Design of ALK Inhibitors for Non-Small Cell Lung Cancer – A Fragment Based Approach

Anika Jain, Nivya James, Veerapapillai Shanthi, Karuppasamy Ramanathan*

Department of Biotechnology, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, INDIA.

ABSTRACT

Background: The increasing cases of non-small cell lung cancer pose a relentless threat to human health. Therefore, using fragment-based drug discovery as the modus operandi, the present study aimed to design small molecule inhibitors for Anaplastic Lymphoma Kinase (ALK) positive Non-Small Cell Lung Cancer (NSCLC) and for its secondary mutation (F1174L). **Materials and methods:** A total of 12 ALK inhibitors (both FDA and clinical) reported in the literature was utilised in the present study to design the novel and potent ALK inhibitors. Fragment script and BREED of Schrödinger suite was used to generate novel structural combinations. GLIDE algorithm was employed to identify the fragments with high binding affinity to the native and mutant forms of ALK. Screened out fragments were subjected to ADMET analysis. Additionally, the impact of F1174L mutation in the stability of ALK protein was assessed using SDM and I-Mutant tools. RING algorithm was employed to compare the pattern of intramolecular interaction between the native and F1174L ALK protein. **Results:** A novel hybrid molecule LF16 was identified with better binding, superior CNS activity, higher drug score and low toxicity. Annihilation of intramolecular interactions with Ile 1179, Ile 1170 and Phe1098 were found to be the cause of the destabilizing effect of the mutant protein. Of note, these residues were found to play important roles in the formation of F-core as well as in stabilizing αC-helix, which is important for maintaining the inactive form of ALK. Moreover, **Conclusion:** We believe that these results could foster the designing and development of novel ALK inhibitors towards the management of crizotinib resistance in the near future.

Key words: ALK, SDM, I-Mutant, RIN, GLIDE, ADMET.

INTRODUCTION

The recent revolutionary changes in the field of molecular biology have led to the apprehension that two-thirds of Non-Small Cell Lung Cancer (NSCLC) patients harbor an oncogenic driver mutation. These driver mutations were found to be therapeutically targetable, leading to an improved survival and safety profiles when compared to chemotherapy. One such activating mutation found in about 3-5% of the patient population, is Anaplastic Lymphoma Kinase (ALK) mutations.[1] ALK-positive NSCLCs were discovered in 2007 and is most often seen in non-smoker[1]. The mutation in ALK was found to be due to an oncogenic inversion and fusion of two genes namely; ALK and EML4 (echinoderm microtubule-associated protein-like 4) found on chromosome 2p.[2] This fusion oncogene codes for a protein tyrosine kinase that has cancerous property, making it a diagnostic molecular marker as well as a therapeutic target.[3] The currently available standard of care for patients with ALK involve the treatment with ALK Tyrosine Kinase (TK) inhibitors. Even though TKI treatment has led to unprecedented progression-free survival in patients, these inhibitors are struggling to prevail over secondary mutations and other pharmacological resistances occurring within the patients.[4] Of note, F1174L mutation is one of the more frequent mutations and is resistant to crizotinib in both *in vitro* and *in vivo* studies.[5] Importantly, F1174L

mutation is considered to be one of the most recurrent mutations in ALK-positive NSCLC. It has also been found to be a common ALK mutation in Neuroblastoma. Most importantly, F1174L mutation considered to be a lethal variant amongst crizotinib resistant mutations in ALK protein reported in the literature.[6-8] Therefore, there is an immense need to surmount this resistance mutation in ALK-positive NSCLC patients.

Computer-Aided Drug Designing (CADD) methods are increasing in popularity due to their reliability, accuracy and quickness as compared to traditional drug designing.[9,10] For instance, some clinically approved drugs discovered by CADD methods include Saquinavir in 1995: Human Immunodeficiency Virus (HIV) inhibitor, Zanamivir in 1999: Neuraminidase inhibitor for treating influenza, Raltegravir in 2007: Human Immunodeficiency Virus (HIV) inhibitor. Therefore, in the present study, fragment-based drug designing was applied to identify new combinations of the available drug molecules that are more active with both native and F1174L mutant proteins.

MATERIALS AND METHODS

Dataset preparation

The 3D structures of the native ALK protein and its mutated form (F1174L) was retrieved from Protein Data Bank (PDB).[11] The PDB codes corresponding to the native and mutant forms were 2XP2 and 2YJR respectively with a resolution of 1.9 Å. Importantly, these structures having R_{free} values of <0.30 depicts that absence of any structural defects in the retrieved proteins. Subsequently, the 3D conformers of 12 ALK inhibitors (FDA approved as well as clinical trials) namely; Alectinib, Brigatinib, TAE-684, AZD-3463, ASP-3026, CEP-28122, CEP 37440, Ceritinib, Crizotinib, Entrectinib, Lorlatinib and X-396[12] were retrieved from PubChem in SDF file format and was taken for the fragment generation process.

Prior to fragment generation, the molecules were cleaned using LigPrep module of Schrodinger Suite.[13] LigPrep performs 2D to 3D conversion, corrects bond orders, assigns proper ionization states using Epik (at pH 7±2) and generates most probable conformers of the molecules.[13] Finally keeping all the other options as default, the inhibitors were energy minimized using OPLS (Optimized Potentials for Liquid Simulations) - 2005 force field.[13] Figure 1 provides a representation of the steps followed in this present perspective.

Protein preparation, grid generation and molecular docking

The native and mutant ALK structures were prepared using Protein Preparation Wizard of Schrödinger inter-
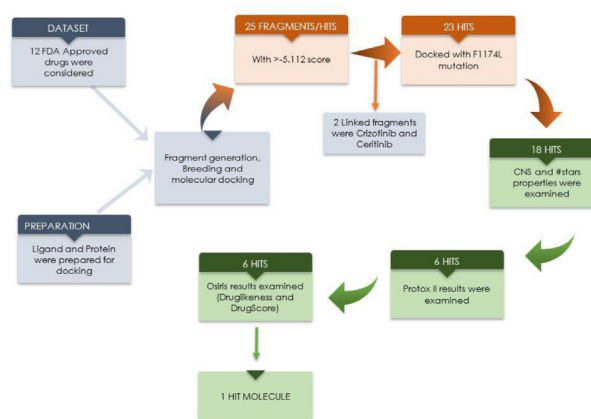
Figure 1: Workflow representation depicting the steps followed in the design of the hit molecule, LF16.

face.[13] During pre-processing, hydrogen atoms were added, disulfide bonds were created, bond orders for amino acid residues were assigned and crystallized waters were removed from the protein structures.[13] Finally, the protein structures underwent geometry refinement with restrained energy minimization and optimization processes using Impact Refinement (impref) module and OPLS-2005 force field with a default Root Mean Square Deviation (RMSD) cut off of 0.3 Å.[13] Subsequently, a grid was generated around the crystallized crizotinib, for the native ALK protein. On the other hand for the mutant ALK protein, its active site was predicted using the Site Map module of Schrödinger interface.[14] During sitemap analysis, energetically favorable binding sites were determined and were ranked based on their site score.[14] Finally, a grid was generated around the binding site having the highest site score. Further using the generated grids, molecular docking protocols were carried out in Extra Precision (XP) mode. Glide XP mode is recommended for small ligand libraries, where elimination of false positives is accomplished by a more extensive sampling and advanced scoring function.[15]

Fragment generation and breeding

Fragment generation and breeding processes were performed with the known ALK inhibitors using the Schrödinger suite. Fragment-based drug design involves fragmenting a drug or small molecule into smaller pieces and then developing or identifying new leads by sequentially combining these molecules. The principle behind this method is that combining different fragments will produce a novel molecule that has a binding affinity equal to the sum of individual interactions of each fragment.[16] Therefore, in the present study, the fragments were generated from the prepared ALK inhibitors using the fragment.py script available in the Maestro interface. The generated fragments were

then used as input to "BREED" option available in Schrödinger suite. This option will create novel hybrid molecules by finding overlapping bonds from all possible pairs of selected fragments and swapping the two sides of the fragments.[17]

Protein Structural Stability Analysis

The effect of F1174L mutation on the stability of ALK protein was primarily investigated using I-Mutant 2.0[18] and SDM[19] computational algorithms. In particular, the effects of mutation on protein stability can be measured by free energy stability change ($\Delta\Delta G$)[20] using the following equation:

$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wild}$, where ΔG is the protein unfolding energy.

A mutation usually destabilizes the protein by reducing the free energy gap between folded and unfolded states. In general, a $\Delta\Delta G$ below zero means that the mutation causes destabilization; otherwise, it induces stabilization.[20] Therefore, I-Mutant 2.0 was utilized for analyzing the $\Delta\Delta G$ values for F1174L mutation. It can predict the stability change of the mutated protein structure with the help of a neural network-based algorithm.[18] The results obtained from I-Mutant 2.0 is further validated using the SDM2 tool which functions in a similar method as the former.[19]

Further a Residue Interaction Network (RIN) was generated using RING server to analyze the interactions of wild and mutant residues in the ALK protein.[21] Initially, RINs were generated for both the wild type and mutant complexes. Subsequently, a comparison network was generated by the superimposition of the two RINs using RINalyzer plugin of Cytoscape. Of note, an amino acid in a protein structure is depicted as the node and the interactions between the nodes are represented as edges. These nodes and edges of both native and mutant protein were compared to find the differences and similarities between native and F1174L ALK protein.

ADMET analysis

Qikprop of Schrodinger interface was employed to predict the pharmacological and physicochemical properties of the lead molecules.[22] A total of 51 descriptors were predicted for the lead molecule and we have considered eight descriptors that were essential for eliminating the molecules that would not be an effective drug. One of the main descriptors that were taken into consideration was the Central Nervous System (CNS) activity. Since our drug molecule should be capable of penetrating the Blood-Brain Barrier (BBB) to combat the issue of brain metastases in NSCLC,[23,24] the CNS descriptor was considered. A value in the range

between -2 (inactive) to +2 (active), determines the CNS activity and the potential to cross the BBB. Further, #stars descriptor that indicates the number of property descriptors which falls outside the optimum range of values for 95% of known drugs, were also considered. Additionally, once ingested, its metabolism in the liver would depend upon its absorption. Therefore, Human Oral Absorption (HOA) descriptor was considered. The scale for HOA is as follows; 1 = low, 2 = medium and 3 = high absorption.[25] Further, Caco-2 Cell permeability (QPPCaco; Range: < 25-poor to >500-great), a key factor in governing a drug's metabolism and its access to biological membranes (gut-blood barrier) was also analyzed. Additionally, QPlogHERG descriptors used for determining the blockage of HERG channels (human cardiac K+ channels) by drugs, were calculated. Furthermore, the molecular weight (MW) (MW should be in the range of 130.0 – 725.0), QPlogPo/w (predicts the octanol/water partition coefficient - range –2.0 to 6.5) and QPlogS (Aqueous solubility- range –0.5 to 0.5) were also analyzed for the hit molecules.

Subsequently, the toxicity values for the molecules were explored using Osiris property explorer and Protox-II algorithm. Osiris program was used to estimate the possible side effects, such as mutagenic, tumorigenic, or reproductive risks of the hits (https://www.organic-chemistry.org/prog/peo/). Whereas ProTox-II algorithm was used for the prediction of their oral toxicities (LD_{50}).[26] The prediction method is based on the analysis of the similarity of compounds with known median lethal doses (LD_{50}) and also identifies toxic fragments in a molecule, thus representing a novel approach in toxicity prediction.

RESULTS

Fragmentation and docking analysis

Fragmentation of the 12 known ALK drug molecules performed using the fragment. py script. This process resulted a total of 177 fragments which were then XP docked into the native ALK protein. Prior to fragment docking, the reference ligand crizotinib was XP docked to the native protein and the respective glide score (-5.112 kcal/mol) was taken as the threshold value. Further, the 177 fragments were docked into the native protein and those fragments having lower scores (better binding affinity) than the reference ligand were eliminated. Thus, 40 fragments were selected for the consequent breeding process.

Breeding of Fragments and SMILES Retrieval

The 40 fragments were linked together using BREED program to produce novel molecules with potent ALK

S. No.	Compound	Glide Gscore *Native*	Glide Energy *Native* (kcal/mol)	Glide Gscore F1174L mutant	Glide Energy F1174L mutant (kcal/mol)
		Table 1: Docking scores of 25 hit molecules with native and mutant ALK protein.			
1	Crizotinib	-5.112	-43.376	-5.062	-42.584
2	LF1	-8.719	-59.366	-8.377	-52.012
3	LF2	-7.268	-37.723	-7.901	-48.042
4	LF3	-7.232	-42.333	-6.146	-42.584
5	LF4	-6.722	-38.336	-7.911	-33.147
6	LF5	-6.680	-39.478	-3.728	-45.514
7	LF6	-6.609	-36.859	-7.044	-40.814
8	LF7	-6.560	-37.874	-5.422	-50.169
9	LF8	-6.450	-35.519	-7.707	-46.087
10	LF9	-6.216	-43.635	-8.553	-40.053
11	LF10	-6.176	-36.258	-6.44	-46.019
12	LF11	-6.120	-42.482	-4.365	-37.049
13	LF12	-6.079	-45.734	-7.515	-47.705
14	LF13	-6.014	-50.882	-5.254	-46.785
15	LF14	-5.982	-34.714	-4.666	-31.832
16	LF15	-5.972	-35.29	-6.395	-43.463
17	LF16	-5.949	-35.609	-6.321	-53.147
18	LF17	-5.939	-41.928	-6.78	-43.578
19	LF18	-5.899	-41.137	-8.312	-46.121
20	LF19	-5.899	-41.137	-8.312	-46.121
21	LF20	-5.736	-47.974	-4.856	-41.769
22	LF21	-5.700	-41.805	-7.944	-46.278
23	LF22	-5.646	-41.286	-8.664	-48.459
24	LF23	-5.554	-48.469	-6.907	-42.927
25	LF24	-5.533	-47.172	-6.596	-44.603
26	LF25	-5.502	-56.303	-4.327	-28.714

inhibitory activity than each individual fragment and the structures. This protocol produced 58 compounds which were then XP docked into the native protein. Further analysis and elimination of hits based on docking scores, resulted in a final set of 25 hits (Table 1). These resultant compounds are named as 'LF1' through to 'LF25' (LF: Linked Fragment). Subsequently, the SMILES of the 25 hit molecules were generated using PubChem chemical structure sketcher. These SMILES were then rechecked with PubChem database[27] to see if the known drugs have been regenerated during the BREED process. The results highlights that two linked fragments (LF3 and LF5) had the same structure as crizotinib and ceritinib. Hence, these molecules were eliminated from our further analysis. The structures, fragment constituents and the smiles of the generated fragments are provided in Table S1.

Stability Analysis of Mutation and Molecular Docking with F1174L

The stability analysis of the F1174L mutation revealed it has a destabilizing effect on the structure of ALK protein (Table 2). Therefore, to obtain a holistic view of the cause of instability, a RIN was constructed using the PDB files of both native and mutant forms (Figure 2). It was observed that 1174 is interacting with 7 neighboring amino acids (Phe1271, Ile1179, Leu1240, Phe1245, Tyr1239, Ile1170, Phe 1098) in the native form. On the contrary, 1174 was able to maintain 4 interactions in the mutant structure. Note that Ile 1179, Ile 1170 and Phe1098 were not able to maintain the interaction in the mutant form of ALK. Thus, the destruction of the intramolecular interactions was found to be the cause for the destabilizing effect and resistance of F1174L mutant protein towards crizotinib.

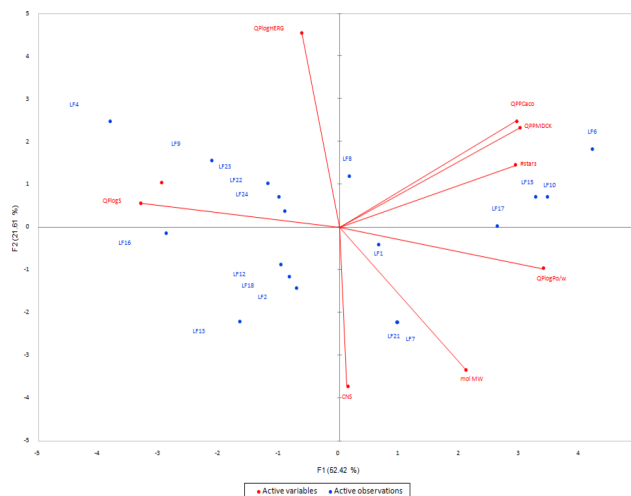| Table 2: F1174L mutant protein stability analysis using I-Mutant and SDM.				
S. No	Mutation	I-Mutant DDG value (kcal/mol)	SDM DDG value (kcal/mol)	Overall stability
1	F1174L	-2.05	-1.46	Destabilizing
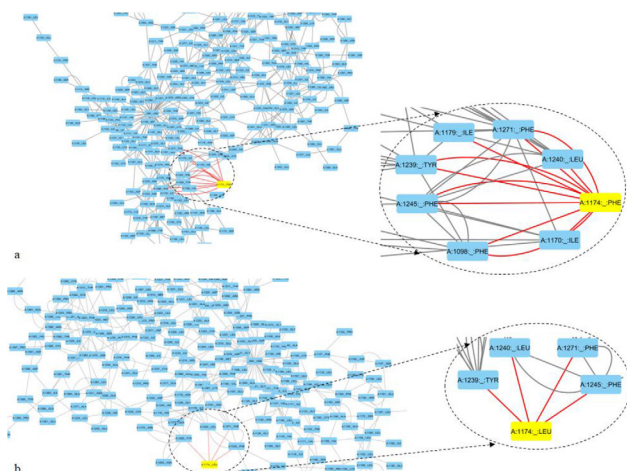
Figure 2: The RINs of both (a) Wild Type and (b) Mutant proteins of ALK. The illustrative segment shows the interactions at the 1174 position before and after mutation.

Figure 3: Principle Component Analysis (PCA) plots depicting bi-plot of the ADME properties of linked fragments.

Further to analyze the efficacy of the generated 23 LFs with the mutant protein, XP docking was carried out. It is understood from the results that 18 LFs were found to possess higher glide scores than the reference molecule (-5.062 kcal/mol). Therefore, these 18 LFs were screened out for ADMET analysis.

ADMET analysis

The hit molecules were further analyzed for their ADME properties using QikProp module of the Schrödinger interface and the results are summarized using a Principle Component Analysis (PCA) plot (Figure 3). The red color lines depict the different descriptors (active variables) and the blue dots (active observations) depicts the hit molecules. In the PCA plot, narrow angles indicate a positive correlation (higher value) and large angles depict a negative correlation (lower value). Therefore, from the plot, it can be visualized that only 6 hit molecules namely; LF2, LF7, LF13, LF16, LF18 and LF21 were found to have higher CNS and lower #stars values. Additionally, the remaining properties such as Molecular weight, QPlogPo/w, QPPCaco, QPPMDCK, QPlogH-ERG and QPlogS were also found to be in the recommended ranges for the 6 hit molecules.

Further, Protox-II server was employed to predict the oral toxicity of the 6 hit molecules and crizotinib. The results are depicted as a box plot in Figure 4a. In the plot, the edges of the box indicate upper and lower quartile and the whiskers are called upper and lower whisker. It can be observed that the plot has a longer box and a small lower whisker, demonstrating that all the hit molecules are having a higher range of LD_{50} values (Class IV and Class V range). This indicates that they are significantly non-fatal if ingested.
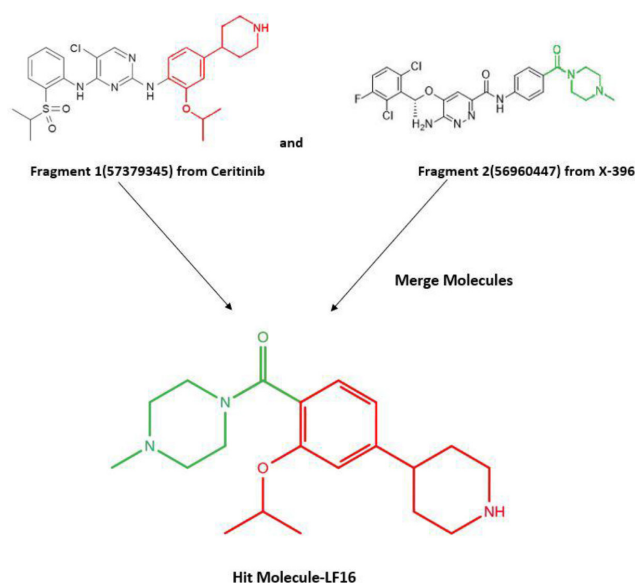
Additionally, other multiple toxicity factors such as the mutagenicity, irritant factor, tumorigenic capability, drug score and drug-likeness of the hit molecules along with crizotinib were analyzed using Osiris (Table 3). From the results, it is evident that only LF16 comes out clean without any toxic effects as well as higher drug score (Table S1 and Figure 4b). These results essentially indicate the non-toxic nature of LF16.

DISCUSSION

ALK-EML4 fusion oncogene is caused by an inversion of chromosome 2 that results in the joining of 1-13 exons of EML4 gene with 20-29 exons of ALK. The resulting fusion protein has an EML4 derived N-terminus and an ALK derived C-terminus, that contains the entire tyrosine kinase domain.[28] Targeted therapy of ALK-positive tumors with ALK tyrosine kinase inhibitors (TKI) has shown a great amount of success in the past due to its high sensitivity to these TKIs. The reference molecule used in our study, Crizotinib is the first-line treatment for ALK-positive NSCLC in many

Fragment	Mutagenic	Tumorigenic	Irritant	Reproductive Effect	Druglikeness	Toxicity Class
LF2	NIL	NIL	Yes	NIL	-14.15	Class 5
LF7	NIL	NIL	Yes	NIL	-10.7	Class 5
LF13	Yes	Yes	Yes	Yes	-6.91	Class 4
LF16	NIL	NIL	NIL	NIL	6.73	Class 4
LF18	NIL	NIL	Yes	NIL	-10.2	Class 5
LF21	NIL	NIL	Yes	NIL	-11.78	Class 5

Table 3: Toxicity analysis using Osiris Property explorer and ProTox-II webserver.
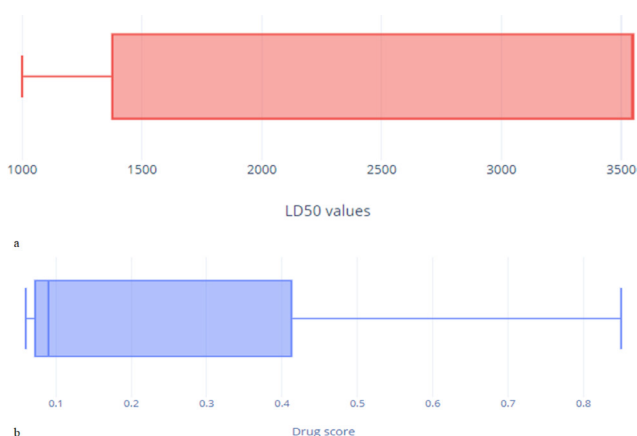
Figure 4: Box plots depicting the results obtained from (a) ProTox-II server (b) Osiris Property Explorer.

Figure 5: Illustration of the formation of the hit fragment, LF16. Fragment 1 (Fragment ID: 57379345) is derived from Ceritinib while Fragment 2 (Fragment ID: 56960447) is derived from X-396.

prolonged use of crizotinib. A study found the presence of brain metastases in about 23.8% at the time of diagnosis and in about 58.4% patients after 3 years of continued treatment with crizotinib.[30] This necessitates the constant discovery of novel inhibitors.

Our study involved a dataset of 12 known FDA approved ALK inhibitors as well as inhibitors undergoing clinical trials. Fragmentation of these known inhibitors resulted in a large library of fragments which was then narrowed down based on their Glide XP score. The best fragments of each compound thus obtained were linked together using the BREED feature. This algorithm generates a total of 25 novel compounds denoted as LF1 to LF25. These fragments were analysed for their efficacy towards F1174L mutation.

The mutational analysis revealed the disruption of three intra-molecular interactions with the 1174th residue of the mutant protein. The structural analysis of wild type ALK reveals that Ile 1170 is present in the αC helix, Ile 1179 is in the β4 strand of the N-lobe and Phe 1098 is present in the anti-parallel β sheet that packs against the αC helix.[31] Of note, it was found that Phe 1098 residue forms hydrophobic stacking interactions with Phe 1174, Phe 1271 (in the activation loop) and Phe 1245 residues (in the αE helix), forming an F-core which is essential for maintaining the inactive conformation to which crizotinib binds.[32-35] Additionally, we found that disruption of interaction with Ile1170 residue could be the reason for the destabilization of the αC helix of the activation loop during mutation.[32-35] These changes in the protein lead to an active conformation of the protein thus hindering crizotinib binding. We hope that this insight structural information is of immense importance for the experimental biologist in designing the novel and potent ALK inhibitor.

Further, from our study, the novel hybrid molecule LF16 was found to have a better binding affinity towards native and mutant ALK protein (Figure 5). LF16 forms interactions with Ile1119 and Leu1122 of the native ALK protein (Figure 6). Of note, the structural analysis of ALK illustrates that Leu1122 is a Glycine-rich loop residue which anchors the α/β phos-

countries and was approved by the FDA in November 2013. In spite of its effectiveness and initial benefit, drug resistance soon develops due to its toxicity effects.[29] Of note, brain metastases are a characteristic of NSCLC that has been difficult to treat even with
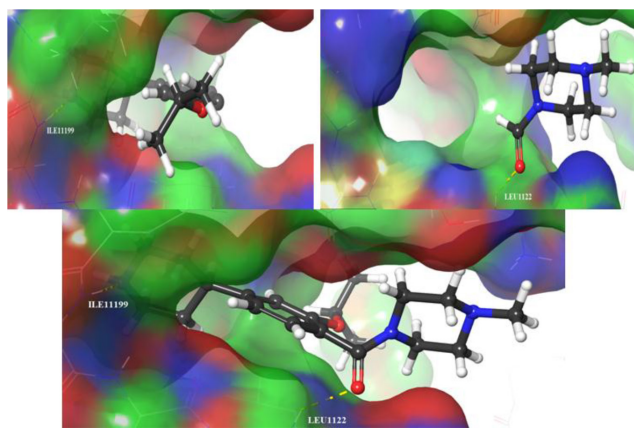
Figure 6: The superimposition of the fragments and their interacting residues with native ALK protein.
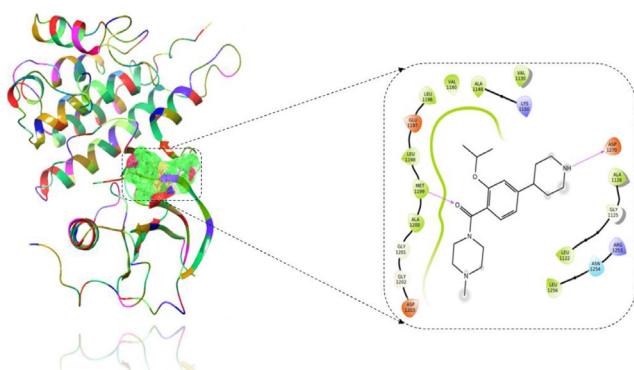
Figure 7: The LID of the hit molecule with F1174L mutant ALK protein.

phates of ATP for the appropriate positioning of the γ phosphate for the catalysis.[36] Furthermore, LF16 had interactions with Met1199 and Asp1270 residues of the F1174L mutant protein (Figure 7). Met1199 is the hinge region residue that makes up the binding pocket of ATP and is commonly targeted residue of kinase inhibitors.[36] Additionally, Asp1270 residue is present in the DFG motif of the activation loop, which is the regulatory portion that assists in substrate binding and influences the catalytic efficiency of the protein. Thus, binding LF16 to these residues will disrupt the ATP catalysis and halt the downstream signaling processes leading to the management of tumors in ALK-positive NSCLC.

CONCLUSION

Developing small molecule inhibitors against ALK-positive NSCLC is of immense significance, in the quest for tackling lung cancer cases. Many ALK inhibitors have been FDA approved and many are in their pursuit towards FDA approval. All these known inhibitors were analyzed for the possibility of new combinations, so as to design a new lead for native and

F1174L mutation of ALK protein. Thus, fragment-based drug designing in combination with molecular docking, ADMET analysis and interaction profiling revealed LF16 as a possible hit. We hope this study may pave the way for the designing and development of novel ALK inhibitors for the management of drug resistance in NSCLC in the near future.

ACKNOWLEDGEMENT

CONFLICT OF INTEREST

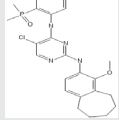The authors declare no conflict of interest.

ABBREVIATIONS
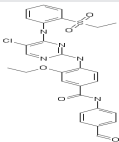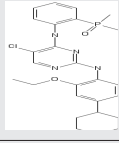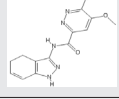
ALK: Anaplastic Lymphoma Kinase; **NSCLC:** Non-Small Cell Lung Cancer; **GLIDE:** Grid-Based Ligand Docking with Energetics; **XP:** Extra Precision; **SP:** Standard Precision; **ADMET:** Absorption, Distribution, Metabolism, Excretion and Toxicity; **SDM:** Site-Directed Mutator; **LF:** Linked Fragment; **PCA:** Principle Component Analysis; **RIN:** Residue Interaction Network.
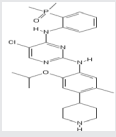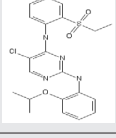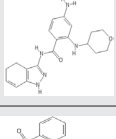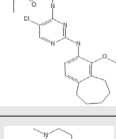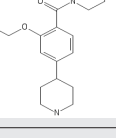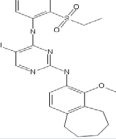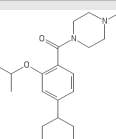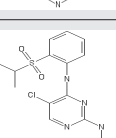
REFERENCES
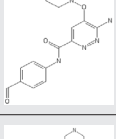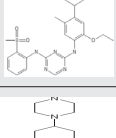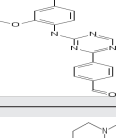
1. Solomon B, Wilner KD, Shaw AT. Current Status of Targeted Therapy for Anaplastic Lymphoma Kinase–Rearranged Non–Small Cell Lung Cancer. Clin Pharmacol Ther. 2013;95(1):15-23.

2. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, *et al*. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561-6.

3. Sakamoto H, Tsukaguchi T, Hiroshima S, Kodama T, Kobayashi T, Fukami TA, *et al*. CH5424802, a Selective ALK Inhibitor Capable of Blocking the Resistant Gatekeeper Mutant. Cancer Cell. 2011;19(5):679-90.

4. Holla VR, Elamin YY, Bailey AM, Johnson AM, Litzenburger BC, Khotskaya YB, *et al*. ALK: A tyrosine kinase target for cancer therapy. Mol Case Stud. 2017;3(1):a001115.

5. Bresler SC, Wood AC, Haglund EA, Courtright J, Belcastro LT, Plegaria JS, *et al*. Differential inhibitor sensitivity of anaplastic lymphoma kinase variants found in neuroblastoma. Sci Transl Med. 2011;3(108):108ra114.

6. Ziogas DC, Tsiara A, Tsironis G, Lykka M, Liontos M, Bamias A, *et al*. Treating ALK-positive non-small cell lung cancer. Ann Transl Med. 2018;6(8):141.

7. Bresler SC, Weiser DA, Huwe PJ, Park JH, Krytska K, Ryles H, *et al*. ALK Mutations Confer Differential Oncogenic Activation and Sensitivity to ALK Inhibition Therapy in Neuroblastoma. Cancer Cell. 2014;26(5):682-94.

8. Liu Q, Pan P, Yu H, Liu Q, Kong X, Chen H, *et al*. Combating Drug-Resistant Mutants of Anaplastic Lymphoma Kinase with Potent and Selective Type-I1/2 Inhibitors by Stabilizing Unique DFG-Shifted Loop Conformation. ACS Cent Sci. 2017;3(11):1208-20.
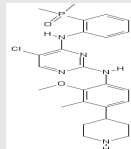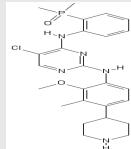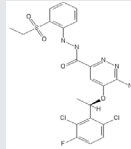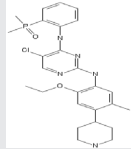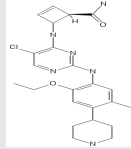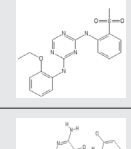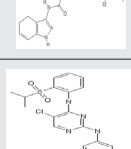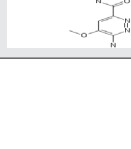
9. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. Pharmacological Reviews. 2014;66(1):334-95.

10. Yu W, JrMacKerell AD. Computer-Aided Drug Design Methods. Methods Mol Biol. 2017;1520:85-106.

11. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.

12. Awad MM, Shaw AT. ALK inhibitors in non-small cell lung cancer: Crizotinib and beyond. Clin Adv Hematol Oncol. 2014;12(7):429-39.

13. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013;27(3):221-34.

14. Halgren TA. Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model. 2009;49(2):377-89.

15. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, *et al*. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem. 2006;49(21):6177-96.

16. Wang L, Xie Z, Wipf P, Xie XQ. Residue preference mapping of ligand fragments in the protein data bank. J Chem Inf Model. 2011;51(4):807-15.

17. Pierce AC, Rao G, Bemis GW. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, P38 and HIV protease. J Med Chem. 2004;47(11):2768-75.

18. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33(Web Server issue):W306-10.

19. Worth CL, Preissner R, Blundell TL. SDM--a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res. 2011;39(Web Server issue):W215-22.

20. Bromberg Y, Rost B. Correlating protein function and stability through the analysis of single amino acid substitutions. BMC Bioinformatics. 2009;10(S8):S8.

21. Martin AJM, Vidotto M, Boscariol F, DiDomenico T, Walsh I, Tosatto SCE. RING: Networking interacting residues, evolutionary information and energetics in protein structures. Bioinformatics. 2011;27(14):2003-5.

22. Jorgensen WL, Duffy EM. Prediction of drug solubility from structure. Adv Drug Deliv Rev. 2002;54(3):355-66.

23. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, *et al*. Crizotinib versus Chemotherapy in Advanced ALK -Positive Lung Cancer. N Engl J Med. 2013;368:2385-94.

24. Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA, *et al*. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: A retrospective analysis. Lancet Oncol. 2011;12(11):1004-12.

25. Ntie-Kang F. An *in silico* evaluation of the ADMET profile of the StreptomeDB database. Springerplus. 2013;2(1):353.

26. Drwal MN, Banerjee P, Dunkel M, Wettig MR, Preissner R. ProTox: A web server for the *in silico* prediction of rodent oral toxicity. Nucleic Acids Res. 2014;42(Web Server issue):W53-8.

27. Ihlenfeldt WD, Bolton EE, Bryant SH. The PubChem chemical structure sketcher. J Cheminform. 2009;1(1):20.

28. Shaw AT, Solomon B. Targeting anaplastic lymphoma kinase in lung cancer. Clin Cancer Res. 2011;17(8):2081-6.

29. Tsuji T, Ozasa H, Aoki W, Aburaya S, Funazo T, Furugaki K, *et al*. Alectinib Resistance in ALK-Rearranged Lung Cancer by Dual Salvage Signaling in a Clinically Paired Resistance Model. Mol Cancer Res. 2019;17(1):212-24.

30. Rangachari D, Yamaguchi N, VanderLaan PA, Folch E, Mahadevan A, Floyd SR, *et al*. Brain metastases in patients with EGFR-mutated or ALK-rearranged non-small-cell lung cancers. Lung Cancer. 2015;88(1):108-11.

31. Bossi RT, Saccardo MB, Ardini E, Menichincheri M, Rusconi L, Magnaghi P, *et al*. Crystal structures of anaplastic lymphoma kinase in complex with ATP competitive inhibitors. Biochemistry. 2010;49(32):6813-25.

32. Azarova AM, Gautam G, George RE. Emerging importance of ALK in neuroblastoma. Semin Cancer Biol. 2011;21(4):267-75.

33. Karabencheva TG, Lee CC, Black GW, Donev R, Christov CZ. How does conformational flexibility influence key structural features involved in activation of anaplastic lymphoma kinase?. Mol Biosyst. 2014;10(6):1490-5.

34. Sasaki T, Okuda K, Zheng W, Butrynski J, Capelletti M, Wang L, *et al*. The neuroblastoma-associated F1174L ALK mutation causes resistance to an ALK kinase inhibitor in ALK-translocated cancers. Cancer Res. 2010;70(24):10038-43.

35. Jiang C-H, Huang CX, Chen YJ, Chuang YC, Huang BY, Yang CN. Molecular Modeling for Structural Insights Concerning the Activation Mecanisms of F1174L and R1275Q Mutations on Anaplastic Lymphoma Kinase. Molecules. 2018;23(7):E1610.

36. Roskoski R. Anaplastic Lymphoma Kinase (ALK): Structure, oncogenic activation and pharmacological inhibition. Pharmacol Res. 2013;68(1):68-94.

S.No	Name of fragment	SMILES	2D structure
		Table S1: The structures, fragment constituents and the smiles of the generated LFs.	
1	LF1	CCOC1=C(C=CC(=C1)C(=O)NC2=CC=C(C=C2)C=O)NC3=NC(=C(C=N3)Cl)NC4=C(C=CC=C4)[S](=O)(=O)CC	
2	LF2	CCOC1=C(C=CC(=C1)C2CCNCC2)NC3=NC(=C(C=N3)Cl)NC4=C(C=CC=C4)[P](C)(C)=O	
3	LF4	O(C1=CC(=NN=C1N([H])[H])C(=O)N(C2=N[N]([H])C3=C2CCC=C3)[H])C	
4	LF6	COC2=C1CCCCCC1=CC=C2NC3=NC(=C(C=N3)Cl)NC4=C(C=CC=C4)[P](C)(C)=O	

5	LF7	CC1=C(C=C(C(=C1)N(C2=NC=C(Cl)C(=N2)N(C3=CC=CC=C3[P](C)(C)=O)[H])[H])OC(C)C)C4CCN(CC4)[H]	
6	LF8	CC[S](=O)(=O)C1=CC=CC=C1NC2=C(C=NC(=N2)NC3=C(C=CC=C3)OC(C)C)Cl	
7	LF9	N(C1=CC=(C(C=C1)C(=O)N(C2=N[N]([H])C3=C2CCC=C3)[H])N(C4CCOCC4)[H])([H])[H]	
8	LF10	COC2=C1CCCCCC1=CC=C2NC3=NC(=C(C=N3)Cl)NC4=C(C=CC=C4)[S](=O)(=O)C(C)C	
9	LF11	CN3CCN(OC1=C(N=NC(=C1)C(=O)NC2=CC=C(C=C2)C=O)N)CC3	
10	LF12	CCOC1=C(C=C(C(=C1)C2CCNCC2)C)NC3=NC(=NC=N3)NC4=C(C=CC=C4)[S](C)(=O)=O	
11	LF13	COC1=C(C=CC(=C1)N3CCC(N2CCNCC2)CC3)NC4=NC(=NC=N4)C5=CC=C(C=C5)C=O	
12	LF14	CCOC1=C(C=CC(=C1)C2CCNCC2)C(=O)N3CCN(C)CC3	
13	LF15	CC[S](=O)(=O)C1=C(C=CC=C1)NC2=C(C=NC(=N2)NC3=C(C4=C(C=C3)CCCC4)OC)Cl	
14	LF16	CC(C)OC1=C(C=CC(=C1)C2CCNCC2)C(=O)N3CCN(C)CC3	
15	LF17	CCOC1=C(C=CC=C1)NC2=NC(=C(C=N2)Cl)NC3=C(C=CC=C3)[S](=O)(=O)C(C)C	

16	LF18	C1=C(C(=C(C(=C1)C2CCN(CC2)[H])C)OC)N(C3=NC(=C(C=N3)Cl)N(C4=C(C=CC=C4)[P](C)(C)=O)[H])[H]	
17	LF19	C1=C(C(=C(C(=C1)C2CCN(CC2)[H])C)OC)N(C3=NC(=C(C=N3)Cl)N(C4=C(C=CC=C4)[P](C)(C)=O)[H])[H]	
18	LF20	CC[S](=O)(=O)C1=C(C=CC=C1)NNC(=O)C2=NN=C(C(=C2)O[C@H](C)C3=C(C(=CC=C3Cl)F)Cl)N	
19	LF21	CCOC1=C(C=C(C(=C1)C2CCNCC2)C)NC3=NC(=C(C=N3)Cl)NC4=C(C=CC=C4)[P](C)(C)=O	
20	LF22	CCOC1=C(C=C(C(=C1)C2CCNCC2)C)NC3=NC(=C(C=N3)Cl)NC4C=C[C@H]4C(N)=O	
21	LF23	CCOC1=C(C=CC=C1)NC2=NC(=NC=N2)NC3=C(C=CC=C3)[S](C)(=O)=O	
22	LF24	C[C@H](OC1=CC(=NN=C1N([H])[H])C(=O)N(C2=N[N]([H])C3=C2CCC=C3)[H])C4=C(C(=CC=C4Cl)F)Cl	
23	LF25	COC1=C(N=NC(=C1)C(=O)NC2=CC=C(C=C2)NC3=NC(=C(C=N3)Cl)NC4=C(C=CC=C4)[S](=O)(=O)C(C)C)N	

PICTORIAL ABSTRACT
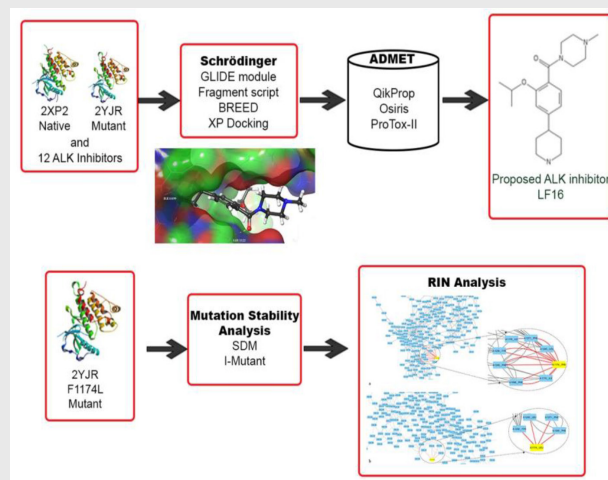
SUMMARY

- The best docking fragments of known ALK inhibitors were combined to create a novel inhibitor, LF16, with a high drug score, good CNS activity, and low toxicity.

- LF16 has interactions with Ile1119 and Leu1122 of the native ALK protein and Met1199 and Asp1270 residues of the F1174L mutant protein. These residues have a significant role to play in tumorigenesis.

- The cause of destabilization of F1174L mutation was found to be due to the disruption of three intra-molecular interactions with the 1174th residue, namely, Ile 1179, Ile 1170 and Phe1098.

- Fragment Based Drug Discovery method shows great promise for the discovery of drugs that are highly efficient and can avoid inactivity due to mutations.

About Authors

Anika Jain, B.Tech Student, Department of Biotechnology, School of BioSciences and Technology, Vellore Institute of Technology, Vellore-632014, Tamil Nadu.

Nivya James, Research Associate, Medical Informatics Lab, Department of Biotechnology, School of BioSciences and Technology, Vellore Institute of Technology, Vellore-632014, TamilNadu.

Dr. Veerapapillai Shanthi, Associate Professor, Department of Biotechnology, School of BioSciences and Technology, Vellore Institute of Technology, Vellore-632014, Tamil Nadu.

Dr. Karuppasamy Ramanathan, Professor, Department of Biotechnology, School of BioSciences and Technology, Vellore Institute of Technology, Vellore-632014, Tamil Nadu.