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

Integrated Sequence and Structural Bioinformatics Analysis of EGFR (T790M/V948R) in Non-Small Cell Lung Cancer: Structural Characterization and Molecular Docking of Targeted Inhibitors

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

Corresponding Author: Uma Kumari(uma27910@gmail.com)

ABSTRACT

Epidermal growth factor receptor (EGFR) is play an important role in the development of therapeutic resistance in non-small cell lung cancer (NSCLC). This study shows the structural and functional properties of the EGFR (T790M/V948R) mutant through an integrated computational approach. The protein structure (PDB ID: 8GB4) was examined using sequence similarity searching, functional domain annotation, structural validation and molecular visualization techniques. Sequence Similarity search revealed 99.4–100% sequence identity with highly significant alignment scores (E-value: 0.0). Structural model assessment using save Server produced,Excellent quality factor of 99.544, while Ramachandran analysis identified approximately 90% of residues within favoured conformational regions. Signature sequence analysis highlighted the conserved protein tyrosine kinase domain and functionally important residues. B-factor evaluation indicated differences in structural flexibility between the kinase core and surface loops. Molecular docking provided preliminary insights into lazertinib binding, while KEGG analysis identified associated oncogenic signalling pathways. These findings support further investigation of mutant EGFR as a potential therapeutic target in NSCLC.

Keywords: EGFR, NSCLC, T790M/V948R, Structural Bioinformatics, Protein Structure Validation, Molecular Docking, Targeted Drug Discovery.

1. INTRODUCTION

Lung cancer is still one of the most common causes of morbidity and mortality from cancer in the world. Non–small-cell lung cancer (NSCLC) is the most common type of lung cancer and accounts for the majority of cases of lung cancer. NSCLC is a heterogeneous group of malignancies with different molecular alterations, genetic heterogeneity and different patterns of disease progression [1,2]. Progress in molecular characterisation has demonstrated that the definition of specific oncogenic alterations is critical for the understanding of tumour biology and the development of personalised treatments [2,3]. In NSCLC, among the molecular drivers, mutations of the epidermal growth factor receptor (EGFR) have received much attention because they are directly relevant for targeted therapeutic approaches.

EGFR is a member of the ErbB receptor tyrosine kinase family and is critical in regulating cellular proliferation, survival, differentiation and intracellular signalling. Activating mutations of EGFR, especially deletions within exon 19 and the L858R substitution in exon 21, have been linked to sensitivity to EGFR-directed tyrosine kinase inhibitors (TKIs) [3,4]. But a major limitation of targeted therapy is the development of acquired resistance. A prototypical resistance-associated

mutation is the T790M substitution, the replacement of threonine by methionine at residue 790 in the EGFR kinase domain. This substitution changes the biochemical and structural properties of the ATP-binding pocket, leading to decreased sensitivity to the first- and second-generation EGFR inhibitors [4,5]. In addition, other signalling pathways and further molecular changes may also be involved in disease progression and therapeutic resistance, highlighting the need to understand mutation-related structural changes in EGFR [5,6].

As an important approach for detecting genomic alterations and characterising molecular heterogeneity in NSCLC, next-generation sequencing (NGS) has emerged. Profiling of circulating cell-free DNA by sequencing, in particular, holds promise for studying tumour-associated mutations and tracing the emergence of genetic alterations associated with resistance during targeted treatment [6,7]. Integrating genomic information with computational bioinformatics approaches can help to identify disease-associated molecular targets, conserved protein domains and possible therapeutic interactions. The combination of NGS, sequence analysis, molecular docking and network pharmacology has been shown to be useful for investigating cancer-associated targets and identifying possible treatment candidates [8]. Furthermore, machine learning, artificial intelligence and BioPython-based computational drug-discovery pipelines have demonstrated the potential of integrating multi-omics information with structure-based drug discovery for the exploration of disease-associated molecular targets [9,10].Structural bioinformatics provides another framework for understanding how amino acid substitutions affect protein architecture, functional domains and ligand-binding interactions. Three-dimensional protein structures are available from the Protein Data Bank (PDB) [11] to study sequence conservation, structural organisation, and molecular interactions. BLAST-based sequence similarity searches identify homologous sequences and allow to evaluate evolutionary conservation, while multiple sequence alignment offers insight into conserved and variable amino acid positions [12,13]. Functional annotation platforms like InterProScan and PROSITE also help to identify conserved domains, protein signatures and residues that are important for function [14,15]. Further structural validation with tools like ERRAT and Ramachandran analysis provides information about non-bonded interactions and stereochemical quality that supports the reliability of structural interpretations [16,17]. In addition, temperature factor (B-factor) analysis may give information about the relative atomic displacement and conformational variability of different regions of a protein.Molecular docking is an important computational approach to study the interactions of therapeutic drugs with disease-related protein targets. Docking studies provide preliminary views on binding-site features, molecular recognition and structure-activity relationships by predicting probable ligand binding orientations and molecular interactions [18,19]. Third-generation inhibitors have improved the management of T790M-associated resistance in EGFR-mutant NSCLC. Lazertinib is a leading example of a targeted EGFR inhibitor, and its structural interaction with mutant EGFR provide a useful framework to explore inhibitor recognition and binding site properties [19]. The experimentally determined protein structure deposited under PDB ID 8GB4 was selected as the primary structural model and compared with additional EGFR structures for structural comparison [20,21]. Sequence conservation and functional features were analysed using BLAST-based sequence similarity analysis, multiple sequence alignment, phylogenetic analysis, InterProScan and PROSITE annotation. ERRAT analysis, Ramachandran plot assessment and B-factor visualisation were done later for the analysis of protein structural quality and flexibility. Molecular docking and active-site identification were performed to investigate the possible interactions between lazertinib and EGFR. Moreover, KEGG pathway analysis was employed to contextualise EGFR-associated signalling networks in NSCLC. Overall, this study presents an integrated computational characterisation of EGFR and preliminary structural insights into the inhibitor interactions that may contribute to further investigations of targeted therapeutic approaches in NSCLC.

2. MATERIAL AND METHODS

PDB is a repository for the three-dimensional structures of nucleic acids, proteins as well as other biological macromolecules that have been experimentally identified. Currently, the archive contains experimental datasets, accompanying metadata, and three-dimensional atomic models generated through well-established structural biology techniques, namely X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy. [24,25,26] The primary aim of Entrez's 3D-structure database is to enable convenient access to structural insights and functional annotations for molecular biology research. [27,28,29], The basic algorithm is simple and robust; it can be implemented in a number of ways and applied in a variety of contexts including straightforward DNA and protein sequence database searches, motif searches, gene identification searches, and in the analysis of multiple regions of

similarity in long DNA sequences. Basic Local Alignment Search Tool (BLAST) is an essential algorithm that researchers use for sequence alignment analysis. The NCBI-BLAST application is the most popular implementation of the BLAST algorithm. [30,31,32] ERRAT is a protein structure verification algorithm, which runs by statistical analysis of non-bonded interactions between different types of atoms. [33,34]. It generates a single output plot showing the error value to the residue window. By statistical data comparison with highly evaluated structures, it generates error values to yield the confidence limits. [35,36]. ERRAT tool finds the overall quality factor of the proteins. ERRAT plot gives statistics of error in the structure for each residue in the 3D protein structure model. [37,38]InterProScan was used for scanning the sample sequence and to determine different families, domains and important sites present within the protein for further analysis.Warren Lyford DeLano developed and initially distributed PyMOL, an open-source molecular visualization software designed for the graphical representation of macromolecular structures. Implemented in the Python programming language, PyMOL supports a wide range of molecular visualization styles, including ribbon diagrams, cartoon (caricature) representations, surface models, spheres, sticks, and line formats, facilitating detailed structural analysis of biomolecules.PyMOL facilitates diverse structural bioinformatics applications, including sequence analysis, molecular docking, molecular dynamics simulations, structure-function relationship studies, protein structure prediction, and virtual screening, thereby supporting comprehensive computational investigations in structural biology and drug discovery. [39,40]

3.RESULTS AND DISCUSSION

3.1 PROTEIN DATA BANK

The initial step in our analysis was to download the three-dimensional structure of the protein of interest. The PDB file corresponding to PDB ID: 8GB4 was successfully retrieved from the Protein Data Bank. This model is a high-resolution image of the three-dimensional structure of the protein, from which the amino acid sequence can be obtained for further analysis. The protein structure (PDB ID: 8GB4) was submitted to the NCBI Molecular Modeling Database (MMDB), and its corresponding amino acid sequence was retrieved in FASTA format. The extracted sequence served as the basis for subsequent sequence-based analyses. [41,42]

3.2 Sequence Similarity Search Tool

After Sequence extraction the obtained FASTA sequence was used as a query to perform a similarity search in the BLAST, Basic Local Alignment Search Tool database. The alignment results highlighted several highly similar protein sequences, providing insights into conserved regions, potential functional domains, and evolutionary relationships [43,44].The FASTA sequence retrieved from the MMDB database of PDB structure 8GB4 was run on BLAST p by changing against the non-redundant protein database to identify homologous sequences and understand evolutionary and functional relationships. [45] The BLAST results revealed that all top hits correspond to the Epidermal Growth Factor Receptor (EGFR) Chain A, from *Homo sapiens*.[46] The maximum alignment score (Max Score) observed was 688, indicating an exceptionally high level of sequence match between the query and subject sequences.[47] The total alignment scores ranged from 688 to 682, as shown in **Figure.1**. demonstrating that the sequences in multiple structural entries are nearly identical.[48] The E-values for all results were 0.0, which depicts an extremely significant match, with virtually no chance of the alignment occurring randomly. The percentage identity ranged from 100% to 99.4%, further confirming the near-complete sequence identity. [49,50]

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Chain A, Epidermal growth factor receptor [Homo sapiens]	Homo sapiens	688	688	100%	0.0	100.00%	331	4ZSE_A
Chain A, Epidermal growth factor receptor [Homo sapiens]	Homo sapiens	684	684	100%	0.0	99.40%	331	7K1H_A
Chain A, Epidermal growth factor receptor [Homo sapiens]	Homo sapiens	683	683	100%	0.0	99.40%	331	4ZJV_A
Chain A, Epidermal growth factor receptor [Homo sapiens]	Homo sapiens	682	682	99%	0.0	99.70%	333	6TFU_A
Chain A, Epidermal growth factor receptor [Homo sapiens]	Homo sapiens	682	682	100%	0.0	99.40%	331	3IKA_A

Figure1. Sequence similarity search of sample 8GB4

The query coverage was between 99% and 100%, indicating that the entire length of the query sequence was aligned. [51]Based on the above observations, it can be concluded that the query protein sequence is identical to the human EGFR (Chain A). From blast p results, we got some detailed three-dimensional protein structures from the PDB database, such as 4ZSE_A, 7K1H_A, and 4ZJV_A, share the same or highly similar protein sequence and structure. [52]

3.3 Phylogenetic analysis of BLAST P Cladogram

The phylogenetic tree (cladogram) was generated using the BLAST tree view method to analyze the evolutionary relationship between the query protein sequence and the other related sequences in the database. The query sequence (Chain B, Receptor tyrosine-protein kinase erbB-4 from *H. sapiens*), highlighted in yellow, is positioned near a well-defined green cluster representing closely related primate sequences, indicating a close evolutionary relationship with other *H. sapiens* Epidermal Growth Factor Receptor (EGFR) sequences as shown in **Figure.1**.



Figure 2. Blast P Phylogenetic analysis

Additionally, the phylogenetic tree results show an extensive multiple-organism cluster (highlighted in yellow, containing 37 leaves), reflecting a heterogeneous group of homologous sequences from various non-primate organisms.; Green: Primate-specific cluster;Grey: Other sequence clusters from non-primate organisms,Yellow: Query/reference sequence.The phylogenetic analysis clearly indicates that the ErbB4 receptor tyrosine kinase (Human) is highly conserved within primates, clustering closely with other epidermal growth factor receptor (EGFR/Erb family) sequences from *H. sapiens* and related primate species.

3.4 Structural Quality Assessment

The structural quality of Chain C from PDB ID: 8GB4 was evaluated using ERRAT software which analyzes non-bonded interactions between different atom types to detect regions of potential structural errors. The overall quality factor was calculated as 99.544, indicating a very high-quality model. The error plot, residues with error values above the 95% confidence threshold are highlighted in yellow. Scores above 95% are typically considered excellent, signifying that the structural model is highly reliable for further computational studies such as docking or dynamic simulations as shown in Figure 3.

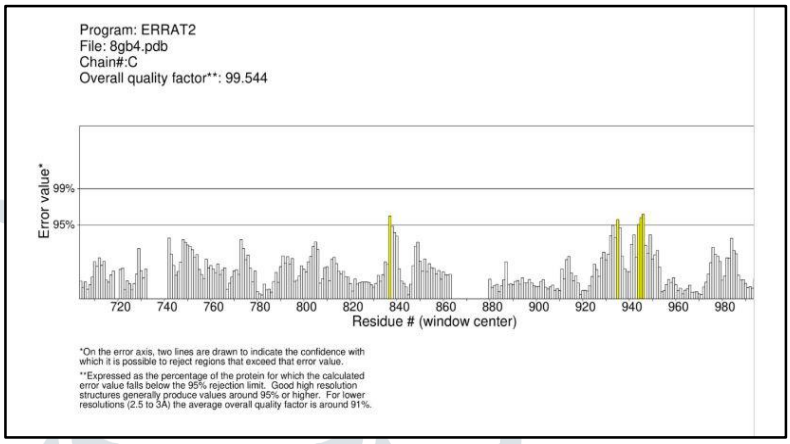

Figure.3 Structure Validation analysis with score 99.5 (8GB4)

3.5 Ramachandran Plot

From the above plot, the yellow-colored bars represent residues with error scores exceeding the 95% confidence threshold.Scores above 95% are typically considered excellent, signifying that the structural model is highly reliable. The yellow highlighted regions represent amino acid residues as shown in Figure 5.

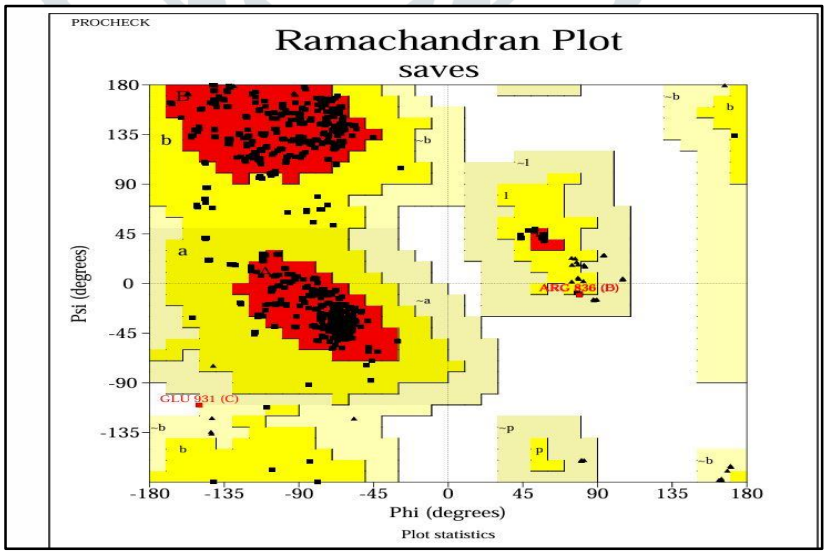

Figure. 4. Ramachandran plot analysis of 8GB4

where the calculated error exceeds the rejection limit. However, the overall high score (>99%) demonstrates that most of the structure is well-resolved.

The Ramachandran plot analysis was performed to assess the stereochemical quality of the protein structure of 8GB4. The Ramachandran plot illustrates the distribution of phi (φ) and psi (ψ) torsion angles for each amino acid residue in the protein.The red regions represent the most core allowed conformational region, while the yellow regions indicate additional allowed regions.The white regions represent the most disallowed conformations that are generally not preferred due to steric hindrance or Irregular backbone geometry.

3.6 Functional Domain Prediction with Sequence Scanning

In sequence scanning analysis was performed to predict the functional domains, conserved residues, active sites and key structural features of the protein 8GB4.Identifying conserved domains and families confirms that the protein is a member of the kinase superfamily, specifically involved in signalling pathways regulated by phosphorylation events. From the above figure domain analysis, it concluded that the presence of PTKc domain (Protein Tyrosine Kinase catalytic domain), shown by the green bar region a large portion of the sequence. This domain acts as the primary functional region responsible for catalytic activity, including phosphorylation of substrate proteins. The domain consists of essential structural features, such as activation loops and the ATP binding site, which is critical for kinase activity and regulation.

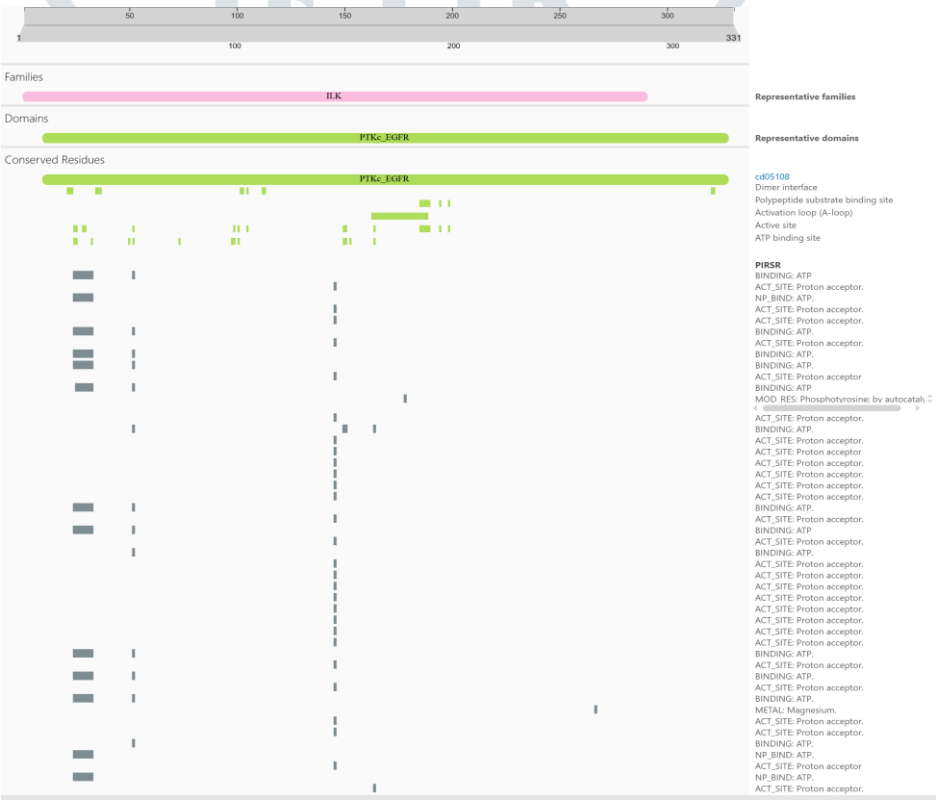

Figure.5 Sequence Scanning Analysis

The vertical markers represent amino acid residues that are highly conserved and essential across the entire kinase family. These residues are essential for ATP binding, catalytic function and substrate recognition, indicating their critical role in maintaining the structural stability and proper functional activity of the kinase domain.InterProScan predicts multiple ATP-binding sites and proton acceptor/catalytic residues within the protein tyrosine kinase domain. This analysis proves a comprehensive view of the functional domains, conserved key residues, and active site architecture of the 8GB4 protein structure. The identification of these conserved and functional regions is critical in drug discovery and in the design of targeted treatments against specific protein domains associated with malignancy. The molecular visualization is made in PyMOL to clearly distinguish four chains within the protein complex. Chain A is yellow, Chain B is green, Chain C is red and Chain D is blue. This color coding facilitates easy identification of each individual chain and thus structural analysis and assessment of inter-chain interactions.

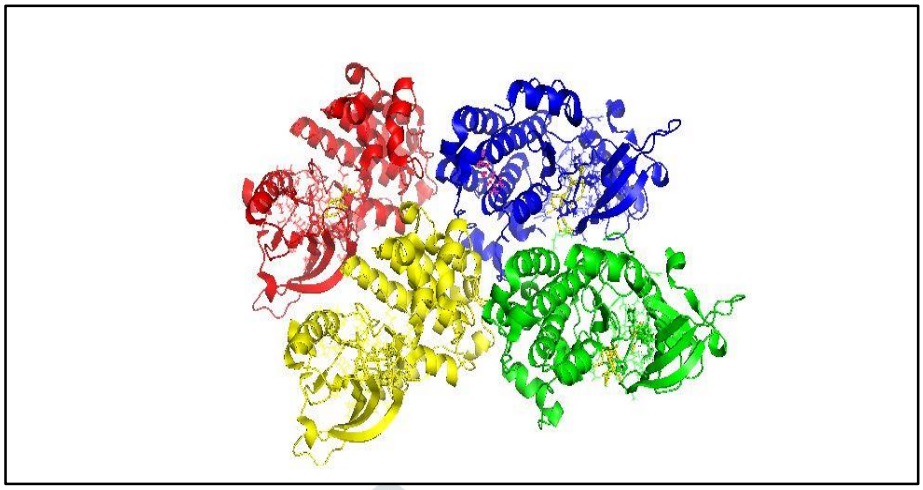

Figure 6. Chain identification in Py-Mol sample chain A (green), Chain B (yellow), Chain C (red), Chain d (blue)

The color-coded representation enables identification of each individual chain, and assessment of inter chain interactions. Such visual differentiation is essential for advanced structural studies, supporting the investigation of functional domains, active sites, and potential inter-chain binding interactions. Importantly, this structural insight contributes to understanding the protein's role in lung cancer-related signaling pathways, aiding the identification of key regions for further computational docking and functional studies.

3.7 B-Factor Analysis of the Proteomic Sample (Non-Small Cell Lung Cancer)

The B-factor (temperature factor) analysis was performed on the protein structure of PDB ID: 8GB4 to assess the flexibility and dynamic behavior of individual residues. The B-factor reflects atomic displacement or vibration within the protein structure, providing insight into regions of structural stability and flexibility.The B factor (temperature factor) analysis was performed on the protein structure of PDB ID: 8GB4 to check the flexibility and dynamic behavior of individual residues. The B-factor measures the degree of atomic movement or disorder in the protein structure, providing valuable information about structural rigidity and dynamic regions.

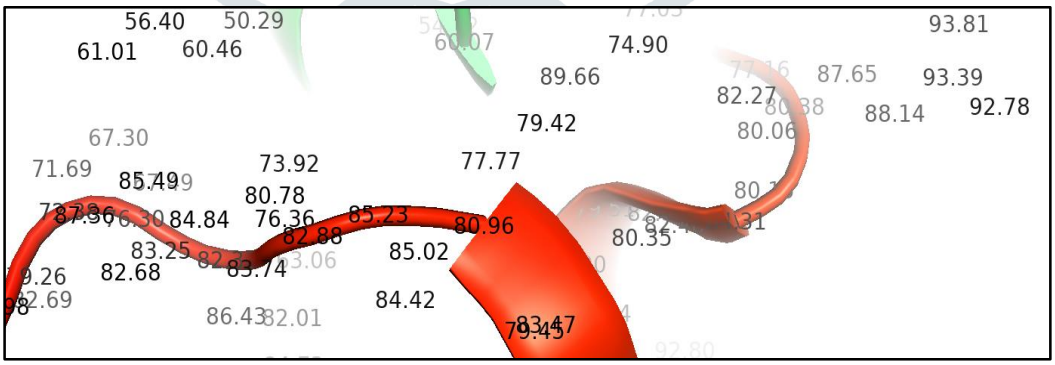

Figure 7. B-factor analysis to entire Structure (red loop) more than 70 plus

In the visual representation, the protein ribbon is color-coded by B-factor values:From the above figure, the protein ribbon is color coded by B-factor values:Red regions indicate areas of higher flexibility (higher B-factor).Green regions represent more stable and well-ordered parts of the structure (lower B-factor) as shown in **Figure.8.**

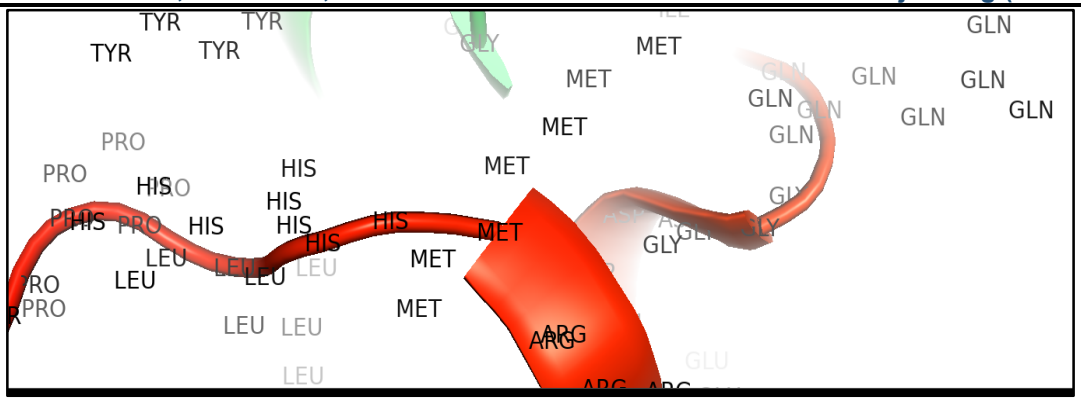

Figure 7. B-factor analysis to entire structure (red loop) showing residue name

The B factor values of amino acid residues mostly range between 50-95.

The residues range between 80-95 range are considered highly flexible or less ordered, mostly located in the surface loops and activation regions.Other than the residues with B factors range between 50-70 correspond to more stable regions of the structure, typically forming the catalytic core of the kinase domain.The catalytic core of the kinase domain maintains relative stability with B-factors around 50-70, indicating a well-resolved and rigid structure.In contrast, surface exposed loops and activation regions show B-factors exceeding 80-90, Reflecting significant structural flexibility and potential intrinsic disorder Figure 9

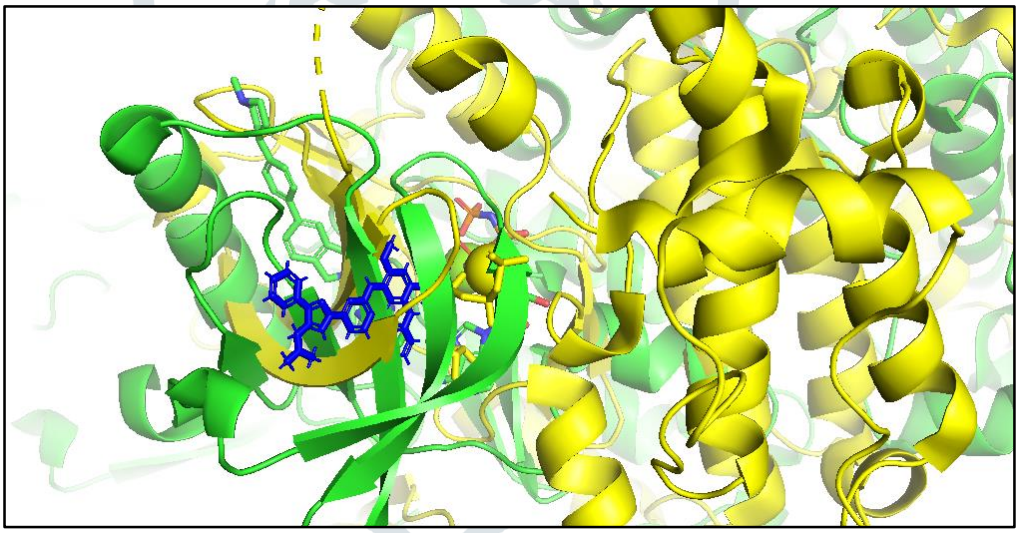

Figure 8. Molecular Docking (Protein (8gb4(green) 4ZSE(yellow) ligand(blue) Interaction)

The structural visualisation represents the ligand's spatial orientation and its association with the protein-binding region, providing insights into potential protein–ligand interactions relevant to the structural analysis of PD-L1 in NSCLC as in **Figure10**.

3.8 KEGG PATHWAY ANALYSIS

The KEGG pathway analysis of NSCLC revealed a complex network of abnormally regulated molecular mechanisms that collectively drive tumor development, progression, and resistance to therapy. The pathway highlighted variations across several critical signaling cascades, including EGFR, PI3K-Akt, MAPK, JAK-STAT, and p53 signaling pathways, each contributing to distinct but interconnected cellular processes Fig.11.

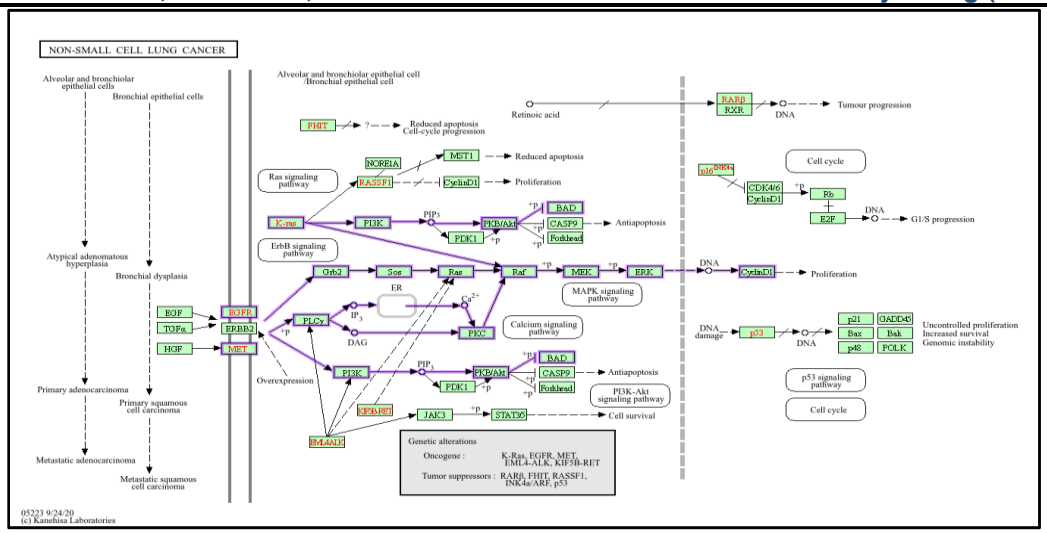

Figure 9. KEGG Pathway analysis of NSCLC

Unusual activation of receptor tyrosine kinases such as EGFR, ALK, and MET were observed, Highlighting their critical role in uncontrolled cell proliferation, survival, and metastatic potential. This aligns with clinical observations, where mutations or overexpression of these kinases are frequently detected in NSCLC patients and serve as actionable targets for tyrosine kinase inhibitors. Abnormal regulation of the PI3K-Akt and MAPK pathways further reinforced

4.CONCLUSION

This study reports an integrated computational analysis of EGFR (T790M/V948R) associated with non-small cell lung cancer (NSCLC). Sequence similarity analysis reconfirmed the high similarity to human EGFR. ERRAT and Ramachandran analysis supported the overall structural quality of the selected protein. InterProScan and PROSITE were used to retrieve information on conserved domains and functional residues. Active site identification and molecular docking provided a framework to study the interactions of lazertinib with mutant EGFR. In addition, KEGG pathway analysis showed that EGFR-related signaling pathways participated in the progression of NSCLC. To summarize, this study presents a computational basis for understanding the mutant EGFR and exploring the potential targeted therapeutic approaches, which deserves further experimental validation.

References

1. Testa, U., Castelli, G., & Pelosi, E. (2018). Lung Cancers: Molecular Characterization, Clonal Heterogeneity and Evolution, and Cancer Stem Cells. *Cancers*, 10(8), 248. https://doi.org/10.3390/cancers10080248

2. Schiller, J. H. (2001). Current standards of care in small-cell and non-small-cell lung cancer. *Oncology*, 61(Suppl. 1), 3–13. https://doi.org/10.1159/000055386

3. Yoneda, K., Imanishi, N., Ichiki, Y., & Tanaka, F. (2019). Treatment of Non-small Cell Lung Cancer with EGFR-mutations. *Journal of UOEH*, 41(2), 153–163. https://doi.org/10.7888/juoeh.41.153

4. Murtuza, A., et al. (2019). Novel Third-Generation EGFR Tyrosine Kinase Inhibitors and Strategies to Overcome Therapeutic Resistance in Lung Cancer. *Cancer Research*, 79(4), 689–698. https://doi.org/10.1158/0008-5472.CAN-18-1281

5. Shiba-Ishii, A., Takemura, N., Kawai, H., & Matsubara, D. (2024). Histological transformation of non-small-cell lung cancer in response to tyrosine kinase inhibitors. *Cancer Science*, 115(7), 2138–2146. https://doi.org/10.1111/cas.16192

6. Esposito Abate, R., et al. (2020). Next Generation Sequencing-Based Profiling of Cell Free DNA in Patients with Advanced Non-Small Cell Lung Cancer: Advantages and Pitfalls. *Cancers*, 12, 3804. https://doi.org/10.3390/cancers12123804

7. Tsui, D. W. Y., Murtaza, M., Wong, A. S. C., et al. (2018). Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in non-small cell lung cancer. *EMBO Molecular Medicine*. https://doi.org/10.15252/emmm.201707945

8. Singh, N., & Kumari, U. (2026). Mult-omics-based drug discovery pipelines incorporating machine learning, AI, and BioPython in Parkinson's disease: A case study of 5A polymorph of alpha-synuclein. *JETIR*, 13(2), b617–b634. https://doi.org/10.56975/jetir.v13i2.575492

9. Mandra, D., Kumari, U., Singh, N., & Marouthu, A. (2026). Integrated Multi-Omics and BioPython Approaches for Target Identification of α-Synuclein Polymorph in Parkinson's Disease. *IOSR Journal of Biotechnology and Biochemistry*, 12(2), 12–21. https://doi.org/10.9790/264X-1202011221

10. Singh, N., Kumari, U., Mandra, D., & Marouthu, A. (2026). Integrated Machine Learning-based Toxicity Prediction with Molecular Docking for Safer Drug Candidate Screening in Parkinson's Disease. *Computational Biology and Bioinformatics*, 14(1), 41–53. https://doi.org/10.11648/j.cbb.20261401.14

11. Burley, S. K., et al. (2017). Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods in Molecular Biology*, 1607, 627–641. https://doi.org/10.1007/978-1-4939-7000-1_26

12. Ye, J., McGinnis, S., & Madden, T. L. (2006). BLAST: Improvements for better sequence analysis. *Nucleic Acids Research*, 34, W6–W9. https://doi.org/10.1093/nar/gkl164

13. Pearson, W. R. (2013). BLAST and FASTA Similarity Searching for Multiple Sequence Alignment. *Methods in Molecular Biology*, 1079, 75–101. https://doi.org/10.1007/978-1-62703-646-7_5

14. Zdobnov, E. M., & Apweiler, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 17(9), 847–848. https://doi.org/10.1093/bioinformatics/17.9.847

15. Sigrist, C. J. A., et al. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41(Database issue), D344–D347. https://doi.org/10.1093/nar/gks1067

16. Laskowski, R. A., Thornton, J. M., Moss, D. S., & MacArthur, M. W. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. https://doi.org/10.1107/S0021889892009944

17. Li, L., Wang, B., Liu, S., et al. (2023). An Updated Review on Developing Small Molecule Kinase Inhibitors Using Computer-Aided Drug Design Approaches. *International Journal of Molecular Sciences*, 24, 13953. https://doi.org/10.3390/ijms241813953

18. Heppner, D. E., et al. (2022). Structural Basis for Inhibition of Mutant EGFR with Lazertinib (YH25448). *ACS Medicinal Chemistry Letters*, 13(12), 1856–1863. https://doi.org/10.1021/acsmedchemlett.2c00213

19. Beyett, T. S., & Eck, M. J. (2024). EGFR(T790M/V948R) kinase in complex with benzimidazole allosteric inhibitor. Protein Data Bank, 8GB4. https://doi.org/10.2210/pdb8GB4/pdb

20. Yan, X. E., & Yun, C. H. (2016). Crystal structure of EGFR 696–1022 T790M/V948R, crystal form II. Protein Data Bank, 4ZSE. https://doi.org/10.2210/pdb4ZSE/pdb

21. Pham, C. D., & Heppner, D. E. (2022). EGFR(T790M/V948R) in complex with Lazertinib (YH25448). Protein Data Bank, 7UKW. https://doi.org/10.2210/pdb7UKW/pdb

22. Li L, Wang B, Liu S, et al (2023) An Updated Review on Developing Small Molecule Kinase Inhibitors Using Computer-Aided Drug Design Approaches. IJMS 24:13953. https://doi.org/10.3390/ijms241813953

23. Iverson DB, Eisenmesser EZ, Ahn NG, et al (2020) Activation Loop Dynamics Are Coupled to Core Motions in Extracellular Signal-Regulated Kinase-2. Biochemistry 59:2698–2706. https://doi.org/10.1021/acs.biochem.0c00485

24. Cheng M, R Ott G (2010) Anaplastic Lymphoma Kinase as a Therapeutic Target in Anaplastic Large Cell Lymphoma, Non-Small Cell Lung Cancer and Neuroblastoma. ACAMC 10:236–249. https://doi.org/10.2174/1871520611009030236

25. 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

26. Burley, S. K., Nakamura, H., Markley, J. L., Berman, H. M., Velankar, S., & Kleywegt, G. J. (2017). Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods in Molecular Biology (Clifton, N.J.)*, *1607*, 627–641. https://doi.org/10.1007/978-1-4939-7000-1_26

27. Noguchi T (2003) PDB-REPRDB: a database of representative protein chains from the Protein Data Bank (PDB) in 2003. Nucleic Acids Research 31:492–493. https://doi.org/10.1093/nar/gkg022

28. Madej, T., Addess, K. J., Fong, J. H., Geer, L. Y., Geer, R. C., Lanczycki, C. J., ... & Bryant, S. H. (2012). MMDB: 3D structures and macromolecular interactions. *Nucleic acids research*, *40*(D1), D461-D464. https://doi.org/10.1093/nar/gkr1162

29. Wang Y, Addess KJ, Chen J, et al (2006) MMDB: annotating protein sequences with Entrez's 3D-structure database. Nucleic Acids Research 35:D298–D300. https://doi.org/10.1093/nar/gkl952

30. Chen J, Anderson JB, Deweese-Scott C, et al (2003) MMDB: Entrez's 3D-structure database. Nucleic acids research 31:474–477. https://doi.org/10.1093/nar/gkg086

31. Ye J, Mcginnis S, Madden TL (2006) BLAST: improvements for better sequence analysis. Nucleic Acids Research 34:W6–W9. https://doi.org/10.1093/nar/gkl164

32. Albà MM, Laskowski RA, Hancock JM (2002) Detecting cryptically simple protein sequences using the SIMPLE algorithm. Bioinformatics 18:672–678. https://doi.org/10.1093/bioinformatics/18.5.672

33. Mcginnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Research 32:W20–W25. https://doi.org/10.1093/nar/gkh435

34. Sarma, R., Verma, A., Chetia, H., & Sharma, D. K. (2013). 3D structure prediction of aging related proteins of Silurana tropicalis Gray, 1864. *journal of pharmacy research*, *7*(8), 762-765. https://doi.org/10.1016/j.jopr.2013.08.019

35. Singh, S. O. N. I., & Jha, A. L. O. K. (2018). Structural characterisation of 5-hydroxytryptamine2a receptor in homo sapiens by in-silico method. *Asian Journal of Pharmaceutical and Clinical Research*, *11*, 81-85. http://dx.doi.org/10.22159/ajpcr.2018.v11s2.28588

36. Bajaj, T., & Kumari, U. (2025). Ngs Data Analysis And Active Site Identification of Alpha-1-Acid Glycoprotein Bound To Potent Anti Tumor Compound UCN-01 In Malignant Brain Tumor. http://doi.one/10.1729/Journal.44259

37. Zdobnov, E. M., & Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*(9), 847-848. https://doi.org/10.1093/bioinformatics/17.9.847

38. Kelly, R. J., Vincent, D. E., & Friedberg, I. (2010). IPRStats: visualization of the functional potential of an InterProScan run. *BMC Bioinformatics*, *11*(Suppl 12). https://doi.org/10.1186/1471-2105-11-s12-s13

39. Laskowski, R. A., Thornton, J. M., Moss, D. S., & Macarthur, M. W. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/10.1107/s0021889892009944

40. Kumari, U., & Tripathi, K. (2023). Computational Analysis and Molecular Docking Approach for Liver Cancer. *International Journal for Research in Applied Science and Engineering Technology*, *11*, 1828. https://doi.org/10.22214/ijraset.2023.54927\

41. Rosignoli, S., & Paiardini, A. (2022). Boosting the full potential of PyMOL with structural biology plugins. *Biomolecules*, *12*(12),1764 https://doi.org/10.3390/biom12121764

42. Kumari, U., & Bajaj, T. (2025). NGS data analysis and active site identification of alpha-1-acid glycoprotein bound to potent antitumor compound UCN-01 in malignant brain tumor. International Journal of Emerging Technologies and Innovative Research, 12(3), g107–g116. https://doi.org/10.1729/Journal.44259

43. Pearson, W. R. (2013). *BLAST and FASTA Similarity Searching for Multiple Sequence Alignment* (Vol. 1079, pp. 75–101). Humana. https://doi.org/10.1007/978-1-62703-646-7_5

44. Mount, D. W. (2007). Using the Basic Local Alignment Search Tool (BLAST). *Cold Spring Harbor Protocols*, *2007*(7), pdb.top17. https://doi.org/10.1101/pdb.top17

45. Samal, K. C., Sahoo, J. P., Behera, L., & Dash, T. (2021). Understanding the BLAST (Basic Local Alignment Search Tool) Program and a Step-by-step Guide for its use in Life Science Research. *Bhartiya Krishi Anusandhan Patrika*, *Of*. https://doi.org/10.18805/BKAP283

46. 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

47. Gotoh, O. (1995). Accuracy of multiple sequence alignments as assessed by reference to structural alignments. *Genome Informatics*, *6*(0), 124–125. https://doi.org/10.11234/gi1990.6.124
48. Collins, J. F., Coulson, A. F. W., & Lyall, A. (1988). The significance of protein sequence similarities. *Bioinformatics*, *4*(1), 67–71. https://doi.org/10.1093/bioinformatics/4.1.67
49. Mott, R. (2005). Alignment: Statistical Significance. In *Encyclopedia of Life Sciences*. Wiley. https://doi.org/10.1038/npg.els.0005264
50. Agrawal, A., Brendel, V., & Huang, X. (2016). *Pairwise Statistical Significance Versus Database Statistical Significance for Local Alignment of Protein Sequences* (pp. 50–61). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79450-9_6
51. Swain, M. T., & Vickers, M. (2022). Interpreting alignment-free sequence comparison: what makes a score a good score? *NAR Genomics and Bioinformatics*, *4*(3), lqac062. https://doi.org/10.1093/nargab/lqac062
52. Mirza, Z., Schulten, H. J., Farsi, H. M., Al-Maghrabi, J. A., Gari, M. A., Chaudhary, A. G., & Karim, S. (2015). Molecular interaction of a kinase inhibitor midostaurin with anticancer drug targets, S100A8 and EGFR: transcriptional profiling and molecular docking study for kidney cancer therapeutics. *PLoS One*, *10*(3), e0119765. https://doi:10.1371/journal.pone.0119765.t004