IOSR Journal of Biotechnology and Biochemistry (IOSR-JBB)
ISSN: 2455-264X, Volume 12, Issue 3 Series 1 (May. – June. 2026), PP 25-34
www.iosrjournals.org

NGS And Structure Based Drug Designing Of Natural Compound Against Drug Resistant EGFR In Non-Small Cell Lung Cancer

Riyaz Gurung[1], Uma Kumari[*1]
[1]*Project Trainee At Bioinformatics Project And Research Institute, Noida - 201301, India*
[*1]*Senior Bioinformatics Scientist, Bioinformatics Project And Research Institute, Noida - 201301, India*

Abstract
Mutations in Epidermal Growth Factor Receptor drives disease progression and treatment resistance in Non-small Cell Lung Cancer. EGFR is a well-documented therapeutic target for management of NSCLC. Our research combines Next Generation Sequencing (NGS) and Structure Based Drug Design (SBDD) approaches to identify natural compounds as potential therapeutical. In our study, we targeted the triple mutated EGFR kinase domain (L858R T790M/C797S) using its crystal structure (PDB ID: 9D3W) to design EGFR tyrosine kinase inhibitors. NGS examined mutations in tumour samples from patients, thereby identifying mutation associated biomarkers and new drug targets. Building on this genomic evidence, we applied SBDD methods including molecular docking and ADMET prediction to identify natural compounds with strong binding affinity and superior safety profiles. Astragaloside IV, kaempferol, furostanol and luteolin emerged as a promising target. ADMET prediction classified astragaloside IV as class 6 (non-toxic) and kaempferol, furostanol and luteolin as class 5 (practically non-toxic). Toxicity predictions confirmed that all compounds demonstrated no hepatotoxicity and cytotoxicity where as astragaloside IV showed mild cardiotoxicity. Our research underscores the therapeutic capacity of astragaloside IV, kaempferol, furostanol and luteolin as novel EGFR inhibitors for NSCLC treatment. This research highlights modern holistic approach to drug discovery. Additionally, in vitro and in vivo studies are essential to validate their efficacy against EGFR mutated NSCLC.

Keywords *– Next-Generation Sequencing (NGS), Structure Based Drug Design (SBDD), Natural products, EGFR Mutants, Non-small Cell Lung Cancer (NSCLC), Molecular Docking, Targeted Therapy, Precision Medicine*

Date of Submission: 26-04-2026 Date of Acceptance: 06-05-2026

I. Introduction

Cancer is a disease that arises when biological alterations lead to unchecked cell growth and division. Any body tissue can acquire cancer, and each variety of cancer has its own distinct characteristics. Cancer starts when a cell breaks free from the usual constraints on cell division and starts to proliferate on its own. (Rajput, 2023)

The first hints of cancer involve genetic and epigenetic changes in specific cells, some of which can spread and metastasize throughout the body. (Piña-Sánchez et al., 2021) The activation of oncogenes occurs in two ways: either by tumour virus infection of cells or alteration of cellular proto-oncogenes. (Bindod et al., 2025) A fundamental aspect of cancer pathophysiology is inflammation. Cancer causes an inflammatory process, and the inflammation can promote cancer growth. Innate immune cells, in fact, are often present in tumours that are immunologically cold — where adaptive immunity is not yet apparent. (Laura Maiorino, n.d.) Cell adhesion and metastasis are one of the key participants in basic processes regulating cancer formation. These adhesion molecules may allow cancer cells to bind more effectively with the ECM. The reciprocal interaction between extracellular matrix and cancer cells, not only fosters local invasion or tumour progression. But it also prepares cancer cells for the subsequent dissemination of the disease. Metastasis occurs when tumour cells travel from the location of the tumour to other organs. (Yayan et al., 2024)

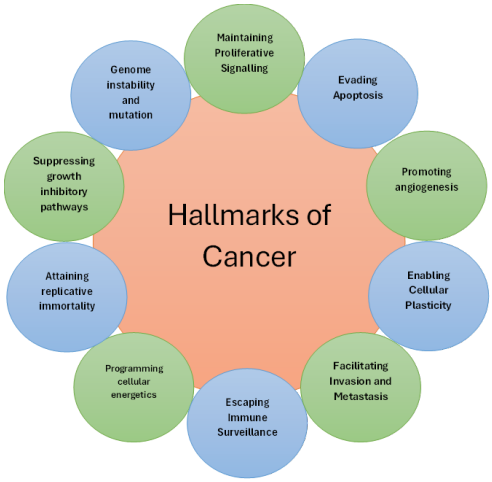


Figure 1 – Hallmarks of Cancer (Hanahan, 2026)

Lung cancer has continued to be the major cause of cancer deaths worldwide and is considered a major public health concern globally. (Jeon et al., 2025) Lung cancer is classified into two major histological subtypes: small cell lung cancer and non-small cell lung cancer (NSCLC). NSCLC accounts for the majority of lung cancer cases, ~85%. (Hendriks et al., 2024) Molecular profiling has become a cornerstone in the management of NSCLC and has revolutionized the way we diagnose, prognose, and treat this disease. Next-generation sequencing-based molecular tests have helped us identify critical driver mutations and genetic changes, including EGFR, ALK, ROS1, BRAF, MET, RET, NTRK, NRG, KRAS, and ERBB2, which are essential in selecting target therapy for our patients. (Jeon et al., 2025) Furthermore, the discovery of PD-L1 expression and tumour mutational burden as biomarkers for NSCLC has also expanded the scope for using immunotherapy in this disease, although this area also warrants further research. (Jeon et al., 2025) Nevertheless, this is not only used for therapeutic decision-making but also helps in understanding drug resistance, thus paving the way for new therapeutic options. (Ferro et al., 2024) As technology in molecular diagnostics advances, its incorporation into practice is necessary for better prognostic tests, therapeutic outcomes, and progress in precision medicine for NSCLC. Furthermore, new evidence points to the detection of minimal residual disease using ctDNA, which needs further investigation. (Jeon et al., 2025)

EGFR is part of the erbB family of receptor tyrosine kinases. Upon binding to a particular ligand, such as epidermal growth factor, the normal EGFR receptor functions by changing conformation and resulting in the phosphorylation of the intracellular domain, which activates signal transduction pathways such as P13K/Akt, Raf1-extracellular signal-regulated kinase and transcription activation pathways.(Inamura et al., 2010) The DNA mutations of EGFR, as identified through polymerase chain reaction, may be located at regions of the EGFR protein which correspond to the extracellular or intracellular domains. In NSCLC, the overexpression of EGFR or the presence of mutations of the intracellular domain of EGFR has been noted in 43-89% of cases represents EGFR overexpression and its intracellular domain mutations. (Gupta et al., 2009) Others found that 75% represents mutations in tyrosine kinase domain of EGFR. (Gillian Bethune, 2010) These mutations can activate signal transduction pathways in a constant manner and promotes cell proliferation. There are two other mutations in exons 18 and 21. It is noteworthy that EGFR and KRAS mutations are mutually exclusive. (Massarelli et al., 2007)

Gefitinib and Erlotinib are oral EGFR TKIs, and they have been tested in a phase III trial involving ~1700 patients with advanced lung cancer and have failed to show a survival advantage over the placebo but have shown a significant survival advantage in non-smoking individuals with bronchioloalveolar cell carcinoma. The 2004 landmark trials showed a 65-90% response rate in EGFR mutation-positive lung adenocarcinomas. The phase III trials have confirmed the survival advantage of TKIs in combination with chemotherapy in EGFR mutation-positive NSCLC cases. (Gillian Bethune, 2010) The EGFR mutation may not always be associated with a survival advantage, as a study showed no association and a poor survival in exon 19 deletions, although the responses are strong, the study showed only 10% radiologic responses in NSCLC patients using gefitinib and suggested that EGFR activation may play a minor role in tumorigenesis, and the mutation in exons 19 and 21 increases the affinity of the drug, but acquired resistance occurs, and T790M (threonine-to-methionine substitution at codon 790) occurs in 50% of the resistant cases. Gefitinib and Erlotinib are used for second- or third-line advanced lung cancer treatment. (Gillian Bethune, 2010)

Patients respond to these therapies but develop resistance within a year of treatment. The development of resistance mutations, for example, T790M, C797S, and L858R, has been cited as a major cause of acquired resistance. These mutations either alter the drug-binding site or create alternative pathways for cell signalling,

making the treatment ineffective. In this regard, there is a constant need for the development of new and more potent therapies. A special feature of this methodology is the source of the potential drug compounds. Compared to the synthetic drug compounds, natural compounds have consistently shown impressive diversity in its structure and bioactivity. Nature as a constant source of medicine provides compounds which has been recorded to form the basis of many current cancer therapies. These compounds have complex three-dimensional structures and unique functional groups. These structures have the ability to interact with proteins in a manner that is difficult for synthetic compounds to achieve. Our research aims to highlight the capacity of phytochemicals to develop a novel compound against the mutant EGFR and a growing need for drug discovery in environmentally conscious and sustainable manner.

However, this potential also has to be balanced with the equally important need to conduct rigorous scientific analysis. This is because there are so many potential compounds that need to be evaluated using robust computational techniques to identify those that have the best potential to succeed. In this study, we will utilize a multi-step approach in computational drug design that includes molecular docking to identify molecule based on their binding affinities followed by an ADMET prediction. This approach will enable us to arrive at a short list of candidate compounds that exhibit favourable interaction modes and pharmacological properties. Using NGS on publicly available NSCLC tumor database, the genomic data is utilised specifically those that exhibit the L858R/T790M/C797S mutations. Such high-resolution genomic data enables to predict accurate structures of the mutant EGFR protein so that we can conduct computational drug design in contexts that are biologically relevant. Translation of in silico studies into clinical realities requires multiple validation steps such as in vitro and in vivo which ultimately leads to human clinical trials. Since oncology is a dynamic and unpredictable field, promising candidates may face obstacles in terms of its efficacy.

The introduction of Next-Generation Sequencing (NGS) techniques has changed our way of deciphering the molecular basis of cancer. Traditional Sanger sequencing was found to be deficient in terms of throughput and were not able to identify mutations in complex tumor genomes. NGS techniques have been able to revolutionize this field due to its ability to sequence millions of DNA fragments in parallel. This analyses the tumour genome in a comprehensive manner. The ability to identify a wide range of genomic changes such as includes single nucleotide variants (SNVs), small insertions/deletions (indels), copy number variations (CNVs), and gene fusions in a single test is one of the major breakthrougsh in NGS techniques. Another advantage of NGS techniques is its ability to analyse multiple tumour samples in parallel using its multiplexing potential. This increases its efficiency in terms of data generation and analysis. The application of NGS in clinical practice has also been made easier due to the development of bioinformatics tools that can analyze the vast amounts of data generated using this technique. This includes aligning of sequencing data to a reference genome and annotating of mutations identified using this technique. However, it is also important to note that accurate interpretation of data generated using this technique requires a multidisciplinary approach.

In summary, NGS has revolutionized the field of precision medicine in oncology by providing a comprehensive information about the genomic alterations associated with the progression of the disease and the development of resistance. As NGS technology advances to become even more sensitive, rapid, and cost-effective, it will undoubtedly have a greater impact in helping to design individualized treatment strategies to manage this formidable clinical entity. Computer-Aided Drug Design (CADD) refers to computer-based methods of drug discovery based on molecular docking and ADMET analysis for the prediction of favorable interaction of potential drug compounds with target macromolecules. CADD allows researchers to identify compounds that have the ability to interact with mutant EGFR protein, including those associated with resistance.

II. Materials And Methodology:

Our research involves intensive in silico workflow to evaluate the potential interacting partners of EGFR kinase domain (PDB ID 9D3W) with triple mutation (L858R/T790M/C797S) as suitable target for designing EGFR TKI. The Next Generation Sequencing pipeline (NGS) includes fetching sequence from NCBI, BLAST, CLUSTAL omega, PPI, Sequence scanning and Pathway analysis whereas Computer Aided Drug Design (CADD) includes fetching target protein and ligand structure from Protein Data Bank (PDB) and PubChem respectively, visualizing them which is followed by B-factor analysis and RMSD calculation using PyMOL and RasMol, Molecular Docking analysis and ADMET prediction. Basic Local Alignment Search Tool was employed for searching sequence similarity. BLAST aligns our query sequences against the Molecular Modelling Database for identification of homologous sequences, evolutionary relationships and potential functional similarities.(Bandbe et al., n.d.) For multiple sequence alignment, CLUSTAL omega was performed. Multiple sequence alignment is an important tool for structure and function prediction of protein, phylogenetic analysis and sequence analysis. (Verma et al., 2025) Potential interactions between proteins were explored using STRING database. Integrating predicted and observed associations, Protein network analysis constructs networks interaction which provides insights about functional pathways and partners of our protein of

interest.(Kumari, Johri, et al., 2024) For functional annotation search of protein sequence, InterProScan was employed. InterProScan is a widely used bioinformatics software tool that helps identify protein domains, families, and functional sites by comparing a protein sequence against multiple signature databases.(Kumari et al., n.d.) KEGG database was used to obtain insights into the biological and functional pathways of the proteins. Vital information on pathways such as metabolic and signalling associated with the protein of interest was provided by KEGG pathways.(Mehrotra & Kumari, 2025) We used RasMol and PyMOL to visualize and analyze the 3-D structures of proteins. Molecular visualization tool was used for detailed protein-ligand interactions study.(Kumari & Agrawal, 2023) Further, B-factor analysis was used to demonstrate stable and flexible regions within three-dimensional structure of protein. The distribution of mobility is important for understanding protein stability, molecular interactions, and potential ligand-binding regions, which are critical aspects in structure-based drug design and computational docking studies.(Parashar & Kumari, 2025) Root Mean Square Deviation is a fundamental structural bioinformatics technique used to quantitatively measure the average distance between the atoms of two superimposed proteins or ligands as a numerical value (in) indicating how similar or different two structures are, making it indispensable for structural comparison, validation, and analysis. (Kumari, Santosh Kulkarni, et al., 2024)For docking analysis and examination of potential binding sites, Server Docking was employed. CB-Dock utilizes detection of cavity to predict binding pockets and automates protein-ligand docking. Using AutoDock Vina, CB-Dock performs blind docking and identifies cavities that lies within the protein thus providing an efficient method to examine protein-ligand interactions.(Bajaj & Kumari, 2025) This will be followed by an analysis of ADME toxicity using Protex 3.0. This approach will enable us to arrive at a short list of candidate compounds that exhibit favourable interaction modes and pharmacological properties.(Kumari & Gupta, 2023)

III. Results And Discussion:

Sequence Similarity Search

The BLAST analysis confirms that the query protein sequence shows almost complete similarity (99 to 100 %) with experimentally verified EGFR data. An E-value of 0.0 indicate the match is extremely significant. Several important functional regions such as active site, ATP binding site, polypeptide substrate binding site, dimer interface and activation loop were identified. Additionally, BLAST hit distribution represents strong sequence conservation across the multiple homologous protein.

Multiple Sequence Alignment

CLUSTAL omega showed that sequences corresponding to 9D3W, 8WD4 and 8HV4 are highly conserved (>95% similarity) structurally and functionally. Presence of ****** lines indicate identical amino acid represent strong evolutionary and functional similarity. Overall, it suggests that all three structures represent homologous protein with conserved biochemical roles as shown in the figure 2.

```
CLUSTAL O(1.2.4) multiple sequence alignment


sequence3     -GSGEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATS    59
sequence1     GSMGEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATS    60
sequence2     -SMGEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATS    59
               . ************************************************************

sequence3     PKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLIMQLMPFGSLLDYVREHKDNIGSQ    119
sequence1     PKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLIMQLMPFGSLLDYVREHKDNIGSQ    120
sequence2     PKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLIMQLMPFGSLLDYVREHKDNIGSQ    119
              ************************************************************

sequence3     YLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGRAKLLGAAAAEYHAE    179
sequence1     YLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGRAKLLGAAAAEYHAE    180
sequence2     YLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGRAKLLGAAAAEYHAE    179
              ************************************************************

sequence3     GGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGER    239
sequence1     GGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGER    240
sequence2     GGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGER    239
              ************************************************************

sequence3     LPQPPICTIDVYMIMVKCWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSP    299
sequence1     LPQPPICTIDVYMIMVKCWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSP    300
sequence2     LPQPPICTIDVYMIMVKCWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSP    299
              ************************************************************

sequence3     TDSNFYRALMDEEDMDDVVDADEYLIPQQGNS 331
sequence1     TDSNFYRALMDEEDMDDVVDADEYLIPQQG-- 330
sequence2     TDSNFYRALMDEEDMDDVVDADEYLIPQQG-- 329
              ******************************
```

Figure 2: Multiple Sequence Alignment with three proteomic sample using CLUSTAL omega

Protein-protein interaction analysis

STRING analysis illustrated EGFR as core node (red) with high connectivity. In figure 3, each node represents interacting proteins and confidence score of ~0.999 (very strong interaction). The major interacting proteins includes ERBB3 and ERBB2; EGF, HBEGF and EREG (bound to EGFR); PIK3CA (controls cell survival growth and metabolism); GABI (adaptor protein); CDH1 (cell adhesion molecule) and DCN (Decorin, tumor suppressor protein). STRING confirms that EGFR is highly connected signalling hub with a strong functional association to ligand receptor family members. Further, bubble plot in figure 4 represents enriched biological processes associated with EGFR dataset. The pathways show: highest signal values (~5.5–7.5); largest bubble sizes (higher gene counts) and very low FDR (~10^{-1} to 10^{-1}) which indicates highly significant.

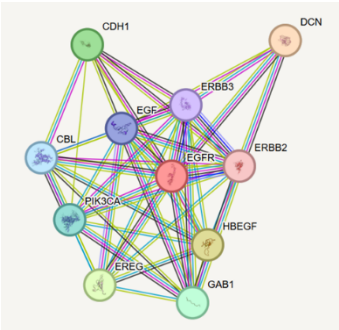


Figure 3: Protein-Protein Interaction (PPI) Network in protein enrichment analysis

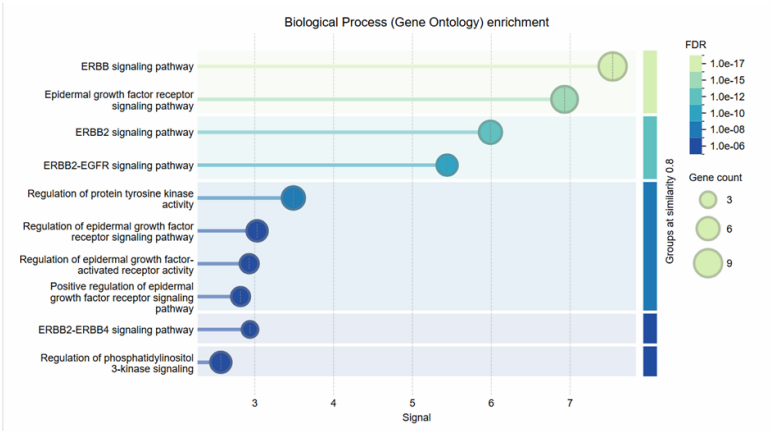


Figure 4: Bubble plot in protein enrichment analysis using STRING

KEGG pathway

KEGG pathway analysis suggests that NSCLC is driven by multiple interconnected signalling pathways not a single gene. Some key drivers include EGFR mutations, PI3K-AKT activation, MAPK pathway overactivation and loss of p53 function as shown in the figure 5.

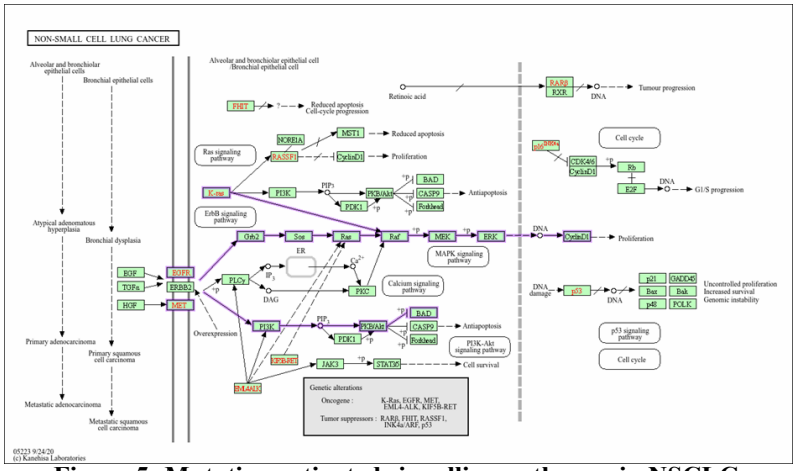


Figure 5: Mutation activated signalling pathways in NSCLC

B-Factor Analysis

The B-factor (temperature factor) distribution across protein residues is visualized with the red ribbon tracing the backbone structure as shown in the figure 6. Labelled residues—such as HIS (histidine), ASN (asparagine), MET (methionine), LEU (leucine), and ARG (arginine)—highlight key sites. Higher B-factor values (120–130) indicate greater atomic displacement and flexibility in those segments, while lower values (~70–95) denote structurally rigid regions.

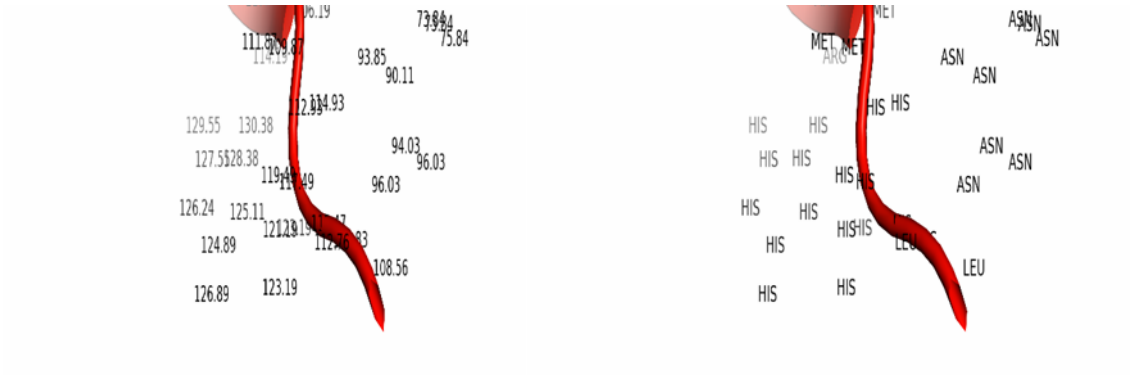
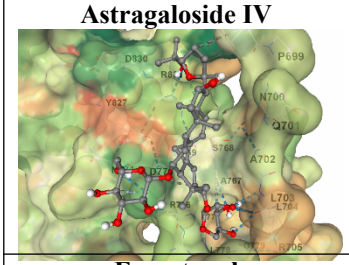
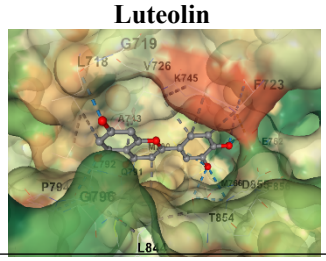
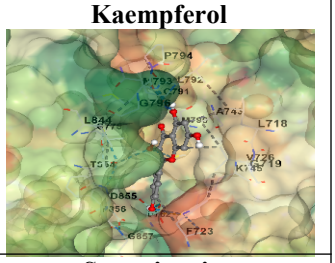


Figure 6: B-factor analysis of 9D3W protein using PyMol

Molecular docking

The 3D structures of all the phytoconstituents and the target protein were retrieved from PubChem and PDB databases respectively. Molecular docking results showed docking scores ranging from −7.4 to −12.5 kcal/mol as shown in the table 1 and visualized in figure 7. In molecular docking complex, natural compound molecule is bound with the active binding pocket of the protein structure PDB ID 6S9C, protein in ball & stick structure and ligands in space representation at the centre with well-defined binding cavity. Blue dashes line indicates hydrogen bond between ligand and amino acid residues with hydrophobic interactions.

Astragaloside IV	**Luteolin**	**Kaempferol**
		
Furostanol	**Ophiopogonin B**	**Sanguinarine**

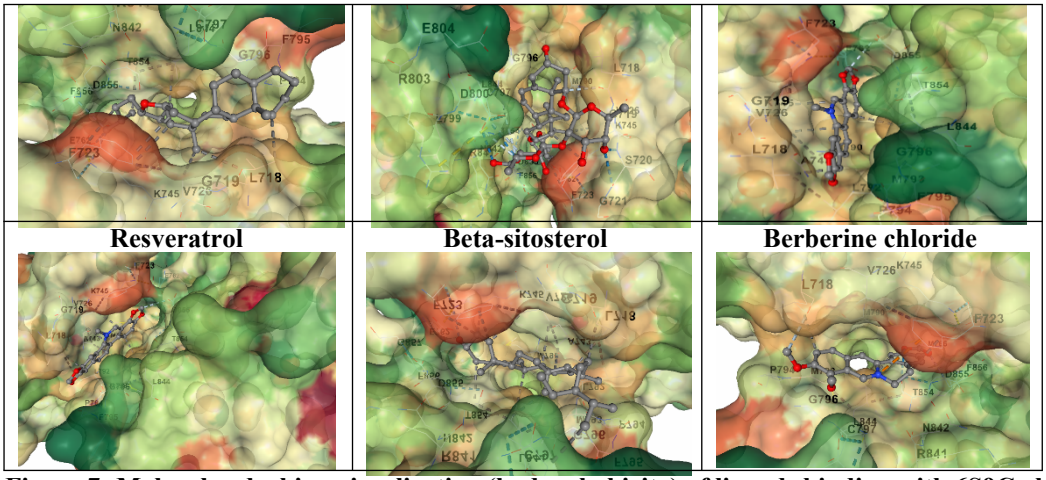


Figure 7: Molecular docking visualization (hydrophobicity) of ligands binding with 6S9C along with their key residues using CB-Dock

PubChem ID	Drug	CurPocket ID	Vina score (kcal/mol)	Cavity volume (Å³)	Center (x,y,z)	Remarks
13943297	Astragaloside IV	C2	-7.4	650	-31,-64,6	Good binding
5280445	Luteolin	C1	-9.1	4361	-30,-56,2	Strong binding
5280863	Kaempferol	C1	-8.1	4361	-30, -56, -2	Strong binding
22295359	Furostanol	C1	-8.6	4361	-30, -56, -2	Strong binding
46173857	Ophiopogonin B	C1	-12.5	4361	-30, -56, -2	Excellent binding
5154	Sanguinarine	C1	-10.1	4361	-30, -56, -2	Excellent binding
445154	Resveratrol	C1	-10.1	4361	-30, -56, -2	Excellent binding
222284	Beta-Sitosterol	C1	-9.9	4361	-30, -56, -2	Excellent binding
12456	Berberine chloride	C1	-10.7	4361	-30, -56, -2	Excellent binding

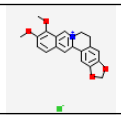
						

Table 1: 6S9C-ligand Molecular docking analysis using CB-Dock

ADMET prediction

Based on their anticancer potential report in the literature, favourable ADMET profiles and binding compatibility with the EGFR binding site, astragaloside IV (CID 13943297), kaempferol (CID 528063), furostanol (CID 22295359) and luteolin (CID 5280445) were selected based on their safety profiles and favorable molecular docking results. ProTex 3.0 classifies astrogaloside IV as Class 6 (Non-toxic) and kaempferol, furostanol and luteolin as Class 5 (practically non-toxic). The predicted lethal dose 50 (LD50) of astragaloside IV, kaempferol, furostanol and luteolin are 23000 mg/kg, 3919 mg/kg, 5000 mg/kg and 3919 mg/kg respectively. ProTox 3.0 predictions indicate that astragaloside IV, kaempferol, furastanol, and luteolin generally show low toxicity across most categories (e.g., hepatotoxicity, carcinogenicity in most cases), with favorable metabolic profiles. All exhibit estrogen receptor (ER) activation, suggesting endocrine disruption risks, alongside moderate mitochondrial dysfunction. However, each has distinct toxicity concerns requiring experimental validation before therapeutic use.

- **Astragaloside IV**: Significant immunotoxicity and mitochondrial dysfunction; moderate cardiotoxicity. Focus validation on immune response, cardiac safety and mitochondrial integrity.
- **Kaempferol**: Significant respiratory toxicity and mitochondrial dysfunction; moderate mutagenicity. Focus validation on respiratory safety and mitochondrial integrity.
- **Furastanol**: Significant immunotoxicity; moderate respiratory toxicity and cardiotoxicity. Focus validation on immune response, respiratory safety and mitochondrial integrity.
- **Luteolin**: Significant respiratory toxicity and mitochondrial dysfunction; moderate carcinogenicity and mutagenicity. Additional AhR/ER signaling concerns. Focus validation on respiratory safety, carcinogenicity, mutagenicity, AhR/ER pathways and mitochondrial integrity.

Parameters	Astragaloside IV	Kaempferol	Furostanol	Luteolin
Hepatotoxicity	Inactive (0.92)	Inactive (0.68)	Inactive (0.71)	Inactive (0.69)
Respiratory toxicity	Inactive (0.54)	Active (0.83)	Inactive (0.52)	Active (0.83)
Cardiotoxicity	Active (0.52)	Inactive (0.91)	Inactive (0.60)	Inactive (0.99)
Carcinogenicity	Inactive (0.74)	Inactive (0.72)	Inactive (0.71)	Active (0.68)
Immunotoxicity	Active (0.99)	Inactive (0.96)	Active (0.98)	Inactive (0.97)
Mutagenicity	Inactive (0.67)	Inactive (0.52)	Inactive (0.88)	Active (0.51)
Cytotoxicity	Inactive (0.67)	Inactive (0.98)	Inactive (0.82)	Inactive (0.99)
Aryl hydrocarbon Receptor (AhR)	Inactive (0.99)	Active (1.0)	Inactive (0.98)	Active (0.91)
Androgen Receptor (AR)	Inactive (0.91)	Inactive (0.99)	Inactive (0.88)	Inactive (0.99)
Estrogen Receptor Alpha (ER)	Active (0.60)	Active (1.0)	Inactive (0.80)	Active (0.87)
PPAR-Gamma	Active (0.51)	Inactive (0.95)	Inactive (0.97)	Inactive (0.98)
Heat shock factor response element	Inactive (0.96)	Inactive (0.99)	Inactive (0.96)	Inactive (0.99)
Mitochondrial Membrane Potential	Active (0.77)	Active (1.0)	Inactive (0.70)	Active (1.0)
Phosphoprotein p53	Active (0.58)	Inactive (0.92)	Inactive (0.95)	Inactive (0.97)
Thyroid hormone receptor alpha (THRα)	Inactive (0.83)	Inactive (0.90)	Inactive (0.81)	Inactive (0.90)
Thyroid hormone receptor beta (THRß)	Inactive (0.9)	Inactive (0.78)	Inactive (0.94)	Inactive (0.78)
Achetylcholinesterase (AChE)	Inactive (0.60)	Inactive (0.68)	Inactive (0.79)	Inactive (0.69)
NADH-quinone oxidoreductase	Active (0.58)	Inactive (0.97)	Inactive (0.62)	Inactive (0.97)
Voltage gated sodium channel (VGSC)	Inactive (0.90)	Inactive (0.95)	Inactive (0.93)	Inactive (0.94)
Cytochrome CYP3A4	Inactive (0.95)	Inactive (0.65)	Inactive (0.73)	Inactive (0.79)

Table 2: Toxicity profile model report analysis for astragaloside IV, kaempferol, furostanol and luteolin using ProTex 3.0

A	B

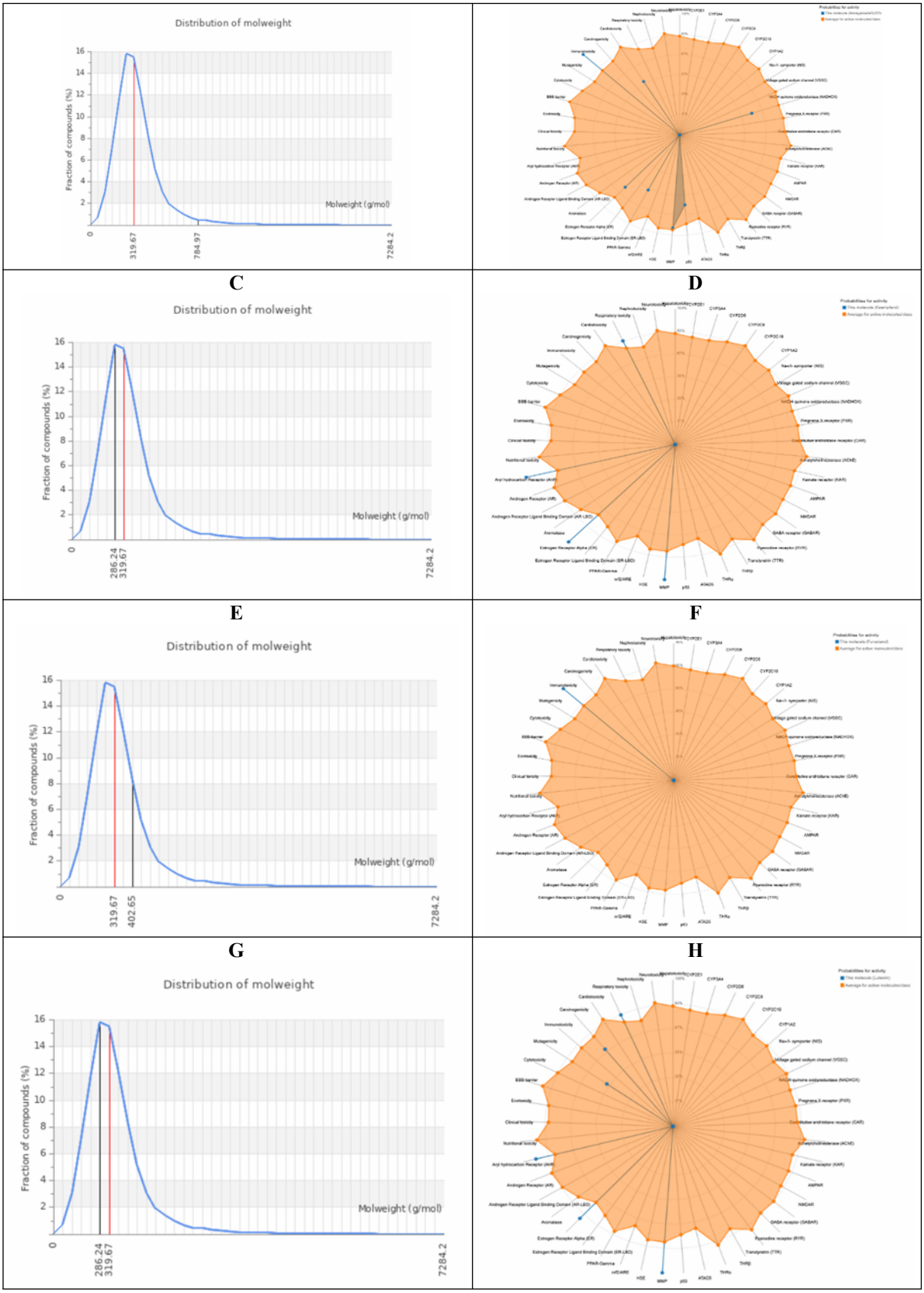


Figure 8 (A-H): A – Distribution of molecular weight for astragaloside IV; B – Radar chart form of toxicity report for astralgaloside IV; C - Distribution of molecular weight for kaempferol; D – Radar chart form of toxicity report for kaempferol; E - Distribution of molecular weight for furostanol; F –

Radar chart form of toxicity report for furostanol; G - Distribution of molecular weight for luteolin and H – Radar chart form of toxicity report for luteolin

In the molecular weight distribution plots, astragaloside IV (784.97 Da) falls far to the right of the dataset mean, kaempferol (286.24 Da) and luteolin (286.24 Da) lie near the right, and furostanol (402.65 Da) near the left. Blue dots/lines mark our compounds, while the orange shaded region shows the average activity range of known toxic/active compounds across toxicity endpoints or biological targets on each axis. A larger orange region signals higher toxicity probability; blue points closest to the centre predict low toxicity as shown in figure 8.

IV. Conclusion

EGFR mutations are well-documented oncogenic drivers and possess strong clinical evidence; representing major clinical relevance and justifies our selection for this research. In our study, a crystal structure of mutated EGFR kinase domain (9D3W) with triple mutation (L858R/T790M/C797S) was selected as suitable target for designing EGFR Tyrosine Kinase inhibitors. Next Generation Sequencing (NGS) workflow confirms that 9D3W, 9D3V, 8WD4, 6S9C, 8HV4 and 6LUB represent homologous protein with conserved biochemical roles. Compared to synthetic tyrosine kinase inhibitors, our research involves a solid strategy for discovering safer and more sustainable alternatives such as natural EGFR inhibitors. Natural compounds such as astragaloside IV, kaempferol, furostanol and luteolin were identified as the most promising candidates among the screened natural products based on their safety profiles, drug-likeness and docking results. Our comprehensive computational workflow assesses genomic analysis and pharmacological properties that lays a solid foundation for in vitro and in vivo validation to evaluate the therapeutic potential of the promising compounds against mutant EGFR in NSCLC.

References

[1]. Bajaj, T., & Kumari, U. (2025). Issue 3 Www.Jetir.Org (ISSN-2349-5162). In JETIR2503614 Journal Of Emerging Technologies And Innovative Research (Vol. 12). Www.Jetir.Org

[2]. Bandbe, T., Saikia, J., & Kumari, U. (N.D.). NGS Analysis Human Papillomavirus Type 18 E2 DNA-Binding Domain Bound To Its DNA Target With Biopython.

[3]. Bindod, H. V., Hatwar, P. R., Bakal, R. L., & Dafe, V. N. (2025). A Comprehensive Review Of Cancer: Types, Pathophysiology, Diagnosis And Treatments. Journal Of Drug Delivery And Therapeutics, 15(5), 114–122. Https://Doi.Org/10.22270/Jddt.V15i5.7104

[4]. Ferro, A., Marinato, G. M., Mulargiu, C., Marino, M., Pasello, G., Guarneri, V., & Bonanno, L. (2024). The Study Of Primary And Acquired Resistance To First-Line Osimertinib To Improve The Outcome Of EGFR-Mutated Advanced Non-Small Cell Lung Cancer Patients: The Challenge Is Open For New Therapeutic Strategies. In Critical Reviews In Oncology/Haematology (Vol. 196). Elsevier Ireland Ltd. Https://Doi.Org/10.1016/J.Critrevonc.2024.104295

[5]. Gillian Bethune, D. B., N. R., Z. X. (2010). Epidermal Growth Factor Receptor (EGFR) In Lung Cancer An Overview And.

[6]. Gupta, R., Dastane, A. M., Forozan, F., Riley-Portuguez, A., Chung, F., Lopategui, J., & Marchevsky, A. M. (2009). Evaluation Of EGFR Abnormalities In Patients With Pulmonary Adenocarcinoma: The Need To Test Neoplasms With More Than One Method. Modern Pathology, 22(1), 128–133. Https://Doi.Org/10.1038/Modpathol.2008.182

[7]. Hanahan, D. (2026). Hallmarks Of Cancer-Then And Now, And Beyond. Cell. Https://Doi.Org/10.1016/J.Cell.2025.12.049

[8]. Hendriks, L. E. L., Remon, J., Faivre-Finn, C., Garassino, M. C., Heymach, J. V., Kerr, K. M., Tan, D. S. W., Veronesi, G., & Reck, M. (2024). Non-Small-Cell Lung Cancer. Nature Reviews Disease Primers, 10(1). Https://Doi.Org/10.1038/S41572-024-00551-9

[9]. Inamura, K., Ninomiya, H., Ishikawa, Y., & Matsubara, O. (2010). Is The Epidermal Growth Factor Receptor Status In Lung Cancers Reflected In Clinicopathologic Features? In Arch Pathol Lab Med (Vol. 134). Https://Aplm.Kglmeridian.Com

[10]. Jeon, H., Wang, S., Song, J., Gill, H., & Cheng, H. (2025). Update 2025: Management Of NonSmall-Cell Lung Cancer. In Lung (Vol. 203, Number 1). Springer. Https://Doi.Org/10.1007/S00408-025-00801-X

[11]. Kumari, U., & Agrawal, N. (2023). NGS And Mutational Profile Analysis Of Non-Small-Cell Lung Carcinoma (NSCLC). International Journal For Research In Applied Science And Engineering Technology, 11(4), 3090–3094. Https://Doi.Org/10.22214/Ijraset.2023.50880

[12]. Kumari, U., Belokar, P., Esompalli, A., Deshpande, S., Kumkar, M., & Tripathi, A. (N.D.). Next-Generation Sequencing To Investigate The P53 Cancer Mutant Y234C For Targeted Cancer Therapies. Issue 4 Ser, 20, 63–72. Https://Doi.Org/10.9790/3008-2004016372

[13]. Kumari, U., & Gupta, S. (2023). NGS And Sequence Analysis With Biopython For Prospective Brain Cancer Therapeutic Studies. International Journal For Research In Applied Science And Engineering Technology, 11(4), 3318–3329. Https://Doi.Org/10.22214/Ijraset.2023.50885

[14]. Kumari, U., Johri, V., Dhopate, S., & Jha, T. (2024). Structure Based Drug Designing For The Prediction Of Epitope For Targeting Malignant Brain Tumor (Vol. 11). JETIR. Www.Jetir.Orgb689

[15]. Kumari, U., Santosh Kulkarni, S., Kaur, G., & Chaudhary, R. (2024). Issue 6 Www.Jetir.Org (ISSN-2349-5162). In JETIR2406827 Journal Of Emerging Technologies And Innovative Research (Vol. 11). Www.Jetir.Org

[16]. Ladanyi, M., & Pao, W. (2008). Lung Adenocarcinoma: Guiding EGFR-Targeted Therapy And Beyond. Modern Pathology, 21, S16–S22. Https://Doi.Org/10.1038/Modpathol.3801018

[17]. Laura Maiorino, J. D.-P. L. S. And M. E. (N.D.). Innate Immunity And Cancer Pathophysiology.

[18]. Massarelli, E., Varella-Garcia, M., Tang, X., Xavier, A. C., Ozburn, N. C., Liu, D. D., Sbekele, B. N., Herbst, R. S., & Wistuba, I. I. (2007). KRAS Mutation Is An Important Predictor Of Resistance To Therapy With Epidermal Growth Factor Receptor Tyrosine Kinase

Inhibitors In Non-Small Cell Lung Cancer. Clinical Cancer Research, 13(10), 2890–2896. Https://Doi.Org/10.1158/1078-0432.CCR-06-3043

[19]. Mehrotra, K., & Kumari, U. (2025). Issue 2 Www.Jetir.Org (ISSN-2349-5162). In JETIR2502274 Journal Of Emerging Technologies And Innovative Research (Vol. 12). JETIR. Www.Jetir.Orgc630

[20]. Parashar, R., & Kumari, U. (2025). Issue 3 Www.Jetir.Org (ISSN-2349-5162). In JETIR2503613 Journal Of Emerging Technologies And Innovative Research (Vol. 12). JETIR. Www.Jetir.Orgg98

[21]. Piña-Sánchez, P., Chávez-González, A., Ruiz-Tachiquín, M., Vadillo, E., Monroy-García, A., Montesinos, J. J., Grajales, R., Gutiérrez De La Barrera, M., & Mayani, H. (2021). Cancer Biology, Epidemiology, And Treatment In The 21st Century: Current Status And Future Challenges From A Biomedical Perspective. In Cancer Control (Vol. 28). SAGE Publications Ltd. Https://Doi.Org/10.1177/10732748211038735

[22]. Rajput, J. M. (2023). CANCER: A COMPREHENSIVE REVIEW. In Cancer: A Comprehensive Review Page42 International Journal Of Research In Pharmacy And Allied Science (Vol. 1, Number 3). Https://Idealpublication.In/Ijrpas/

[23]. Verma, G., Kumari, U., & Murugesan, A. (2025). Issue 2 Www.Jetir.Org (ISSN-2349-5162). In JETIR2502318 Journal Of Emerging Technologies And Innovative Research (Vol. 12). Www.Jetir.Org

[24]. Yayan, J., Franke, K. J., Berger, M., Windisch, W., & Rasche, K. (2024). Adhesion, Metastasis, And Inhibition Of Cancer Cells: A Comprehensive Review. In Molecular Biology Reports (Vol. 51, Number 1). Springer Science And Business Media B.V. Https://Doi.Org/10.1007/S11033-023-08920-5