinal compounds, with plant-derived molecules forming the basis of many frontline cancer therapies, including paclitaxel and vincristine. These natural products often possess complex three-dimensional architectures and unique functional groups that interact with biological targets in ways synthetic compounds may not easily replicate. Recognizing this potential, our research will involve screening a carefully selected library of plant-based compounds against the mutant EGFR protein. By harnessing the structural and functional diversity inherent in phytochemicals, we hope to uncover novel inhibitors capable of effectively binding and modulating the activity of EGFR variants that drive resistance. This emphasis on natural products not only aligns with historical precedent but also taps into a growing recognition of the need for more sustainable and environmentally conscious drug discovery practices [6, 7, 8].

However, the promise of natural products must be balanced with rigorous scientific evaluation. The sheer number of potential compounds necessitates robust computational screening to identify those with the highest likelihood of success. In our study, we will employ a multi-step CADD pipeline beginning with molecular docking to estimate binding modes and affinities. This will be followed by molecular dynamics simulations, which provide a dynamic view of protein-ligand interactions over time, capturing subtle conformational changes that static docking models might miss. By combining these techniques, we aim to generate a shortlist of candidate compounds with favorable interaction profiles and pharmacological properties [9, 10, 11]. The genetic data supporting this screening process will be generated through NGS of patient-derived NSCLC tumor samples, with a specific focus on those harboring the LRTM mutation. This high-resolution genetic information will enable us to construct accurate structural models of the mutant EGFR protein, ensuring our computational predictions are grounded in biologically relevant contexts. Furthermore, by analyzing the broader mutational landscape, we aim to identify co-occurring alterations that may modulate EGFR behavior and influence drug sensitivity or resistance [9, 10, 11]. It's essential to acknowledge the challenges inherent in this ambitious endeavor. Translating computational predictions into clinical realities involves multiple layers of validation, from in vitro assays to animal models and eventually human clinical trials. Moreover, the dynamic and often unpredictable nature of cancer biology means that even the most promising compounds may encounter hurdles in terms of bioavailability, toxicity, and efficacy in real-world settings [11, 12, 13]. Nevertheless, we believe that the convergence of advanced genomic profiling, cutting-edge computational modeling, and the rich pharmacological potential of medicinal plants offers a powerful framework for discovery. Our hope is that by combining these elements, we can uncover new therapeutic avenues that extend and improve the quality of life for patients battling this devastating disease [11, 12, 13]. The introduction of these first-generation TKIs marked a significant advance in the treatment of NSCLC, offering a targeted approach that resulted in improved progression-free survival and higher response rates compared to conventional chemotherapy. However, the benefits of these therapies are typically transient, as most patients develop resistance to the drugs within a year of starting treatment. This phenomenon of acquired resistance represents a major clinical challenge. One of the most common mechanisms behind this acquired resistance is the emergence of the T790M point mutation within exon 20 of the EGFR gene. This mutation involves the substitution of threonine with methionine at position 790. It is particularly noteworthy because it increases the affinity of the ATP-binding site of EGFR for ATP itself, effectively outcompeting the TKIs that are designed to inhibit this site. As a result, the inhibitory effect of drugs like gefitinib and erlotinib is significantly reduced. Clinical studies suggest that the T790M mutation is responsible for more than half of the cases of acquired resistance in EGFR-mutant NSCLC [14, 15, 16].

The advent of Next-Generation Sequencing (NGS) has dramatically transformed our understanding of the molecular landscape of cancer, including Non-Small Cell Lung Cancer (NSCLC). Traditional sequencing methods, such as Sanger sequencing, were limited by their relatively low throughput and inability to detect low-frequency variants within complex tumor samples. NGS, by contrast, enables massively parallel sequencing of millions of DNA fragments simultaneously, allowing for a comprehensive and detailed analysis of the tumor genome [17, 18, 19]. One of the most significant breakthroughs in NGS is its ability to detect a broad spectrum of genomic alterations in a single assay. This includes single nucleotide variants (SNVs), small insertions and deletions (indels), copy number variations (CNVs), and structural rearrangements such as gene fusions. The multiplexing capability of NGS also permits the simultaneous sequencing of multiple samples, greatly improving the efficiency of data generation and analysis [17, 18, 19]. The integration of NGS into routine clinical practice has been facilitated by the development of robust bioinformatics pipelines that can handle the vast volumes of data generated. These pipelines enable the accurate alignment of sequencing reads to reference genomes, variant calling, and annotation of identified mutations. Importantly, the clinical interpretation of these data requires a multidisciplinary approach, involving molecular pathologists, oncologists, and bioinformaticians to ensure that the results are translated into actionable insights for patient care [20, 21, 22]. In summary, NGS has ushered in a new era of precision oncology in NSCLC by providing a comprehensive understanding of the genomic alterations that drive disease progression and resistance. Its ability to detect both primary driver mutations and emerging resistance mechanisms in real time makes it an indispensable tool in the modern management of EGFR-mutant NSCLC. As the technology continues to evolve, with increasing sensitivity, speed, and affordability, its role in guiding personalized therapeutic strategies will only expand, ultimately improving outcomes for patients with this challenging disease [20, 21, 22, 23, 24, 25]. Computer-Aided Drug Design (CADD) encompasses computational methods such as molecular docking, virtual screening, and molecular dynamics simulations to predict the interaction between potential drug candidates and target proteins. In the context of EGFR-mutant NSCLC, CADD enables the identification of compounds that can effectively bind to mutant forms of the EGFR protein, including those with resistance-conferring mutations [20, 21, 22, 23, 24, 25].

MATERIAL AND METHODS

In this study, a comprehensive in silico workflow was employed to investigate the structure, function, and potential interaction partners of target proteins. To visualize and analyze the three-dimensional structures of proteins, we used RasMol and PyMOL. RasMol, a lightweight molecular graphics program, was employed for rapid visualization of macromolecular structures, providing clear images of protein backbones and secondary structural elements. PyMOL, a more advanced molecular visualization tool, was used for detailed examination of protein-ligand interactions and structural modifications, facilitating the generation of publication-quality images [26, 27]. For sequence similarity searches, BLAST (Basic Local Alignment Search Tool) was employed. BLAST enabled the identification of homologous sequences by aligning our query sequences against the NCBI database, helping to infer evolutionary relationships and potential functional similarities [28]. To gain insights into the biological pathways and functional contexts of the proteins, the KEGG (Kyoto Encyclopedia of Genes and Genomes) database was used. KEGG pathways provided valuable information on metabolic and signaling pathways associated with the proteins of interest [29]. Potential protein-protein interactions were explored using the STRING database. This resource integrates known and predicted associations, allowing the construction of interaction networks that could offer clues about functional partners and pathways relevant to the proteins under investigation [30]. To identify conserved domains and functional motifs, we utilized PROSITE and InterProScan. PROSITE was used for the identification of functional sites and signatures within the protein sequences, while InterProScan provided a broader annotation by combining predictive models from multiple databases, thus ensuring comprehensive domain identification [31, 32]. The structural quality of modeled proteins was assessed using Errat, a tool designed to analyze the statistics of non-bonded interactions in protein models. Errat provided a quantitative measure of overall structural reliability, helping to validate the accuracy of our computational models [33]. PDBsum was employed to generate detailed summaries of protein structures. PDBsum provided schematic representations of secondary structure elements and ligand-binding sites, offering a quick yet thorough overview of structural features [34]. For the evaluation of potential binding sites and docking analysis, CB-Dock was utilized. CB-Dock uses cavity detection to predict binding pockets and automates the docking of ligands, offering a streamlined approach for protein-ligand interaction studies. CB-Dock automatically identifies cavities within the protein structures and performs blind docking simulations using AutoDock Vina, providing an efficient means to evaluate protein-ligand interactions [35].

RESULT AND DISCUSSION

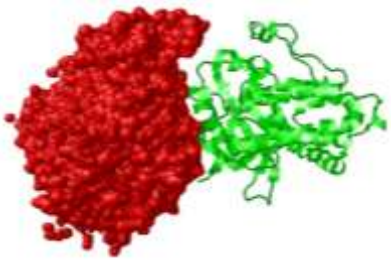

Figure 1: Visualization of chains of protein

Chain A has been shown by red color and chain B has been shown by green color. Utilizing command line in RasMol, the atoms included in chain A and chain B have been calculated. Chain A has 2895 atoms and chain B has 2883 atoms.

Table 1: Analysis of secondary structure through RasMol

Secondary Structure	Atoms included
Helix	2649
Sheet	887
Turns	0
Loops	0

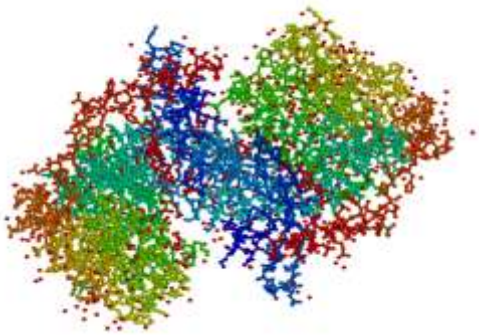

Figure 2: Visualization of hydrophobic regions of protein through Rasmol

The three-dimensional structure of the protein, identified by PDB ID 8D76, was successfully visualized and analyzed using RasMol. Employing the color-by-group scheme allowed for the clear distinction of different structural domains and subunits within the protein complex. Each chain or functional domain is represented by a unique color palette, including variations of red, blue, green, yellow, and orange. This visualization approach provided immediate insights into the spatial arrangement of individual components and their interactions.

The presence of distinct color-coded regions indicates well-defined domains, which is crucial for understanding potential binding interactions and active site accessibility. Notably, the red spheres, representing oxygen atoms in the structure, are scattered throughout the chains, particularly in peripheral regions

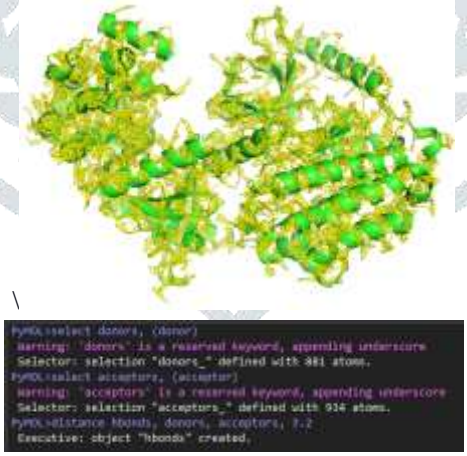

Figure 3: Visualization of hydrogen-bonds through PyMol

The protein's secondary structure, depicted as green helices, is complemented by the yellow dashed lines representing the hydrogen bond interactions. This visualization showcases an intricate hydrogen bond network stabilizing the protein's tertiary structure. By adjusting the visual representations, we have achieved a clearer and more detailed image that enhances the understanding of the protein's conformational stability and intermolecular interactions. The high-quality rendering provides a clear perspective of how hydrogen bonds contribute to the protein's structural integrity.

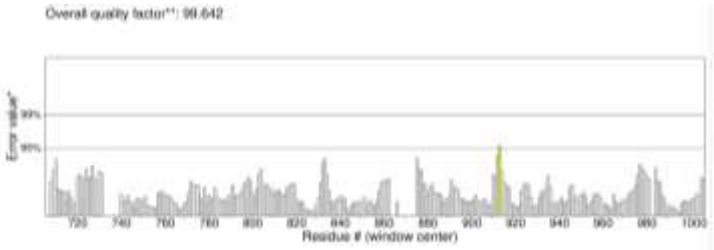

Figure 4: Structure Validation Analysis

The ERRAT2 plot generated for the protein structure (PDB ID: 8d76, Chain A) reveals an overall quality factor of 99.642%, which is indicative of a well-refined model.

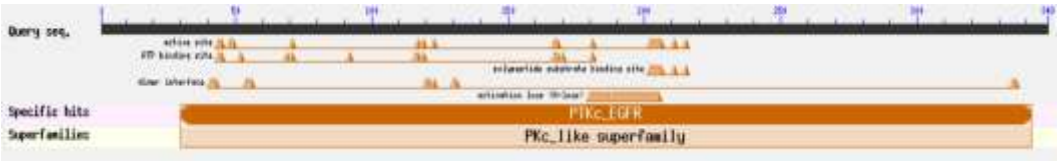

Figure 5: BLAST Result

In Alignment search tools output reveals detailed domain architecture for the protein sequence analyzed. The sequence features several critical functional elements, including active sites, ATP-binding motifs, and polypeptide substrate binding regions, primarily concentrated within the first 250 residues. The activation loop (A-loop) is also highlighted, spanning a central region of the protein. Domain annotations identify the sequence as part of the PTKc_EGFR family, nested within the PKc_like superfamily. This classification underscores the presence of conserved kinase domains essential for catalytic activity. The presence of these domains and motifs strongly supports the protein's functional role as a kinase, likely contributing to signal transduction pathways or other phosphorylation-mediated processes. This analysis confirms the structural and functional homology of the sequence with other kinases in the PTKc_EGFR family, reinforcing its relevance in related biological contexts.

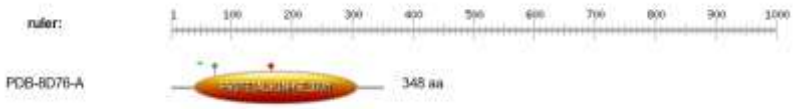

Figure 6: Prosite Result

The Prosite analysis identified a significant domain match within the protein sequence PDB-8D76-A, which is composed of 348 amino acids. The identified profile hit corresponds to the PROTEIN_KINASE_DOM domain, suggesting that this protein sequence exhibits a conserved protein kinase domain architecture. The highlighted sequence region containing the conserved kinase domain includes characteristic motifs such as the ATP-binding site and the activation loop, which are typical features of protein kinases. This conserved domain architecture reinforces the functional identity of this protein as a kinase, potentially involved in signal transduction pathways through phosphorylation events.

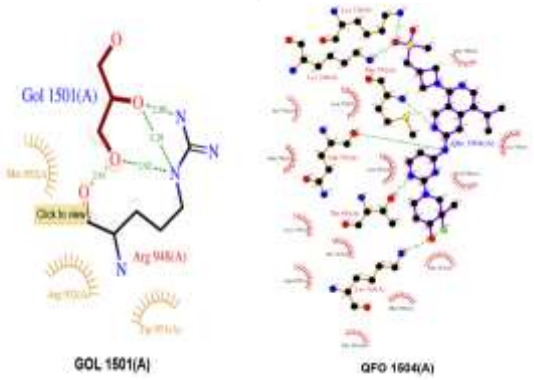

Figure 7: Ligand Analysis through PDBSum

Table 2: Analysis of ligand QFO involved in hydrogen bonding through PDBSum

No.	Atom 1 (no., name, res., res. no., chain)	Atom 2 (no., name, res., res. no., chain)	Distance (Å)
1	128, NZ, LYS, 716, A	5090, O1, QFO, 1504, A	3.07
2	212, NZ, LYS, 728, A	5090, O1, QFO, 1504, A	2.83
3	344, NZ, LYS, 745, A	5109, O3, QFO, 1504, A	2.81
4	692, O, GLN, 791, A	5098, N4, QFO, 1504, A	3.06
5	706, N, MET, 793, A	5096, N3, QFO, 1504, A	2.97
6	1218, OG1, THR, 854, A	5102, N6, QFO, 1504, A	2.81

A total of six hydrogen bonds were identified between the protein chain A residues (LYS, GLN, MET, and THR) and the ligand QFO (residue 1504, chain A). The distances for these interactions ranged from 2.81 Å to 3.07 Å, which indicates strong and potentially significant interactions contributing to the stability of the protein-ligand complex.

Table 3: Analysis of ligand GOL involved in hydrogen bonding through PDBSum

No.	Atom 1 (no., name, res., res. no., chain)	Atom 2 (no., name, res., res. no., chain)	Distance (Å)
1	1965, O, ARG, 948, A	5065, O3, GOL, 1501, A	2.61
2	1969, NE, ARG, 948, A	5063, O2, GOL, 1501, A	3.29
3	1969, NE, ARG, 948, A	5065, O3, GOL, 1501, A	2.92
4	1972, NH2, ARG, 948, A	5063, O2, GOL, 1501, A	2.89

This analysis reveals four hydrogen bonds formed between ARG (residue 948, chain A) and the GOL ligand (residue 1501, chain A). These interactions are characterized by distances ranging from 2.61 Å to 3.29 Å, suggesting robust interactions that could play a role in the stability and functionality of the protein-ligand complex.

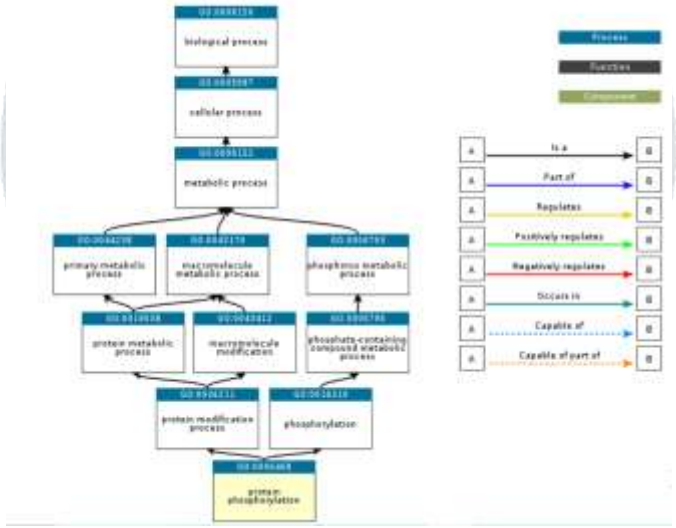

Figure 8: Biological Ontology obtained through InterProScan

Functional annotation conducted using InterProScan revealed the involvement of the identified proteins in a hierarchy of biological processes, as defined by the Gene Ontology (GO) consortium. The primary annotation mapped to the broad category of biological process (GO:0008150), which encompasses all biological objectives to which the gene product contributes. From this overarching classification, the annotation followed a hierarchical path through cellular process (GO:0009987) and further refined to metabolic process (GO:0008152), suggesting active participation in cellular metabolic pathways. This metabolic function diverged into three key branches: primary metabolic process (GO:0044238), macromolecule metabolic process (GO:0043170), and phosphorus metabolic process (GO:0006793). Subsequent specialization highlighted involvement in the protein metabolic process (GO:0019538) and macromolecule modification (GO:0034412), with convergence on phosphate-containing compound metabolic process (GO:0006796) an essential route in post-translational regulation.

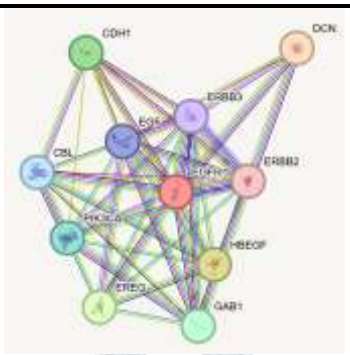

Figure 9: STRING OUTPUT

To elucidate the functional landscape and interconnectivity among key signaling molecules, a protein-protein interaction (PPI) network analysis. The resulting interaction map reveals a dense and highly interconnected network centering around EGFR (Epidermal Growth Factor Receptor), a pivotal node in receptor tyrosine kinase signaling pathways. EGFR exhibits strong direct interactions with multiple ligands and adaptor proteins, including EGF, HBEGF, EREG, and GAB1, suggesting its central role in ligand-mediated signal transduction. Additionally, EGFR demonstrates significant associations with co-receptors such as ERBB2 and ERBB3, indicating potential heterodimerization events that modulate downstream signaling diversity and intensity. The network further highlights interactions with intracellular effectors and signaling molecules, including PIK3CA and CBL, implicating involvement in PI3K/AKT signaling and ubiquitination-dependent regulatory mechanisms. The presence of CDH1 and DCN suggests possible links to cell adhesion and extracellular matrix modulation, reflecting the broader biological implications of EGFR signaling in processes such as migration, proliferation, and tissue remodeling.

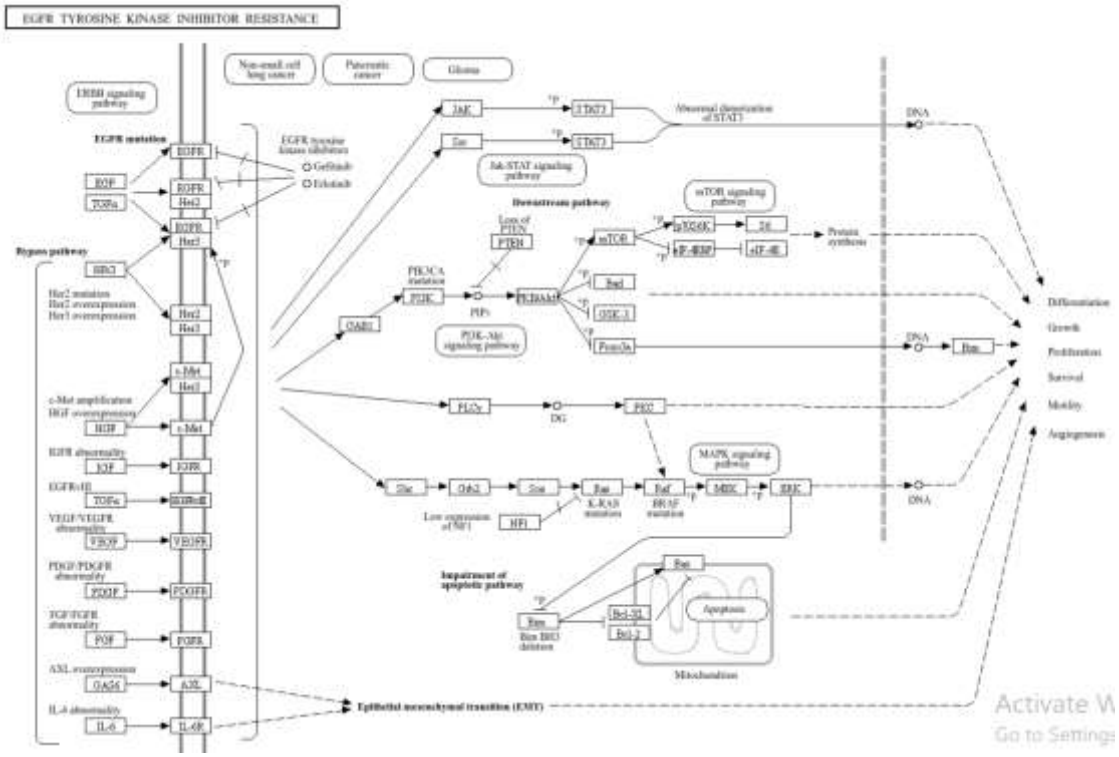

Figure 10: KEGG Pathway

The KEGG pathway analysis of EGFR tyrosine kinase inhibitor resistance reveals a complex signaling network involving multiple molecular alterations that contribute to therapeutic evasion, particularly in cancers such as non-small cell lung cancer (NSCLC), pancreatic cancer, and glioma. At the core of this resistance mechanism lies aberrant EGFR signaling, which can arise from mutations in the EGFR gene itself or through compensatory activation of bypass signaling pathways. Ligands such as EGF and TGF-α activate EGFR and its dimerization partners (ERBB2, ERBB3), which initiate downstream signaling cascades.

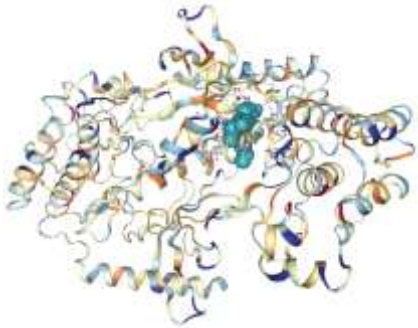

Figure 11: Docked image of 8D76 with Curcumin (PubChem ID - 969516)

Table 4: Docking Score of 8D76 with Curcumin

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center Coordinates (x, y, z)	Docking Box Size (x, y, z)
C3	-8.7	1186	-12, 13, 0	26, 26, 26
C1	-8.6	6695	-14, 27, -11	26, 35, 35
C2	-8.6	1197	-33, 45, -10	26, 26, 26
C4	-6.7	535	2, 32, 9	26, 26, 26
C5	-6.5	501	-13, 34, -30	26, 26, 26

Among all analyzed pockets, C3 exhibited the most favorable binding affinity with a Vina score of -8.7 kcal/mol, suggesting it as the most likely site for stable ligand interaction. Although C1 had a slightly lower binding energy (-8.6 kcal/mol), it possessed the largest cavity volume (6695 Å³), indicating a potentially flexible or accommodating binding region.

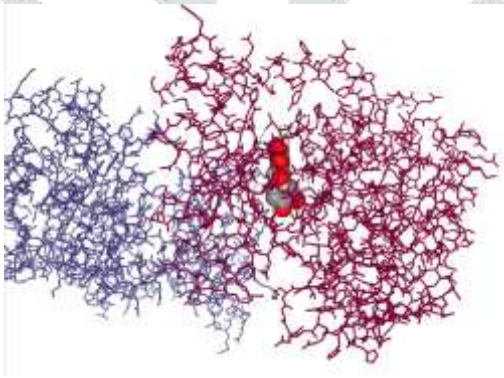

Figure 12: Docked image of 8D76 with Quercetin (PubChem ID - 5280343)

Table 5: Docking Score of 8D76 with Quercetin

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center Coordinates (x, y, z)	Docking Box Size (x, y, z)
C1	-8.8	6695	-14, 27, -11	27, 35, 35
C2	-8.4	1197	-33, 45, -10	21, 28, 21
C3	-8.3	1186	-12, 13, 0	21, 28, 21
C4	-6.6	535	2, 32, 9	21, 21, 21
C5	-6.5	501	-13, 34, -30	21, 27, 21

Among the analyzed binding sites, C1 exhibited the strongest binding affinity with a Vina score of -8.8 kcal/mol and the largest cavity volume (6695 Å³), suggesting this site as the most favorable for ligand interaction. Pockets C2 and C3 also demonstrated appreciable binding energies of -8.4 and -8.3 kcal/mol, respectively, accompanied by moderate cavity sizes, making them viable alternative sites. In contrast, C4 and C5 displayed significantly lower binding affinities and smaller cavity volumes, indicating reduced binding potential. Taken together, pocket C1 stands out as the most promising candidate for subsequent structural and functional analyses due

to its superior docking score and ample binding space, which may accommodate flexible ligand binding and potentially enhance interaction stability.

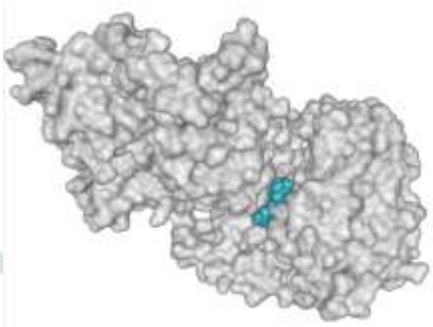

Figure 13: Docked image of 8D76 with Epigallocatechin gallate (PubChem ID - 65064)

Table 6: Docking Score of 8D76 with Epigallocatechin gallate

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center Coordinates (x, y, z)	Docking Box Size (x, y, z)
C3	-9.4	1186	-12, 13, 0	22, 28, 22
C2	-9.3	1197	-33, 45, -10	22, 28, 22
C1	-9.2	6695	-14, 27, -11	22, 35, 35
C4	-7.6	535	2, 32, 9	22, 22, 22
C5	-7.0	501	-13, 34, -30	22, 22, 22

Among the analyzed binding pockets, C3 exhibited the strongest binding affinity, with a Vina score of -9.4 kcal/mol, closely followed by C2 and C1, with scores of -9.3 and -9.2 kcal/mol, respectively. The results highlight C3 as the optimal pocket for docking, offering both strong binding affinity and suitable pocket geometry. These findings support further investigation into C3 as the preferred site for molecular interaction and structure-based drug design.

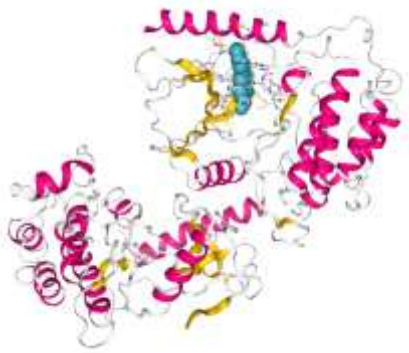

Figure 14: Docked image of 8D76 with Berberine (PubChem ID - 2353)

Table 7: Docking Score of 8D76 with Berberine

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center (x, y, z)	Docking Size (x, y, z)
C2	-9.0	1197	-33, 45, -10	22, 28, 22
C3	-9.0	1186	-12, 13, 0	22, 28, 22
C1	-8.9	6695	-14, 27, -11	22, 35, 35
C4	-7.1	535	2, 32, 9	22, 22, 22
C5	-6.6	501	-13, 34, -30	22, 22, 22

In this round of molecular docking analysis, five potential binding pockets (C1–C5) were evaluated based on their binding affinity, pocket geometry, and docking space parameters. Pockets C2 and C3 showed the most favorable binding affinity, each achieving a Vina score of -9.0 kcal/mol, indicating strong potential for ligand binding. These pockets also featured relatively compact cavity volumes of 1197 Å³ and 1186 Å³, respectively, which may contribute to tighter ligand interactions. C1 followed closely with a Vina score of -8.9 kcal/mol, although it had a significantly larger cavity volume **(**6695 Å³**)** and a larger docking box dimension, suggesting it might accommodate bulkier ligands or provide multiple binding orientations. In contrast, C4 and C5 exhibited weaker binding affinities, with Vina scores of -7.1 and -6.6 kcal/mol, and relatively small cavity volumes, suggesting these sites are less favorable for effective molecular interaction. Overall, the findings identify C2 and C3 as the most promising candidates for further ligand optimization and structural refinement studies, based on their strong affinity and favorable pocket characteristics.

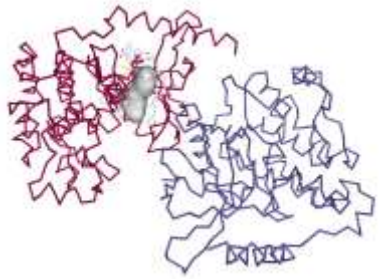

Figure 15: Docked image of 8D76 with Resveratrol (PubChem ID - 445154)

Table 8: Docking Score of 8D76 with Resveratrol

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center Coordinates (x, y, z)	Docking Box Size (x, y, z)
C1	-7.5	6695	-14, 27, -11	27, 35, 35
C2	-7.5	1197	-33, 45, -10	21, 28, 21
C3	-7.5	1186	-12, 13, 0	21, 28, 21
C4	-6.1	535	2, 32, 9	21, 21, 21
C5	-5.7	501	-13, 34, -30	21, 27, 21

The docking analysis yielded five putative binding pockets, with C1, C2, and C3 showing the highest binding affinity, all scoring -7.5 kcal/mol in AutoDock Vina evaluations. Although these pockets share equal binding scores, they differ significantly in cavity dimensions and location. C1 presents the largest cavity volume at 6695 Å³, suggesting a more spacious binding environment possibly accommodating larger or flexible ligands. The broader docking box dimensions (27 × 35 × 35) further support this, making it suitable for extended ligand conformations.

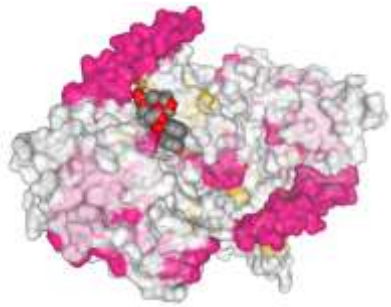

Figure 16: Docked image of 8D76 with Taxol (Paclitaxel) (PubChem ID - 36314)

Table 9: Docking Score of 8D76 with Taxol (Paclitaxel)

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center Coordinates (x, y, z)	Docking Box Size (x, y, z)
C2	-9.2	1197	-33, 45, -10	27, 27, 27
C1	-8.3	6695	-14, 27, -11	27, 35, 35
C3	-8.3	1186	-12, 13, 0	27, 27, 27

C4	-7.2	535	2, 32, 9	27, 27, 27
C5	-6.5	501	-13, 34, -30	27, 27, 27

Among the evaluated binding pockets, C2 exhibited the strongest ligand-receptor interaction with a Vina score of -9.2 kcal/mol, indicating the most favorable binding affinity. Despite its relatively modest cavity volume of 1197 Å³, its symmetrical docking dimensions (27 × 27 × 27) suggest a balanced and compact binding space, potentially optimizing ligand fit and orientation. C1 and C3 followed with identical binding energies of -8.3 kcal/mol. However, C1 presented a significantly larger cavity (6695 Å³), implying more spatial allowance for ligands with extended conformations or bulkier side chains. In contrast, C3, with a smaller pocket volume, may cater to more rigid or small molecule ligands. The remaining sites, C4 and C5, demonstrated lower binding affinities, registering Vina scores of -7.2 and -6.5 kcal/mol, respectively. These cavities were also the smallest in volume, with both under 550 Å³, potentially limiting their druggability for larger compounds.

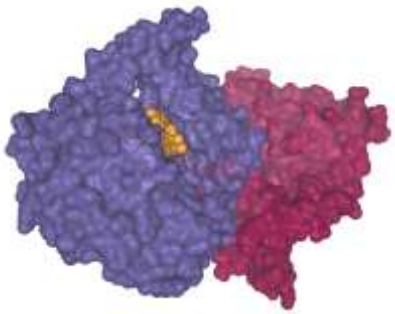

Figure 17: Docked image of 8D76 with Kaempferol (PubChem ID - 5280863)

Table 10: Docking Score of 8D76 with Kaempferol

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center (x, y, z)	Docking Box Size (x, y, z)
C2	-8.8	1197	-33, 45, -10	21, 28, 21
C1	-8.2	6695	-14, 27, -11	27, 35, 35
C3	-8.2	1186	-12, 13, 0	21, 28, 21
C4	-6.4	535	2, 32, 9	21, 21, 21
C5	-6.2	501	-13, 34, -30	21, 27, 21

The binding pocket C2 yielded the most favorable binding affinity with a Vina score of -8.8 kcal/mol**,** making it the top candidate for potential ligand interaction. Although the cavity volume of 1197 Å³ is relatively modest, its compact docking dimensions (21 × 28 × 21) may provide an optimized environment for stable ligand accommodation.

CONCLUSION

This study has provided a comprehensive analysis of the protein sequence PDB-8D76-A, affirming its identity as a member of the PTKc_EGFR family and highlighting its potential role in phosphorylation-driven signal transduction pathways. Structural and domain-level analyses confirmed the presence of conserved kinase motifs, including ATP-binding sites and the activation loop, reinforcing its functional characterization as a kinase. Ligand-binding studies further revealed robust hydrogen-bond interactions with key amino acid residues, suggesting a potential influence on structural stability and functional activity. Functional annotation using InterProScan and STRING PPI networks mapped the protein's involvement in phosphorylation-related signaling and complex cellular regulatory processes, underscoring its biological significance. Docking analyses with bioactive compounds such as curcumin, quercetin, and epigallocatechin gallate identified promising binding pockets, with C3 and C1 demonstrating the most favorable interactions. These findings lay the groundwork for future explorations of the protein's therapeutic relevance, particularly in overcoming EGFR tyrosine kinase inhibitor resistance, and offer insights that could guide the development of next-generation inhibitors or synergistic combination therapies.

REFERENCES

1. Pennell NA, Arcila ME, Gandara DR, West H. Biomarker Testing for Patients With Advanced Non-Small Cell Lung Cancer: Real-World Issues and Tough Choices. *Am Soc Clin Oncol Educ Book*. 2019;39:531-542. doi:10.1200/EDBK_237863

2. Baran K, Brzeziańska-Lasota E. Proteomic biomarkers of non-small cell lung cancer patients. *Adv Respir Med*. 2021;89(4):419-426. doi:10.5603/ARM.a2021.0089

3. Neumann JM, Freitag H, Hartmann JS, et al. Subtyping non-small cell lung cancer by histology-guided spatial metabolomics. *J Cancer Res Clin Oncol*. 2022;148(2):351-360. doi:10.1007/s00432-021-03834-w

4. Zang X, Zhang J, Jiao P, Xue X, Lv Z. Non-Small Cell Lung Cancer Detection and Subtyping by UPLC-HRMS-Based Tissue Metabolomics. *J Proteome Res*. 2022;21(8):2011-2022. doi:10.1021/acs.jproteome.2c00316

5. Song X, Xiong A, Wu F, et al. Spatial multi-omics revealed the impact of tumor ecosystem heterogeneity on immunotherapy efficacy in patients with advanced non-small cell lung cancer treated with bispecific antibody [published correction appears in J Immunother Cancer. 2023 Jul;11(7):e006234corr1. doi: 10.1136/jitc-2022-006234corr1.]. *J Immunother Cancer*. 2023;11(2):e006234. doi:10.1136/jitc-2022-006234

6. Viñal D, Martínez D, Higuera O, de Castro J. Genomic profiling in non-small-cell lung cancer in young patients. A systematic review. *ESMO Open*. 2021;6(1):100045. doi:10.1016/j.esmoop.2020.100045

7. Shi W, Cheng Y, Zhu H, Zhao L. Metabolomics and lipidomics in non-small cell lung cancer. *Clin Chim Acta*. 2024;555:117823. doi:10.1016/j.cca.2024.117823

8. Chen PH, Cai L, Huffman K, et al. Metabolic Diversity in Human Non-Small Cell Lung Cancer Cells. *Mol Cell*. 2019;76(5):838-851.e5. doi:10.1016/j.molcel.2019.08.028

9. Bhattarai A, Shah S, Abu Serhan H, Sah R, Sah S. Genomic profiling for non-small cell lung cancer: Clinical relevance in staging and prognosis. *Medicine (Baltimore)*. 2023;102(47):e36003. doi:10.1097/MD.0000000000036003

10. Yu H, Pang Z, Li G, Gu T. Bioinformatics analysis of differentially expressed miRNAs in non-small cell lung cancer. *J Clin Lab Anal*. 2021;35(2):e23588. doi:10.1002/jcla.23588

11. Hamouz M, Hammouz RY, Bajwa MA, Alsayed AW, Orzechowska M, Bednarek AK. A Functional Genomics Review of Non-Small-Cell Lung Cancer in Never Smokers. *Int J Mol Sci*. 2023;24(17):13314. Published 2023 Aug 28. doi:10.3390/ijms241713314

12. Shi Y, Zhu S, Yang J, et al. Investigation of Potential Mechanisms Associated with Non-small Cell Lung Cancer. *J Comput Biol*. 2020;27(9):1433-1442. doi:10.1089/cmb.2019.0081

13. Yang S, Huang Y, Zhao Q. Epigenetic Alterations and Inflammation as Emerging Use for the Advancement of Treatment in Non-Small Cell Lung Cancer. *Front Immunol*. 2022;13:878740. Published 2022 Apr 20. doi:10.3389/fimmu.2022.878740

14. Yu H, Li SB. Role of LINC00152 in non-small cell lung cancer. *J Zhejiang Univ Sci B*. 2020;21(3):179-191. doi:10.1631/jzus.B1900312

15. Levy B, Hu ZI, Cordova KN, Close S, Lee K, Becker D. Clinical Utility of Liquid Diagnostic Platforms in Non-Small Cell Lung Cancer. *Oncologist*. 2016;21(9):1121-1130. doi:10.1634/theoncologist.2016-0082

16. Cai X, Lin L, Zhang Q, Wu W, Su A. Bioinformatics analysis of the circRNA-miRNA-mRNA network for non-small cell lung cancer. *J Int Med Res*. 2020;48(6):300060520929167. doi:10.1177/0300060520929167

17. Lee JW, Lee H, Chun YS, et al. Characterization of chemoresistant human non-small cell lung cancer cells by metabolic and lipidomic profiling. *Metabolomics*. 2023;19(9):80. Published 2023 Sep 10. doi:10.1007/s11306-023-02045-3

18. Bar J, Leibowitz R, Reinmuth N, et al. Biological insights from plasma proteomics of non-small cell lung cancer patients treated with immunotherapy. *Front Immunol*. 2024;15:1364473. Published 2024 Feb 29. doi:10.3389/fimmu.2024.1364473

19. Feng L, Cai W, Jin S, et al. Integrated bioinformatics analysis of microarray data from non-small cell lung cancer. *Cell Mol Biol (Noisy-le-grand)*. 2023;69(7):218-224. Published 2023 Jul 31. doi:10.14715/cmb/2023.69.7.35

20. Ding C, Huang H, Wu D, et al. Pan-cancer analysis predict that FAT1 is a therapeutic target and immunotherapy biomarker for multiple cancer types including non-small cell lung cancer. *Front Immunol*. 2024;15:1369073. Published 2024 May 24. doi:10.3389/fimmu.2024.1369073

21. Yu L, Liang X, Wang J, et al. Identification of Key Biomarkers and Candidate Molecules in Non-Small-Cell Lung Cancer by Integrated Bioinformatics Analysis. *Genet Res (Camb)*. 2023;2023:6782732. Published 2023 Jan 3. doi:10.1155/2023/6782732

22. Ouyang C, Yu X, Wang H, Zeng P. Multidimensional bioinformatics perspective on smoking-linked driver genes and immune regulatory mechanisms in non-small cell lung cancer. *J Transl Med*. 2025;23(1):330. Published 2025 Mar 14. doi:10.1186/s12967-025-06301-z

23. Planchard D, Remon J, Nowak F, Soria JC. Future Genetic/Genomic Biomarker Testing in Non-Small Cell Lung Cancer. *Am Soc Clin Oncol Educ Book*. 2017;37:12-17. doi:10.1200/EDBK_100007

24. Mei T, Wang T, Zhou Q. Multi-omics and artificial intelligence predict clinical outcomes of immunotherapy in non-small cell lung cancer patients. *Clin Exp Med*. 2024;24(1):60. Published 2024 Mar 30. doi:10.1007/s10238-024-01324-0

25. Soo RA, Adjei AA. Predicting Clinical Outcomes Using Proteomics in Non-Small Cell Lung Cancer-The Past, Present, and Future. *J Thorac Oncol*. 2017;12(4):602-606. doi:10.1016/j.jtho.2017.01.025

26. Biswas S, Manekar S, Bakshi SR. A Case Study on PPM1D and 9 Other Shared Germline Alterations in a Family. *Asian Pac J Cancer Prev*. 2023;24(6):2129-2134. Published 2023 Jun 1. doi:10.31557/APJCP.2023.24.6.2129

27. Xu L. Identification of Autophagy-Related Targets of Berberine against Non-Small Cell Lung Cancer and Their Correlation with Immune Cell Infiltration By Combining Network Pharmacology, Molecular Docking, and Experimental Verification. *Crit Rev Immunol*. 2023;43(2):27-47. doi:10.1615/CritRevImmunol.2023049923

28. Ogata H, Okamoto I, Yoshimoto G, et al. Chronic myelomonocytic leukemia blast crisis in a patient with advanced non-small cell lung cancer treated with EGFR tyrosine kinase inhibitors. *Respir Investig*. 2017;55(2):181-183. doi:10.1016/j.resinv.2016.12.002

29. Li Z, Zhou B, Zhu X, et al. Differentiation-related genes in tumor-associated macrophages as potential prognostic biomarkers in non-small cell lung cancer. *Front Immunol*. 2023;14:1123840. Published 2023 Mar 9. doi:10.3389/fimmu.2023.1123840

30. Sultana A, Alam MS, Liu X, et al. Single-cell RNA-seq analysis to identify potential biomarkers for diagnosis, and prognosis of non-small cell lung cancer by using comprehensive bioinformatics approaches. *Transl Oncol*. 2023;27:101571. doi:10.1016/j.tranon.2022.101571

31. Protopopov A, Kashuba V, Zabarovska VI, et al. An integrated physical and gene map of the 3.5-Mb chromosome 3p21.3 (AP20) region implicated in major human epithelial malignancies. *Cancer Res*. 2003;63(2):404-412.

32. Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: protein domains identifier. *Nucleic Acids Res*. 2005;33(Web Server issue):W116-W120. doi:10.1093/nar/gki442

33. Elrashedy A, Nayel M, Salama A, Salama MM, Hasan ME. Bioinformatics approach for structure modeling, vaccine design, and molecular docking of Brucella candidate proteins BvrR, OMP25, and OMP31. *Sci Rep*. 2024;14(1):11951. Published 2024 May 25. doi:10.1038/s41598-024-61991-7

34. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. *Protein Sci*. 2018;27(1):129-134. doi:10.1002/pro.3289

35. Liu Y, Yang X, Gan J, Chen S, Xiao ZX, Cao Y. CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res*. 2022;50(W1):W159-W164. doi:10.1093/nar/gkac394

36. Altschul, S. F. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410.

37. Balourdas, D. I. (2024). Structural basis of p53 inactivation by cavity-creating mutations. . Cell Death & Disease, 15(1), 18–29., 9.

38. Berman, H. M. (2000). Nucleic Acids Research, 28(1),235-2342. The Protein Data Bank.

39. Cock, P. J. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 1422--1423.

40. Colovos, C. &. (1193). Verification of protein structures: Patterns of nonbonded atomic interactions. Wiley.

41. DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.

42. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science \& Engineering, 3.

43. Joerger, A. C. (2008). Structural biology of the tumor suppressor p53 and cancer-associated mutants. Annual Review of Biochemistry, 25.

44. Kufareva, I. &. (2012). Methods of protein structure comparison. In Protein Structure Prediction. Humana Press, (pp. 231–257).

45. Laskowski, R. A. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. International Union of Crystallography.

46. Muller, P. A. (2014). Mutant p53 in cancer: New functions and therapeutic opportunities. Cancer Cell, 25(3), 304-317.

47. Pal, A. B. (2023). Multidimensional quantitative phenotypic and molecular profiling of TP53 mutant cancers. npj Breast Cancer, 9, Article 33.

48. Sayle, R. &.-W. (1995). RasMol: Biomolecular graphics for all. Trends in Biochemical Sciences, 20(9), 374.

49. Sievers, F. a. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, 1189--1191.

50. Sievers, F. a. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Sievers, Fabian and Wilm, Andreas and Dineen, David and Gibson, Toby J and Karplus, Kevin and Li, Weizhong and Lopez, Rodrigo and McWilliam, Hamish and Remmert, Michael and Söding, Johannes and Thompson, Julie D and Higgins, Desmond G, 539.

51. Soussi, T. &. (2015). TP53: An oncogene in disguise. Cell Death and Differentiation, 22(8), 1239–1249.

52. Tornesello, M. L. (2024). TP53 mutations in cancer: Molecular features and therapeutic strategies. . Frontiers in Oncology,14, Article 1198765.

53. Vousden, K. H. (2009). Blinded by the light: The growing complexity of p53. Cell, 137(3), 413-431.

54. Waterhouse, A. M. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, 1189--1191.

55. Uma kumari, Devanshi Gupta, Computational analysis of Glioma Transcriptome and proteome with Bio python ;2023Volume-13, Issue-1, Pp 1-14,IJBTR

56. Kumari Uma & Tripathi, Kartik. (2023). Computational Analysis and Molecular Docking Approach for Liver Cancer. International Journal for Research in Applied Science and Engineering Technology. 11. 1828 to 1633. 10.22214/ijraset.2023.54927

57. Uma Kumari,Gurpreet Kaur *et al*,"Biopython/Network Of Protein Identification And NGS Analysis Of Glioma Cancer ATP Competitive Type III C-MET Inhibitor : Volume 11, Issue 2 : 27-Jun-2024 :pp 41-51 : http://doi.one/10.1729/Journal.40229

58. 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.

59. 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.

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

61. 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

62. 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

63. 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

64. 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