© August 2024 | IJIRT | Volume 11 Issue 3 | ISSN: 2349-6002

Molecular Docking and Dynamic Studies Human Anaplastic Lymphoma Kinase Domain

Sambit Majumdar[1], Sudarshan Gopinath[1] Uma Kumari[1*]
[1]*Trainee at Bioinformatics Project and Research Insitute, Noida - 201301, India*
[1]*Trainee at Bioinformatics Project and Research Insitute, Noida - 201301, India*
[1*]*Senior Bioinformatics Scientist, Bioinformatics Project And Research Insitute, Noida - 201301, India*
Corresponding Author Uma kumari (uma27910@gmail.com)

***Abstract*- Lung cancer is one of the most common diseases in the world today. Early diagnosis of these conditions will facilitate more effective treatment plans for the patient. The 7R7R gene in ALK-positive non-small cell lung cancer is the subject of this study. In order to examine the activity and similarities of the 7R7R gene within the ALK gene in ALK-positive non-small cell lung cancer, the study currently offers a structure-based drug design strategy. Human ALK-positive cancer generally starts as lung cancer, but it can also start in the brain or breast, among many other regions of the body. About 5% of lung cancer patients also have ALK-positive lung cancer. As compared to lung cancer patients overall, approximately half of those with ALK-positive lung cancer receive a diagnosis before the age of 50. Many of these patients are in their 30s and 40s, but some are even in their teens and twenties. In this paper, we have used ERRAT tool to validate our structure. Also BioPython has been used to simplify our data and finally, the molecular docking study, to observe strong binding affinity of the active site of human anaplastic lymphoma kinase with a docking score.**

***Keywords*-Human Anaplastic Lymphoma Kinase, Structure analysis, Homology modelling, Biopython, CB-Dock, H-Dock**

I. INTRODUCTION

Cancer is a generic term for a large group of diseases that can affect any part of the body that is characterized by uncontrolled cell growth and proliferation. It is considered as one of the most common diseases in the world. It remains to be a leading cause of mortality worldwide and despite drastic advancements in the medical field, it proves to be a challenge to medical practitioners and patients suffering from this disease. Global statistics show that cancer claims the lives of almost 7 million people . India ranked third after China and the United States of America. It has been noted that the majority of males suffer from cancers of the prostate, lungs, bronchus, colon, rectum, bladder etc, On the other hand cancers of the breast, lungs and bronchus, colon and rectum, uterine corpus, thyroid etc. can affect woman. Gene mutation and alterations in the cellular processes are the causes of it. There are numerous chemicals found in the environment have cancer causing qualities [1,2,3,4].

Anaplastic lymphoma kinase (ALK) is a transmembrane receptor tyrosine kinase that belongs to the insulin receptor superfamily. It is made by the anaplastic lymphoma kinase (ALK) gene, which may be changed in some types of cancer, such as anaplastic large cell lymphoma (ALCL), neuroblastoma, and non-small cell lunger cancer (NSCLC). Irregular activation of ALK, often through gene rearrangements or mutations, has been implicated in the pathogenesis of several cancers, including ALK-positive NSCLC, anaplastic large cell lymphoma (ALCL), and neuroblastoma. The discovery of ALK rearrangements in NSCLC has led to the development of targeted therapies designed to inhibit the aberrant activity of the ALK fusion protein. The incidence of ALK-positive NSCLC is estimated to be around 7% of all NSCLC cases, translating to tens of thousands of new cases worldwide each year. It has been identified that lorlatinib, a tyrosine kinase inhibitor, is used in a variety of therapy for the majority of patients with anaplastic lymphoma kinase positive lung cancer. The overall survival of patients with advanced ALK-positive NSCLC remains limited, highlighting the need for continued research and development of new therapeutic strategies [5,6,7,8].

CADD (Computer-aided drug design) may facilitate the manufacturing process by transferring detailed diagrams of a products materials, processes, tolerances and dimensions with specific conventions for the product in question. It is an in-silico approach to drug development that simulates and optimizes chemical compounds, analyzes molecular processes, and creates new compounds. Drug discovery becomes more effective with these methods. These methods include molecular docking, QSAR modelling, and molecular dynamics simulations. In the case of lung cancer patients with ALK gene alterations, CADD has been essential in identifying and enhancing ALK inhibitors. By constructing a 3-dimensional model of the ALK kinase and analysing its interactions

with potential drugs, scientists have scientists have developed medications that precisely target the kinase's active site, leading to increased effectiveness and reduced side effects. Drugs relevant to cancer are often designed using two processes: ligand-based drug design and structure-based drug design. Drug design becomes more challenging when dealing with complicated structures or pathways because of the increased efficiency of the procedures. CADD is not only limited to the development of targeted therapies but is also being increasingly applied in the design of combination therapies, immune-therapeutics, and novel drug delivery systems. As the field of oncology continues to evolve, CADD is expected to play an increasingly central role in the development of new cancer treatments, offering the potential for more personalized and effective therapeutic options [9,10,11,12].

II. METHOD AND MATERIAL

The Protein Data Bank (PDB) is a US data center for global protein. It consists of a 3D structure of large data molecules, such as DNA, RNA, and proteins, used for research and education. It also contains the data of some of the experimental techniques such as X-ray crystallography, NMR spectroscopy and cryo-electron microscopy. It is freely available data and can be visualized by many free software programs such as PyMOL, RasMol, Chimera, and VMD. The structure of the 'Human Anaplastic Lymphoma Kinase Domain' PDB ID (7R7R) was taken from this database.

PyMOl which is an integrated python interpreted open-source license software. It is been used to visualize the structure by various angle. To validate our structure the ERRATof SAVESv6.1 server is used.

The BLAST software (Basic Local Alignment Search Tool) is a bioinformatics tool that is used to compare the nucleotide or protein sequences to databases based on the functional, structural or evolutionary relationships between sequences. It is used to predict the homologs of the structure using percentage identical. Using those homologs RMSD (Root Mean Square Deviation) value is predicted by aligning the two structures in PyMOL software. Bio Python was used to plot the distance matrix and the residue position graph. It is a software for making the work easier for scientist and researchers by using computational biology libraries that provides tools for reading, writing, and analyzing biological data. It makes numerous bioinformatics tasks, including data visualization, machine learning, phylogenic tree analysis, structural biology, and sequence analysis, easier. For predicting the binding affinity with the homologous structure H-Dock docking tool is used. It is a hybrid docking technique to predict binding complexes between two molecules. Certain evaluation measures are employed in the interpretation of H-Dock results, including docking score, confidence score, ligand RMSD and interface residues [13,14,1.

For predicting the binding affinities of various drugs which could be potentially be used for the treatment of non-small cell lung cancer CB-Dock 2 by Cao Lab is used. It is a protein-ligand docking tool that uses AutoDockVina to execute molecular docking after automatically determining the binding sites by blind docking; it also calculates the size and center and customizes the docking box size based on the query ligands. Desired drugs were used for docking are crizotinib, ceritinib, alectinib, brigatinib, Entrectinib and lorlatinib [18].

III. RESULT AND DISCUSSION

For estimation of the quality of the structure ERRAT server is used. Its result displays the distribution of amino acids in both error and non-error regions. Misfolded regions, represented by black bars, are situated distantly from the active site. Errorregions, depicted by grey bars, fall between the 95% and 99% confidence levels. White bars represent regions with a lower error rate in protein folding. Yellow regions on the structure can be rejected at the 95% confidence level, as it is expected that 5% of a well-constructed protein structure may have an error value surpassing this threshold. The red regions can be rejected at the 99% confidence level.

The overall quality factor of our sample is **96.2963** which shows that our structure quality is perfect.

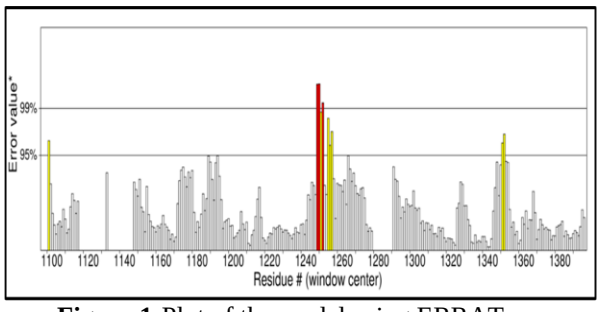

Figure 1: Plot of the model using ERRAT

RMSD Value was also predicted which gives the average deviation between two proteins. The smaller the RMSD Value, the more similar the structures.The surface representation setting available on PyMOL to focus while using the default ribbon display. After aligning both Chain A (7R7R) and Chain B (2XP2), the RMSD Value is 0.248 and after aligning both Chain A (7R7R) and Chain B (5A9U) the RMSD Value is 0.203(Figure 2). On aligning the chain 2XP2 and 5A9U, it can be interpreted that they are having high degree of structural similarity.

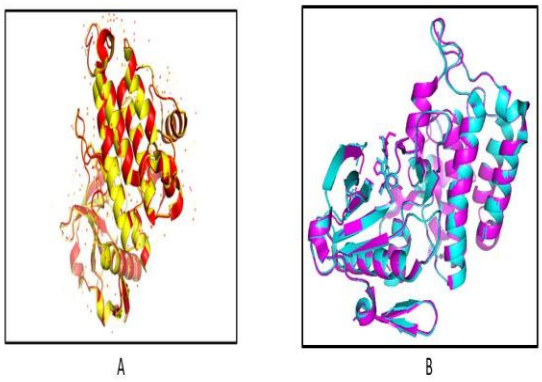

Figure 2:RMSD Value of**A.**7R7R with 2XP2; **B.**7R7R with 5A9U

Bio python which is used to interpret numerous bioinformatics tasks, including data visualization, machine learning, phylogenic tree analysis, structural biology, and sequence analysis, easier. Using this we have plotted the distance matrix which is a mathematical representation used in Machine Learning to measure the similarity or dissimilarity between objects or data points. In, figure 6 explains about the interaction between the amino acids within the protein of 7R7R.

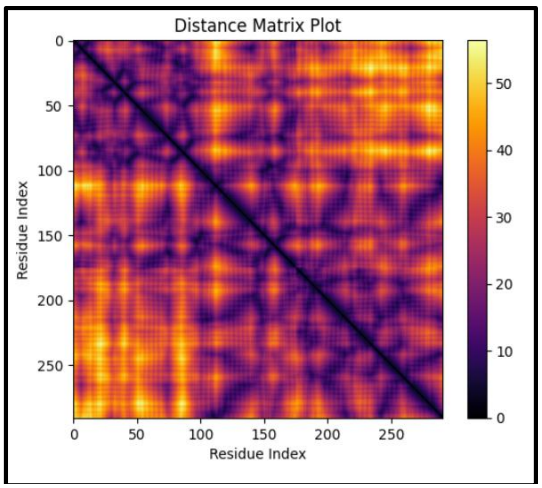

Figure 3: Distance Matrix Plot of 7R7R

The molecular docking analysis on CB-Dock2 to check the binding affinities of various drugs which could be potentially used for the treatment of non-small cell lung cancer.Negative binding energy is better than positive binding energy. The higher negativity of the binding the more stable the protein-ligand complex.Effective drugs used for docking are alectinib, ceritinib and crizotinib(Table 1).Drugalectinibis used for the treatment of the non-small cell lung cancer (NSCLC), it is kinase inhibitor drug.It was observed the docking score of alectinibis -8.9 which indicate the strong docking between the molecules(Figure 3).For alectinib contact residues are LEU1118 LEU1122 HIS1124 GLY1128 GLU1129 VAL1130 TYR1131 ALA1148 LYS1150 THR1151 LEU1152 PRO1153 GLU1154 VAL1155 CYS1156 ASP1160 ASP1163 PHE1164 GLU1167 PRO1191 PHE1193 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1202 ASP1203 ASP1249 ARG1253 ASN1254 LEU1256 GLY1269 ASP1270 GLY1272 MET1273 ARG1275 ASP1276 GLY1286 ALA1289 MET1290. Ceritinib is used to treat metastatic (cancer that has already spread) non-small cell lung cancer in patients who have a certain type of abnormal anaplastic lymphoma kinase (ALK) gene.It was observed the docking score of ceritinibis -8.8 which indicate the strong docking between the molecules(Figure 4).For Ceritinibcontact residues are ARG1120 GLY1121 LEU1122 GLY1123 HIS1124 VAL1130 GLU1132 ALA1148 LYS1150 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 SER1206 PHE1207 GLU1210 ARG1253 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270. Crizotinib is a receptor tyrosine kinase inhibitor used to treat metastatic non-small cell lung cancer (NSCLC) where the tumours have been confirmed to be anaplastic lymphoma kinase (ALK), or ROS1-positive.It was observed the docking score of crizotinib is -8.3 which indicate the strong docking between the molecules(Figure 5).For Crizotinib contact residues are ARG1120 LEU1122 GLY1123 HIS1124 VAL1130 GLU1132 ALA1148 LYS1150 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 SER1206 PHE1207 GLU1210 ARG1253 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270.

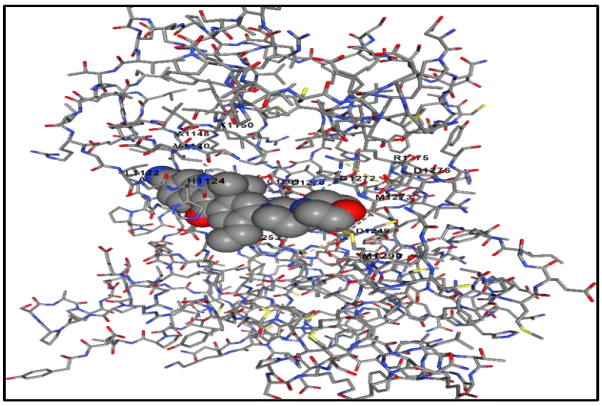

Figure 4: Docking by CB Dock 2 position of Alectinibdrug in 7R7R structure

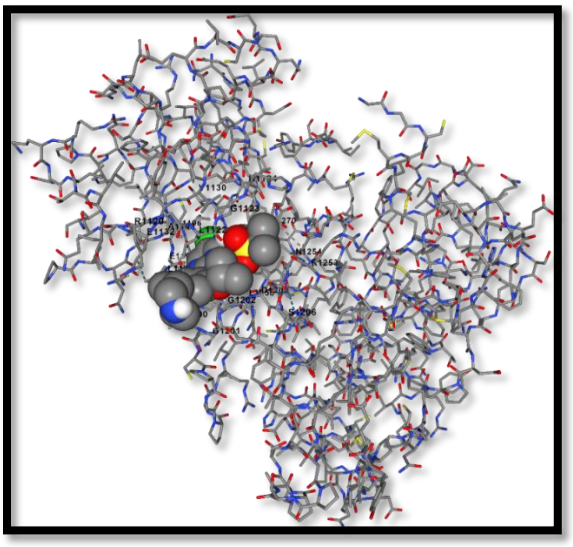

Figure 5: Docking by CB Dock 2 position of Ceritinibdrug in 7R7R structure

Table 1: Molecular docking of the effective drugs

Drugs	CurPocket Id	Vina Score	Cavity Volume (Å^3)	Center			Docking Size		
				X	Y	Z	X	Y	Z
Alectinib	C4	-8.9	166	39	37	3	26	26	26
Ceritinib	C2	-8.8	736	31	44	9	26	26	26
Crizotinib	C2	-8.3	736	31	44	9	23	23	23

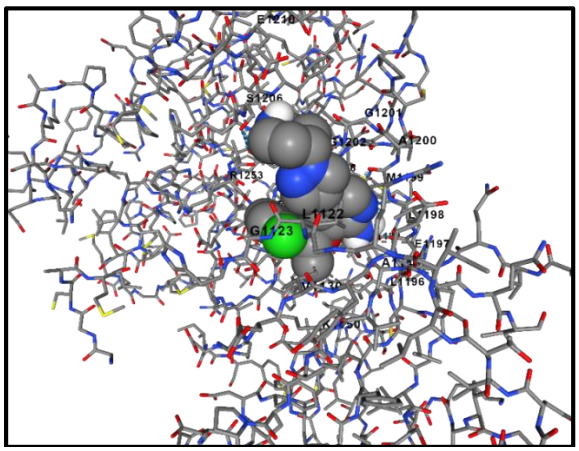

Figure 6: Docking by CB Dock 2 position of Crizotinibdrug in 7R7R structure

For studying the protein- protein interaction, H-Dock is used which is also known as hybrid docking. This server is used to predict the interaction of the two proteins. H-Dock shows the protein interaction, where one is taken as a ligand and one as a receptor to identify the disease mechanism Model 0 (left) is predicted by template-based homology modeling. The built model of "Model 0" based on the above PDB complex template has a high confidence(Figure 7).Model 0 and Model 3 overlap each other and have high similarity and confidence(Figure 8). It represents the docking score, confidence score and RMSD calculations for predicting the protein-protein interaction (Table 2).

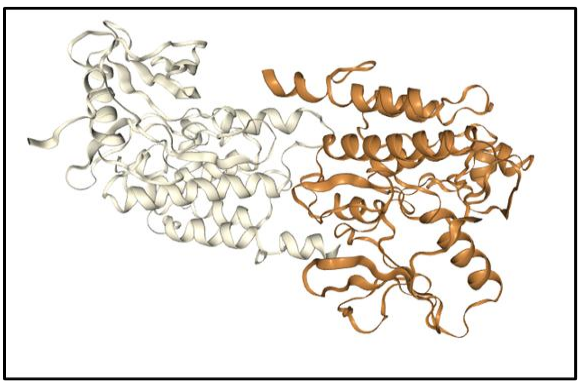

Figure 7: H-Dock Protein- Protein Interaction

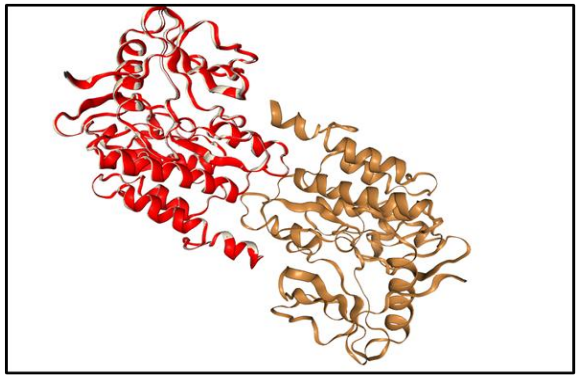

Figure 8: Model 3 overlapping Model 0

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-250.90	-241.92	-238.88	-236.18	-234.95	-233.63	-233.42	-230.39	-277.78	-221.91
Confidence Score	0.8827	0.8628	0.8554	0.8486	0.8454	0.8419	0.8414	0.8331	0.8257	0.8082
Ligand RMSD (Å)	82.33	38.70	0.40	43.45	66.38	56.86	75.41	42.14	55.69	41.95

Table 2: Representing the docking score, confidence score and RMSD calculations. To predict protein protein interaction studies

IV. CONCLUSION

Based on the molecular docking study, we observe strong binding affinity to the active site of human anaplastic lymphoma kinase with a docking score (-8.9) showing a strong binding affinity for a biological sample. In protein-protein interactions (involving 5A9U and 2XP2), with the help of computational analysis, we observe the same kind of structural similarity where Root Mean Square Deviation (RMSD) value near to zero represent identical similarity between the structures. In the ligand, a superior docking result compared to new ALK inhibitors is shown, considering it a promising candidate for further treatment strategy for cancer patients. The strong and specific interactions with a key residue in the binding site, particularly involved in the ALK catalytic activity, suggest that the ligand could effectively inhibit the ALK.

Bio-python Drug Designing (for analysis in proteomics) and its ability to handle large data set with various bioinformatics tools and provide customizable workflow helps researchers and aids in integration with an external database simplifying the experimental proteomics data with reference sequence.

Using the Save Server protein analysis, we performed a critical assessment of the quality and accuracy of the protein model to identify higher quality of protein which is observed.

REFERENCE

[1]. Grover, A., Grover, M., & Sharma, K. (2016). A practical overview of quantitative structure-activity relationship. World J. Pharm. Pharm. Sci, 5, 427-437.

[2]. Yl, C. (2008). Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. Cancer Res, 68, 4971-4976.

[3]. Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M.& Bray, F. (2020). Global cancer observatory: cancer today. Lyon: International agency for research on cancer, 20182020.

[4]. Kwak, E. L., Bang, Y. J., Camidge, D. R., Shaw, A. T., Solomon, B., Maki, R. G.,&Iafrate, A. J. (2010). Anaplastic lymphoma kinase inhibition in non–small-cell lung cancer. *New England Journal of Medicine*, *363*(18), 1693-1703.

[5]. Miller, K. D., Siegel, R. L., Lin, C. C., Mariotto, A. B., Kramer, J. L., Rowland, J. H., ... &Jemal, A. (2016). Cancer treatment and survivorship statistics, 2016. *CA: a cancer journal for clinicians*, *66*(4), 271-289.

[6]. Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L., & Look, A. T. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science, 263(5151), 1281-1284.

[7]. Naseer, R., Cheema, N. H., &Ikram, Q. (2023). Multiple Malignancies in Relation to Vitamin D Levels.

[8]. Opeyemi, I., Paul, O., Olawale, F., Olorunfemi, B., Elekofehinti, O. O., Kikiowo, B., ...&Shityakov, S. (2023). Computer-aided drug design in anti-cancer drug discovery: what have we learnt and what is the way forward??. *Informatics in Medicine Unlocked*, 101332.

[9]. Poon, S. L., McPherson, J. R., Tan, P., Teh, B. T., &Rozen, S. G. (2014). Mutation signatures of carcinogen exposure: genome-wide detection and new opportunities for cancer prevention. Genome medicine, 6, 1-14.

[10]. Kore, P. P., Mutha, M. M., Antre, R. V., Oswal, R. J., &Kshirsagar, S. S. (2012). Computer-aided drug design: an innovative tool for modeling.

[11]. Shiba-Ishii, A., Johnson, T. W., Dagogo-Jack, I., Mino-Kenudson, M., Johnson, T. R., Wei, P., ...&Hata, A. N. (2022). Analysis of lorlatinib analogs reveals a roadmap for targeting diverse compound resistance mutations in ALK-positive lung cancer. Nature cancer, 3(6), 710-722.

[12]. Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., ...& Mano, H. (2007). Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. Nature, 448(7153), 561-566.

[13]. Tang, Y., Zhu, W., Chen, K., & Jiang, H. (2006). New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. Drug discovery today: technologies, 3(3), 307-313.

[14]. Yang, X., Liu, Y., Gan, J., Xiao, Z. X., & Cao, Y. (2022). FitDock: protein–ligand docking by template fitting. Briefings in bioinformatics, 23(3), bbac087.

[15]. Colovos, C., &Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*, *2*(9), 1511-1519.

[16]. Kukreja, Vinita &Kumari, Uma. (2023). Data Analysis of Brain Cancer with Biopython. 8. 2146-2154

[17]. 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.

[18]. Liu, Y., Yang, X., Gan, J., Chen, S., Xiao, Z. X., & Cao, Y. (2022). CB-Dock2: Improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. Nucleic acids research, 50(W1), W159-W164.