

A systematic in-silico functional and structural analysis reveals deleterious missense nsSNPs in the human *CSF1R* gene

Purvi Malhotra¹, Aaryan Jaitly¹, Harshil Walia¹, Ojasvi Dutta¹, Deepanshi Rajput¹,
Mujtaba Husaini¹, Chander Jyoti Thakur^{1,*,#}, Sandeep Saini^{1,2,*,#}

1) Department of Bioinformatics, Goswami Ganesh Dutta Sanatan Dharma College, Sector 32, Chandigarh, 160030, India

2) Department of Biophysics, Panjab University, Sector-25, Chandigarh, 160014, India

ABSTRACT

Colony Stimulating Factor-1 Receptor (CSF1R) is a tyrosine kinase transmembrane receptor that plays a vital role in innate immunity and neurogenesis and controls the differentiation and maintenance of most tissue-resident macrophages. *CSF1R* mutations have been linked with many neurodegenerative diseases. In this work, we aim to identify the functional and structural impact of deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) mutations on CSF1R, which could help understand the consequences of these mutational changes. A consensus-based prediction approach was used to screen the missense SNPs using six in-silico tools: SIFT, PROVEAN, PMut, MutPred, MISSENSE 3D, and FATHMM. SNPs found to be deleterious by more than five out of six tools were subjected to further analysis, such as protein secondary structure and domain architecture analysis by PSIPRED and NCBI-CDD, respectively. Mutant models of highly deleterious SNPs were modeled using PyMol, followed by energy minimization and Root Mean Square Deviation (RMSD) analysis and molecular dynamic (MD) simulation by YASARA, TM-ALIGN, and WebGro, respectively. Out of 780 missense SNPs screened, we found the four most deleterious SNPs (L301S, A770P, I775N, and F849S) that decreased the protein stability because of their presence in the conserved regions of wild-type CSF1R. Structural and functional studies revealed that these mutations could disrupt the protein's core and surface interactions, leading to destabilization and functional impairment. Moreover, the mutated proteins exhibited enhanced conformational flexibility and instability, as confirmed by MD simulation analysis.

Keywords: SNP; bioinformatics; conservation; domain analysis; mutation

INTRODUCTION

Colony stimulating factor-1 receptor (CSF1R), belonging to the type III protein tyrosine kinase receptor family, is recognized as the cell surface receptor for the macrophage colony-stimulating factor 1 (CSF-1). The proto-oncogene c-fms, which is located on chromosome

#These authors have equally contributed to this work.

*Corresponding Authors: Department of Bioinformatics, Goswami Ganesh Dutta Sanatan Dharma College Sector 32-C, Chandigarh, India; Tel: +91 86 997 76533, +91 76 966 06035; Fax: +91 1722661077;

Email: chander.jyoti@ggdsd.ac.in AND sandeep.saini@ggdsd.ac.in

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

5q33.3, is known to encode it [1]. CSF1R consists of extracellular domains, a transmembrane domain, and an intracellular tyrosine kinase domain [2]. CSF1R, known for its multifaceted roles, is considered to play a fundamental function in innate immunity by governing the proliferation of various cell types, including tissue macrophages, osteoclasts, Langerhans cells in the skin, Paneth cells in the small intestine, and microglia in the cerebrum. Predominantly expressed in microglia within the central nervous system, CSF1R is believed to be crucial for their proliferation and maintenance under normal conditions [3]. Moreover, its broad tissue expression pattern has been described as pivotal in various pathological conditions such as neoplastic, inflammatory, and neurological diseases [2].

The discovery and purification of CSF-1 led to the recognition of CSF1R and the demonstration of its intracellular tyrosine kinase domain activity [4-6]. Studies have shown the human CSF1R has been shown to share around 75% and 84% overall homology with the mouse and feline versions, respectively [7, 8]. Besides this, it has been demonstrated in several studies that genetic excision or loss of *CSF1R* function results in microglial depletion across species, highlighting the dependence of microglial proliferation and development on CSF1R [9, 10].

Genetic variations of the *CSF1R* gene have been implicated in several neurodegenerative diseases. For instance, loss-of-function mutations in the *CSF1R* gene are the major cause of adult-onset leukoencephalopathy with axonal spheroids and pigmented glia (ALSP) [11]. Single nucleotide polymorphisms (SNPs) associated with the *CSF1R* gene are associated with inhibitor development in hemophilia A [12]. Furthermore, these mutations have also been linked to hereditary diffuse leukoencephalopathy with spheroids (HDLS) and pigmented orthochromatic leukodystrophy (POLD) [13, 14].

SNPs are regarded as the most prevalent form of chromosomal variations, occurring once every 100–300 base pairs, and are known to have a crucial impact on disease outcomes [15, 16]. Among them, non-synonymous single nucleotide polymorphisms (nsSNPs) are particularly noted for their potential to alter the structure and function of the protein due to modifications in the amino acid sequence. Such modifications can significantly influence the disease's development and progression [17].

Given the large amounts of variation data generated through genome sequencing efforts, it is considered challenging to study the effect of these mutations on gene function and their encoded proteins through experimental methods alone [18]. Therefore, a large number of previous studies have been conducted by research community to screen these large variation datasets using in-silico approaches [19-28].

Considering the involvement of *CSF1R* mutations in several diseases and the extensive SNP dataset that may not be feasibly analysed through experimental approaches alone, this study was designed to screen the missense SNPs of the *CSF1R* gene and to investigate their damaging effects on protein stability using in-silico tools. Furthermore, the screened high-risk SNPs were analysed for their effects on domain architecture and for changes in the modeled mutant's secondary and tertiary protein structures.

MATERIALS AND METHODS

Retrieval of dataset: The SNP data for the *CSF1R* gene was retrieved from NCBI SNP database (dbSNP) (<https://www.ncbi.nlm.nih.gov/snp>) [29]. Only missense SNPs were selected, as they are potentially capable of altering protein structure and function. The sequence of the CSF1R protein (UniProt accession no. P07333) was also retrieved from the UniProt database (<https://www.uniprot.org/>) [30]. A summary of the steps followed in the study is provided in Fig. 1.

Prediction of deleterious nsSNPs: The missense SNPs retrieved from dbSNP were screened using six bioinformatics tools, namely, SIFT (Sorting Intolerant From Tolerant, <https://sift.bii.a-star.edu.sg/>) [31], PROVEAN (PROtein Variation Effect Analyzer, <http://>

provean.jcvi.org/index.php) [32], FATHMM (Functional Analysis Through Hidden Markov Model, <http://fathmm.biocompute.org.uk/>) [33], PMut (<http://mmb.irbbarcelona.org/PMut/>) [34], Missense3D-DB (<http://missense3d.bc.ic.ac.uk:8080/>) [35] and MutPred2 (<http://mutpred.mutdb.org/>) [36].

SIFT classifies nsSNPs as deleterious or neutral. The scoring system ranges from 0 to 1. An amino acid substitution is classified as deleterious if the score is ≤ 0.05 , and tolerated if >0.05 [31]. Moreover, PROVEAN classifies single or multiple amino acid substitutions, in-frame insertions, and deletions as deleterious if the score is ≤ -2.5 and neutral if >-2.5 [32].

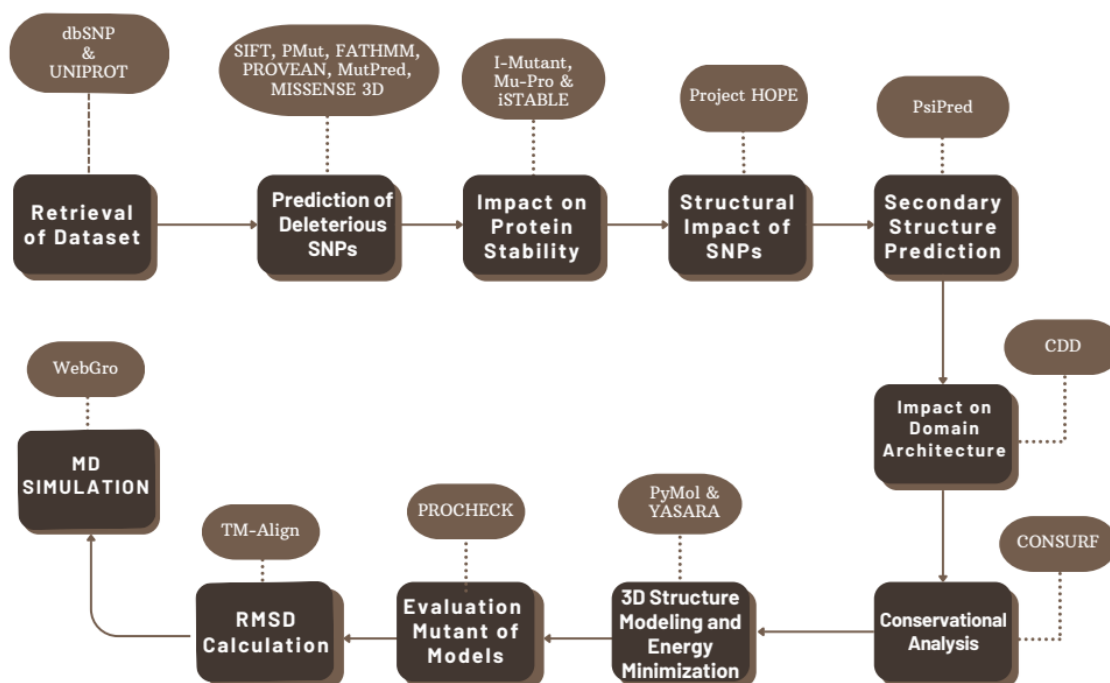


Figure 1: Flowchart of methodology showing the steps followed during analysis

To predict the functional implications of missense mutations, FATHMM was employed, which integrates sequence conservation via Hidden Markov Models and “pathogenicity weights” to assess tolerance to mutations. Variants scoring below the threshold of 0.75 are identified as potentially cancer-associated [33]. Furthermore, PMut was used to classify mutations as disease-causing or neutral. Scores above 0.5 are considered disease-causing, while scores below indicate neutrality [34].

Moreover, Missense3D was utilized to assess structural effects of amino acid substitutions using 16 structural parameters, categorizing them as “Damaging” or “Neutral” [35]. Besides this, MutPred2 was applied to predict whether substitutions are pathogenic or benign, using a threshold score of 0.5. Substitutions scoring above 0.5 are interpreted as likely affecting protein function [36].

Predicting the protein stability: We used three web tools, namely, I-Mutant 2.0 (<https://folding.biofold.org/cgi-bin/i-mutant2.0.cgi>) [37], iSTABLE (<http://predictor.nchu.edu.tw/iStable/>) [38], and MuPro (<http://mupro.proteomics.ics.uci.edu/>) to predict the stability of mutant protein sequences [39]. I-Mutant 2.0 utilizes a support vector machine (SVM) algorithm to predict the effect of amino acid mutations on protein stability. It calculates the energy change ($\Delta\Delta G$) value by subtracting the unfolding Gibbs free energy of the wild type from the unfolding Gibbs free energy of the mutated protein [37]. Similarly, iSTABLE is an integrated predictor constructed using sequence information and prediction results from different predictors to provides output in the form of $\Delta\Delta G$ values [38].

Additionally, MuPro, a machine learning (ML) classifier was used to predict the effect of amino acid change on protein stability. It predicts $\Delta\Delta G$ and classifies mutations with a confidence score based on the effects on stability, using both SVM and neural networks [39].

Analysis of structural impacts of point mutations: The project HOPE (Have (y)Our Protein Explained) (<https://www3.cmbi.umcn.nl/hope/input/>) server was used to analyze the structural effect of mutations on proteins and gain insights into the differences in properties between wild-type and mutant amino acids at specific positions. Information from both 2D and 3D structures, along with sequence annotation data from various protein structural and sequence analysis algorithms, is integrated to provide comprehensive predictions [40].

Secondary structure prediction: PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) was utilized to predict the secondary structure of the mutant proteins. The FASTA sequences were submitted as input and resulting annotations were provided using color codes: helices in pink, strands in yellow, and transmembrane regions in grey [41].

Analysis of domain architecture: The NCBI Conserved Domain Database (CDD) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) was utilized to predict the impact of mutations on the domains of the CSF1R protein. CDD is based on reverse position-specific (RPS) BLAST, a variant of PSI-BLAST, to scan the query protein against a set of pre-calculated Position Specific Scoring Matrices (PSSM) [42, 43]. The wild-type UniProt ID and mutant FASTA sequences were used for the domain analysis.

Conservational analysis: The ConSurf server (<https://consurf.tau.ac.il/>) was used to assess the evolutionary conservation scores for each residue using a Bayesian method. Conservation was visualized using a 1–9 scale, where 9 represents the most conserved residue and 1 indicates the least conserved [44]. FASTA sequences of the wild-type and mutant proteins were used for the analysis.

Tertiary structure modeling and energy minimization: The PDB (Protein Data Bank) ID: 6WXJ was used as a structural template to generate mutant models (A770P, I775N, and F849S) in PyMol using the “Mutagenesis Wizard.” The most stable rotamers were selected, and models were refined using YASARA (Yet Another Scientific Artificial Reality Application) (<http://www.yasara.org/minimizationserver.html>), a web-server for protein model refinement [45]. Discovery Studio (DS) visualizer was used for model visualization [46].

Evaluation of mutant models: The refined mutant models were evaluated using SAVES v6.0 (<https://saves.mbi.ucla.edu/>), a collection of six web-based tools for protein structure evaluation. PROCHECK within SAVES was used to generate Ramachandran plots, which were examined for allowed conformational space [47].

RMSD value calculation: TM-Align (<https://seq2fun.dcmf.med.umich.edu/TM-align/>) was used to compare the wild-type and mutant models. RMSD (Root Mean Square Deviation) and TM-scores (Template Model-Score) were computed to assess global structural deviation. TM-scores range from 0–1 (higher indicating better alignment), and smaller RMSD values indicate closer structural similarity [48, 49].

Molecular dynamics (MD) simulation: Molecular dynamics simulations were conducted using the GROMACS package via the WebGro server (<https://simlab.uams.edu/>) [50]. A well-established simulation protocol [51, 52] was followed, involving the simple point charge (SPC) water model in a triclinic periodic box to solvate the system. In addition, the GROMOS96 43a1 force field was applied to optimize the system. Moreover, the system was equilibrated at 300 K and 1.0 bar. Simulations were run for 50 ns, with 1000 frames recorded. To analyze the

simulation findings, we calculated the RMSD and root mean square fluctuation (RMSF) of each atom. Additionally, radius of gyration (Rg) and solvent-accessible surface area (SASA) were calculated to investigate the effects of mutation.

RESULTS

The CSF1R gene SNP data was retrieved from the NCBI SNP database, dbSNP. The entire set was comprised of 780 missense SNPs, 442 synonymous SNPs, and 21,094 intronic SNPs. Given the potential of missense SNPs to impact protein structure and function, we conducted an initial screening using the 780 missense SNPs.

The 780 nsSNPs were subjected to initial screening through six web-based bioinformatics tools. SIFT predicted 51 out of 780 SNPs as deleterious and 77 as tolerated, while the rest were not reported. These 51 deleterious SNPs were further analyzed using PROVEAN and MutPred. PROVEAN predicted 29 SNPs as deleterious, and the remainder were found to be neutral. MutPred predicted 24 SNPs with scores above the threshold, suggesting potential structural disruption or functional loss. Furthermore, FATHMM categorized 44 SNPs as cancerous, PMut predicted 70 as disease-causing, and Missense3D classified 85 SNPs in the damaging category.

To reduce false positives, a consensus-based approach was implemented to filter out the deleterious, damaging, cancerous, and disease-causing SNPs. The consensus demonstrating the comparative outputs from the six web-based tools is listed in Supplementary Table S1. SNPs predicted as deleterious, damaging, cancerous, and disease-causing by at least five of the six tools were selected for further analysis. A total of seven SNPs (rs1801271, rs121913390, rs281860269, rs281860271, rs281860273, rs281860277, and rs200489778) were identified that follow this filtering criteria, as shown in Table 1.

Table 1. List of deleterious and damaging SNPs predicted by mutation analysis tools

rs IDs	Amino Acid Change	SIFT	PROVEAN	FATHMM	PMut	MISSENSE 3D	MutPred
rs1801271	Y969C	Deleterious	Deleterious	Cancer	Disease	-	Damaging
rs121913390	L301S	Deleterious	Deleterious	Cancer	-	Damaging	Damaging
rs281860269	E633K	Deleterious	Deleterious	Cancer	-	Damaging	Damaging
rs281860271	A770P	Deleterious	Deleterious	Cancer	-	Damaging	Damaging
rs281860273	I775N	Deleterious	Deleterious	Cancer	-	Damaging	Damaging
rs281860277	F849S	Deleterious	Deleterious	Cancer	-	Damaging	Damaging
rs200489778	T663M	Deleterious	Deleterious	Cancer	-	Damaging	Damaging

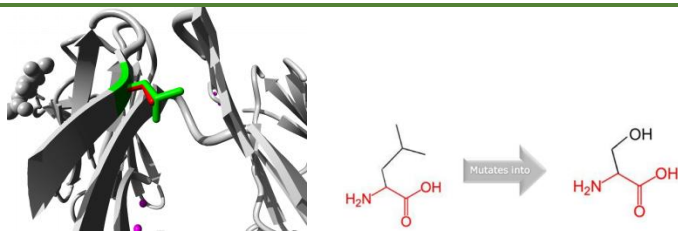
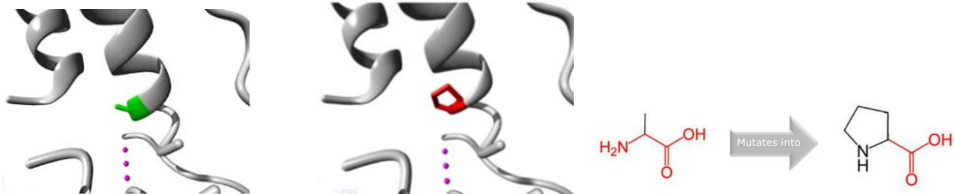
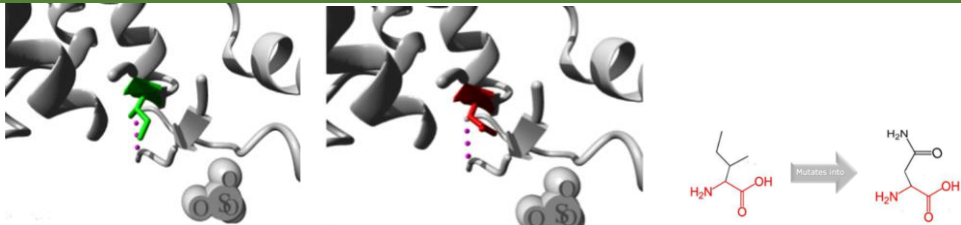
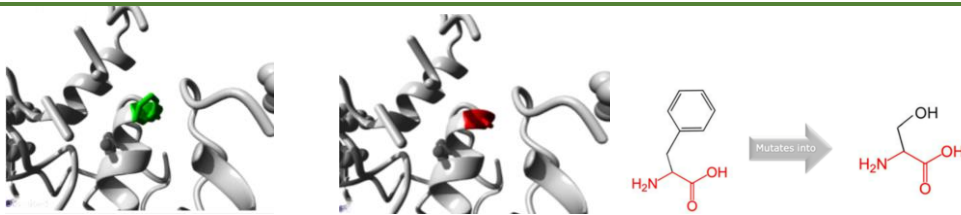
Following consensus filtering, protein stability analysis was conducted by MuPro, I-Mutant, and iSTABLE algorithms. MuPro and I-Mutant 2.0 are ML-based algorithms, while iSTABLE is based on meta-approach that determines protein stability based on sequence data. Again, a consensus-based approach was used in which the confidence score from iSTABLE and $\Delta\Delta G$ values from the MuPro and I-Mutant were compared (Supplementary Table S2) to filter out those nsSNPs that decrease protein stability. We have found that four of the seven nsSNPs (rs121913390, rs281860271, rs281860273, rs281860277) were found to decrease protein stability as shown in Table 2.

Table 2: Protein stability predictions by I-Mutant, iSTABLE, and MuPro

rs IDs	AMINO ACID MUTATIONS	MuPro ($\Delta\Delta G$ value)	I-Mutant ($\Delta\Delta G$ value)	iSTABLE (Score)
rs121913390	L301S	-1.8671494 (Decrease)	-3.14 (Decrease)	0.89314 (Decrease)
rs281860271	A770P	-1.3311997 (Decrease)	-2.97 (Decrease)	0.781124 (Decrease)
rs281860273	I775N	-2.1269065 (Decrease)	-1.11 (Decrease)	0.857828 (Decrease)
rs281860277	F849S	-1.3760635 (Decrease)	-2.53 (Decrease)	0.872933 (Decrease)

The deleterious nsSNPs that predicted to affect protein stability were analysed further using Project HOPE. For the rs ID: rs121913390 (L301S), the mutant amino acid residue (serine) was observed to be smaller than the wild-type residue (lysine). This mutation was found to create a void in the protein core and to cause a loss of hydrophobic interactions. The mutation was located in the Ig-like C2-type 4 domain. Furthermore, substitution of wild-type alanine residue with the proline residue (rs ID: rs281860271 (A770P)) introduce a larger mutant residue. Since the alanine is located on the protein's surface, this mutation may disrupt interactions with other molecules. The addition of proline can also destabilize the α -helix, potentially leading to significant structural alterations in the protein. This mutation was found to be occurred within the kinase domain of CSF1R. In the case of isoleucine to asparagine substitution (rs281860273 (I775N)), the mutant residue was found to be bigger. The wild-type residue isoleucine was buried within the protein core, and the bulkier asparagine cannot be accommodated in this space, thus resulting in the loss of the hydrophobic interactions. In the instance of rs281860277 (F849S), the mutant residue serine has been observed to be smaller than the wild-type phenylalanine. This mutation causes a loss of hydrophobic interactions in the core of the protein and it also causes a void in the core of the protein (Table 3).

Table 3: Summary of structural consequences observed for each mutation using Project HOPE, highlighting alterations in amino acid size, hydrophobicity, and spatial accommodation

Mutation	STRUCTURE
L301S	
A770P	
I775N	
F849S	

Secondary structure prediction was performed using PSIPRED. The FASTA sequences both wild-type and mutant proteins were submitted as inputs, with mutations manually inserted. In L301S (rs121913390), the loss of beta-strands at positions 48-50, 341-342, 350-352, 436-437, 471-474, and 953-955 was predicted (Fig. S1). Moreover, the new strand was found to be

introduced at positions 358-360, 562-63,575-577,724-725 and 775-776 (Fig. S1B). Beside this, a helix was detected at position 701–702, and a helix-to-strand transition was observed at position 168–170 (Fig. S1B).

The PSIPRED analysis of A770P revealed strand additions at 6-7, 22-23, and 576-577, and deletions at 48-50 and 953-955 (Fig. S1C). A strand at 618-619 was replaced by a helix. Strands were lost at 456-457, 784-786, and 797-802, while new helices emerged at 573-574, 699-700, and 702-704 (Fig. S1C). For I775N, strands were found at positions 23-24, 109-110, 576-577, 775-777, and 791-794 (Fig. S1D), and a helix was predicted at 618-619. In F849S, strand deletion was observed at 953-955, with helix formation at 700-702, and strand introduction at 576-577 (Fig. S1E). These variations were compared with the native structure (Table 4).

Table 4: The change in number of strands and helices between the mutants and the wild-type (native state) structure of CSF1R, as predicted by PSIPRED

Amino acid change	No. of helix	No. of strands
Native State	20	53
L301S	21	54
A770P	22	55
I775N	21	58
F849S	21	56

Domain architecture analysis via CDD indicated the presence of four domains in the wild-type protein: Ig-like domain, Ig3-CSF1R-like domain, Ig-3 domain, and PTKc-CSF1R domain. The PTKc-CSF1R domain contained active-site, ATP-binding, substrate-binding, and activation loop sites. In the A770P, I775N, and F849S mutants, the loss of active, substrate-binding, and activation loop regions was noted (Fig. 2). Surprisingly, no change in domain architecture was observed for the mutant L301S.

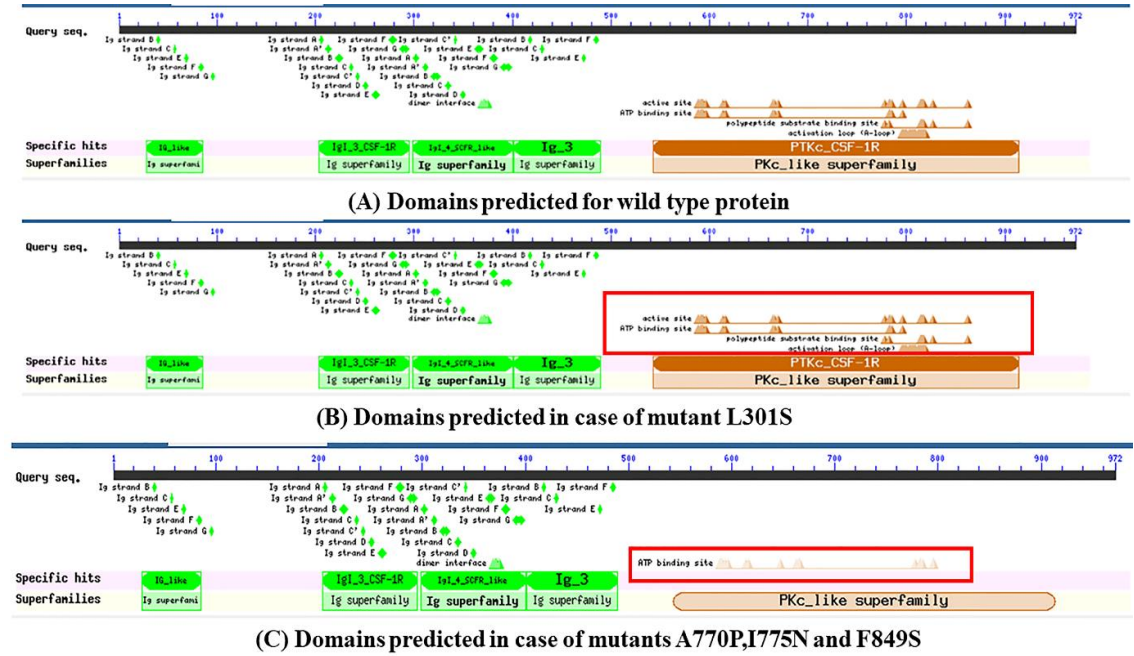


Figure 2: Domains predicted by CDD showed the loss of the active site, polypeptide substrate binding site, and the activation loop (marked in the red box) in mutants A770P, I775N, and F849S (C) compared to wild-type (A) while mutant L301S (B) did not show any loss of these signature elements (marked in the red box).

Conservation analysis was conducted using ConSurf. Mutation A770P was found in a moderately conserved region (score 8), and I775N and F849S were located in highly conserved

regions (score 9) (Fig. S2). These findings suggest that structural and functional alterations may result from mutations in evolutionarily conserved residues. Table 5 lists the mutations occurring in regions with conservation scores from 7 to 9.

Table 5: Conservational score for the amino acid positions as predicted by ConSurf

rs ID	AMINO ACID CHANGE	CONSURF SCORE
rs281860271	A770P	8/conserved
rs281860273	I775N	9/conserved
rs281860277	F849S	9/conserved

3D structure prediction was carried out using PyMol. Mutagenesis was performed using the Mutagenesis Wizard. The 6WXJ template was used to model A770P, I775N, and F849S 3D structures. These mutants were selected for modeling due to their loss of domain architecture and the alignment of their sequence coordinates within a single PDB template (6WXJ). The L301S mutation was excluded due to domain preservation and a different template. During modeling of the 3D structure of mutants, the water molecules and the heteroatoms present in the template were removed. Multiple rotamers were generated for each mutation, and the one with the highest confidence was selected for structure modeling.

Furthermore, the geometry optimization of the 3D mutant models was performed by energy minimizations through the YASARA server. Mutant and energy-refined models are shown in Figure 3. Model evaluation was conducted using SAVES v6.0, with PROCHECK used for quality assessment via Ramachandran plot (Fig. 4). The percentage of residues in allowed regions is reported in Table 6.

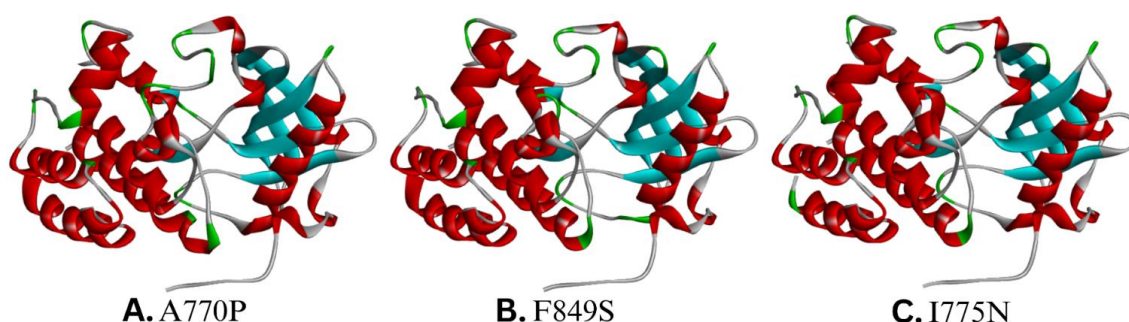


Figure 3: The predicted and energy-refined mutant models generated by PyMOL and YASARA, respectively were visualized using Discovery Studio Visualiser.

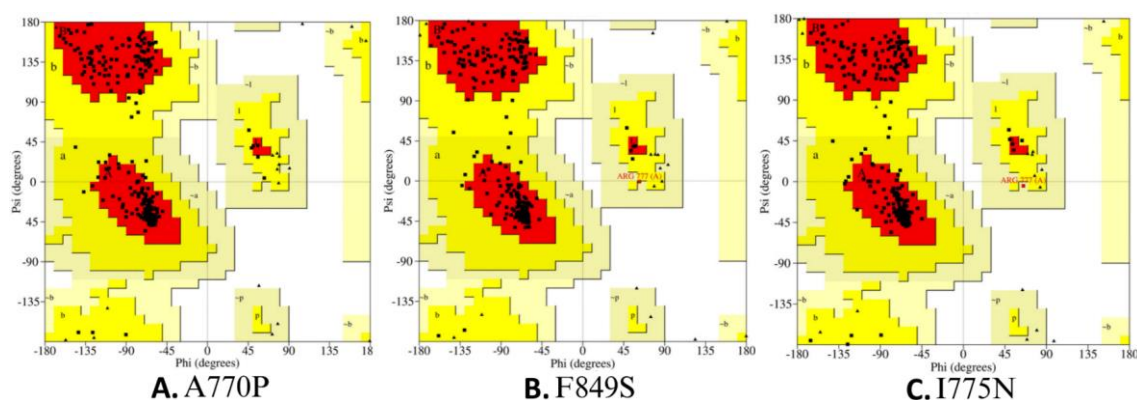


Figure 4: Ramachandran Plots of the mutant models, (A) (A770P), (B) (F849S) and (C) (I775N) generated by PROCHECK. The plots show the clustering of amino acids in the allowed quadrant of the plot.

Table 6: Percentage of residues in the allowed regions for three mutants

rs ID	Amino acid change	Residues in allowed regions
rs281860271	A770P	91.8%
rs281860273	I775N	92.7%
rs281860277	F849S	94.3%

The energy-minimized mutant models were analysed using TM-Align for RMSD and TM-score. The TM-Score obtained describes the topological similarity, and the RMSD value determines the deviation of the mutant model from the wild-type protein. Structural similarity and deviation were determined by comparing mutants to the wild-type. Table 7 depicts the values obtained from TM-Align.

Table 7: The TM-Scores and the RMSD values of mutant models compared to the wild-type template predicted by TM-Align

rs ID	Amino acid change	TM-SCORE	RMSD value
rs281860271	A770P	0.99623	0.38
rs281860273	I775N	0.99728	0.32
rs281860277	F849S	0.99623	0.36

MD simulations were conducted to examine the initial configurations of mutant proteins for structural flexibility, stability, hydrogen bonding, and solvation over 50 ns in a triclinic box. RMSD was computed to analyze conformational changes. The average RMSD of the native structure was 0.28 nm, while mutants A770P, F849S, and I775N showed increased values of 0.36 nm, 0.31 nm, and 0.36 nm, respectively (Fig. 5A).

To assess dynamic changes, RMSF values were computed. We have observed highest RMSF (0.5773 nm) for ASP917 in the native protein. However, mutant A770P and F849S showed increased RMSF values of 0.653 nm and 0.61 nm, respectively. Surprisingly mutant I775N had the least RMSF value of 0.4683 nm at the same position. Overall, the total RMSF value of the mutant A770P and F849S differed considerably from the native, while mutant I775N showed a similar level of flexibility compared to the native (Fig. 5B).

Protein stability was further assessed by analysing total hydrogen bonds. Notably, the mutants A770P and F849S displayed fewer overall hydrogen bonds. Furthermore, we calculated the SASA values for both the native and mutants. It is worth noting that the F849S mutant exhibited a considerably higher average SASA value of 133.67 nm² in contrast to the native protein (131 nm²). Conversely, I775N and A770P showed lower average SASA values of 129.61 nm² and 125.22 nm², respectively, compared to native (Fig. 5C). The Rg was evaluated to assess compactness. The native average Rg value was recorded as 1.85 nm, while the values for I775N, F849S, and A770P were found to be 1.84 nm, 1.87 nm, and 1.85 nm, respectively. F849S displayed the greatest Rg fluctuation (Fig. 5D).

DISCUSSION

The emergence of high-throughput genome sequencing technology has made it possible to identify a large number of SNPs. This emphasizes the necessity of thorough investigations to evaluate the clinical significance of these SNPs [53, 54]. Furthermore, in the advancing age of precision medicine, analyzing vast amounts of SNP data can provide valuable insights into the structural and functional variations in gene products that can impact several physiological and pathological processes [55, 56].

Moreover, in the last few years, various studies have determined the effect of CSF1R SNPs on various pathophysiological conditions [57-61]. Still, given the rise of SNP data, there is an increasing need and scope for more analysis. Thus, the present study adds to this objective by comprehensively evaluating CSF1R SNPs and elucidating their potential influence on protein

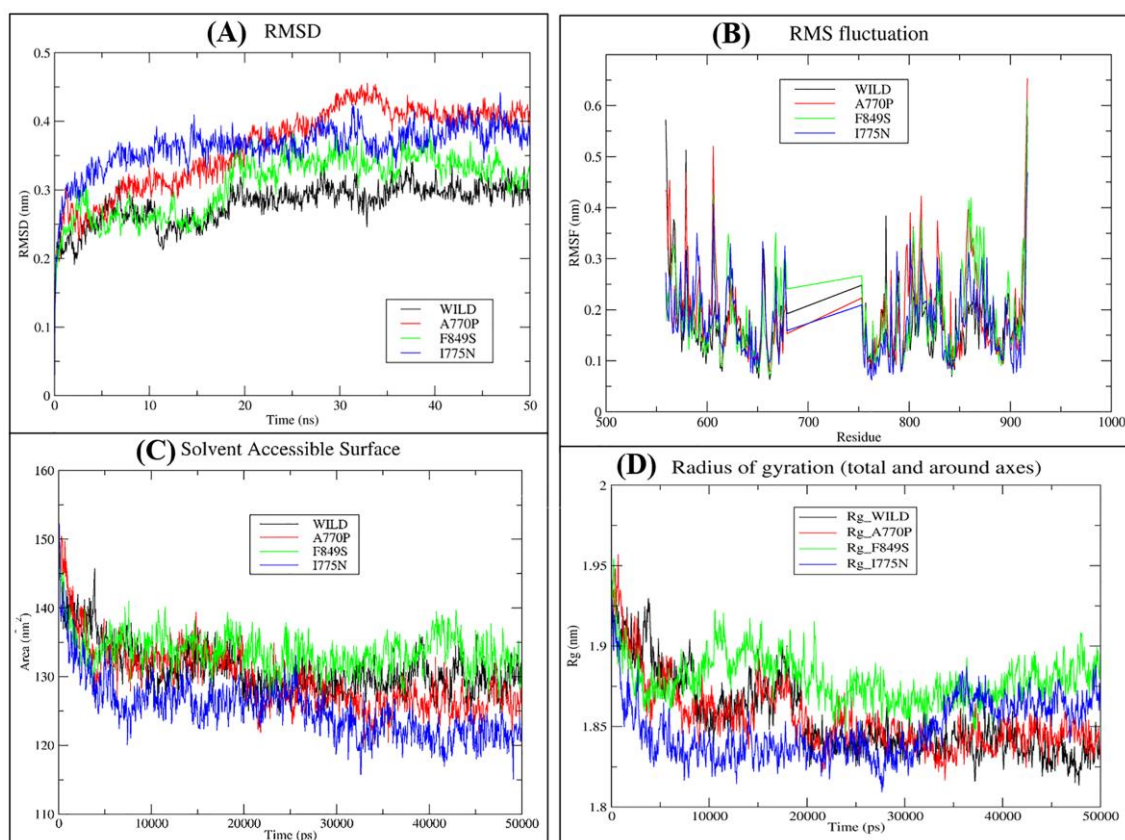


Figure 5: The RMSD (A), RMSF (B), SASA (C) and Radius of gyration (D) plots of the mutant (A770P, F849S, and I775N) and wild-type protein models obtained by GROMACS simulation package, accessible through the WebGro server.

structure and function. The study used a combination of bioinformatics and molecular dynamics simulation tools to predict the high-risk deleterious missense SNPs. In this work, we analyzed a large dataset of 780 missense SNPs of the *CSF1R* gene to identify deleterious SNPs. A consensus-based approach was followed, which included selecting those SNPs that were predicted deleterious or disease-causing across at least five of the six tools used. A consensus-based approach minimized the chances of false-positive predictions, as demonstrated by numerous prior studies that used multiple tool consensus to enhance predictive accuracy [62-65]. We have identified seven missense SNPs (Y969C, L301S, E633K, A770P, I775N, F849S, and T663M) in the category of deleterious, damaging or disease-causing by at least five tools. These filtered SNPs were considered for further downstream analysis.

Predicting changes in protein stability due to polymorphism in genetic data is of great significance because it impacts personalized medicine and diagnostics. With the increased sensitivity and specificity of the ML-based tools for predicting protein stability upon point mutations, a large volume of SNP data has been screened [62, 66, 67]. In our analysis, we found four of the seven identified SNPs (rs121913390 (L301S), rs281860271 (A770P), rs281860273 (I775N), and rs281860277 (F849S)) were predicted to decrease protein stability, as indicated by the prediction results of MuPro, I-Mutant, and iSTABLE. This finding is crucial because reduced protein stability is often found to be associated with changes in protein functions; for instance, a work by Gerasimavicius et al. deciphered the different roles of loss-of-function, gain-of-function, and dominant-negative mutations on the protein [68].

In addition, the structural analyses performed using Project HOPE and PSIPRED provided more insights into the molecular mechanisms that underlie the impacts of these SNPs. The L301S and F849S mutations have been predicted to create voids inside the protein core due to

the substitution of larger wild-type residues with smaller mutant residues. These voids may likely disrupt the hydrophobic core, resulting in an overall decrease in structural integrity.

Previous research had also shown a similar pattern in the Fibroblast Growth Factor Receptor 1 (*FGFR1*) gene, resulting in a decrease in stability upon substitution (P722S) of larger wild-type residues with smaller mutant residues [69, 70]. Besides this, the A770P mutation would be predicted to hinder the protein interactions on the surface, thus possibly affecting the protein's ability to interact with other molecules or substrates. In contrast, the I775N mutation resulted in the insertion of a bulkier residue within the protein core, which could lead to steric clashes and further destabilization of the protein structure. These findings are consistent with previously published work that has identified steric hindrance as a key factor in protein destabilization due to SNPs [71].

The secondary structure predictions by PSIPRED revealed considerable alterations in the protein's secondary structure elements due to these mutations. For example, the mutants exhibit beta-strand and alpha-helices changes, indicating that these SNPs cause substantial conformational changes, which could alter the protein's functional domains. The secondary structure analysis of the A770P mutant revealed an unexpected increase in both beta-strands and alpha-helices. Although proline is classically known to disrupt alpha-helices, the A770P mutation showed an increase in both strands and helices. This apparent contradiction may be due to compensatory structural rearrangements, where regions adjacent to the disrupted site reconfigure into stable secondary elements. Moreover, some coil regions may have adopted more ordered structures as a consequence of local conformational stress redistribution induced by the proline substitution. Such behavior, though less common, has been documented in previous studies involving proline mutations [72]. These findings are especially significant considering the role of CSF1R protein in signaling pathways related to cell proliferation and differentiation..

Moreover, the domain analysis also found that mutations such as A770P, I775N, and F849S resulted in the loss of critical functional sites, including the active and ATP binding sites. The structural and functional aberrations noted in this work could lead to loss of CSF1R activities which can result in the lack of microglia and disturbance in brain development that were observed by Erblich et al. in their study on homozygous mouse bearing null mutation of *CSF1R* gene [73].

During the evolutionary conservation analysis, these mutations were found to occur within highly conserved regions of the protein, underscoring their potential to disrupt essential functions. The high conservation scores of 8 (A770P) and 9 (I775N, F849S) associated with these SNPs suggest that these residues are critical for maintaining the structural and functional integrity of CSF1R, and mutations in these are expected to have deleterious effects. The profound impact of SNPs in modifying evolutionary conservation regions of genes has been a subject of tremendous significance that helps in understanding the structural and functional changes occurring over time [74-79].

Lastly, the molecular dynamics simulations validated the destabilizing effects of the identified deleterious SNPs. The higher RMSD for the mutant structures than the native structure indicates that these SNPs lead to greater conformational flexibility and instability. Specifically, the mutations A770P and I775N showed the highest RMSD values, indicating significant deviations from the wild-type protein. Furthermore, the RMSF analysis supported these findings, with mutations A770P and F849S causing increased flexibility at the residue level, notably at position ASP917. Interestingly, the I775N mutation displayed a lower RMSF value at this position, indicating a complex interplay between local flexibility and overall protein stability.

The present study comprehensively screened the deleterious nsSNPs based on functional and structural stability analysis. However, due to its dependance on computational methods, it prompted future research that should aim to decipher the biological implications of these mutations using functional in-vitro and in-vivo assays.

In conclusion, this study provides an in-depth analysis of the functional and structural impacts of deleterious nsSNPs in the *CSF1R* gene. By employing a consensus-based in-silico approach, four highly deleterious mutations, such as L301S, A770P, I775N, and F849S, were identified. Furthermore, structural analysis revealed that these mutations not only destabilize the protein, particularly within conserved regions, but also lead to substantial alterations in secondary structure and domain architecture. Specifically, the A770P, I775N, and F849S mutations resulted in the gain or loss of beta-strands and alpha-helices and the disruption of essential functional domains such as active sites and ATP-binding regions. Moreover, MD simulations highlighted increased conformational flexibility and instability in the mutated proteins. Further in vitro and in vivo instigations are required to validate these predictions and explore therapeutic interventions targeting these mutations, potentially contributing to the development of treatments for diseases associated with *CSF1R* dysfunction.

Acknowledgements: The authors would like to acknowledge Simlab for providing GROMACS's simulation facilities in the form of an open-access server, WebGro. They are also thankful to DST (Department of Science and Technology) and DBT (Department of Biotechnology), Government of India, for providing the infrastructural facility.

Conflict of Interest: The authors declare no conflicts of interest.

Authors' Contribution: PM: Methodology, Writing-original draft, Data Curation, Formal analysis. AJ: Methodology, Data Curation, Formal analysis. HW: Methodology, Data Curation, Formal analysis. OD: Data Curation, Formal analysis. DR: Data Curation, Formal analysis. MH: Data Curation, Formal analysis. CJTh: Conceptualization, Methodology, Formal analysis, Data curation, Writing-original draft, Writing-review & editing. SS: Conceptualization, Methodology, Formal analysis, Data curation, Writing-original draft, Writing-review & editing.

REFERENCES

1. Hampe A, Shamooin BM, Gobet M, Sherr CJ, Galibert F. Nucleotide sequence and structural organization of the human FMS proto-oncogene. *Oncogene Res* 1989;4:9-17.
2. Stanley ER, Chitu V. CSF-1 receptor signaling in myeloid cells. *Cold Spring Harb Perspect Biol* 2014;6:a021857.
3. Han J, Chitu V, Stanley ER, Wszolek ZK, Karrenbauer VD, Harris RA. Inhibition of colony stimulating factor-1 receptor (CSF-1R) as a potential therapeutic strategy for neurodegenerative diseases: opportunities and challenges. *Cell Mol Life Sci* 2022;79:219.
4. Stanley ER, Cifone M, Heard PM, Defendi V. Factors regulating macrophage production and growth: identity of colony-stimulating factor and macrophage growth factor. *J Exp Med* 1976;143:631-647.
5. Guilbert LJ, Stanley ER. Specific interaction of murine colony-stimulating factor with mononuclear phagocytic cells. *J Cell Biol* 1980;85:153-159.
6. Yeung YG, Jubinsky PT, Sengupta A, Yeung DC, Stanley ER. Purification of the colony-stimulating factor 1 receptor and demonstration of its tyrosine kinase activity. *Proc Natl Acad Sci USA* 1987;84:1268-1271.
7. Rothwell VM, Rohrschneider LR. Murine c-fms cDNA: cloning, sequence analysis and retroviral expression. *Oncogene Res* 1987;1:311-324.
8. Woolford J, McAuliffe A, Rohrschneider LR. Activation of the feline c-fms proto-oncogene: multiple alterations are required to generate a fully transformed phenotype. *Cell* 1988;55:965-977.
9. Oosterhof N, Chang IJ, Karimiani EG, Kuil LE, Jensen DM, Daza R, Young E, Astle L, van der Linde HC, Shivaram GM, Demmers J, Latimer CS, Keene CD, Loter E, Maroofian R, van Ham TJ, Hevner RF, Bennett JT. Homozygous mutations in *CSF1R* cause a pediatric-

- onset leukoencephalopathy and can result in congenital absence of microglia. *Am J Hum Genet* 2019;104:936-947.
10. Ginhoux F, Greter M, Leboeuf M, Nandi S, See P, Gokhan S, Mehler MF, Conway SJ, Ng LG, Stanley ER, Samokhvalov IM, Merad M. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* 2010;330:841-845.
11. Hu B, Duan S, Wang Z, Li X, Zhou Y, Zhang X, Zhang YW, Xu H, Zheng H. Insights into the role of CSF1R in the central nervous system and neurological disorders. *Front Aging Neurosci* 2021;13:789834.
12. Zhao M, Zhang Y, Liu Y, Sun G, Tian H, Hong L. Polymorphisms in *MAPK9* (rs4147385) and *CSF1R* (rs17725712) are associated with the development of inhibitors in patients with haemophilia A in North China. *Int J Lab Hematol* 2019;41:572-577.
13. Nicholson AM, Baker MC, Finch NA, Rutherford NJ, Wider C, Graff-Radford NR, Nelson PT, Clark HB, Wszolek ZK, Dickson DW, Knopman DS, Rademakers R. CSF1R mutations link POLD and HDLS as a single disease entity. *Neurology* 2013;80:1033-1040.
14. Rademakers R, Baker M, Nicholson AM, Rutherford NJ, Finch N, Soto-Ortolaza A, Lash J, Wider C, Wojtas A, DeJesus-Hernandez M, Adamson J, Kouri N, Sundal C, Shuster EA, Aasly J, MacKenzie J, Roeber S, Kretschmar HA, Boeve BF, Knopman DS, Petersen RC, Cairns NJ, Ghetti B, Spina S, Garbern J, Tselis AC, Uitti R, Das P, Van Gerpen JA, Meschia JF, Levy S, Broderick DF, Graff-Radford N, Ross OA, Miller BB, Swerdlow RH, Dickson DW, Wszolek ZK. Mutations in the colony stimulating factor 1 receptor (*CSF1R*) gene cause hereditary diffuse leukoencephalopathy with spheroids. *Nat Genet* 2011;44:200-205.
15. Fareed MM, Ullah S, Aziz S, Johnsen TA, Shityakov S. In-silico analysis of non-synonymous single nucleotide polymorphisms in human β -defensin type 1 gene reveals their impact on protein-ligand binding sites. *Comput Biol Chem* 2022;98:107669.
16. Das SC, Rahman MA, Das Gupta S. In-silico analysis unravels the structural and functional consequences of non-synonymous SNPs in the human IL-10 gene. *Egyptian J Med Human Genet* 2022;23:10.
17. Venkata Subbiah H, Ramesh Babu P, Subbiah U. Determination of deleterious single-nucleotide polymorphisms of human *LYZ* C gene: an in silico study. *J Genet Eng Biotechnol* 2022;20:92.
18. Mah JT, Low ES, Lee E. In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery. *Drug Discov Today* 2011;16:800-809.
19. Scotti C, Olivieri C, Boeri L, Canzonieri C, Ornati F, Buscarini E, Pagella F, Danesino C. Bioinformatic analysis of pathogenic missense mutations of activin receptor like kinase 1 ectodomain. *PLoS One* 2011;6:e26431.
20. Marjan MN, Hamzeh MT, Rahman E, Sadeq V. A computational prospect to aspirin side effects: aspirin and COX-1 interaction analysis based on non-synonymous SNPs. *Comput Biol Chem* 2014;51:57-62.
21. Fazel-Najafabadi E, Vahdat Ahar E, Fattahpour S, Sedghi M. Structural and functional impact of missense mutations in *TPMT*: An integrated computational approach. *Comput Biol Chem* 2015;59 Pt A:48-55.
22. Solayman M, Saleh MA, Paul S, Khalil MI, Gan SH. In silico analysis of non-synonymous single nucleotide polymorphisms of the human adiponectin receptor 2 (*ADIPOR2*) gene. *Comput Biol Chem* 2017;68:175-185.
23. Seifi M, Walter MA. Accurate prediction of functional, structural, and stability changes in PITX2 mutations using in silico bioinformatics algorithms. *PLoS One* 2018;13:e0195971.
24. Falahi S, Karaji AG, Koohyanizadeh F, Rezaeiemanesh A, Salari F. A comprehensive in Silico analysis of the functional and structural impact of single nucleotide polymorphisms (SNPs) in the human IL-33 gene. *Comput Biol Chem* 2021;94:107560.
25. Halder SK, Rafi MO, Shahriar EB, Albogami S, El-Shehawi AM, Daullah SMMU, Himel MK, Emran TB. Identification of the most damaging nsSNPs in the human *CFL1* gene and their functional and structural impacts on cofilin-1 protein. *Gene* 2022;819:146206.

26. Kalmari A, Hosseinzadeh Colagar A, Heydari M, Arash V. Missense polymorphisms potentially involved in mandibular prognathism. *J Oral Biol Craniofac Res* 2023;13:453-460.
27. Bouqdayr M, Abbad A, Baba H, Saih A, Wakrim L, Kettani A. Computational analysis of structural and functional evaluation of the deleterious missense variants in the human CTLA4 gene. *J Biomol Struct Dyn* 2023;41:14179-14196.
28. Sivakumar K, Subbiah U. Computational analysis of non-synonymous SNPs in the human LCN2 gene. *Egypt J Med Human Genet* 2024;25:94.
29. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308-311.
30. UniProt Consortium T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2018;46:2699.
31. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073-1081.
32. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015;31:2745-2747.
33. Shihab HA, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum Genomics* 2014;8:11.
34. López-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpí JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res* 2017;45:W222-W228.
35. Khanna T, Hanna G, Sternberg MJE, David A. Missense3D-DB web catalogue: an atom-based analysis and repository of 4M human protein-coding genetic variants. *Hum Genet* 2021;140:805-812.
36. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, Mort M, Cooper DN, Sebat J, Lakoucheva LM, Mooney SD, Radivojac P. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun* 2020;11:5918.
37. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* 2005;33:W306-W310.
38. Chen CW, Lin J, Chu YW. iStable: off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics* 2013;14(Suppl 2):S5.
39. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* 2006;62:1125-1132.
40. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 2010;11:548.
41. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292:195-202.
42. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* 2004;32(Web Server issue):W327-W331.
43. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* 2004;20:1322-1324.
44. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 2016;44:W344-W350.
45. Land H, Humble MS. YASARA: A tool to obtain structural guidance in biocatalytic investigations. *Methods Mol Biol* 2018;1685:43-67.
46. Biovia DS. Discovery studio visualizer. San Diego, CA, USA 2017;936:240-249.
47. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996;8:477-486.

48. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;33:2302-2309.
49. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci* 2001;10:1470-1473.
50. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1:19-25.
51. Feroz T, Islam MK. A computational analysis reveals eight novel high-risk single nucleotide variants of human tumor suppressor LHPP gene. *Egyp J Med Human Gene* 2023;24:47.
52. Pandey S, Maurya N, Avashthi H, Katara P, Singh S, Gautam B, Singh DB. Comprehensive analysis of non-synonymous SNPs related to Parkinson's Disease and molecular dynamics simulation of PRKN mutants. *Results in Chemistry* 2023;5:100817.
53. Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, Wang J. SNP detection for massively parallel whole-genome resequencing. *Genome Res* 2009;19:1124-1132.
54. Shastri BS. SNPs in disease gene mapping, medicinal drug development and evolution. *J Hum Genet* 2007;52:871-880.
55. Shastri BS. SNPs: impact on gene function and phenotype. Single nucleotide polymorphisms: *Methods Mol Biol* 2009;578:3-22.
56. Katara P. Single nucleotide polymorphism and its dynamics for pharmacogenomics. *Interdiscip Sci* 2014;6:85-92.
57. Kang HG, Lee SY, Jeon HS, Choi YY, Kim S, Lee WK, Lee HC, Choi JE, Bae EY, Yoo SS, Lee J, Cha SI, Kim CH, Lee MH, Kim YT, Kim JH, Hong YC, Kim YH, Park JY. A functional polymorphism in CSF1R gene is a novel susceptibility marker for lung cancer among never-smoking females. *J Thorac Oncol* 2014;9:1647-1655.
58. Chang KH, Wu YR, Chen YC, Wu HC, Chen CM. Association between CSF1 and CSF1R Polymorphisms and Parkinson's Disease in Taiwan. *J Clin Med* 2019;8:1529.
59. Shin EK, Lee SH, Cho SH, Jung S, Yoon SH, Park SW, Park JS, Uh ST, Kim YK, Kim YH, Choi JS, Park BL, Shin HD, Park CS. Association between colony-stimulating factor 1 receptor gene polymorphisms and asthma risk. *Hum Genet* 2010;128:293-302.
60. Soares MJ, Pinto M, Henrique R, Vieira J, Cerveira N, Peixoto A, Martins AT, Oliveira J, Jerónimo C, Teixeira MR. CSF1R copy number changes, point mutations, and RNA and protein overexpression in renal cell carcinomas. *Mod Pathol* 2009;22:744-752.
61. Kang WS, Kim YJ, Paik JW. PM453. Association between CSF1R gene polymorphism and the risk of schizophrenia in Korean population. *Int J Neuropsychopharmacology* 2016; 19(Suppl 1):64-65.
62. Bahia W, Soltani I, Abidi A, Mahdhi A, Mastouri M, Ferchichi S, Almowi WY. Structural impact, ligand-protein interactions, and molecular phenotypic effects of TGF- β 1 gene variants: In silico analysis with implications for idiopathic pulmonary fibrosis. *Gene* 2024; 922:148565.
63. Hasnain MJU, Shoaib M, Qadri S, Afzal B, Anwar T, Abbas SH, Sarwar A, Talha Malik HM, Tariq Pervez M. Computational analysis of functional single nucleotide polymorphisms associated with SLC26A4 gene. *PLoS One* 2020;15:e0225368.
64. Emadi E, Akhoundi F, Kalantar SM, Emadi-Baygi M. Predicting the most deleterious missense nsSNPs of the protein isoforms of the human HLA-G gene and in silico evaluation of their structural and functional consequences. *BMC Genet* 2020;21:94.
65. Thakur CJ, Saini S, Notra A, Chauhan B, Arya S, Gupta R, Thakur J, Kumar V. Deciphering the functional role of hypothetical proteins from *Chloroflexus aurantiacus* J-10-f1 using bioinformatics approach. *Mol Biol Res Commun* 2020;9:129-139.
66. Fang J. A critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. *Brief Bioinform* 2020;21:1285-1292.
67. Ho DSW, Schierding W, Wake M, Saffery R, O'Sullivan J. Machine Learning SNP Based Prediction for Precision Medicine. *Front Genet* 2019;10:267.

68. Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nat Commun* 2022;13:3895.
69. Doss CG, Rajith B, Garwasis N, Mathew PR, Raju AS, Apoorva K, William D, Sadhana NR, Himani T, Dike IP. Screening of mutations affecting protein stability and dynamics of FGFR1-A simulation analysis. *Appl Transl Genom* 2012;1:37-43.
70. Xiong D, Lee D, Li L, Zhao Q, Yu H. Implications of disease-related mutations at protein-protein interfaces. *Curr Opin Struct Biol* 2022;72:219-225.
71. Petrosino M, Novak L, Pasquo A, Chiaraluce R, Turina P, Capriotti E, Consalvi V. Analysis and interpretation of the impact of missense variants in cancer. *Int J Mol Sci* 2021;22:5416.
72. Li SC, Goto NK, Williams KA, Deber CM. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc Natl Acad Sci USA* 1996;93:6676-6681.
73. Erblich B, Zhu L, Etgen AM, Dobrenis K, Pollard JW. Absence of colony stimulation factor-1 receptor results in loss of microglia, disrupted brain development and olfactory deficits. *PLoS One* 2011;6:e26317.
74. Plyler ZE, McAtee CW, Hill AE, Crowley MR, Tindall JM, Tindall SR, Joshi D, Sorscher EJ. Relationships between genomic dissipation and de novo SNP evolution. *PLoS One* 2024;19:e0303257.
75. Fadason T, Farrow S, Gokuladhas S, Golovina E, Nyaga D, O'Sullivan JM, Schierding W. Assigning function to SNPs: considerations when interpreting genetic variation. *Semin Cell Dev Biol* 2022;121:135-142.
76. Hu G, Hovav R, Grover CE, Faigenboim-Doron A, Kadmon N, Page JT, Udall JA, Wendel JF. Evolutionary conservation and divergence of gene coexpression networks in gossypium (cotton) seeds. *Genome Biol Evol* 2016;8(12):3765-3783.
77. Jordan DM, Ramensky VE, Sunyaev SR. Human allelic variation: perspective from protein function, structure, and evolution. *Curr Opin struct Biol* 2010;20:342-250.
78. Saini S, Jyoti-Thakur C, Kumar V, Suhag A, Jakhar N. In silico mutational analysis and identification of stability centers in human interleukin-4. *Mol Biol Res Commun* 2018;7:67-76.
79. Sharma D, Singh H, Arya A, Choudhary H, Guleria P, Saini S, Thakur CJ. Comprehensive computational analysis of deleterious nsSNPs in *PTEN* gene for structural and functional insights. *Mol Biol Res Commun* 2025;14:219-239.