



Research paper

# Evaluation of Detrimental Missense SNPs of Human *CXCL6* Gene by Combining Algorithms, Homology Modeling, and Molecular Docking

Hamna Tariq <sup>a\*</sup> , Mehmooda Asif <sup>a</sup>, Muhammad Saleem <sup>a</sup>, Kainat Ramzan <sup>b\*</sup> ,  
Moeen Zulfiqar <sup>a</sup>, Aniqah Amir <sup>a</sup>, Abdur Rehman Asif <sup>c</sup>

<sup>a</sup> Department of Molecular Biology, University of Okara, Punjab, 53600, Pakistan

<sup>b</sup> Department of Biochemistry, University of Okara, Punjab, 53600, Pakistan

<sup>c</sup> Department of General Medicine, Semey Medical University, East Kazakhstan, 070000, Kazakhstan

## ARTICLE INFO

## ABSTRACT

### Keywords

*CXCL6*  
Cancer  
Inflammation  
SNP

Cancer is marked by uncontrolled cell proliferation, often driven by genetic mutations, including alterations in the *CXCL6* gene on chromosome 4q21. *CXCL6* is a crucial gene involved in immune responses and inflammation and has been implicated in promoting tumor growth and metastasis when overexpressed. This study utilized various bioinformatics tools to investigate the pathogenicity of missense nsSNPs within the *CXCL6* gene, identified through the NCBI-SNP database and assessed their impact on protein stability and structural integrity. A total of 22 nsSNPs were identified as potentially harmful, with mutations such as L47M, C51R, L92M, L92V, P73S, and Q104K demonstrating significant structural alterations. Mutation clustering was performed using Mutation3D, and molecular docking studies were conducted with PyRx to evaluate ligand interactions with both wild-type and mutant protein forms. Five of the 25 ligands examined displayed promising docking scores, suggesting their potential as therapeutic inhibitors. Additionally, the Hope algorithm provided further insights into the structural consequences of these mutations. This study highlights the role of *CXCL6* mutations in cancer progression and underscores the potential of *CXCL6* expression as a predictive biomarker across various cancer types. These findings offer a foundation for exploring *CXCL6* as a therapeutic target, contributing to personalized approaches in cancer treatment.



DOI  
[10.5281/ib-1040724](https://doi.org/10.5281/ib-1040724)

### \*Corresponding author

Hamna Tariq  
Kainat Ramzan

### ✉ Email

[hamnatariq@uo.edu](mailto:hamnatariq@uo.edu)  
[kainatramzan54@gmail.com](mailto:kainatramzan54@gmail.com)



## 1. Introduction

Cancer is characterized by the uncontrolled growth of aberrant cells that can penetrate neighboring organs and spread to other regions of the body, a process referred to as metastasis (Kurma & Alix-Panabières, 2023). Cancer is the most prevalent disease worldwide and is influenced by environmental agents categorized into chemical carcinogens, physical

carcinogens, and individual genetic factors. Chemical carcinogens, including substances such as asbestos, cigarette smoke, aflatoxins, and arsenic, contribute significantly to cancer risk, as do physical carcinogens such as ultraviolet and ionizing radiation. In recent decades, cancer incidence has surged, with over 100 million people currently affected globally. Between 2008 and 2018, cancer mortality rose to 9.7 million,

underscoring its growing health impact. By 2030, it is projected that annual new cancer cases may reach 23.6 million worldwide, highlighting a critical need for strengthened prevention and treatment approaches (Adhikari et al., 2024). Colon, lung, breast, lymph, brain, kidney, and blood cancers are among the most well-known forms of cancer. Early identification is key to successfully treating many forms of cancer, although other lifestyle factors raise the risk (Sung et al., 2021).

Radiotherapy, surgery, and systemic medicines such as hormone treatments, targeted biological therapies, chemotherapy, and targeted immunotherapies are the most popular and approved ways to treat cancer (Ghorbanpour et al., 2024). Mutations in tumor suppressor genes are a key hereditary factor in cancer development. These genes typically regulate cell division and differentiation, preventing uncontrolled growth (Gerke et al., 2024). Chromosomal translocations can lead to the formation of fusion genes, such as BCR-ABL in leukemia and EML4-ALK in lung cancer, which are known to drive cancer progression. Additionally, epigenetic abnormalities, including changes in DNA methylation and histone modifications, can disrupt normal gene expression patterns, contributing to cancer development. These epigenetic changes, by altering regulatory mechanisms without changing the DNA sequence, underscore the complexity of gene regulation in cancer and present potential targets for therapeutic intervention (Zhang & Li, 2023). Furthermore, germline mutations in *BRCA1/2*, *TP53*, and *PTEN*, and somatic mutations including *PIK3CA*, *AKT1*, *CTNNB1*, and *CXCL6*, can increase cancer risk (Chen et al., 2023). Cancers of the breast, lungs, colon, leukemia, and other organs can acquire these genetic abnormalities (Mohd Abas et al., 2024).

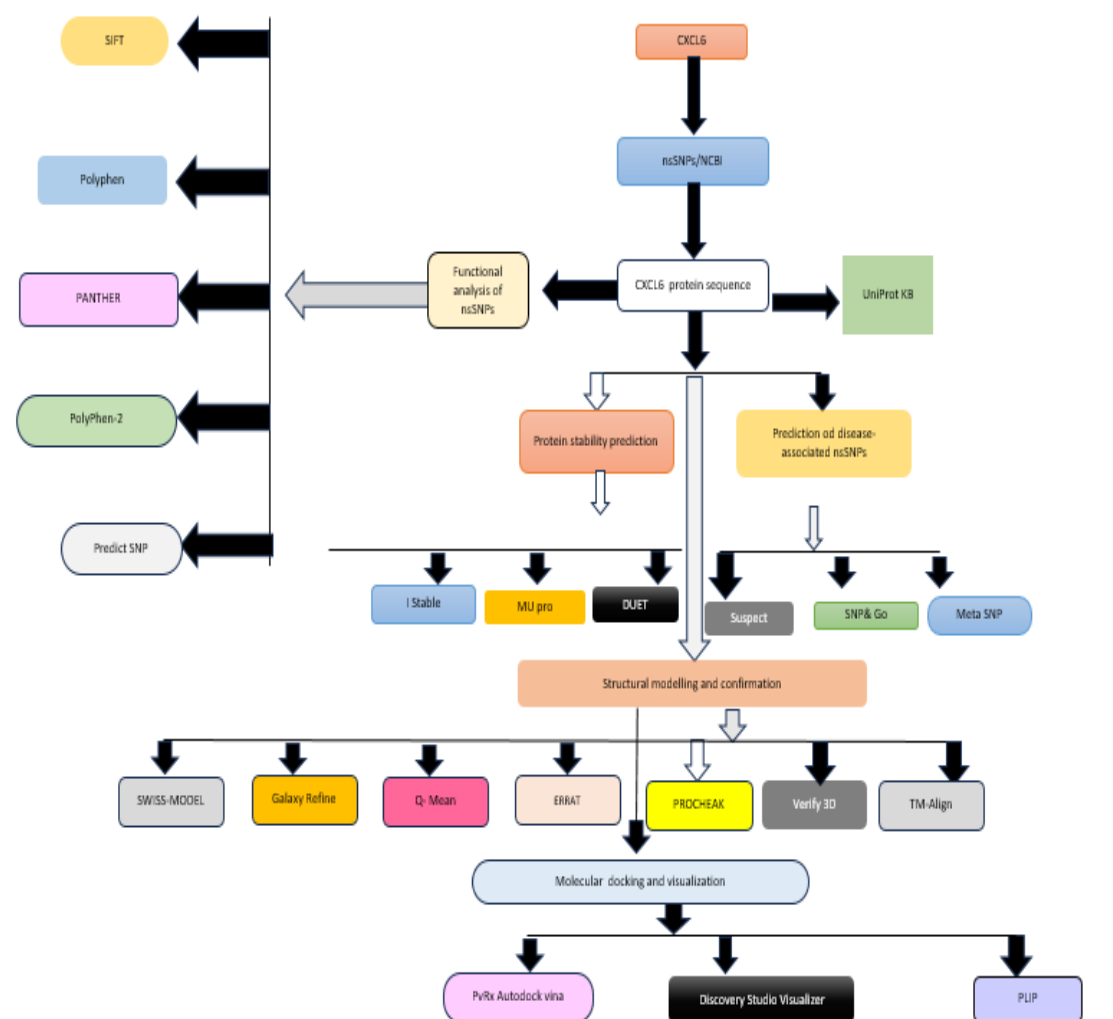
An increased number of single-nucleotide polymorphisms (SNPs) can elevate the risk of cancer, as certain SNPs are associated with genetic susceptibility to the disease. For instance, specific SNPs in the *TP53* gene, such as rs1042522, have been linked to a higher probability of cancer development. In certain populations, mutations cause the production of numerous alleles, accounting for over 99% of the variation in the human genome. The variant with the lowest frequency is 1% and are approximately 300-400 base pairs (Hoffman et al., 2024). There are two kinds of single nucleotide polymorphisms: genome-encoding and non-coding. Coding SNPs, which can influence protein sequences, include both synonymous and nonsynonymous variants. Mutations affecting amino acids cannot be caused by synonymous SNPs or any other non-coding SNP. SNPs that do not code for genes can be categorized into three groups: intronic, UTR, and intergenic (Yang & Nam, 2020).

In 2023, Studies found that single nucleotide polymorphisms (SNPs) in genes that encode inflammatory factors and chemokines can affect the development and progression of disease. Genetic variants in chemokine genes (*CCL3*, *CCL4*, and *CXCL8*) can impact a variety of diseases and conditions, including cancer, heart disease, and asthma in the bronchi (Tatarkova et al., 2022). Chemotactic and inducible small-molecule peptides, or chemokines, are present everywhere and have a significant impact on both short-term and long-term inflammation. One important inflammatory cytokine is soluble *CXCL6* C-X-C Motif chemokine, which binds to the receptors *CXCR1* and *CXCR2* and brings inflammatory cells to the site of inflammation. In 2022, Komolafe and Pacurari discovered that *CXCL6* levels are significantly elevated in individuals with idiopathic pulmonary fibrosis. This increase in *CXCL6* may play a role in the progression of lung fibrosis, suggesting its potential as a biomarker and a possible therapeutic target in managing the disease (Komolafe and Pacurari, 2022).

The *CXCL6* gene, also known as *SCYB6*, *CKA-3*, and *GCP-2*, encodes a protein belonging to the CXC chemokine family. This family is characterized by a conserved C-X-C motif, where a single amino acid separates two cysteine residues. This motif is crucial for functions such as chemotaxis, angiogenesis, and inflammatory responses (NCBI, 2022). *CXCL6* is crucial for immune defense, exhibiting activity against both Gram-positive and Gram-negative bacteria while attracting neutrophil granulocytes. This gene influences a range of physiological processes, including inflammation, immune response, cell proliferation, metastasis, and tissue repair. Its multifaceted role underscores its importance in maintaining homeostasis and responding to pathological conditions (Wuyts et al., 2003). In this study, we employed the non-synonymous SNPs that affect *CXCL6* genetic structure and function, even though *CXCL6* has well-established involvement in cancer formation and progression. Research on deleterious SNPs in the *CXCL6* gene, they affect protein stability and binding affinity, and whether or not they contribute to cancer formation and progression is particularly present. A lack of understanding in this area prevents the creation of efficient cancer treatments that target *CXCL6* and its variations.

## 2. Materials and Methods

Fig. 1 shows the overall methodology of the present study by employing multiple computational tools.



**Fig. 1** Overall flowchart of the present study

## 2.1 Data collection

To find *CXCL6* missense nsSNPs, the NCBI-SNP database (<https://www.ncbi.nlm.nih.gov/>) was queried. Through the UniProt web service (<https://www.uniprot.org/uniprotkb>), the FASTA sequence of the *CXCL6* protein (accession number: P80162) was retrieved.

## 2.2 Screening of deleterious nsSNPs

SNPNexus (<https://www.snp-nexus.org>) combines the PolyPhen and SIFT tools for mutation analysis (Waheed et al., 2024). The sorting Intolerant from Tolerant (SIFT) tool is employed to differentiate between disease-causing mutations and neutral polymorphisms by predicting the impact of amino acid substitutions on protein function. It generates a score ranging from 0 to 1, reflecting the likelihood that a novel amino acid will be tolerated at a specific site in the protein (Azmi et al., 2023). Polymorphism Phenotyping (PolyPhen) is utilized to predict the potential functional impacts of nsSNPs by integrating phylogenetic, structural, and sequence annotation data. Protein structure and function can be examined

using PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2>) (Sultana et al., 2024).

Moreover, the PredictSNP tool (<https://loschmidt.chemi.muni.cz/predictsnp>) provides more accurate and efficient consensus predictions by evaluating the effects of modifications to individual amino acids (Mohkam et al., 2022). Hidden Markov Models (HMMs) and evolutionary linkages are utilized in the Protein Analysis Through Evolutionary Relationship (PANTHER) database (<https://www.pantherdb.org>) (Dakshitha et al., 2024). Protein sequence, location, amino acid substitution, and Homo sapiens are the inputs used for study of the *CXCL6* gene.

## 2.3 Estimation of disease-associated nsSNPs

The SNPs & GO tool (<https://snps.biofold.org/snps-and-go/snps-and-go.html>) analyzes sequence, evolution, and Gene Ontology (GO) terms to predict disease-associated mutations in proteins. In nsSNPs disease classifications, variants with a probability greater than 0.5 are considered. (Azmi et al., 2023). MetaSNP (<https://snps.biofold.org/meta-snp/>) uses results from PANTHER, PhD-SNP, SIFT, and SNAP, applying a random forest classifier to discriminate

disease-related from polymorphism nsSNVs (Panchal et al., 2024). Mutations with frequencies over 0.5 are projected to be disease-related, with MetaSNP reaching 79% accuracy (Waheed et al., 2024). Suspect (<http://www.sbg.bio.ic.ac.uk/t/>) estimates the phenotypic impact of missense mutations by integrating sequence, structure, and system biology features, scoring mutations on a scale from 0 to 100, with 50 as the minimum score (Sultana et al., 2024).

#### 2.4 Prediction of Protein Stability Change

The *CXCL6* protein's stability and functional characteristics were examined with the use of i-Stable (<http://predictor.nchu.edu.tw/iStable/>), a program that determines how different amino acid changes affect protein stability (Akhtar et al., 2020). By entering each substitution and protein sequence, variants were examined. Because disease-causing mutations can impair protein structure and function, MUpPro (<http://mupro.proteomics.ics.uci.edu/>) was utilized to forecast the impact of mutations on protein stability. The SVM approach analyzes the protein sequence and structural features, which are represented by  $\Delta\Delta G$  values, to forecast changes in protein stability caused by mutations in a single amino acid (Laskar et al., 2023). We provided the locations of wild and mutant residues after verifying the *CXCL6* protein sequence. Applying three-dimensional models of both wild-type and mutant *CXCL6* proteins as input, the integrated computational tool DUET (<http://bleoberis.bioc.cam.ac.uk/duet/>) was utilized to assess the effect of nonsynonymous mutations on protein stability (Aljindan et al., 2021).

#### 2.5 Predicting mutation clusters

The clustering of amino acid replacements was evaluated using Mutation3D (<http://www.mutation3d.org/about.shtml/>), which identifies locations where mutations could have a major impact on protein function (Yasmin, 2022). We utilized homology modeling, which is extensively employed in pharmaceutical research and discovery due to its efficacy in predicting protein structure and function, to forecast the three-dimensional structure of the *CXCL6* protein as an experimental structure was not available (Harihar et al., 2024). To create three-dimensional models of wild-type and mutant *CXCL6* proteins, we used the SWISS-MODEL server (<https://swissmodel.expasy.org/>) (Banerjee et al., 2022). Based on the sequence of amino acids, the server can automatically anticipate the protein's three-dimensional structure. Finding structural templates, aligning the desired sequence with these templates, building the model, and evaluating its quality are all steps in the process (Kumar et al., 2024).

#### 2.6 Structural validation and refinement

We used the GalaxyRefine server (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE/>), which passed major evaluations for predicting protein structures, to modify the models. Molecular dynamics simulations were used for structural refinement (Rehman et al., 2024). To determine the best strategy for prediction, the updated models were tested using different validation tools. For more information about Qualitative Model Energy Analysis (QMEAN), visit (<https://swissmodel.expasy.org/qmean/help/>) (Khamlich et al., 2023). The clustering algorithm's z-score determines the reliability of the model. The SAVES Structure Validation Server, accessible at (<https://saves.mbi.ucla.edu/>), was utilized to evaluate structural flaws and the z-score of the chosen model to validate the 3D protein models (Azmi et al., 2023). The ERRAT program examined the nonbonded interactions between different atomic residues (Khamlich et al., 2023), while PROCHECK assessed the stereochemical features of the homology models (Goswami et al., 2024), such as the Ramachandran plot and dihedral angles ( $\varphi$  and  $\psi$ ) for potential amino acid configurations (An et al., 2023). To validate the model, a minimum of 65% of the amino acids must have a score higher than 0.2 (Arega et al., 2024). We also used the Verify 3D application to make sure the amino acid sequence was compatible with the 3D structure of the protein.

#### 2.7 Structural Comparison Analysis

To assess the updated model, we superimposed protein structures using the TM-align tool (<http://zhanglab.dcmf.med.umich.edu/TM-align/>) to find structural similarities. This program calculates the root mean square deviation (RMSD) and the template modeling score (TM score) (Rozario et al., 2021).

#### 2.8 Predicting the phenotypic effects of the nsSNPs

We analyzed the structural and functional implications of point mutations using the HOPE server, which can be available at (<https://www3.cmbi.umcn.nl/hope/>). As a result of each mutation, this the most advanced automated program produces a report that describes the protein's altered size, charge, bonding pattern, and interactions with other molecules (Yasmin, 2022).

#### 2.9 Ligand preparation

From the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/compound/>), we obtained the three-dimensional chemical blueprints of chosen compounds. Afterwards, we decreased ligand energies and optimized their 3D geometry using specialized algorithms (Shah et al., 2020). We changed the ligands

from SDF to PDB after downloading them so we could use them in future docking investigations.

### 2.10 Pyrx Autodock Vina

Molecular docking plays a crucial role in the fields of drug discovery and biomolecular interaction research. Finding a new drug candidates and learning how they bind to targets are both made easier with its help. Docking technology has recently advanced by facilitating studies in biomolecular interactions and structure-based drug design (Chen et al., 2020). We docked the chosen ligands with *CXCL6* using the PyRx program (<https://pyrx.sourceforge.io/>), which is part of the Lamarckian genetic algorithm (LGA) that allows for virtual ligand screening with AutoDock and Autodock Vina (Ni et al., 2024). The five most significant values for each ligand were determined by altering the grid center size (XYZ axis) (Huang et al., 2024).

### 2.11 PLIP Visualization Interaction

Discover and visualize protein-ligand interaction patterns in 3D structures using the Protein Data Bank (PDB) or user-uploaded data with the help of the

Protein-Ligand Interaction Profiler (PLIP), a free program attainable at ([projects.biotec.tu-dresden.de/plip-web/](https://projects.biotec.tu-dresden.de/plip-web/)) (Adasme et al., 2021). In addition to creating XML and text files, each binding site also generates 3D interaction diagrams that may be analyzed live using JSmol or offline with PyMOL (Kantelis et al., 2022). You may find several kinds of interactions, like hydrogen bonds, hydrophobic contacts,  $\pi$ -stacking,  $\pi$ -cation interactions, salt bridges, water bridges, metal complexes, and halogen bonds, with this tool (Adasme et al., 2021).

## 3. Results

### 3.1 Retrieval of the nsSNP dataset

The NCBI-SNP database was used to extract the nsSNPs of the *CXCL6* gene. Fig. 2 displays the total number of SNPs, which was 22,83. Fig. 2 displays that of the 2283 nsSNPs, 117 are synonymous, 250 are missense or non-synonymous, 632 are coding, 437 are non-coding, 636 are intronic, 66 are exonic, 410 are in the UTR, 284 are in the 3' UTR, 126 are in the 5' UTR, 427 are in the 3' downstream, and 377 are in the 5' upstream.

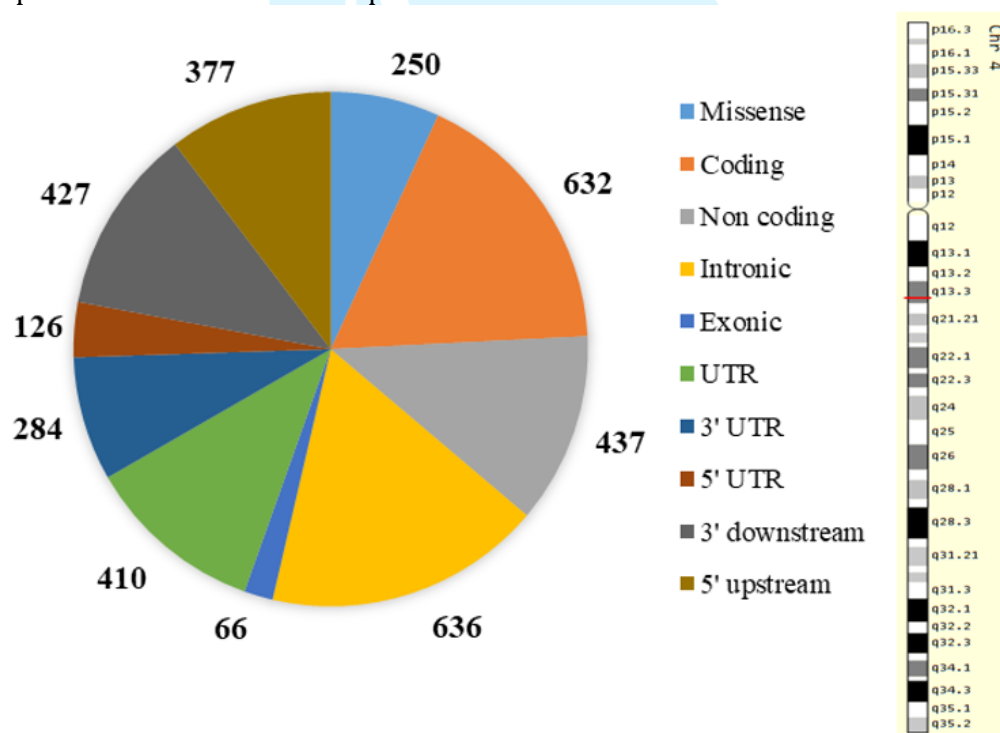


Fig. 2 Distribution of nsSNPs and banding location mapped on human chromosome in *CXCL6* gene

We chose missense nsSNPs for further study because a change in the coding sequence could alter the protein sequence, modifying the protein structure and increasing the risk of many disorders. Though some missense substitutions do alter protein structure and function, others are neutral or have no effect at all. It is therefore critical to distinguish between harmful SNPs and neutral ones. We obtained the sequence of protein P80162 for this investigation from the UniProt

database. Moreover, 114 amino acids make up the *CXCL6* gene in humans. It is situated on chromosome 4. Granulocyte chemotactic protein 2 (*GCP-2*) and small-inducible cytokine *B6* (*SCYB6*) are alternative names for the *CXCL6* gene.

### 3.2 Deleterious SNP prediction by SIFT and Polyphen

The SIFT predicts a list of nsSNPs based on sequence homology and amino acid physical properties. Using the tolerance index (TI), which is inversely related to the functional impact of amino acid substitutions, SNPnexus identified 234 deleterious substitutions among 2,283 nsSNPs predicted by SIFT. PolyPhen analysis indicated that out of the assessed nsSNPs, 53 were predicted to be likely harmful, 38 potentially

harmful, and 120 benign. A comparative evaluation of the findings from SIFT and PolyPhen revealed an overlap of 22 nsSNPs, which were classified as harmful or likely harmful, highlighting the significance of these genetic variations. Additionally, the PANTHER analysis identified 22 common nsSNPs, as summarized in Table 1. This convergence of results across multiple analytical methods underscores the potential impact of these nsSNPs on protein function and disease risk.

**Table 1** Prediction of Effects and consequences of SNPs by using bioinformatics tools

Variation ID	AA	SIFT		Polyphen		PANTHER		Polyphen2		Predict SNP	SuSpect	
		Score	Effect	Score	Effect	Score	Effect	Score	Effect	Effect	Score	Effect
rs752824015	M1V	0	D	0.935	PD	0.57	PD	0.993	PD	D	8	N
rs1403759603	M1I	0	D	0.956	PD	0.57	PD	0.993	PD	D	7	N
rs367612330	L47M	0	D	0.999	PD	0.57	PD	1	PD	D	32	D
rs565771057	R48C	0	D	0.999	PD	0.57	PD	1	PD	D	44	D
rs1373342928	R48H	0	D	0.999	PD	0.57	PD	1	PD	D	33	D
rs1191410166	C51R	0	D	0.999	PD	0.57	PD	1	PD	D	99	D
rs757247423	C51Y	0	D	0.999	PD	0.57	PD	1	PD	D	97	D
rs4608774	G72S	0	D	1	PD	0.57	PD	1	PD	D	12	N
rs896104907	P73L	0	D	0.999	PD	0.57	PD	1	PD	D	25	D
rs1213029852	C75G	0	D	0.998	PD	0.57	PD	1	PD	D	95	D
rs762611604	V78M	0	D	0.989	PD	0.19	PB	0.999	PD	D	41	D
rs140549348	L92M	0	D	0.998	PD	0.57	PD	1	PD	D	61	D
rs140549348	L92V	0	D	0.998	PD	0.57	PD	0.999	PD	D	51	D
rs975429889	P94L	0	D	0.999	PD	0.57	PD	1	PD	D	66	D
rs1239375597	P97T	0	D	0.984	PD	0.57	PD	1	PD	D	18	D
rs1471923911	P97R	0	D	0.986	PD	0.57	PD	0.999	PD	D	13	N
rs146538128	K101N	0.01	D	0.979	PD	0.5	PD	0.999	PD	D	38	D
rs771400203	P73S	0.02	D	0.999	PD	0.57	PD	1	PD	D	11	N
rs774289824	A37T	0.03	D	0.992	PD	0.19	PB	1	PD	D	20	D
rs1218826120	P39S	0.04	D	0.998	PD	0.19	PB	1	PD	N	22	D
rs768861968	Q104K	0.04	D	0.987	PD	0.19	PB	0.997	PD	N	13	N
rs774289824	A37S	0.05	D	0.992	PD	0.19	PB	0.997	PD	N	17	D

\*D = Deleterious, PD\* = Probably damaging, PB\* = Probably benign, in Suspect D\*= Disease, N\*= Neutral

### 3.3 Prediction of functional impact of mutation

Table 1 illustrates that among the 22 nsSNPs analyzed, 17 were predicted to be certainly detrimental. Conversely, several nsSNPs were classified as probably benign, including V78M (0.999), A37T (1), P39S (1), Q104K (0.997), and A37S (0.997). The PolyPhen-2 service employs protein sequences, phylogenetic trees, and structural characteristics to differentiate the effects of amino acid substitutions. Notably, according to Table 1, PolyPhen-2 indicated that all 22 nsSNPs were likely harmful, reinforcing the potential significance of these variants concerning protein function and disease susceptibility. The Predict SNP revealed that P39S (rs1218826120), Q104K (rs768861968), and A37S (rs774289824) were found to be neutral, and M1V (rs7528244015), M1I(rs1403759603), L47M (RS367612330), R48C (rs565771057), R48H (rs1373342928), C51R (rs1191410166), C51Y (rs757247423), G72S (rs4608774), P73L (rs896104907), C75G (rs1213029852), V78M (rs762611604), L92M (rs140549348), L92V (rs140549348), P94L (rs975429889), P97T (rs1239375597), P97R

(rs1471923911), K101N (rs146538128), P73S (rs771400203), A37T (rs774289824) were found to be deleterious effect as shown in Table 1.

### 3.4 Estimation of disease-associated nsSNPs

According to SNP&Go, R48C (rs565771057), R48H (rs1373342928), C51R (rs1191410166), C51Y (rs757247423), G72S (rs4608774), C75G (rs1213029852), L92V (rs140549348), P94L (rs975429889), and P97R (rs1471923911) were associated with diseases remaining 14 variants such as MIV (rs7528244015), M1I (rs1403759603), L47M (RS367612330), P73L (rs896104907), V78M (rs762611604), L92M (rs140549348), P97T (rs1239375597), K101N (rs146538128), P73S (rs771400203), A37T (rs774289824), P39S (rs1218826120), Q104K (rs768861968) and A37S (rs774289824) were neutral as shown in Table 2. Moreover, in MetaSNP, 10 nsSNPs such as MIV (rs7528244015), M1I (rs1403759603), V78M (rs762611604), L92V (rs140549348), K101N (rs146538128), P73S (rs771400203), A37T (rs774289824), P39S (rs1218826120), and Q104K

(rs768861968) were predicted as neutral, and 12 nsSNPs including L47M (RS367612330), R48C (rs565771057), R48H (rs1373342928), C51R (rs1191410166), C51Y (rs757247423), G72S (rs4608774), P73L (rs896104907), C75G

(rs1213029852), L92M (rs140549348), P94L (rs975429889), P97T (rs1239375597), and P97R (rs1471923911) were disease causing as shown in [Table 2](#).

**Table 2** Prediction of disease-associated nsSNPs and Protein stability prediction

Variation ID	Mutation	SNP & Go		Meta SNP		I Stable		Mu- Pro		DUET	
		Score	Effect	Score	Effect	Stability	Score	Stability	Score	Stability	Score
rs752824015	M1V	9	N	7	N	Increase	0.51605	Decrease	-0.2513237	DS	-0.071
rs1403759603	M1I	8	N	6	N	Decrease	0.759847	Decrease	-0.055192	DS	-0.086
rs367612330	L47M	3	N	3	D	Decrease	0.689869	Decrease	-1.0733077	DS	-0.226
rs565771057	R48C	3	D	4	D	Decrease	0.663238	Decrease	-0.2265095	DS	-0.106
rs1373342928	R48H	3	D	3	D	Decrease	0.602831	Decrease	-0.7868968	DS	-0.501
rs1191410166	C51R	6	D	9	D	Decrease	0.527775	Decrease	-1.1259131	S	0.3
rs757247423	C51Y	6	D	9	D	Decrease	0.533731	Decrease	-0.7310219	DS	-0.601
rs4608774	G72S	3	D	1	D	Decrease	0.75212	Decrease	-0.71767034	DS	-0.249
rs896104907	P73L	2	N	2	D	Increase	0.588488	Decrease	-1.562131	S	0.238
rs1213029852	C75G	4	D	8	D	Decrease	0.785371	Decrease	-1.562131	DS	-0.92
rs762611604	V78M	6	N	2	N	Decrease	0.817122	Decrease	-1.1247352	DS	-0.449
rs140549348	L92M	1	N	0	D	Decrease	0.830624	Decrease	-1.1321535	DS	-1.34
rs140549348	L92V	1	D	1	N	Decrease	0.857348	Decrease	-1.1321535	DS	-1.87
rs975429889	P94L	1	D	1	D	Increase	0.515991	Increase	0.33773694	S	0.065
rs1239375597	P97T	0	N	2	D	Decrease	0.785386	Decrease	-1.264408	DS	-0.243
rs1471923911	P97R	1	D	1	D	Decrease	0.74334	Decrease	-1.0604488	S	0.332
rs146538128	K101N	2	N	2	N	Increase	0.644964	Decrease	-0.0503005	DS	-0.069
rs771400203	P73S	3	N	2	N	Decrease	0.776627	Decrease	-0.4147564	DS	-0.089
rs774289824	A37T	5	N	6	N	Decrease	0.770042	Decrease	-0.67437669	DS	-0.579
rs1218826120	P39S	3	N	7	N	Decrease	0.795753	Decrease	-0.8979973	DS	-0.429
rs768861968	Q104K	5	N	7	N	Decrease	0.709301	Decrease	-1.2778399	S	0.209
rs774289824	A37S	6	N	7	N	Increase	0.620896	Decrease	-0.6154484	DS	-0.613

\*N= Neutral, \*D = Disease. In DUET \*DS= Destabilizing, \*S= Stability

To forecast the phenotypic impact of a non-synonymous alteration, suspects were utilized. Expected scores ranged from 0 to 100, with different colors denoting different degrees of deleterious. At the neutral end of the spectrum, it was blue, and at the disease-causing end, it was red. In prediction, the 9 nsSNPs MIV (rs7528244015), M1I (rs1403759603), G72S (rs4608774), P97T (rs1239375597), and P97R (rs1471923911), P73S (rs771400203), A37T (rs774289824), Q104K (rs768861968), and A37S (rs774289824) were considered to be neutral effects of nsSNPs, and the remaining 13 nsSNPs L47M (RS367612330), R48C (rs565771057), R48H (rs1373342928), C51R (rs1191410166), C51Y (rs757247423), P73L (rs896104907), C75G (rs1213029852), V78M (rs762611604), L92M (rs140549348), L92V (rs140549348), P94L (rs975429889), K101N (rs146538128), P39S (rs1218826120) were identified as disease-causing nsSNPs as shown in [Table 2](#).

### 3.5 Analysis of Protein Stability

Our results found i-Stable identified 5 mutations, including MIV (rs7528244015), P73L (rs896104907), P94L (rs975429889), K101N (rs146538128), and A37S (rs774289824), that were expected to increase the stability of the mutant protein, whereas the remaining 17 nsSNPs (M1I (rs1403759603), L47M (RS367612330), R48C (rs565771057), R48H

(rs1373342928), C51R (rs1191410166), C51Y (rs757247423), G72S (rs4608774), C75G (rs1213029852), V78M (rs762611604), L92M (rs140549348), L92V (rs140549348), P97T (rs1239375597), P97R (rs1471923911), P73S (rs771400203), A37T (rs774289824), P39S (rs1218826120), Q104K (rs768861968), and (rs774289824) were predicted to decrease the stability of the protein, hence lowering protein activity.

Protein stability changes can be assessed by the MUpro server using sequence-related information or a combination of sequence and tertiary structure data. In this analysis, 21 nsSNPs MIV (rs7528244015), M1I (rs1403759603), G72S (rs4608774), P97T (rs1239375597), and P97R (rs1471923911), P73S (rs771400203), A37T (rs774289824), Q104K (rs768861968), A37S (rs774289824), L47M (RS367612330), R48C (rs565771057), R48H (rs1373342928), C51R (rs1191410166), C51Y (rs757247423), P73L (rs896104907), C75G (rs1213029852), V78M (rs762611604), L92M (rs140549348), L92V (rs140549348), K101N (rs146538128), P39S (rs1218826120) were found with decreased stability, and 1 nsSNPs P94L (rs975429889) were found with increased stability as shown in [Table 2](#).

While the DUET server predicted 17 nsSNPs M1I (rs1403759603), MIV (rs7528244015), L47M (RS367612330), R48C (rs565771057), R48H

(rs1373342928), C51Y (rs757247423), G72S (rs4608774), C75G (rs1213029852), V78M (rs762611604), L92M (rs140549348), L92V (rs140549348), P97T (rs1239375597), A37S (rs774289824), K101N (rs146538128), P73S (rs771400203), A37T (rs774289824), P39S (rs1218826120) as destabilizing and 5 nsSNPs C51R (rs1191410166), P73L (rs896104907), P94L (rs975429889), P97R (rs1471923911), Q104K (rs768861968) as stabilizing as shown in [Table 2](#).

### 3.6 Predicting Mutation Clusters

The mutation 3D web server was used to study the *CXCL6* protein. With all 22 nsSNPs provided as input, a mutation involving amino acid residues at locations MIV and M1I was uncovered mutations. According to [Table 3](#), a cluster of mutations L47M, R48H, R48C, G72S, P73L, C75G, P73S, A37T, P39S, and A37S. Another set of covered mutations included C51R, C51Y, V78M, L92M, L92V, P94L, P97T, P97R, K101N, and Q104K.

**Table 3** Functional prediction and visualization of *CXCL6* gene by mutation 3D

Variation ID	Mutation	Prediction	Variation ID	Mutation	Prediction
rs752824015	M1V	Uncovered mutation	rs1213029852	C75G	Clustered mutation
rs1403759603	M1I	Uncovered mutation	rs762611604	V78M	Covered mutation
rs367612330	L47M	Clustered mutation	rs140549348	L92M	Covered mutation
rs565771057	R48C	Clustered mutation	rs140549348	L92V	Covered mutation
rs1373342928	R48H	Clustered mutation	rs975429889	P94L	Covered mutation
rs1191410166	C51R	Covered mutation	rs1239375597	P97T	Covered mutation
rs757247423	C51Y	Covered mutation	rs1471923911	P97R	Covered mutation
rs4608774	G72S	Clustered mutation	rs146538128	K101N	Covered mutation
rs896104907	P73L	Clustered mutation	rs771400203	P73S	Clustered mutation
rs768861968	Q104K	Covered mutation	rs774289824	A37T	Clustered mutation
rs774289824	A37S	Clustered mutation	rs1218826120	P39S	Clustered mutation

### 3.7 SWISS Modeling for the *CXCL6* Protein

We studied 22 highly conserved *CXCL6* mutations according to the prediction score to identify the changes in protein structure these mutations caused. Below is a list of the 22 nsSNPs that have been identified: L47M, R48C, R48H, C51R, C51Y, C51Y, G72S, P73L, C75G, V78M, L92M, L92V, P94L, P97T, P97R, K101N, P73S, P39S, Q104K, and A37S. The produced sequences were chosen for comparative homology modeling based on their identity and similarity levels, which must be greater than 30%. We have obtained fifty templates with the STML ID P80162 and the query sequence must be identified. Using the query sequence as a basis, we constructed a three-dimensional model of the *CXCL6* protein using the AlphaFold DB model of AF-P80162-F1 (Organism: *Homo sapiens*) and the template PDB ID P80162.1 (range: 1-114 aa; coverage: 1.00). [Fig. 3](#) displays the outcomes of creating a mutant model using a template (P80162.1.A) and PyMoL that shows the visual representation of the results obtained from a template using model 05, which was improved using Galaxy Refine. When estimating the model's quality using QMEAN, a QMEAN Z-score of -3.10 was produced for experimental structures of comparable sizes. Protein structural instability is indicated by a negative QMEAN Z-score. [Fig. 3](#) shows the P80162.1.A model as a red star. The full model of the desired gene is not reported in PDB.

We modified the proteins mentioned above using PyMol after downloading their corresponding PDB files. [Table 4](#) shows that the mutants known as L47M, C51R, L92M, L92V, P73S, and Q104K have significant RMSD values. We validated the modeled framework

using SAVES and examined the secondary structure using RAMACHANDRAN plot evaluation. The vast majority of the *CXCL6* protein's amino acid residues (82.6%) were located in an extremely favorable location on a RAMACHANDRAN plot, as seen in [Fig. 3](#). For a full rundown of all the expected outcomes, see [Table 4](#). [Fig. 3](#) displays the ERRAT-identified projected model's total quality, which is 87.9518%.

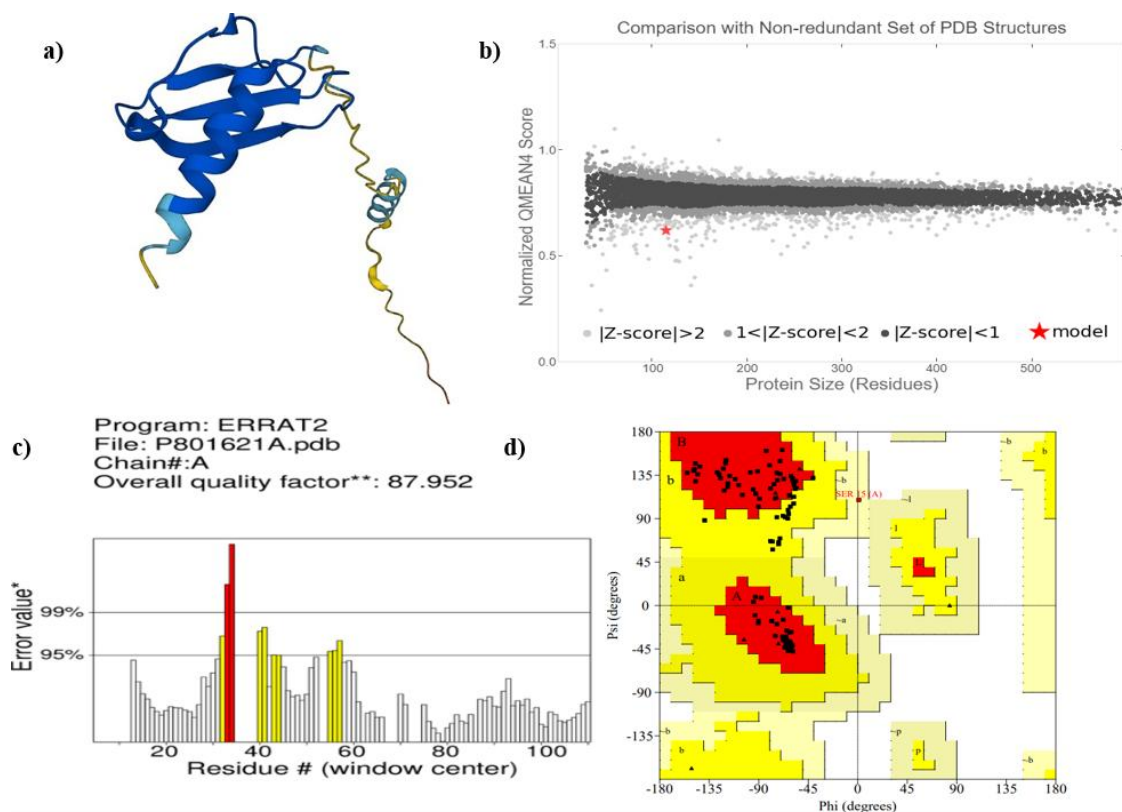
### 3.8 Effect of polymorphism by HOPE analysis

HOPE program is used to analyze the difference between wild-type and mutant amino acids by evaluating their size, charge, hydrophobicity value, conserved location, and how different amino acid residues affect the domain. There is a larger effect of variation at rs367612330 (L47M) compared to the wild-type residue. The wild type is extremely preserved, and this might cause bumps. Because the mutant residue is positioned inside a particular domain, changing it may cause the domain to no longer operate as intended. In the case of rs1191410166 (C51R), This makes the mutant residue smaller and less hydrophobic than the wild-type one. Furthermore, the mutant residue has a positive charge, whereas the wild-type residue has no charge at all. Damage to the protein is probably inevitable due to this mutation. The mutant residue at rs757247423 (C51Y) is bigger and less hydrophobic than the wild-type residue. This mutation is likely harmful to the protein because hydrophobic interactions are lost. The mutant residue at rs1213029852 (C75G) is less hydrophobic and smaller than the wild-type residue. Damage to the protein is probably inevitable due to this mutation. At



residue rs1239375597, P97T is relatively well conserved and, compared with the mutant residue, it shows higher hydrophobicity. According to the

conservation scores reported in Supplementary **Table S1**, it seems likely that this mutation is detrimental to the protein.



**Fig. 3** Structural modeling and Validation (a) Structure of *CXCL6* (b) QMEAN (c) Procheck-RAMACHANDRAN plot (d) ERRAT of the native *CXCL6* predicted Model

**Table 4** Structural validation and comparison by using online tools

Protein	ERRAT	Procheck				Verify	TM Align	
	Score	Core	Allow	Generously	Disallowed	Score	Tm Score	RMSD
P80162.1.A	87.9518	82.60%	16.30%	1.10%	0.00%	19.30%		
M1V	88.4615	82.60%	16.30%	1.10%	0.00%	19.30%	0.97767	0.62
M1I	96.2501	82.60%	16.30%	1.10%	0.00%	19.30%	0.9773	0.63
L47M	94.1176	82.60%	16.30%	1.10%	0.00%	19.30%	0.97491	0.66
R48C	96.3415	82.60%	16.30%	1.10%	0.00%	19.30%	0.97771	0.62
R48H	84.0909	82.60%	16.30%	1.10%	0.00%	19.30%	0.97673	0.63
C51R	87.8049	82.60%	16.30%	1.10%	0.00%	21.93%	0.97367	0.68
C51Y	92.9412	82.60%	16.30%	1.10%	0.00%	19.30%	0.97818	0.61
G72S	98.7654	81.70%	17.20%	1.10%	0.00%	19.30%	0.97667	0.63
P73L	93.75	82.80%	16.10%	1.10%	0.00%	19.30%	0.97772	0.62
C75G	94.8718	82.40%	16.50%	1.10%	0.00%	29.82%	0.978	0.62
V78M	97.2222	82.60%	16.30%	1.10%	0.00%	19.30%	0.97931	0.6
L92M	90.5882	82.60%	16.30%	1.10%	0.00%	19.30%	0.97478	0.66
L92V	91.4634	82.60%	16.30%	1.10%	0.00%	16.67%	0.97478	0.66
P94L	86.8421	82.80%	16.10%	1.10%	0.00%	3.51%	0.976	0.64
P97T	94.5205	82.80%	16.10%	1.10%	0.00%	19.30%	0.97801	0.62
P97R	91.5663	82.80%	16.10%	1.10%	0.00%	7.02%	0.97618	0.65
K101N	97.4026	82.60%	16.30%	1.10%	0.00%	19.30%	0.97811	0.61
P73S	95.7143	82.80%	16.10%	1.10%	0.00%	19.30%	0.97543	0.66
A37T	90.9091	82.60%	16.30%	1.10%	0.00%	19.30%	0.97555	0.65
P39S	90.4635	81.70%	17.20%	1.10%	0.00%	19.30%	0.97901	0.6
Q104K	97.5	82.60%	16.30%	1.10%	0.00%	19.30%	0.975	0.66
A37S	87.3418	82.60%	16.30%	1.10%	0.00%	19.30%	0.97838	0.61

### 3.9 Molecular Docking by PyRx

We employed the PyRx tool to investigate ligand-protein interactions via molecular docking, docking all selected ligands with *CXCL6*. Each ligand was modeled with 10 distinct conformations, with a bind-

ing affinity (-Kcal/mol) serving as the defining metric. Supplementary Table S2 lists the binding affinities of five out of the 22 compounds analyzed, demonstrating a correlation between these affinities and the activity levels of the ligands. To further examine their interactions, we docked five compounds with signific-

ant binding affinities IND24, amphotericin, lectin, pyranomentoflavone, and laminin against all six of our normal and mutant protein complexes. Notably, all selected ligands exhibited binding free energies greater than  $-4$  Kcal/mol, indicating favorable interactions.

The *CXCL6* protein displayed the highest binding energy after docking with amphotericin. Among the several ligand-binding affinities, amphotericin and lectin exhibit the highest, at  $-8.5$  and  $-7.9$  Kcal/mol, respectively. For a comprehensive analysis of the protein's structural properties, we employed PLIP (Protein-Ligand Interaction Profiler), as illustrated in Fig. 4. The findings from PLIP were subsequently

visualized using Discovery Studio, allowing for detailed insights into the interactions and structural characteristics of the protein. Amphotericin B, an antifungal medication, interacts with the cancer cell surface protein *CXCL6* to influence gene expression and cellular processes, ultimately leading to the treatment of cancer. Studies have demonstrated that amphotericin B can induce angiogenesis in oxidative stress-sensitive hepatocellular carcinoma cells by upregulating angiogenic genes (Hayashi et al., 2023). Cancer and immune-related illnesses can be influenced by lectins and *CXCL6* (Faysal Ahmed et al., 2023).

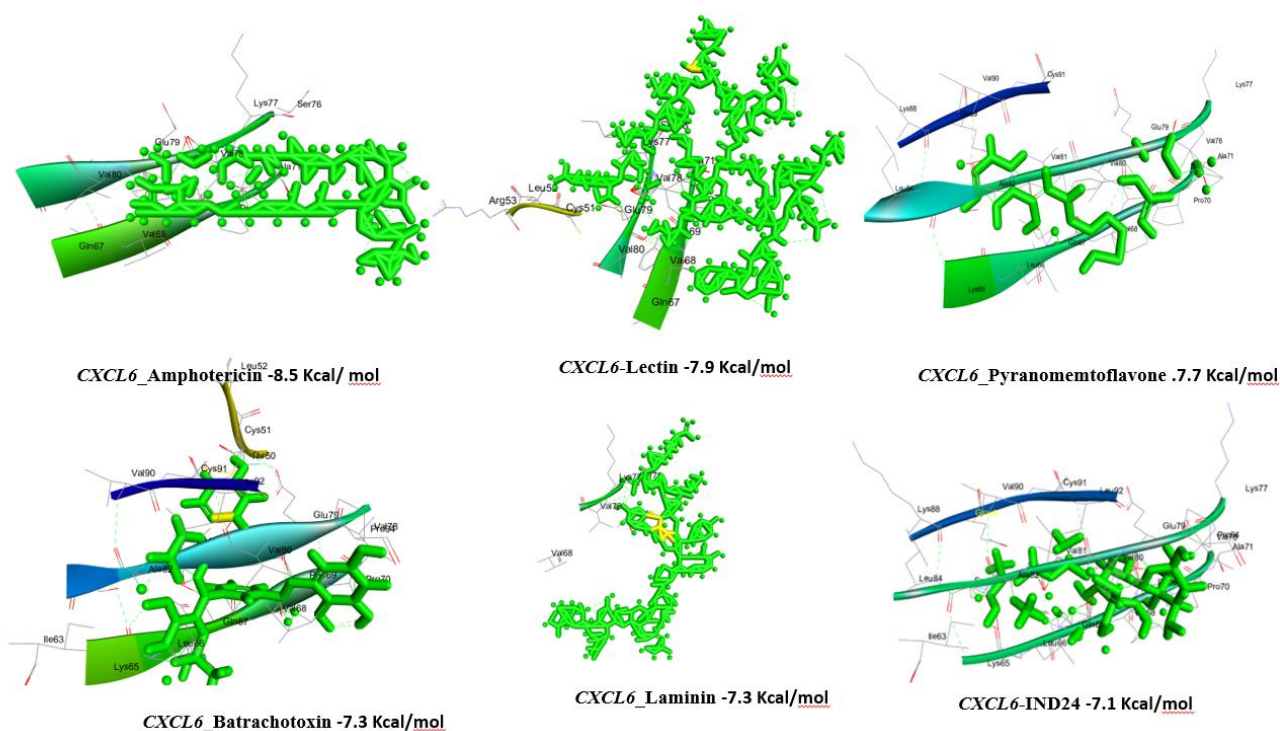


Fig. 4 Protein-ligand binding energy structure of protein

A study conducted by (Gupta, 2020) revealed that A potential strategy for gene therapy targeting cancer cells involves the use of lectins, which can bind selectively to carbohydrates found on the surfaces of cancer cells. The molecular pathways uncovered by docking interactions between lectins and the *CXCL6* gene could potentially lead to the development of targeted therapeutics for cancer and immune-related disorders. The lectin molecule interacted with the wild-type *CXCL6* protein, as shown in Table 5. Pyranomentoflavone and the other synthesized pyrano[3,2-c]pyranones demonstrated potent growth inhibitory effects on the MCF-7 breast cancer cells (Burns & Helsby, 2023). In a selective method, To determine which cytochrome P450 enzymes impede the conversion of procarcinogens to carcinogenic metabolites, researchers have examined pyranomentoflavone and other flavone derivatives (Sak, 2021; Khodair et al., 2023). These studies suggest that pyranomentoflavone is a promising candidate for

cancer treatment due to its molecular interactions with DNA-topoisomerase complexes and its selective growth inhibitory effects on cancer cells. Table 5 presents the docking contacts between the pyranomentoflavone molecule and the wild-type *CXCL6* protein, highlighting the specific interactions that may contribute to its therapeutic potential.

The interaction between laminin and several genes and signaling pathways is critical to the development of cancer. According to research (Rousselle & Scoazec, 2020), laminin-332 in particular interacts with integrin receptors to facilitate the migration and invasion of malignant cells. Also, (Wattanathavorn et al., 2024) say that circRNA circ\_0006089 is very important in gastric cancer because it encourages cancer cells to behave in a not good way through the miR-515-5p/*CXCL6* axis. A complex web of molecular processes facilitates the formation and progression of cancer, as illustrated by the interaction between circ\_0006089 and *CXCL6*.

Finding out how laminin is connected to genes like *CXCL6* is important for creating new cancer treatments because it leads to possible therapeutic targets (Q. Zhu et al., 2012). Table 5 displays the docking interaction between the laminin molecule and the wild-type *CXCL6* protein.

Docking batrachotoxin with cancer-related proteins can offer valuable insights into cancer channelopathies and potential treatment options (Sahayarayan et al., 2021). This study will examine how the toxin interacts with voltage-gated channels. Table 5 displays the docking interaction between the batrachotoxin molecule and the wild-type *CXCL6* protein. Docking experiments showed that the IND24 ligand interacted with the *CXCL6* gene, which may indicate a binding site on the gene. Since *CXCL6* is an important chemokine in cancer cell migration and invasion, this interaction might be critical in regulating these activities. In 2021 Ghosh et al. (2021) further investigation into the IND24-*CXCL6* connection and its possible implications for cancer therapy is required. According to Table 5, the IND24 chemical bound to the *CXCL6* protein in its wild-type form interactions.

**Table 5** Protein ligand interactions with wild-type *CXCL6* gene

Ligands	Interactions
Amphotericin	Lys77,ser76,Glu79,Val80,Val78, Val68,Gln67,Ala71
Lectin	Arg53,Leu52,Cys51,Lys77, Glu79,Val78,Val80, Val68, Gln67, Pro70, Phe69, Ala71
Pyranomentoflavone	Val90, Cys91, Cys51, Ala82,Ile63, Lys65,Leu66, Gln67, Val68, Phe69,Val78, Pro94,Glu79,Leu92,Leu52
Batrachotoxin	Lys88,Gln89, Val90, Cys91, Leu92, Glu79, Lys77, Val80, Val81, Leu84, Ile63, Lys65, Leu66, Pro70, Ala71, Lys77, Glu79
Laminin	Val68, Val78, Lys77, Ser76, Ala71
IND24	Lys88, Val90, Cys91, Gln89, Leu84, Ala82, Val81, Val80, Glu79, Lys77, Val78, Ala71, Pro70, Val68, Gln67, Leu66, Lys65

#### 4. Discussion

The chemokine *CXCL6*, encoded by the *CXCL6* gene, is crucial for regulating the immune response and has been associated with various cancers. A 2014 research study by Aldinucci and Colombatti found that *CXCL6*, also known as granulocyte chemotactic protein-2 (GCP-2), draws immune cells such as neutrophils to regions surrounding tumors. This chemokine influences the formation of new blood vessels, the growth of tumors, metastasis, or the spread of cancer (Aldinucci & Colombatti, 2014). In several cancers, increased *CXCL6* expression is associated with tumor growth and poor prognosis (Do et al., 2020). Gaining a deep understanding of *CXCL6*'s role in cancer is crucial for developing targeted treatments to inhibit its activity, thereby reducing tumor growth and metastasis. Research has linked several single nucleotide polymorphisms

(SNPs) to various cancers, including breast, colon, ovarian, and lung cancers. Sameer and Nissar (2021) report that there are several significant SNPs associated with these conditions. Multiple SNPs and other alterations have been associated with the *CXCL6* gene in recent years. However, most of the SNPs in this region are still unknown. Our research set out to fill this void (Sameer and Nissar, 2021).

We have taken into account all known *CXCL6* SNPs in this analysis. Due to the possibility of insufficient results from a single bioinformatics tool, we selected twelve reliable prediction methods for identifying harmful missense SNPs. Finding the most detrimental nsSNPs in the *CXCL6* gene was facilitated by employing a combined illustration of such variants. This was created by combining multiple computer programs: SIFT, Polyphen, Polyphen2, Predict SNP, MutPred2, SNP &Go, Meta SNP, Suspect, i Stable, Mu-Pro, and DUET. The *CXCL6* gene has 22 nsSNPs that have been found to have adverse, uncertain, or disease-causing consequences, according to functional studies. Based on their predictions, SNP & Go found that nine nsSNPs were associated with disease-causing SNPs, nine nsSNPs were linked to Meta SNP 10, and thirteen nsSNPs were linked to Suspect 13. DUET server, Mu Pro, and i-stable allowed us to predict how these 22 would affect protein stability. To comprehend the functional effects of twenty-two *CXCL6* mutations, we employ the mutation 3D tool. By doing so, we can learn a great deal about gene function, disease mechanisms, and personalised medicine (Nussinov et al., 2019). This approach allows us to predict how mutations might affect protein function, which is key to understanding how they can either worsen or improve disease conditions.

From our analysis, we identified two uncovered mutations out of twenty-two, clustered ten others, and examined ten are as covered, all using 3D mutation modeling. In order to ensure that the 22 nsSNPs are structurally valid, we conduct further analyses on them. To improve the quality of the structure of the targeted proteins, it is essential to validate the experimental model. We used various computational tools such as SWISS-MODEL, PROCHECK, QMEAN, ERRAT, check 3D, and TM-Align to confirm the experimental model. Nevertheless, we submitted our query sequence to the homology-modeling server SWISS-MODEL. The server generated the one hundred and sixteenth template, P801621, out of fifty templates based on the best alignment. *CXCL6* covered the entire sequence of our target protein. The model falls within the amino acid range of 1–114, suggesting potential conservation in this area of the protein sequence. The Ramachandran plot is given top priority among all the verification matrices because it shows the  $\phi$ - $\psi$  torsion angles of the projected models' protein backbone. The

Ramachandran plot is split into four areas: the core, the allowed, the generously allowed, and the disallowed region. PROCHECK shows the stereochemical quality of a protein structure, and SWISS-MODEL shows the favored region. Over 90% of residues in the core region or the most favored region can achieve a favorable structure. The *CXCL6* protein appears to be in a stable and energetically advantageous shape, as 82.6% of its amino acid residues fall inside a highly favorable area, according to a Ramachandran plot analysis. Other computational techniques provide scores for protein model quality estimation. With identically sized experimental structures, the QMEAN Z-score for model quality estimation was -3.10. According to AboElkhair et al. (2021), a model is considered to have low quality if its QMEAN-Z score is -4.0 or lower. On the other hand, a higher score shows that the structure is in a favorable state. Using the TM-align tool, we structurally compared the wild-type and mutant structures.

When the TM score is low and the RMSD value is high, there is a structural dissimilarity. According to the Swiss model, the produced structure is of high quality and suitable for protein-ligand interactions. We proceeded to analyze 1 model and 6 mutant proteins based on the average score of all validation software. We conducted docking for 25 medicines and their five probable targets using the software PyRx Auto Dock Vina 0.4. The software provided docking scores as well as energy minimization values. Out of a total of 25 ligand and 6 protein interactions, the molecular docking studies found the following drug-target interactions to have the best scores. The drug-target interactions with the highest scores were the Model-Amphotericin complex (-8.5), Model-Lectin complex (-7.9), Model-Pyranomentoflavone complex (-7.7), Model-Laminin complex (-7.3), Model-IND24 complex (-7.1), and Model-Batrachotoxin complex (7.3). We used Pyrx and the Discovery Studio Visualizer to see how the ligand-target complex interacted.

We can infer the impact of a harmful mutation on a protein from its evolutionary conservation profile. Negative effects are more likely to be caused by nsSNPs found in highly conserved regions compared to those found in variable regions (Yazar & Özbek, 2021). We examined the possible consequences of the most harmful nsSNPs—L47M, C51R, C51Y, C75G, and P97T using the Hope server. We also ran HOPE on the *CXCL6* protein structure to investigate the impact of polymorphism, size, position, amino acid characteristics, variations, and variant locations of the five most harmful nsSNPs listed above. The hope server predicted that 5 nsSNPs could potentially damage the protein structure. Future research is needed to understand pathogenicity, protein stability, and nsSNPs linked to disease. Genetic testing and personalized treatment for *CXCL6*-related maligna-

ncies will become more accurate with the help of new computational methods and docking studies. With better in silico models, we can learn more about nsSNPs and their effects, which can lead to more efficient drug discovery.

More research into the structural and functional effects of these SNPs utilizing state-of-the-art computational methods and molecular dynamics models is an exciting promise for the future. Enhancing our understanding of the molecular pathways behind *CXCL6*-related malignancies will pave the way for the development of more effective treatment approaches. Experimental validation using wet lab techniques will provide more proof of the SNPs' involvement in cancer genesis and progression. This information will enable the development of targeted medicines that aim to disrupt the *CXCL6* signaling pathway.

## 5. Conclusion

The *CXCL6* gene plays a pivotal role in tumor growth and progression by promoting angiogenesis and metastasis. Its overexpression in breast, lung, and colon cancers is associated with poorer prognosis and shorter survival rates. Targeted therapies that inhibit the *CXCL6* pathway are critical for effective early detection and treatment of cancer. Our study identified 22 nsSNPs that significantly impact the *CXCL6*, underscoring the necessity for future genome association studies and in silico approaches for rapid and cost-effective screening. We modeled the 3D *CXCL6* protein and identified the mutations L47M, C51R, L92M, L92V, P73S, and Q104K, which were docked with 25 compounds. The 2D and 3D interactions of both the wild-type and mutant proteins demonstrated high binding scores with ligands such as Amphotericin, Lectin, Pyranomentoflavone, Batrachotoxin, Laminin, and IND. Our research aims to elucidate the mechanisms by which *CXCL6* contributes to cancer progression, ultimately developing therapeutic targets that enhance patient outcomes. Understanding the *CXCL6* mutations drive cancer development is crucial for advancing treatment strategies.

## Acknowledgment

We all grateful to our supervisor Dr. Hamna Tariq and our mentor Kainat Ramzan for their guidance in this study.

## Declaration of Conflict

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.






## References

1. AboElkhair, M.A., Hasan, M.E., Mousa, A., Moharam, I., Sultan, H., Malik, Y. & Sakr, M.A. (2021) In-silico evidence for enhancement of avian influenza virus H9N2 virulence by modulation of its hemagglutinin (HA) antigen function and stability during co-infection with infectious bronchitis virus in chickens. *VirusDisease*, 32, 548–558. DOI: [10.1007/s13337-021-00688-1](https://doi.org/10.1007/s13337-021-00688-1), PubMed: [34631979](https://pubmed.ncbi.nlm.nih.gov/34631979/).
2. Adasme, M.F., Linnemann, K.L., Bolz, S.N., Kaiser, F., Salentin, S., Haupt, V.J. & Schroeder, M. (2021) PLIP 2021: Expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Research*, 49, W530–W534. DOI: [10.1093/nar/gkab294](https://doi.org/10.1093/nar/gkab294), PubMed: [33950214](https://pubmed.ncbi.nlm.nih.gov/33950214/).
3. Adhikari, S., Nath, P., Das, A., Datta, A., Baildya, N., Duttaroy, A.K. & Pathak, S. (2024) A review on metal complexes and its anti-cancer activities: Recent updates from in vivo studies. *Biomedicine and Pharmacotherapy*, 171, 116211. DOI: [10.1016/j.biopha.2024.116211](https://doi.org/10.1016/j.biopha.2024.116211), PubMed: [38290253](https://pubmed.ncbi.nlm.nih.gov/38290253/).
4. Akhtar, A., Choudhry, S.N., Mateen, R.M. & Hussain, M. (2020) A comprehensive in silico analysis of deleterious SNPs of paraplegin protein associated with hereditary spastic paraplegia through mitochondrial dysfunction. *BioScientific Review*, 2, 1–14. DOI: [10.32350/BSR.0202.01](https://doi.org/10.32350/BSR.0202.01).
5. Aldinucci, D. & Colombatti, A. (2014) The inflammatory chemokine CCL5 and cancer progression. *Mediators of Inflammation*, 2014, 292376. DOI: [10.1155/2014/292376](https://doi.org/10.1155/2014/292376), PubMed: [24523569](https://pubmed.ncbi.nlm.nih.gov/24523569/).
6. Aljindan, R.Y., Al-Subaie, A.M., Al-Ohali, A.I., Kumar D, T., Doss C, G.P. & Kamaraj, B. (2021) Investigation of nonsynonymous mutations in the spike protein of SARS-CoV-2 and its interaction with the ACE2 receptor by molecular docking and MM/GBSA approach. *Computers in Biology and Medicine*, 135, 104654. DOI: [10.1016/j.compbiomed.2021.104654](https://doi.org/10.1016/j.compbiomed.2021.104654), PubMed: [34346317](https://pubmed.ncbi.nlm.nih.gov/34346317/).
7. An, X., Zhang, W., Rong, C. & Liu, S. (2023) Understanding Ramachandran plot for dipeptide: A density functional theory and information-theoretic approach study. *Journal of the Chinese Chemical Society*, 70, 243–252. DOI: [10.1002/jccs.202200444](https://doi.org/10.1002/jccs.202200444).
8. Arega, A.M., Dhal, A.K., Pattanaik, K.P., Nayak, S. & Mahapatra, R.K. (2024) An immunoinformatics-based study of Mycobacterium tuberculosis region of difference-2 uncharacterized protein (Rv1987) as a potential subunit vaccine candidate for preliminary ex vivo analysis. *Applied Biochemistry and Biotechnology*, 196, 2367–2395. DOI: [10.1007/s12010-023-04658-9](https://doi.org/10.1007/s12010-023-04658-9), PubMed: [37498378](https://pubmed.ncbi.nlm.nih.gov/37498378/).
9. Azmi, M.B., Khan, W., Azim, M.K., Nisar, M.I. & Jehan, F. (2023) Identification of potential therapeutic intervening targets by in-silico analysis of nsSNPs in preterm birth-related genes. *PLOS ONE*, 18, e0280305. DOI: [10.1371/journal.pone.0280305](https://doi.org/10.1371/journal.pone.0280305), PubMed: [36881567](https://pubmed.ncbi.nlm.nih.gov/36881567/).
10. Banerjee, A., Kanwar, M., Santra, D. & Maiti, S. (2022) Global conserved RBD fraction of SARS-CoV-2 S-protein with T500S mutation in silico significantly blocks ACE2 and rejects viral spike. *Translational Medicine Communications*, 7, 2. DOI: [10.1186/s41231-022-00109-5](https://doi.org/10.1186/s41231-022-00109-5), PubMed: [35136839](https://pubmed.ncbi.nlm.nih.gov/35136839/).
11. Chen, C., Lin, C.J., Pei, Y.C., Ma, D., Liao, L., Li, S.Y., Fan, L., Di, G.H., Wu, S.Y., Liu, X.Y., Wang, Y.J., Hong, Q., Zhang, G.L., Xu, L.L., Li, B.B., Huang, W., Shi, J.X., Jiang, Y.Z., Hu, X. & Shao, Z.M. (2023) Comprehensive genomic profiling of breast cancers characterizes germline-somatic mutation interactions mediating therapeutic vulnerabilities. *Cell Discovery*, 9, 125. DOI: [10.1038/s41421-023-00614-3](https://doi.org/10.1038/s41421-023-00614-3), PubMed: [38114467](https://pubmed.ncbi.nlm.nih.gov/38114467/).
12. Chen, G., Seukep, A.J. & Guo, M. (2020) Recent advances in molecular docking for the research and discovery of potential marine drugs. *Marine Drugs*, 18, 545. DOI: [10.3390/md18110545](https://doi.org/10.3390/md18110545), PubMed: [33143025](https://pubmed.ncbi.nlm.nih.gov/33143025/).
13. Dakshitha, S., Priya Dharshini, B., Suresh, V. & Dilipan, E. (2024) Computational exploration of single-nucleotide polymorphisms in the human hRAS gene: Implications and insights. *Cureus*, 16, e53119. DOI: [10.7759/cureus.53119](https://doi.org/10.7759/cureus.53119), PubMed: [38420094](https://pubmed.ncbi.nlm.nih.gov/38420094/).
14. Do, H.T.T., Lee, C.H. & Cho, J. (2020) Chemokines and their receptors: Multifaceted roles in cancer progression and potential value as cancer prognostic markers. *Cancers*, 12, 287. DOI: [10.3390/cancers12020287](https://doi.org/10.3390/cancers12020287), PubMed: [31991604](https://pubmed.ncbi.nlm.nih.gov/31991604/).
15. Gerke, M.B., Jansen, C.S. & Bilén, M.A. (2024) Circulating tumor DNA in genitourinary cancers: Detection, prognostics, and therapeutic implications. *Cancers*, 16. DOI: [10.3390/cancers16122280](https://doi.org/10.3390/cancers16122280), PubMed: [38927984](https://pubmed.ncbi.nlm.nih.gov/38927984/).
16. Ghorbanpour, M., Shayanfar, A. & Soltani, B. (2024) Copper pyrazole complexes as potential anticancer agents: Evaluation of cytotoxic response against cancer cells and their mechanistic action at the molecular level. *Coordination Chemistry Reviews*, 498, 215459. DOI: [10.1016/j.ccr.2023.215459](https://doi.org/10.1016/j.ccr.2023.215459).
17. Goswami, V., Patel, D., Rohit, S., Chaube, U. & Patel, B. (2024) Homology modeling, binding site identification, molecular docking and molecular dynamics simulation study of emerging and promising drug target of Wnt signaling–Human Porcupine enzyme. *Results in Chemistry*, 7, 101482. DOI: [10.1016/j.rechem.2024.101482](https://doi.org/10.1016/j.rechem.2024.101482).
18. Harihar, B., Saravanan, K.M., Gromiha, M.M. & Selvaraj, S. (2024) Importance of inter-residue contacts for understanding protein folding and unfolding rates, remote homology, and drug design. *Molecular Biotechnology*, 1–23. DOI: [10.1007/s12033-024-01119-4](https://doi.org/10.1007/s12033-024-01119-4), PubMed: [38498284](https://pubmed.ncbi.nlm.nih.gov/38498284/).
19. Hoffman, J.I., Vendrami, D.L.J., Hench, K., Chen, R.S., Stoffel, M.A., Kardos, M., Amos, W., Kalinowski, J., Rickert, D., Köhrer, K., Wachtmeister, T., Goebel, M.E., Bonin, C.A., Gulland, F.M.D. & Dasmahapatra, K.K. (2024) Genomic and fitness consequences of a near-extinction event in the northern elephant seal. *Nature Ecology and Evolution*. DOI: [10.1038/s41559-024-02533-2](https://doi.org/10.1038/s41559-024-02533-2), PubMed: [39333394](https://pubmed.ncbi.nlm.nih.gov/39333394/).
20. Huang, P.P., Wu, T.T., Tuo, M.Q., Ge, J., Huang, P., Wang, W.Q., Yang, J.P., Pan, H.B. & Lu, J.F. (2024) Supramolecular complexes of Co (II). *Journal of Molecular Structure*, 1307, 138061. DOI: [10.1016/j.molstruc.2024.138061](https://doi.org/10.1016/j.molstruc.2024.138061).
21. Kantelis, K.F., Asteriou, V., Papadimitriou-Tsantarliotou, A., Petrou, A., Angelis, L., Nicopolitidis, P., Papadimitriou, G. & Vizirianakis, I.S. (2022) Graph theory-based simulation tools for protein structure

- networks. *Simulation Modelling Practice and Theory*, 121, 102640. DOI: [10.1016/j.simpat.2022.102640](https://doi.org/10.1016/j.simpat.2022.102640).
22. Khamlich, J. et al. (2023) In silico analysis predicting the structural and functional effects of high-risk nsSNPs in the human GCK gene associated with gestational diabetes. *Physical Chemistry Research*, 11, 653–673.
  23. Komolafe, K., Pacurari, M. & Chemokines, C.X.C. (2022) CXC Chemokines in the Pathogenesis of Pulmonary Disease and Pharmacological Relevance. In: *International Journal of Inflammation*, 2022, 4558159. DOI: [10.1155/2022/4558159](https://doi.org/10.1155/2022/4558159), PubMed: [36164329](https://pubmed.ncbi.nlm.nih.gov/36164329/).
  24. Kumar, N. et al. (2024) Bioinformatics tools to study homology modeling. *Computational Biology in Drug Discovery and Repurposing*, 75.
  25. Kurma, K. & Alix-Panabières, C. (2023) Mechanobiology and survival strategies of circulating tumor cells: A process towards the invasive and metastatic phenotype. *Frontiers in Cell and Developmental Biology*, 11, 1188499. DOI: [10.3389/fcell.2023.1188499](https://doi.org/10.3389/fcell.2023.1188499), PubMed: [37215087](https://pubmed.ncbi.nlm.nih.gov/37215087/).
  26. Laskar, F.S., Bappy, M.N.I., Hossain, M.S., Alam, Z., Afrin, D., Saha, S. & Ali Zinnah, K.M. (2023) An in silico Approach towards Finding the Cancer-Causing Mutations in Human MET Gene. *International Journal of Genomics*, 2023, 9705159. DOI: [10.1155/2023/9705159](https://doi.org/10.1155/2023/9705159), PubMed: [37200850](https://pubmed.ncbi.nlm.nih.gov/37200850/).
  27. Mohd Abas, M.D., Mohd Asri, M.F., Yusafawi, N.A.S., Rosman, N.A.Z., Baharudin, N.A.Z., Taher, M., Susanti, D. & Khotib, J. (2024) Advancements of gene therapy in cancer treatment: A comprehensive review. *Pathology, Research and Practice*, 261, 155509. DOI: [10.1016/j.prp.2024.155509](https://doi.org/10.1016/j.prp.2024.155509), PubMed: [39121791](https://pubmed.ncbi.nlm.nih.gov/39121791/).
  28. Mohkam, M., Golkar, N., Nabavizadeh, S.H., Esmaeilzadeh, H., Berenjian, A., Ghahramani, Z., Golami, A. & Alyasin, S. (2022) In silico evaluation of nonsynonymous SNPs in human ADAM33: The most common form of genetic association to asthma susceptibility. *Computational and Mathematical Methods in Medicine*, 2022, 1089722. DOI: [10.1155/2022/1089722](https://doi.org/10.1155/2022/1089722), PubMed: [36411793](https://pubmed.ncbi.nlm.nih.gov/36411793/).
  29. Ni, B., Wang, H., Khalaf, H.K.S., Blay, V. & Houston, D.R. (2024) AutoDock-SS: AutoDock for multiconformational ligand-based virtual screening. *Journal of Chemical Information and Modeling*, 64, 3779–3789. DOI: [10.1021/acs.jcim.4c00136](https://doi.org/10.1021/acs.jcim.4c00136), PubMed: [38624083](https://pubmed.ncbi.nlm.nih.gov/38624083/).
  30. Nussinov, R., Jang, H., Tsai, C.J. & Cheng, F. (2019) Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLOS Computational Biology*, 15, e1006658. DOI: [10.1371/journal.pcbi.1006658](https://doi.org/10.1371/journal.pcbi.1006658), PubMed: [30921324](https://pubmed.ncbi.nlm.nih.gov/30921324/).
  31. Panchal, N.K., Mohanty, S. & Prince, S.E. (2024) Computational insights into NIMA-related kinase 6: unraveling mutational effects on structure and function. *Molecular and Cellular Biochemistry*, 479, 2989–3009. DOI: [10.1007/s11010-023-04910-0](https://doi.org/10.1007/s11010-023-04910-0), PubMed: [38117419](https://pubmed.ncbi.nlm.nih.gov/38117419/).
  32. Rehman, H.M. et al. (2024) A comprehensive in silico study of the NDB-IL-24 fusion protein for tumor targeting: A promising anticancer therapeutic candidate. *Journal of Biological Regulators and Homeostatic Agents*, 38, 3449–3461.
  33. Rozario, L.T., Sharker, T. & Nila, T.A. (2021) In silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function. *PLOS ONE*, 16, e0252932. DOI: [10.1371/journal.pone.0252932](https://doi.org/10.1371/journal.pone.0252932), PubMed: [34125870](https://pubmed.ncbi.nlm.nih.gov/34125870/).
  34. Sameer, A.S. & Nissar, S. (2021) Toll-like receptors (TLRs): Structure, functions, signaling, and role of their polymorphisms in colorectal cancer susceptibility. *BioMed Research International*, 2021, 1157023. DOI: [10.1155/2021/1157023](https://doi.org/10.1155/2021/1157023), PubMed: [34552981](https://pubmed.ncbi.nlm.nih.gov/34552981/).
  35. Shah, B., Modi, P. & Sagar, S.R. (2020) In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. *Life Sciences*, 252, 117652. DOI: [10.1016/j.lfs.2020.117652](https://doi.org/10.1016/j.lfs.2020.117652), PubMed: [32278693](https://pubmed.ncbi.nlm.nih.gov/32278693/).
  36. Sultana, T., Mou, S.I., Chatterjee, D., Faruk, M.O. & Hosen, M.I. (2024) Computational exploration of SLC14A1 genetic variants through structure modeling, protein-ligand docking, and molecular dynamics simulation. *Biochemistry and Biophysics Reports*, 38, 101703. DOI: [10.1016/j.bbrep.2024.101703](https://doi.org/10.1016/j.bbrep.2024.101703), PubMed: [38596408](https://pubmed.ncbi.nlm.nih.gov/38596408/).
  37. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. & Bray, F. (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71, 209–249. DOI: [10.3322/caac.21660](https://doi.org/10.3322/caac.21660), PubMed: [33538338](https://pubmed.ncbi.nlm.nih.gov/33538338/).
  38. Tatarkova, E.A., Tuguz, A.R., Shumilov, D.S., Muzhenya, D.V., Rudenko, K.A. & Smolkov, I.V. (2022) SNP genes of immune response mediators and predisposition to development of socially significant diseases. *Medical Immunology (Russia)*, 24, 751–764. DOI: [10.15789/1563-0625-SGO-2380](https://doi.org/10.15789/1563-0625-SGO-2380).
  39. Waheed, S., Ramzan, K., Ahmad, S., Khan, M.S., Wajid, M., Ullah, H., Umar, A., Iqbal, R., Ullah, R. & Bari, A. (2024) Identification and in-silico study of non-synonymous functional SNPs in the human SCN9A gene. *PLOS ONE*, 19, e0297367. DOI: [10.1371/journal.pone.0297367](https://doi.org/10.1371/journal.pone.0297367), PubMed: [38394191](https://pubmed.ncbi.nlm.nih.gov/38394191/).
  40. Wuyts, A., Struyf, S., Gijssbers, K., Schutyser, E., Put, W., Conings, R., Lenaerts, J.P., Geboes, K., Opdenakker, G., Menten, P., Proost, P. & Van Damme, J. (2003) The CXC chemokine GCP-2/CXCL6 is predominantly induced in mesenchymal cells by interleukin-1beta and is down-regulated by interferon-gamma: Comparison with interleukin-8/CXCL8. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, 83, 23–34. DOI: [10.1097/01.lab.0000048719.53282.00](https://doi.org/10.1097/01.lab.0000048719.53282.00), PubMed: [12533683](https://pubmed.ncbi.nlm.nih.gov/12533683/).
  41. Yang, H.D. & Nam, S.W. (2020) Pathogenic diversity of RNA variants and RNA variation-associated factors in cancer development. *Experimental and Molecular Medicine*, 52, 582–593. DOI: [10.1038/s12276-020-0429-6](https://doi.org/10.1038/s12276-020-0429-6), PubMed: [32346127](https://pubmed.ncbi.nlm.nih.gov/32346127/).
  42. Yasmin, T. (2022) In silico comprehensive analysis of coding and non-coding SNPs in human mTOR protein. *PLOS ONE*, 17, e0270919. DOI: [10.1371/journal.pone.0270919](https://doi.org/10.1371/journal.pone.0270919), PubMed: [35788771](https://pubmed.ncbi.nlm.nih.gov/35788771/).
  43. Zhang, N. & Li, Y. (2023) Receptor tyrosine kinases: Biological functions and anticancer targeted therapy. *MedicoComm* (2020), 4, e446. DOI: [10.1002/mco2.446](https://doi.org/10.1002/mco2.446), PubMed: [38077251](https://pubmed.ncbi.nlm.nih.gov/38077251/).

## Supplementary Data

**Table S1** Effect of polymorphism as analyzed by HOPE tool.

SNP ID	Amino Acid change	Effect of variation
rs367612330 (L47M)		Variant is bigger than the wild type residue. This might lead to bumps. The wild type is very conserved. The mutated residue is located in a domain. Mutation of the residue might disturb this function.
rs1191410166 (C51R)		The wild residue is higher in magnitude and more hydrophobic than the mutant residue. Wild residue was neutral and mutant residue was positive.. this mutation is probably damaging to protein.
rs757247423 (C51Y)		The mutant residue is bigger than the wild type residue. The wild type residue is more hydrophobic than the mutant residue. Hydrophobic interactions can be lost. This mutation is probably damaging to protein.
rs1213029852 (C75G)		The mutant residue is smaller than the wild type residue. The wild type residue is more hydrophobic than the mutant residue. C75G is probably damaging to protein.
rs1239375597 (P97T)		The residue of wild type is more hydrophobic than the mutant residue. The wild type residue is very conserved. Based on conservation scores this mutation is probably damaging to protein.

**Table S2** Binding affinities of 25 compounds with native and mutant proteins

Ligand	P80162.1.A (Model)	L47M	C51R	L92M	L92V	P73S	Q104K
Amphotericin	-8.5	-9.5	-9.5	-9.5	-9.5	-9.4	-9.4
Apoptozole	-6.4	-6.2	-6.4	-6.2	-6.3	-6.2	-6.4
Batrachotoxin	-7.3	-7.3	-7.3	-7.3	-7.3	-6.3	-7.3
Benzazepinone	-5.1	-5.1	-5.1	-5	-4.5	-5	-5.1
Chromane	-4.1	-4.1	-4.1	-4.1	-4.1	-4	-4.1
Chromophore	-6.2	-6.5	-6.3	-6.3	-6.3	-6.3	-6.6
Citrinin	-5.5	-5.4	-5.3	-5.4	-5.4	-5.4	-5.4
Hispolon	-5.3	-5.3	-5.3	-5.4	-5.3	-5.3	-5.3
Hyaluronic acid	-5.7	-5.7	-5.8	-5.8	-5.7	-6.1	-5.7
Imipramine	-5	-4.8	-4.8	-5	-5	-4.9	-4.8
IND24	-7.1	-6.1	-7.1	-7	-7.1	-7	-6.6
Lacosamide	-5.2	-5.2	-5.2	-4.7	-4.7	-5	-5.1
Laminin	-7.3	-7.4	-7.1	-7.2	-7.7	-7.5	-7.4
Lamotrigine	-6	-6	-6	-6.1	-6	-6	-6
Lectin	-7.9	-8.9	-7.8	-9.1	-8.9	-8.7	-8.8
Ladocaine	-4.6	-4.6	-4.6	-4.8	-4.8	-4.7	-4.8
Melanin	-6.4	-6.3	-6.4	-6.4	-6.4	-6.3	-6.4
PX4	-4	-4.6	-4.5	-4.2	-4.4	-4.6	-4.5
Pyranomentoflavone	-7.7	-7.7	-7.7	-7.7	-7.7	-7.7	-7.7
Quercetin	-6	-6	-6	-6	-6	-6	-6
Tetracycline	-5.4	-6.3	-4.6	-5.4	-5.4	-6.2	-5.4
Saxitoxin	-5.6	-5.6	-5.6	-5.7	-5.6	-5.8	-5.6
Vixotrigine	-6.3	-6	-6	-6.2	-6.2	-6.1	-6.1
Proflavine	-5.3	-5.3	-5.6	-5.3	-5.2	-5.5	-5.1
Eugenol	-4.5	-4.6	-4.6	-4.6	-4.6	-4.6	-4.6