Unveiling the Molecular Mechanisms Behind the Devastating Impact of the ALK Protein on Pediatric Cancers: Insights into Deleterious SNPs through In Silico Predictions, Molecular Docking, and Dynamics Studies

Abdulhadi Almazroea

1 Associate Professor of General Pediatrics, Medical College, Taibah University, Saudi Arabia

Address for correspondence Dr. Abdulhadi Almazroea, MD, Associate Professor of General Pediatrics, Medical College, Taibah University, Saudi Arabia (e-mail: amazroea@taibahu.edu.sa).

Abstract

Introduction Pediatric cancers present significant challenges in terms of diagnosis and treatment, and the anaplastic lymphoma kinase (ALK) protein has emerged as a crucial molecular target in these malignancies. ALK, a receptor tyrosine kinase, plays a vital role in normal cellular processes, but genetic alterations and aberrant activation of the ALK gene have been implicated in various pediatric cancer types. While genetic alterations have been well studied, the precise molecular mechanisms underlying the pathogenicity of the ALK protein in pediatric cancers remain poorly understood.

Objective In this study, the primary objective is to uncover the molecular mechanisms associated with the effects of deleterious single-nucleotide polymorphisms (SNPs) on the structure and functionality of the ALK protein.

Material and Methods Several known point mutations of the ALK protein were taken for the in silico predictions such as PolyPhen-2, SIFT, PANTHER, PredictSNP, etc., residue conservation analysis using Consurf server, molecular docking (AutoDock), and molecular dynamics simulation studies (GROMACS).

Results The computation predictions found that the studied variants are deleterious in different tools. The residue conservation analysis reveals all the variants are located in highly conserved regions. The molecular docking study of wild-type and mutant structures with the crizotinib drug molecule found the variants were modulating the binding cavity and had a strong impact on the binding affinity. The binding energy of the wild-type is –5.896 kcal/mol, whereas the mutants have –9.988 kcal/mol. The specific amino acid Ala1200 of wild-type was found to interact with crizotinib, and Asp1203 residue was found to interact predominantly in the mutant structures.

Conclusion The simulation study differentiates the variants in terms of structural stability and residue fluctuation. Among the studied variants, R1275Q, F1245V, and F1174L had strong deleterious effects, structural changes, and pathogenicity based on the in silico predictions. By elucidating the functional consequences of deleterious mutations within the ALK gene, this research may uncover novel therapeutic targets.
and personalized medicine approaches for the management of pediatric cancers. Ultimately, gaining insights into the molecular mechanisms of the ALK protein’s role in driving response and resistance will contribute to improving patient outcomes and advancing our understanding of this complex disease.

**Introduction**

Pediatric cancers continue to pose significant challenges in terms of diagnosis and treatment. Among the various molecular targets implicated in pediatric cancers, the anaplastic lymphoma kinase (ALK) protein has gained considerable attention due to its devastating impact on tumor development and progression. ALK is a receptor tyrosine kinase that plays a crucial role in normal cellular processes, including neuronal development. However, aberrant activation or genetic alterations in the ALK gene have been associated with the onset and progression of several pediatric cancer types, including neuroblastoma and anaplastic large cell lymphoma.

While genetic alterations, such as chromosomal rearrangements and gene fusions, have been well documented as key events in ALK-driven pediatric cancers, the precise molecular mechanisms underlying the pathogenicity of ALK protein remain largely unexplored. Recent advancements in genomic research have revealed the presence of single-nucleotide polymorphisms (SNPs) within the ALK gene, some of which are predicted to have deleterious effects on the protein structure and function. These deleterious SNPs hold the potential to contribute to the susceptibility and progression of pediatric cancers, but their specific roles and impact on ALK signaling pathways remain elusive.

Since 2011, crizotinib drug molecule approved by the Food and Drug Administration has been used as a first-line treatment for ALK fusion positives and anaplastic large cell lymphomas. Crizotinib resistance was found in the patients with L1196M, I1171N, and F1174L/V/C mutations. From the detailed review analysis, it was found R1275Q/L has 45% of frequency in ALK-mutant tumors, 30 and 12% for the L1196M, I1171N, and F1174L/V/C mutations. These observations suggest that specific molecular contexts play a role in determining varying sensitivity to direct ALK kinase inhibition. This implies that the effectiveness of ALK inhibition and the mechanisms of resistance may be influenced by the specific context in which it is applied, highlighting the context-dependent nature of ALK’s role in driving response and resistance.

Pediatric cancers pose a significant public health concern, and understanding the molecular mechanisms driving their development is crucial for the development of effective treatment strategies. The ALK protein has emerged as a key player in various pediatric cancers, making it a promising target for further investigation. Deleterious SNPs in the ALK gene have been associated with an increased risk of pediatric cancers. Therefore, studying these SNPs and their impact on ALK protein function can offer valuable insights into the underlying molecular mechanisms that drive disease progression.

In silico studies provide a powerful approach for exploring the structural and functional aspects of proteins. By utilizing computational techniques such as molecular docking and dynamics simulations, we can investigate the interactions between the ALK protein and its ligands or potential therapeutic agents. These computational methods yield detailed insights into the molecular mechanisms and potential therapeutic targets. Molecular docking allows for the prediction of protein–ligand interactions, aiding in the identification of potential inhibitors or modulators of ALK protein activity. This information can guide the development of novel therapeutic interventions tailored to combat pediatric cancers associated with ALK dysregulation. On the other hand, molecular dynamics (MD) simulations enable the examination of the dynamic behavior of the ALK protein, providing valuable information about its conformational changes, stability, and interactions over time. These simulations shed light on how deleterious SNPs may affect the structure and function of the ALK protein, thus aiding in the identification of key functional regions and potential molecular targets.

Overall, the aim of this study is to unravel the intricate molecular mechanisms underlying the devastating impact of the ALK protein on pediatric cancers by investigating deleterious SNPs. The integration of in silico studies, molecular docking, and dynamics simulations will provide crucial insights that can ultimately contribute to the development of targeted therapies for pediatric cancer patients.

In this study, the aim is to unveil the molecular mechanisms behind the ALK protein’s devastating impact on pediatric cancers by focusing on the insights provided by deleterious SNPs. By elucidating the functional consequences of deleterious SNPs within the ALK gene, this research holds the potential to uncover novel therapeutic targets and strategies for the management of pediatric cancers. Moreover, it may pave the way for personalized medicine approaches that take individual genetic variations and their impact on the ALK signaling pathways. Ultimately, gaining insights into the molecular mechanisms of the ALK protein’s devastating impact on pediatric cancers will contribute to improving patient outcomes and advancing our understanding of this complex disease.

**Methodology**

**Data Collection**

Genomic and protein data relevant to pediatric cancer patients with ALK gene alterations were obtained from publicly available databases and repositories, including PubMed, RCSB-PDB, Ensembl genome browser, and UniProt. The focus was on identifying SNPs associated with the ALK
Table 1: Complete genetic data of the SNPs included in this study

<table>
<thead>
<tr>
<th>Missense SNP</th>
<th>dbSNP ID</th>
<th>Chr.bp position/alleles</th>
<th>Evidence</th>
<th>Clinical significance</th>
<th>Transcript ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1171N</td>
<td>rs1057519698</td>
<td>2:1057519698-A/G/T</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>F1174L</td>
<td>rs863225281</td>
<td>2:29220829-G/C/T</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>F1174V</td>
<td>rs281864719</td>
<td>2:29220831-A/G/C/T</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>F1174C</td>
<td>rs1057519697</td>
<td>2:29220830-A/C</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>L1196M</td>
<td>rs1057519784</td>
<td>2:29220765-G/T</td>
<td>Cited, disease phenotype</td>
<td>Likely pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>F1245V</td>
<td>rs281864720</td>
<td>2:29213994-A/G/C/T</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>F1245C</td>
<td>rs863225283</td>
<td>2:29213993-A/C</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>R1275Q</td>
<td>rs113994087</td>
<td>2:29209798-C/A/T</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
<tr>
<td>R1275L</td>
<td>rs113994087</td>
<td>2:29209798-C/A/T</td>
<td>Cited, disease phenotype</td>
<td>Pathogenic</td>
<td>ENST00000389048.8</td>
</tr>
</tbody>
</table>

Abbreviations: dbSNP, Single Nucleotide Polymorphism Database; SNP, single-nucleotide polymorphism.

Note: Bold values indicate the allele entries in dbSNP and ensembl genome browser.

A comprehensive bioinformatics analysis was conducted to evaluate the potential deleterious effects of the identified SNPs on the structure and function of the ALK protein. This analysis involved the use of various noncommercial prediction tools and software, including PolyPhen-2, SIFT, PredictSNP, PANTHER, MetaLR, SNAP (screening for nonsynonymous SNPs (nsSNPs) within the coding region of the ALK gene), while MetaLR employs a logistic regression-based ensemble method to predict the pathogenicity of single-nucleotide variants. SNAP, a neural network-based screening tool, integrates various sequence information to predict the functional impact of nsSNPs and provides reliability information for the predictions.16

Amino Acid Conservation Analysis
To assess the conservation and evolutionary importance of the SNPs across various species, multiple sequence alignments were conducted. The Consurf server was employed for this purpose, utilizing the Multiple Alignment using Fast Fourier Transform (MAFFT) algorithm to perform the alignments by searching for homologous sequences in the UNIFREF90 database using the HMMER search algorithm.17,18 The server utilized a Bayesian method to predict conservation scores for each amino acid and determined the most suitable amino acid substitution based on the alignment. The resulting alignment was then presented with a color-coded scheme, distinguishing conserved and variable amino acids, providing insights into the conservation patterns within the protein sequence across different species.19

Structural Modeling and Molecular Docking
A three-dimensional (3D) structural model of the ALK protein (wild-type) and its variants was generated based on the crystal structure (PDB ID: 2ZP2). To assess the binding affinity changes associated with the studied mutations, molecular docking studies were conducted between the ALK protein and the drug molecule crizotinib (Fig. 1). Fig. 1 represents the 3D cartoon structure of the protein molecule (green) and ligand molecule (blue) docked in the active site. Each structure, including the wild-type and mutant variants, was generated using the PyMol software by introducing one amino acid mutation at a time. AutoDock Tools version 4.2.6 was employed to prepare the protein and ligand structures, assigning polar hydrogens, united atom Kollman charges, solvation parameters, and fragmental volumes to the protein.20,21 The prepared structures were saved in PDBQT format. AutoGrid was utilized to create a grid map,
specifying a grid box size of $60 \times 60 \times 60$ xyz points, 0.375 Å of grid spacing, and a designated grid center at coordinates $(x, y, z)$: 27.468, 46.380, and 7.560. To reduce computation time, a scoring grid was calculated from the ligand structure. Docking was performed using AutoDock/Vina, employing an iterated local search global optimizer and treating both the protein and ligands as rigid entities during the docking process. Docked results with a positional root-mean-square deviation (RMSD) below 1.0 Å were clustered, and the cluster representative with the most favorable free energy of binding was selected. The docking pose with the lowest energy of binding or highest binding affinity was extracted and aligned with the protein receptor structure for further analysis. Each mutant and wild-type structure underwent individual docking, and the predicted binding energies were correlated.

Molecular Dynamics Simulation (MDS)
In this study, a MD simulation was conducted to analyze the structural stability and changes caused by the impact of SNPs. The simulation included one wild-type structure and four mutant structures (I1171N, R1275Q, F1174L, and F1245V), while excluding other structures due to minimal changes observed during docking studies and computational efficiency. The flexibility and conformational stability of the ALK protein and mutant complexes were determined by the GROMACS v5.0.6 (Groningen Machine for Chemical Simulations) software. The energy-minimized ligand topologies were generated using the PRODRG server (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg). The ligands were merged into the protein structure and the cubic systems were generated with 1.0 nm distance from the protein–ligand complexes. Furthermore, the systems were added by water molecules. The entire systems were neutralized by adding the appropriate Na$^+$ and Cl$^-$ counterions. The GROMOS96 43a1 force field was used to minimize the structure energy and the electrostatic interaction was utilized by the particle-mesh Ewald and the steepest algorithm methods. In the energy minimization process, the steepest descent (50,000 steps) algorithm was utilized to maintain the solvating system. The system maintained a constant temperature and pressure despite being solvated. In the NVT ensemble, the temperature was maintained using the Berendsen thermostat (0.1 ps). The NPT ensemble was maintained at constant pressure (1 bar). The well-equilibrated complexes were further used for the MDS production run for 500 nanoseconds time period. Finally, the obtained simulation data were analyzed and plotted using Origin Pro 2018 for further structural analysis.

Inclusion & Exclusion Criteria
NA.

Primary & Secondary Outcome
NA.

Statistical Analysis
NA.
Results

Computational Prediction on nsSNPs
Based on the online tools used to predict the functional and structural effects of the studied nsSNPs, it was observed that all the mutations listed in this study had deleterious effects on the protein structure. These mutations played a substantial role in impacting protein function and reducing the overall structural stability. However, it is noteworthy that the mutations I1171N, L1196M, and F1174L were predicted to be neutral according to the PredictSNP, SNAP, and PhD-SNP tools, as indicated in ►Suppl. Table. 1 On the other hand, the remaining variants were determined to have deleterious effects with higher confidence scores by these prediction tools.

Residue Conservation Analysis
The Consurf server was utilized for residue conservation analysis of the ALK protein, using the UniProt database entry (ID: Q9UM73) as the input source. The analysis revealed that all the mutated positions were found to be highly conserved, and a majority of the amino acids were exposed, as depicted in ►Fig. 2. The region spanning from position 1171 to 1275 exhibited complete conservation in the ALK protein, and these mutations within this region are likely the primary cause for the deleterious effects and negative impact on protein function. ►Fig. 2 provided a visualization of the conservation pattern across the entire protein structure, indicating that the helix regions were highly conserved, and interestingly, all the studied mutations were located within these helices.

Molecular Docking
The molecular docking analysis of ALK wild-type and mutant structures was done using AutoDock software. The results revealed that all the mutant structures exhibited better scores compared with the wild-type structure. This indicates that the mutations had an impact on the ligand-binding site. The mutant structures demonstrated increased binding energy and favorable interaction patterns, as observed in ►Suppl. Table 2. Specifically, the amino acids Ala1200, Asp1203, Glu1197, and Met1199 were frequently identified in the docking results for both the wild-type and mutant structures when docked with the crizotinib drug molecule. Notably, the L1196M and I1171N mutants did not exhibit significantly higher binding energy compared with the wild-type, and their binding poses and patterns were similar (►Suppl. Fig. 1). Intriguingly, both the wild-type and the I1171N and L1196M mutants displayed identical binding poses with similar interaction patterns, suggesting that these mutations did not significantly impact the binding affinity of the docked compound. On the other hand, the

Fig. 2 Amino acid conservation analyzed using Consurf server and tertiary structure of anaplastic lymphoma kinase (ALK) protein representing the conserved region shown in dark purple color.
variants R1275Q/L, F1174L/V/C, and F1245V/C exhibited binding energies greater than $-9.458$ to $-9.988$ kcal/mol and displayed strong hydrogen bond interactions with the ALK protein (Suppl. Table 2). This indicates that the mutant structures have a higher affinity for binding the drug molecule within the active site of the ALK protein, thereby increasing the overall binding affinity.

**Molecular Dynamics**

A MD simulation was executed to evaluate the stability and dynamics of the wild-type and mutant ALK protein structures. Molecular docking results revealed no significant changes in the docking score or interactions for the mutants F1174L/V/C and F1245V/C complexes. Consequently, one mutant was selected for each position and subjected to further MD studies. The study included four mutant structures (I1171N, R1275Q, F1174L, and F1245V) along with the wild-type structure. Analysis of the RMSD of the studied complexes indicated that the wild-type structure remained stable throughout the simulation period. However, the mutant structures displayed notable instability and substantial fluctuations, indicating structural alterations compared with the wild-type. During the simulation, the variants F1174L and I1171N demonstrated stable RMSD values throughout the entire duration, as shown in Fig. 3. Initial deviations observed were attributed to the stabilization of the protein's equilibrium state. Conversely, the variants R1275Q and F1245V exhibited unstable RMSD values when compared with the wild-type, as depicted in Fig. 4. Even after 300 nanoseconds of simulation, these mutant structures remained unstable, with a difference of 2 Å in deviation. Analyzing the root mean square fluctuation of ALK protein residues revealed a significant fluctuation in the plot for amino acids 1150 to 1160, which are in proximity to the mutant positions 1171 and 1174 (Suppl. Fig. 2). While other variants showed substantial fluctuations, the I1171N and F1174L mutants displayed lesser fluctuations due to the structural changes they introduce. The remaining residues exhibited stability with an average deviation of 3 Å, with similar changes observed in the loop region of the protein.

**Discussion**

The computational prediction in this study provides valuable insights into the functional and structural effects of the identified mutations in the ALK protein. The analysis revealed that all the studied mutations had deleterious or damaging effects on the protein structure, resulting in reduced stability and potentially affecting protein function. Interestingly, certain prediction tools predicted a few of the studied mutations, namely I1171N, L1196M, and F1174L, to be neutral or benign, in contrast to the deleterious effects predicted by other tools and software. Residue conservation analysis indicated that the mutated positions in the protein sequence were highly conserved, with a Consurf score of 9, suggesting their importance for protein function. Mutations in these positions directly impact function through structural and functional changes. Molecular docking results demonstrated that the mutant structures exhibited better binding scores ($\sim -9.80$ kcal/mol) compared with the wild-type ($-5.89$ kcal/mol), indicating potential alterations in ligand binding sites. The interacting amino acid residues, specifically Glu197 and Met1199, were found to be commonly present in all structures. However, Ala200 was observed only in the wild-type and L1196M mutant structures. Interestingly, the mutant structures with higher docking scores exhibited interactions with the Asp1203 residue. This clearly demonstrates that the interaction with Asp1203 residue enhances the binding affinity. The mutations in and around the active site of the protein disrupt the cavity and create a favorable region for drug binding. Furthermore, MD simulations revealed that the mutant structures displayed instability and substantial fluctuations, suggesting structural alterations. Hence, the promising binding affinity observed in the mutant complexes during the

![Fig. 3](image1.png) Root-mean-square deviation (RMSD) plot of the least deviated protein mutant complex with wild-type obtained from molecular dynamics (MD) simulation study.

![Fig. 4](image2.png) Root-mean-square deviation (RMSD) plot of the most deviated protein mutant complex with wild type obtained from molecular dynamics (MD) simulation study.
During the simulation, this suggests that the binding affinity of the drug molecule in the mutant structure may be compromised in a dynamic environment. Overall, these findings underscore the significance of the studied nsSNPs in the ALK protein’s function and stability, providing a basis for further exploration and understanding of their role in pediatric cancers and potential implications for targeted therapies. Moreover, as a prospective study, it would be beneficial to generate clinical correlation and genetic data through clinical trials or pharmacogenomics studies. This additional information would contribute to a better understanding and management of this condition, enabling more personalized and effective approaches to treatment.

**Conclusion**

Computational prediction and screening of deleterious SNPs consistently offer significant contributions to disease diagnosis and treatment. The variants studied in the ALK protein were predicted to be deleterious and had a significant impact on modulating protein function, according to a series of computational analyses. The pathogenic nature of these variants was further supported by their location in highly conserved regions of the protein. Comparisons with wild-type variants showed changes in binding affinity during molecular docking studies. Additionally, MD simulations revealed destabilization of protein structure and higher fluctuation in variant structures compared with wild-type. Further pharmacogenomics studies can help correlate these findings with experimental data to better understand the molecular mechanisms underlying the devastating impact of ALK protein in pediatric cancers. These results, along with patient clinical data, can be used to draw conclusions regarding the role of specific SNPs in pediatric cancer susceptibility, progression, and potential therapeutic implications.

**Ethics**

None.

**Financial Support**

None.

**Conflict of Interest**

None declared.

**References**


