

Original Research

A Comprehensive *in Silico* Analysis of the Functional and Structural Consequences of the Deleterious Missense Nonsynonymous SNPs in Human *GABRA6* Gene

Tahere Mohammadpour, Reza Mohammadzadeh *

Department of Cellular and Molecular Biology, Faculty of Basic Science, University of Maragheh, Maragheh, Iran; E-Mails: taheremp@stu.maragheh.ac.ir; rmohamadzadeh@maragheh.ac.ir* **Correspondence:** Reza Mohammadzadeh; E-Mail: rmohamadzadeh@maragheh.ac.ir**Academic Editor:** Thomas Liehr*OBM Genetics*

2024, volume 8, issue 2

doi:10.21926/obm.genet.2402227

Received: December 20, 2023**Accepted:** April 08, 2024**Published:** April 15, 2024

Abstract

Epilepsy, a prevalent neurological disorder, affects more than 50 million individuals worldwide and is characterized by recurring seizures. Nonsynonymous single nucleotide polymorphisms (nsSNPs) found within coding regions of epilepsy-related genes are believed to have significant impacts on protein function. This is due to their tendency to cause mutations in the encoded amino acids, which can subsequently lead to pathogenic alterations in protein structure and function. Consequently, nsSNPs have the potential to serve as diagnostic markers for epilepsy and other neuropsychiatric conditions. The primary objective of this study is to evaluate the harmful effects of missense nsSNP mutations on the *GABRA6* gene. The *GABRA6* gene encodes the alpha-6 subunit of the GABA_A receptor, and previous research showed one case substitution mutation in the *GABRA6* gene is associated with childhood absence epilepsy (CAE) and atonic seizures. To achieve this, we employed various computational tools, including SIFT, PolyPhen-2, PROVEAN, Condel, SNPs & GO, PMut, SNAP2, MutPred2, and SnpEff4.0, for predicting missense nsSNPs. Additionally, we used I-Mutant3.0 and MUpPro to analyze protein stability, ConSurf to assess evolutionary conservation, FTSite and COACH to predict ligand binding sites, SOPMA and PSIPRED to analyze protein secondary structures, project HOPE to predict structural changes, and I-TASSER to model the 3D structure. Furthermore, structural validation was conducted using



© 2024 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

the PROCHECK and ERRAT servers. At the same time, molecular dynamics simulations were performed using GROMACS to gain a better understanding of the effects of mutations on protein structure. Among the 451 missense nsSNPs identified within the *GABRA6* gene, three were found to have pathogenic effects on the structure and function of the protein, potentially, there may be a contribution to the development of seizures or other neuropsychiatric disorders. Notably, two of these missense nsSNPs (*W87S* and *W112R*) were located within the ligand-binding domain, while the third (*C310R*) was situated in the transmembrane domain. It is crucial to acknowledge that despite their predicted pathogenicity, these variants are currently classified as of uncertain significance in clinical and genomic databases worldwide due to the lack of correlation with epilepsy in empirical studies. Without experimental data to validate these predictions, caution is warranted in interpreting the findings.

Keywords

Epilepsy; *GABRA6* gene; single nucleotide polymorphism; *in silico* analysis; genetic marker

1. Introduction

Epilepsy is a neurological disorder characterized by recurring seizures [1]. While approximately 20-30% of epilepsy cases are associated with tumors, strokes, or head injuries, the remaining 70-80% are believed to have genetic factors [2]. Genetic epilepsy refers to cases where seizures are a primary symptom of the disorder and are directly caused by presumed or known mutations. It is important to note that the underlying mutations responsible for genetic epilepsies are still largely unknown [3].

In a study conducted by Wang et al. in 2017, 977 genes associated with epilepsy were identified through a comprehensive search of databases such as OMIM, HGMD, Epilepsy Gene, and PubMed publications. These genes were classified into four groups based on the phenotypes of epilepsy manifestation. Among the 977 genes, 84 were designated as epilepsy genes, 73 as neurodevelopment-related genes, 536 as epilepsy-related genes, and 284 as putatively epilepsy-associated genes, which require further investigation [4]. The focus of this study is to evaluate the *GABRA6* gene, which falls under the putatively epilepsy-associated genes category.

GABA_A receptors, which belong to the Cys-loop ion channel superfamily, play a crucial role as the major inhibitory mediators in the mammalian central nervous system (CNS) [5]. These receptors are composed of pentameric subunits, with 19 different subtypes identified ($\rho 1$ - $\rho 3$, $\gamma 1$ - $\gamma 3$, $\beta 1$ - $\beta 3$, $\alpha 1$ - $\alpha 6$, θ , π , ϵ , and δ) [6]. The diverse molecular composition and expression of GABA_A receptors' subunits contribute to the variations in their properties, including agonist binding affinity, conductance, kinetics, and distribution within the brain [7]. These differences may potentially contribute to the pathogenesis of epilepsy and chronic seizure recurrence [8].

The *GABRA6* gene encodes the alpha-6 subunit of the GABA_A receptor. Previous studies have identified connections between variations in the *GABRA6* gene and genetic epilepsy disorders such as CAE. Dibbens et al. reported a mutation (Arg46Trp) in the *GABRA6* gene in a patient with CAE. This mutation is located in a homologous region to the $\gamma 2$ subunit mutation (Arg82Gln) in the

gamma-2 subunit of the GABA_A receptor, which is known to be associated with febrile seizures and CAE in humans. The $\gamma 2$ subunit mutation results in reduced surface receptor levels and receptor currents, indicating impaired assembly and functionality of GABA_A receptors [9]. Furthermore, Hernandez et al. found that the CAE-associated Arg46Trp mutation in the *GABRA6* gene can cause neuronal disinhibition in the GABA_A receptor, leading to an increased predisposition to generalized seizures due to reduced function and expression of $\alpha\beta\gamma$ and $\alpha\beta\delta$ receptors. This mechanism is similar to the $\gamma 2$ subunit mutation (Arg82Gln) linked to human CAE [10].

When comparing genomic DNA sequences from different individuals, it is observed that certain positions can contain two or more bases. These variations, known as single nucleotide polymorphisms (SNPs), are highly prevalent, occurring at a frequency of approximately one out of every 0.3-1 kilobase in the human genome. Additionally, SNPs can have distinct effects on phenotypic levels depending on their location [11]. SNPs with a minimum frequency of 1% in the general population are believed to account for approximately 90% of inter-individual variability and nearly 100,000 amino acid differences. Due to their low mutation rate, SNPs serve as valuable markers for investigating complex genetic characteristics. Consequently, genetic studies increasingly use SNP markers for their numerous advantages. In general, SNPs can be categorized as either intronic or exonic. Intronic SNPs are located in non-coding genome areas and do not impact protein products when translated. On the other hand, nsSNPs are exonic SNPs found in coding regions that result in changes to protein length or amino acid variants [12]. Among the various types of SNPs, nsSNPs in coding regions are considered to have the most significant effect on protein functions, often leading to mutations in encoded amino acids that can detrimentally affect protein structure and function [13].

The emergence of large, complex biological data sets has necessitated rapid advancements in the field of bioinformatics or computational biology. While not a new field, bioinformatics has become increasingly indispensable as biology transitions towards high-throughput methods for analyzing entire genomes. Bioinformatics encompasses the scientific discipline involved in acquiring, storing, distributing, processing, interpreting, and analyzing biological information. Consequently, numerous databases and software applications have expanded to facilitate the analysis of genetic and physical data in order to infer the functions of model organisms and individuals. Computational analysis provides effective and alternative methods for generating new hypotheses, designing appropriate investigations, and interpreting vast amounts of data derived from genome-scale research. Computational or *in silico* analyses supplement traditional experimental biology techniques by significantly enhancing predictive capabilities [14]. The term "*in silico*" refers to experiments conducted using computers and is analogous to the more familiar biological terms "*in vitro*" and "*in vivo*" [15]. Given the exponential number of SNPs, it is impractical to determine the biological significance of each SNP through wet laboratory techniques alone. However, computational tools can be employed to screen potentially deleterious SNPs that may alter protein function and structure before proceeding to wet laboratory analysis [16].

This study aimed to identify and validate the most deleterious missense nsSNPs associated with the *GABRA6* gene and assess their impact on protein function and structure. This approach involved investigating the underlying molecular mechanisms using various *in silico* methods. Predicted missense nsSNPs, along with the native form, were further analyzed through molecular dynamics simulations to understand their effects on protein structure better. This investigation explores the potential impact of missense nsSNPs on epilepsy and other neuropsychiatric disorders, focusing on

the *GABRA6* gene. While our findings suggest a link between these variants and disease susceptibility, further experimental validation is essential to substantiate these observations. Nonetheless, elucidating the role of missense nsSNPs could offer valuable insights into disease mechanisms and inform future diagnostic and therapeutic strategies.

2. Materials and Methods

The harmful nature of missense nsSNPs in the function and structure of the human *GABRA6* gene was predicted by several computational algorithms. Various *in silico* tools were utilized to make predictions to ensure that the results were highly accurate. The approach followed to analyze missense nsSNPs in the present study is shown in Figure 1.

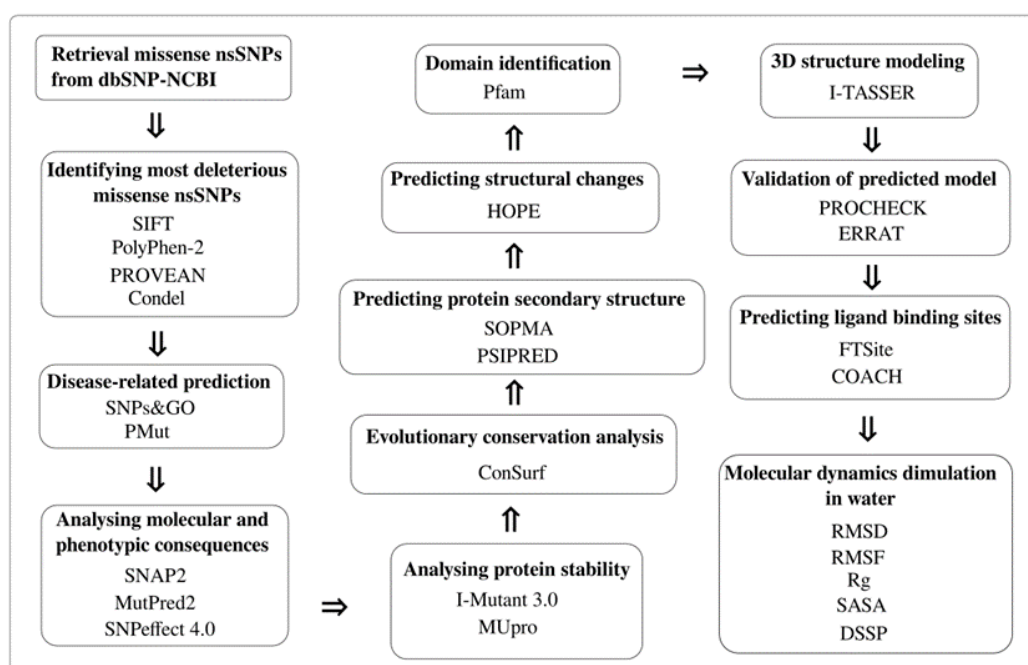


Figure 1 Schematic representation of *in silico* workflow to analyze deleterious missense nsSNPs in human *GABRA6* gene.

2.1 Data Mining

The missense nsSNPs for the human *GABRA6* gene were retrieved from the National Center for Biotechnology Information (NCBI) dbSNP database [17] (<https://www.ncbi.nlm.nih.gov/snp/>). Moreover, the protein sequence encoded by the human *GABRA6* gene was obtained from the UniProt database [18] (<https://www.uniprot.org/>).

2.2 Predicting Deleterious Missense nsSNPs

In silico tools, including SIFT, PolyPhen-2, PROVEAN, and Condel, were utilized to predict deleterious missense nsSNPs.

SIFT (Sorting Intolerant from Tolerant): An algorithm that uses sequence homology and the physical properties of amino acids to predict an amino acid substitution effect on protein function. The algorithm considered evolutionarily conserved regions to be relatively intolerant to mutations,

and therefore, amino acid substitution in these regions is more likely to be expected to affect protein function [19]. In this study, a faster version of SIFT called SIFT 4G (SIFT for genomes) [20] is used (The package is available at <https://github.com/rvaser/sift4g>). SIFT takes a protein sequence and a list of substitutions, which is searched against a protein database for homology searching. The protein database is UniRef90 (16 February 2021 release) [21]. SIFT score ranges from 0 to 1. The amino acid substitution is considered harmful if the score < 0.05 , while the score ≥ 0.05 is tolerated.

PolyPhen-2 (Polymorphism Phenotyping v2): A tool that predicts the structural and functional impacts of an amino acid substitution in human proteins. PolyPhen-2 (The package is available at <http://genetics.bwh.harvard.edu/pph2/dokuwiki/downloads>) uses a Naïve Bayes classifier, which is trained by supervised learning algorithms. HumDiv and HumVar datasets from the UniProt database were used to train PolyPhen-2 prediction models. The classifier that used HumDiv should be used for rare alleles assessment, natural selection analysis, and dense mapping of regions determined through genome-wide association studies. The HumVar classifier should be utilized for Mendelian disease recognition, which requires recognizing mutations with severe effects from all the remaining human variation, including many mildly deleterious alleles [22]. PolyPhen-2 takes protein UniProt ID and substitutions as an input file and identifies homologs of the protein sequences by Basic Local Alignment Search Tool (BLAST) [23] in the UniRef100 (27 February 2021 release) database. The PolyPhen-2 score ranges from 0 (neutral) to 1 (deleterious), and functional prediction is categorized into probably damaging (damaging with high confidence), possibly damaging (damaging with low confidence), and benign (benign with high confidence).

PROVEAN (Protein Variation Effect Analyzer): A tool based on an alignment scoring method that predicts functional effects of single amino acid substitutions, in-frame insertions, deletions, and multiple substitutions [24]. PROVEAN (<http://provean.jcvi.org/index.php>) uses a delta alignment score (or delta score) to measure the effect of a variation. The protein sequence and a list of substitutions were submitted as inputs to the algorithm. PROVEAN's default threshold score is -2.5. A score < -2.5 is considered harmful, while a score > -2.5 is considered neutral [25].

Condel (CONsensus DEleteriousness): A method to predict the outcome of non-synonymous SNPs based on a weighted average of the normalized scores (WAS) of various tools. Condel (<https://bbglab.irbbarcelona.org/fannsd/>) uses a consensus deleteriousness score that integrates the output of five tools (Logre [26], MAPP [27], Mutation Assessor [28], Polyphen2, and SIFT) into a unified classification [29]. Protein UniProt ID and substitutions were submitted as an input file to the algorithm. Condel's score ranges from 0 to 1, which is a higher score that predicts SNPs as deleterious.

2.3 Predicting Disease-Related Variants

To disease-related prediction of selected missense nsSNPs, SNPs & GO and PMut tools were used in this step.

SNPs & GO A support vector machine (SVM) based tool for predicting disease-related SNPs using Gene Ontology (GO) terms (<http://snps.biofold.org/snps-and-go>) [30]. Protein Swiss Prot code and substitutions were submitted as inputs to the algorithm. When the input sequence is a Swiss Prot code, GO terms for the prediction will be retrieved automatically. The result consists of three different algorithms: SNPs & GO, PhD-SNP [31], and PANTHER [32]. SNPs & GO score ranges from 0

to 1, if the score > 0.5 is considered a disease-related variation, while a score ≤ 0.5 is considered benign.

PMut: A tool that predicts pathological properties of single amino acid substitutions. PMut (<http://mmb.irbbarcelona.org/PMut>) uses a neural network-based method that is trained using a manually curated database Swiss Prot (October 2016 release), includes 27,203 deleterious and 38,078 benign mutations for 12,141 proteins [33]. Protein Uni Prot ID and substitutions were submitted as inputs to the algorithm. PMut score ranges from 0 to 1. A score > 0.5 is classified as pathological variation, and a score < 0.5 is classified as neutral.

2.4 Predicting the Molecular and Phenotypic Consequences of Variants

Further investigation was done to predict molecular and phenotypic consequences of selected missense nsSNPs through SNAP2, MutPred2, and SNPeff4.0 tools.

SNAP2: A neural network-based classifier tool that predicts the functional effects of non-synonymous SNPs (<https://roslab.org/services/snap2web/>). SNAP2 uses evolutionary information and structural features like solvent accessibility and predicted secondary structure to predict variants' effects [34]. The protein sequence was submitted as input to the algorithm. SNAP2 score ranges from -100 (firmly neutral) to +100 (enormously influential).

MutPred2: A machine learning-based method that predicts the molecular and phenotypic consequences caused by amino acid substitutions as pathogenic or benign (The package is available at <http://mutpred.mutdb.org/>). It was trained using 53,180 deleterious and 206,946 putatively benign mutations from the Human Gene Mutation Database (HGMD) [35], Swiss Var database [36], dbSNP database, and inter-species pairwise alignments. MutPred2 takes protein Uni Prot ID and substitutions as an input file. MutPred2 gives two output scores: general score (g) and property score (pr). The g score demonstrates the pathogenicity of the substitution and ranges between 0 and 1, a higher score indicates a greater probability of being pathogenic. The pr score is the posterior probability of losing or gaining a certain property because of the substitution and it also ranges between 0 and 1, a higher score indicates more alteration of the property in the molecular mechanism of the disease [37].

SNPeff 4.0: A database that uses 4 tools to predict molecular and structural phenotypic consequences of human protein-coding SNVs (<https://snpeffect.switchlab.org/>). SNPeff 4.0 uses TANGO [38] for aggregation-prone regions prediction, WALTZ [39] for amyloidogenic regions prediction, LIMBO [40] for hsp70 chaperone-binding sites prediction, and FoldX [41] for analyzing the possible impact of protein stability [42]. Protein UniProt ID and substitutions were submitted as inputs to the algorithm. Mutations can increase (dTANGO, dWALTZ, and dWALTZ > 50), decrease (dTANGO, dWALTZ, and dWALTZ < -50), or have no effect (dTANGO, dWALTZ, and dWALTZ between -50 and 50), aggregation propensity, amyloid propensity, and chaperone binding.

2.5 Predicting the Effect of Variants on Protein Stability

Analyzing missense nsSNPs effects on protein stability was accomplished by I-Mutant3.0 and MUpro tools through Gibbs free energy change calculation of mutant protein and wild-type protein.

I-Mutant: A support vector machine (SVM) based web server tool that predicts protein stability changes upon single amino acid substitutions (<https://folding.biocloud.org/i-mutant/>). I-Mutant estimates protein stability by calculating Gibbs free energy change ($\Delta\Delta G$, kcal/mol) of the mutant

protein and wild-type protein ($\Delta\Delta G = \Delta G$ (mutant type)- ΔG (wild-type)) [43]. This server can evaluate the protein stability change by the protein sequence or structure. In this study, protein sequence was used. The protein sequence and substitutions were submitted as inputs to the algorithm. If $0 < \Delta\Delta G$ is considered to increase protein stability, while $\Delta\Delta G < 0$ is considered to decrease protein stability. I-Mutant 3.0 predicts which a single substitution has a low effect ($-0.5 \leq \Delta\Delta G \leq 0.5$ kcal/mol) or extensively increases protein stability ($\Delta\Delta G > 0.5$ Kcal/mol) or extensively decreases protein stability ($\Delta\Delta G < -0.5$ Kcal/mol).

MUpro: A tool based on support vector machine (SVM) and neural network methods, which is used for protein stability prediction resulting from single amino acid substitutions (<http://mupro.proteomics.ics.uci.edu/>) [44]. MUpro also like I-Mutant can predict the stability changes using protein sequence or protein structure. The protein sequence and substitutions were submitted as inputs to the algorithm. When the energy change ($\Delta\Delta G$) is positive, the mutation increases protein stability. Instead, if $\Delta\Delta G$ is negative the mutation decreases protein stability.

2.6 Evolutionary Conservation Analysis

The ConSurf tool was utilized to identify selected missense nsSNPs' positions in evolutionarily conserved regions.

ConSurf: A tool used for evolutionary conservation analysis of amino acids in a protein, DNA, or RNA to reveal important functional and/or structural areas (<https://consurf.tau.ac.il/>). ConSurf uses an empirical Bayesian method to estimate the protein sequence conservation score [45]. The protein sequence was submitted as input to the algorithm. The conservation score is grouped into 9 grades, where 1 indicates most rapidly evolving areas, 5 indicates mildly changing areas, and 9 indicates most evolutionary conserved areas.

2.7 Predicting Protein Secondary Structure

Secondary structures of selected missense nsSNPs were predicted by SOPMA and PSIPRED tools.

SOPMA: A web version of the self-optimized prediction method (SOPM) [46], which is used for secondary structure prediction of amino acids (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). The joint prediction using SOPMA, and a neural network method (PHD) correctly predicts 82.2% of residues for 74% of co-predicted amino acids. The protein sequence was submitted as input to the algorithm.

PSIPRED: A highly accurate method that predicts the secondary structure of proteins (<http://bioinf.cs.ucl.ac.uk/psipred/>). PSIPRED [47] includes two feed-forward neural networks that perform an analysis of the output obtained from PSI-BLAST [23]. The protein sequence was submitted as input to the algorithm.

2.8 Predicting Structural Effects of Variants

The HOPE project predicted structural changes such as size, charge, and hydrophobicity-value between mutant and wild residues.

HOPE (Have (y)Our Protein Explained): A web tool to predict the structural effects of single amino acids substitutions on proteins (<https://www3.cmbi.umcn.nl/hope/>). HOPE uses 3D structures of the proteins available in the Uni Prot database and Distributed Annotation (DAS) servers to simulate

the structural features of mutations on the wild-type protein [48]. However, if necessary, the HOPE server can build homology models independently. HOPE server also predicts structural changes between mutant and wild residues. The protein sequence was submitted as input to the algorithm.

2.9 3D Structure Modeling and Visualizing of Natives and Variants Structures

The 3D structure of the protein is crucial to annotate the protein function of native and mutant structures. Identification of domain regions and locating missense nsSNPs position in domains were done through the Pfam database [49] search (<http://pfam.xfam.org/>). Then, generating 3D structures of each native domain using I-TASSER (Iterative Threading ASS Embly Refinement) was accomplished separately (<https://zhanglab.dcmf.med.umich.edu/I-TASSER/>). I-TASSER is a method for predicting protein structure that consists of 3 steps. In the first step, after the domain sequence was submitted to the I-TASSER server as input, a meta-threading program called LOMETS [50] retrieved structural templates from the PDB library [51]. In the second step, full-length models are constructed by assembling well-aligned continuous fragments excised from the PDB templates, with unaligned regions structures built by ab-initio modeling based on replica-exchange Monte Carlo simulations [52]. I-TASSER simulations generate thousands of conformations named decoys. I-TASSER uses the SPICKER program (a clustering algorithm) [53] to cluster all decoys. Through clustering, SPICKER identifies low free-energy states. Eventually, the functional annotations are obtained by matching the predicted structure models with known proteins in the BiLiP function library [54]. I-TASSER will predict up to the top five models. A confidence score or C-score estimates the quality of predicted models. C-score is typically in the range of [-5.2], wherein a higher value signifies a better-quality model; when C score > -1.5, it indicates the model has a correct fold. I-TASSER takes protein sequence as input, and up to five models are predicted as output. In cases where the output final models are less than five, the similarity of the top templates identified via LOMETS is indicated, and this causes the converging of I-TASSER simulations. The C-score is usually high in such cases and predicts a high-quality model [55-57].

PROCHECK and ERRAT carried out structural validation for the predicted models. PROCHECK by Psi/Phi Ramachandran plot analysis assesses the stereochemical quality of a protein structure [58]. ERRAT is another protein structural verification tool that demonstrates the model's reliability. For the best quality of the predicted model, the ERRAT value must be over 80% [59]. PROCHECK and ERRAT are available at the Structural Analysis and Verification Server (SAVES) (<https://saves.mbi.ucla.edu/>). PDB file was submitted as input to the server for both tools. Modeling of mutated protein structures was performed using the mutagenesis feature in the PyMol [60] visualization tool (2.3.0 version) using the predicted native type by I-TASSER as a reference.

2.10 Predicting Ligand Binding Sites

The FTSite and COACH tools were used to predict the presence of selected missense nsSNPs in ligand-binding sites.

FTSite: An accurate computational method for ligand binding site prediction based on experimental evidence with an accuracy of 94% (<https://ftsite.bu.edu/>) [61]. PDB file was submitted as input to the algorithm. FTSite identifies three potential binding sites in proteins.

COACH (COnsensus approACH): A tool for ligand binding site prediction, which is based on two comparative methods, TM-SITE (binding-specific substructure comparison) and S-SITE (sequence

profile alignment), which identify ligand binding templates from the BioLiP library. These two methods, combined with COFACTOR [62], Con Cavity [63], and FINDSITE [64] methods, generate final predictions [65]. After the protein sequence was submitted to the I-TASSER for generating 3D structures, it will be fed into the COACH pipeline for predicting ligand binding sites.

2.11 Molecular Dynamics Simulation in Water

Molecular dynamics simulations were performed using the GROMACS (2021.2 version) to reveal changes at the atomic level in different time scales for wild and mutant types [66]. The Amber ff99SB protein force field was used for simulations. A cubic box was formed by extending 1 nm on each side of the protein. The system was implemented by adding TIP3P water model molecules and neutralizing them with Cl-ions. The energy minimization step was performed using the steepest descent algorithm before the simulation. Long-range electrostatic interactions were modeled using the particle mesh Ewald (PME) method. The short-range electrostatic and van der Waals interactions cut-off radius was set to 1 nm. Periodic boundary conditions were maintained to eliminate surface effects. The leap-frog algorithm was used for simulations to integrate Newton's equations of motion. A time step interval of 2 fs was used for all simulations. To constrain bonds involving hydrogens the Lincs algorithm. Minimized systems should be equilibrated to reach the desired temperature and pressure before the fundamental dynamics start. Equilibration is conducted in two ensembles: NVT (constant Number of particles, Volume, and Temperature) and NPT (constant Number of particles, Pressure, and Temperature). The system temperature and pressure were coupled to a v-rescale thermostat with a time constant of 0.1-ps at 300 K and a Parrinello-Rahman barostat with a time constant of 2-ps at 1 bar. We first conducted a 200-ps NVT equilibration and then a 1000-ps NPT equilibration. The molecular dynamics simulation of each equilibrated system of the wild and mutant types was run for 100 ns. At last, the MDS trajectory files analyses were calculated by GROMACS build-in programs to get RMSD (root mean square deviations), RMSF (root mean square fluctuations), Rg (radius of gyration), SASA (solvent accessible surface area), and DSSP (definition secondary structure of the protein) analyses. RMSD, RMSF, Rg, and SASA of C α atoms in the trajectories were evaluated. The changes in the secondary structure of the protein during molecular dynamics simulations were also calculated.

3. Results

3.1 Missense nsSNPs Datasets

For the biological consequences study of missense nsSNPs, we retrieved 451 missense nsSNPs with 386 rsIDs in the human *GABRA6* gene mapped to NM_000811.3 RefSeq (Table S1) from the dbSNP database. The protein sequence of the *GABRA6* gene (with Q16445 ID) was retrieved from the UniProt database.

3.2 Prediction of Most Deleterious Missense nsSNPs

This study used four in silico tools, SIFT, PolyPhen-2, PROVEAN, and Condel, to predict deleterious missense nsSNPs. Out of 451 missense nsSNPs, 303 were predicted as deleterious and 148 as tolerated by SIFT, 306 were predicted as deleterious and 145 as neutral by HumDiv PolyPhen-2 (143 benign, 49 possibly damaging, and 259 probably damaging), 290 were predicted as deleterious and

161 as neutral by HumVar PolyPhen-2 (157 benign, 55 possibly damaging, and 239 probably damaging), 225 were predicted as deleterious and 226 as neutral by PROVEAN, and 298 were predicted as deleterious and 153 as neutral by Condel (Table S2 and Figure 2).

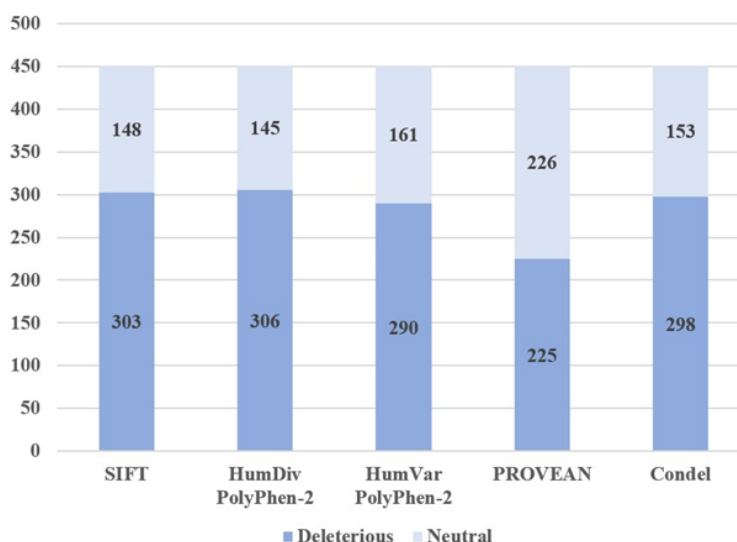


Figure 2 Distribution of predicted deleterious and neutral missense nsSNPs by 4 tools for human *GABRA6* gene. Deleterious missense nsSNPs are shown in dark blue, and neutral nsSNPs are in light blue.

Deleterious missense nsSNPs were predicted with default scores for each tool. High scores across the tools are considered to obtain high confidence deleterious missense nsSNPs. SIFT score = 0, PolyPhen-2 score > 0.99, PROVEAN score < -9, and Condel score > 0.7 are used. After integrating the scores, 3 missense nsSNPs in the *GABRA6* gene (Table 1) were obtained and used for further analysis.

Table 1 The most deleterious missense nsSNPs by SIFT, PolyPhen-2, PROVEAN, and CONDEL tools in the *GABRA6* gene.

rs ID	Transcript change	Substitution	SIFT4G Score	PolyPhen-2 Score		PROVEAN Score	CONDEL Score
				HumDiv probability	HumVar probability		
rs1428649051	c.260G > C	W87S	0	1	1	-12.906	0.709363974
rs1199782347	c.334T > C	W112R	0	1	1	-12.837	0.735214831
rs1317373536	c.928T > C	C310R	0	1	1	-10.964	0.764688222

3.3 Prediction of Disease-Related Variants

To get more accurate results, selected missense nsSNPs were analyzed by SNPs & GO and PMut to predict disease-related missense nsSNPs. SNPs & GO's result consists of three different algorithms; SNPs & GO, PhD-SNP, and PANTHER. Three selected missense nsSNPs in the human *GABRA6* gene were submitted to SNPs & GO and PMut to analyze disease-related missense nsSNPs. The output of SNPs & GO, PhD-SNP, PANTHER, and PMut predicted that all three selected missense nsSNPs are disease-related (Table 2).

Table 2 Lists disease-related missense nsSNPs of *GABRA6* by SNPs & GO and PMut tools.

Substitution	SNPs & GO						PMut				
	SNPs & GO			PhD-SNP			PANTHER			PMut	
	Prediction	RI	Probability	Prediction	RI	Probability	Prediction	RI	Probability	Prediction	Score
W87S	Disease	8	0.919	Disease	8	0.922	Disease	10	0.999	Disease	0.8214
W112R	Disease	9	0.93	Disease	9	0.937	Disease	10	0.979	Disease	0.8214
C310R	Disease	9	0.949	Disease	9	0.97	Disease	9	0.974	Disease	0.8627

RI: Reliability Index.

3.4 Prediction of the Molecular and Phenotypic Consequences of Variants

To predict the molecular and phenotypic consequences of selected missense nsSNPs, we further investigated missense nsSNPs through SNAP2, MutPred2, and SNPeffct4.0 tools. The SNPeffct tool predicts SNPs' consequence on aggregation-prone regions by TANGO, amyloid-forming regions by WALTZ, and hsp70 chaperone binding sites by LIMBO. The output of SNAP2 of three selected missense nsSNPs in the *GABRA6* gene predicted that all three selected missense nsSNPs have damaging effects on protein structure. Since the score > 0.5 suggests pathogenicity in MutPred2, the output of MutPred2 showed that all three selected missense nsSNPs in the *GABRA6* gene have damaging functional and structural effects (Detailed information in Table S3). According to the SNP effect results, TANGO and WALTZ analysis revealed that the *C310R* variant (dTANGO equals 302.70 and dWALTZ equals -553.68) increases the aggregation tendency of the protein and decreases the amyloid propensity of the protein. However, none of the variations does affect the chaperone binding tendency of the protein (Table 3).

Table 3 List of analyzed missense nsSNPs of the *GABRA6* gene by SNAP2, MutPred2, and SNPeffct4.0 tools.

Substitution	SNAP2		MutPred2 score	SNPeffct4.0		
	Prediction	Score		dTANGO	dWALTZ	dLIMBO
W87S	effect	87	0.953	0.00	-0.07	0.00
W112R	effect	87	0.948	0.00	-0.12	0.00
C310R	effect	87	0.964	302.70	-553.68	0.00

(d in dTANGO, dWALTZ, and dLIMBO is the score change between the mutant and wild-type proteins).

3.5 Prediction of the Effect of Variants on Protein Stability

The effects of selected missense nsSNPs on protein stability were analyzed by I-Mutant3.0 and MUpuro tools by calculating the Gibbs free energy change of mutant and wild-type proteins. I-Mutant3.0 analyses showed that *W87S* and *W112R* variants decreased protein stability, and the *C310R* variant increased protein stability. In contrast, all three variants by Mupro prediction showed decreasing protein stability. *W87S* and *W112R* variants have $\Delta\Delta G$ values < -1 kcal/mol in both tools, which are expected to alter the function and structure of the protein (Table 4).

Table 4 List of missense nsSNPs of the *GABRA6* gene, which was analyzed for protein stability by I-Mutant3.0 and MUpro tools.

Substitution	I-Mutant3.0		Mupro		
	Prediction	ΔΔG	RI	Prediction	ΔΔG
W87S	Decrease	-1.69	9	Decrease	-1.0174971
W112R	Decrease	-1.44	9	Decrease	-1.2580621
C310R	Increase	-0.32	2	Decrease	-0.85748451

RI: Reliability Index.

3.6 Evolutionary Conservation Analysis

To analyze the evolutionary conservation of selected missense nsSNPs, the ConSurf tool was used, which grouped amino acids based on conservation scores in 9 grades. As shown in Figure 3, ConSurf results indicated that W87 and W112 residues located in highly conserved regions with conservation scores of 9 and predicted as buried residues and have a structural impact on the protein, and C310 predicted as buried residue with conservation scores of 8.

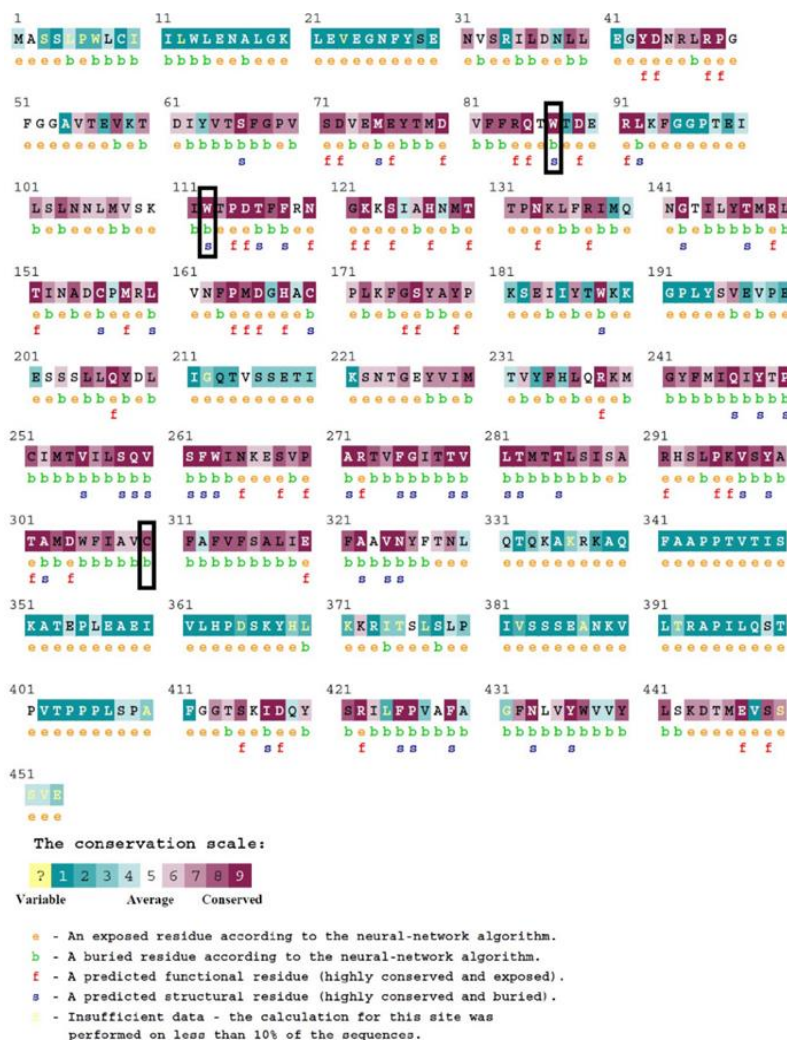


Figure 3 ConSurf analysis of *GABRA6* gene residues. The black boxes indicate the most deleterious missense nsSNPs.

3.7 Prediction of Protein Secondary Structure

Amino acids' secondary structure corresponding to the proteins was predicted by SOPMA and PSIPRED tools. SOPMA prediction showed the distributions of alpha helix, extended strand, beta-turn, and random coil in proteins. PSIPRED predicted the distributions of strands, helices, and coils in proteins and validated the secondary structure of the proteins. SOPMA results showed 135 alpha helices (29.80%), 104 extended strands (22.96%), 11 beta turns (2.43%), and 203 random coils (44.81%) in the predicted secondary structure. In the 3 amino acid residues corresponding to selected missense nsSNPs, W87 and W112 are in random coils, and C310 in alpha helices (Figure 4). PSIPRED results indicated that in the 3 amino acid residues, W87 is located in strands, Y186 in coils, and C310 in helices (Figure S1).

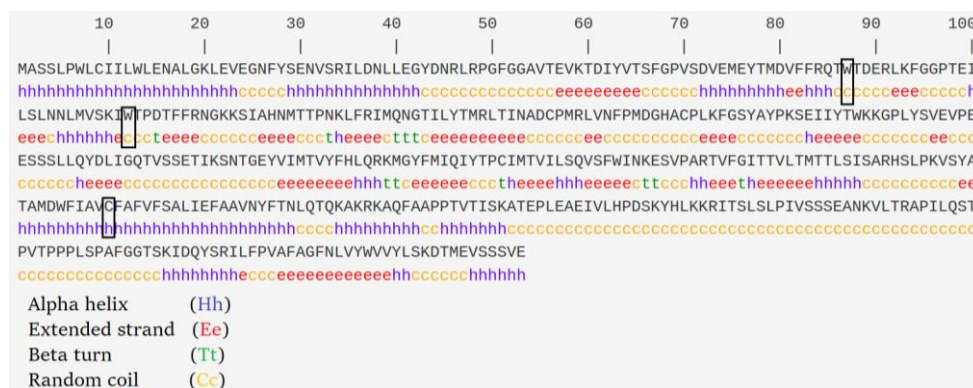


Figure 4 SOPMA analysis of *GABRA6* gene residues. The black boxes indicate the most deleterious missense nsSNPs.

3.8 Prediction of Structural Effects of Variants

HOPE was used to predict structural changes, including size, charge, and hydrophobicity value between mutant and wild residues. Of 3 selected missense nsSNPs in the *GABRA6* gene, Ser and Arg residues in W87S and W112R mutants are more minor than Trp residue in wild-type forms. In contrast, Arg residue in the C310R mutant is more significant than the Cys residue in the wild-type form. The charge of Trp and Cys residues in W112R and C310R mutants was neutral, then turned to Arg residues with a positive charge. There is not any significant charge change in the W87S mutant. The wild-type residues in all three missense nsSNPs are more hydrophobic than the mutant residues. (Table S4). According to the project HOPE results, when the mutant residue is smaller than the wild-type residue, it might lead to a loss of interactions in protein structure. If the mutant residue is more significant than the wild-type residue, it might lead to bumps in protein structure. When the charge of the wild-type residue is lost, it might cause a loss of interactions with other molecules or residues. If the mutation introduces a charge, it might cause the repulsion of ligands or other residues with the same charge. When the hydrophobicity of the wild-type residue is lost or decreased, the hydrophobic interactions will be lost either in the core of the protein or on the surface. If the mutation introduces a more hydrophobic residue, it might cause a loss of hydrogen bonds and/or disturb correct folding.

3.9 3D Structure Modeling and Visualizing of Wild and Mutant Structures

The Pfam server was used to identify domain regions in the *GABRA6* gene and locate selected missense nsSNP positions in different domains. In this study, we need domains that involve deleterious missense nsSNPs (Table 5). Generating 3D structure models was performed by the I-TASSER server. Predicted models' validation was done by PROCHECK and ERRAT servers. PROCHECK predicted the quality of wild models using Ramachandran plot analysis. The ERRAT indicated the overall quality factor of the predicted models. PyMol software was used to visualize wild and mutant protein structures.

Table 5 List domains involving selected deleterious missense nsSNPs in the *GABRA6* gene.

Protein name	Domain name	Start residue	Stop residue	Domain length	Involved missense nsSNPs
GBRA6_HUMAN	Neurotransmitter-gated ion-channel ligand-binding domain	32	240	209 aa	W87S and W112R
GBRA6_HUMAN	Neurotransmitter-gated ion-channel transmembrane domain	247	399	153 aa	C310R

Pfam reported two domains in the *GABRA6* gene, including the neurotransmitter-gated ion-channel ligand-binding domain (32-240), and a neurotransmitter-gated ion-channel transmembrane domain (247-399). Out of three selected missense nsSNPs for further analysis, two (*W87S* and *W112R*) are located in the ligand-binding domain, and one (*C310R*) is in the transmembrane domain.

3D structure prediction of wild types for the ligand-binding domain and the transmembrane domain in the *GABRA6* gene was modeled by I-TASSER. One model predicted a C-score of 1.16 for the ligand-binding domain, which was a high-quality model. And in the transmembrane domain, we selected the first model out of 5 predicted models with a C-score of -1.62 (Figure S2a and Figure S3a). Validation of predicted models' quality was done by PROCHECK and ERRAT servers. The result of PROCHECK for the ligand-binding domain model showed 79.7% of residues in most favored regions, 18.7% in additional allowed regions, 1.6% in generously allowed regions, and 0% in the disallowed areas (Figure S2b). The ERRAT result showed that the overall quality factor for the predicted model was 87.940 (Figure S2c). PROCHECK's result for the transmembrane domain model showed 77.3% of residues in the most favored regions, 12.8% in additional allowed regions, 7.1% in generously allowed regions, and 2.8% in the disallowed areas (Figure S3b). The ERRAT result showed that the predicted model's overall quality factor was 83.916 (Figure S3c).

Finally, the predicted models were used to model mutant types by utilizing the mutagenesis feature in PyMol. Structural models for wild-types and deleterious missense nsSNPs in the ligand-binding domain (*W87S* and *W112R*) and the transmembrane domain (*C310R*) are shown in Figure 5, Figure 6 and Figure 7.

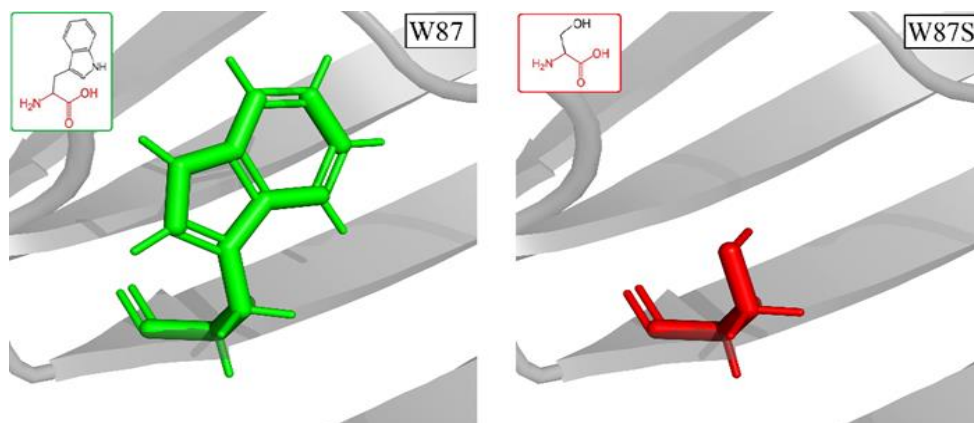


Figure 5 (W87S): The amino acid Tryptophan (green) changed to Serine (red) at position 87 in the ligand-binding domain. Visualization was done by PyMol software and HOPE result.

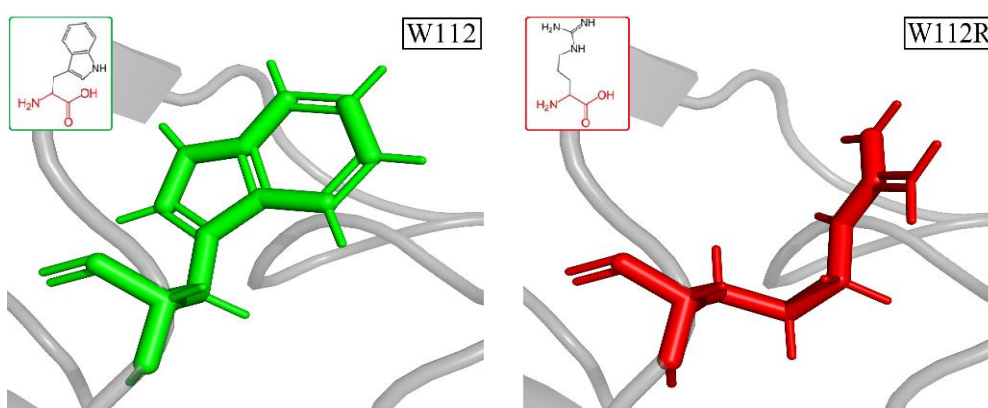


Figure 6 (W112R): The amino acid Tryptophan (green) changed to Arginine (red) at position 112 in the ligand-binding domain. Visualization was done by PyMol software and HOPE result.

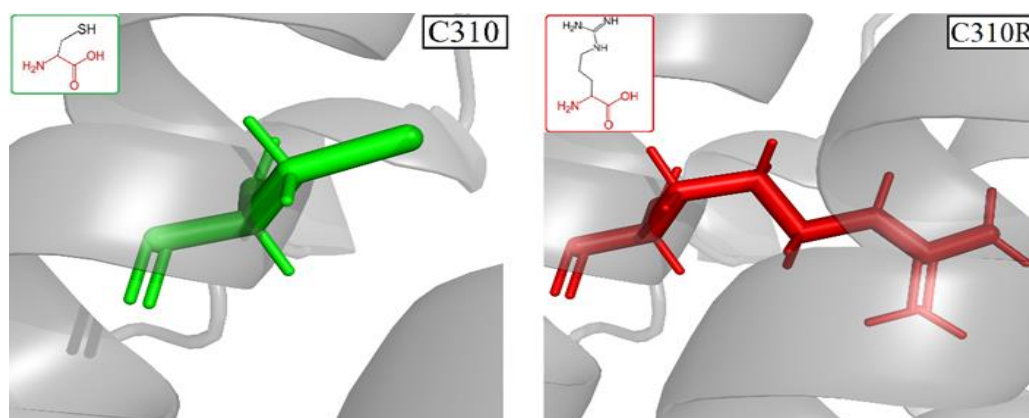


Figure 7 (C310R): The amino acid Cysteine (green) changed to Arginine (red) at position 310 in the transmembrane domain. Visualization was done by PyMol software and HOPE result.

3.10 Prediction of Ligand Binding Sites

To determine the presence of the selected deleterious missense nsSNPs in protein-binding regions of the *GABRA6* gene, we employed the FTSite and COACH algorithms, which are protein-ligand docking tools. FTSite identified three potential binding sites in the ligand-binding domain of the protein.

FTSite predicted three potential binding sites in the ligand-binding domain. Site 1 contains 15 binding residues, site 2 includes 4, and site 3 contains 12 (Figure 8a and Table S5). Additionally, COACH predicted 20 amino acid residues as potential binding sites (Table S6).

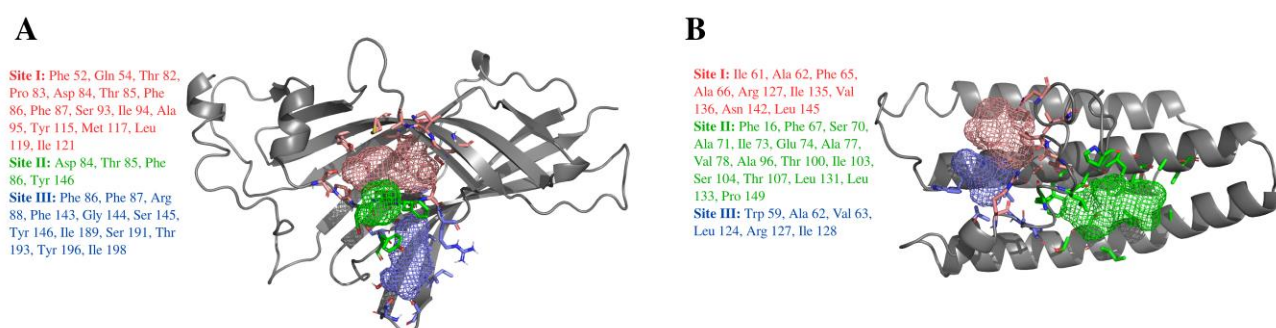


Figure 8 Illustration of *GABRA6*, ligand-binding domain (A), and transmembrane domain (B), with ligand-binding site predictions. FTSite were annotated the ligand-binding pockets. Site 1 is illustrated in red, while sites 2 and 3 are described in green and blue, respectively.

Similarly, in the transmembrane domain, FTSite predicted three potential binding sites. Site 1 contains 9 binding residues, site 2 includes 16 binding residues, and site 3 contains 6 (Figure 8b and Table S5). COACH predicted 44 amino acid residues as potential binding sites (Table S6).

3.11 Molecular Dynamics Simulation of WT and Mutant Types

To comparatively study the conformational changes of the WT and mutant types in physiological environments, we performed 100 ns molecular dynamic simulations for each domain. Various parameters, such as root mean square deviations (RMSD), root mean square fluctuations (RMSF), the radius of gyration (Rg), and solvent accessible surface area (SASA), were analyzed using the time-dependent function of molecular dynamic simulation. These parameters were calculated for C α atoms during the molecular dynamics simulations, with reference to their WT structures.

The RMSD analysis of the WT and mutant types revealed significant deviations in their structural stability. The *W87S* and *W112R* mutants in the ligand-binding domain exhibited similar RMSD values. Compared to the WT structure, the mutant types in the ligand-binding domain showed higher fluctuation, as depicted in Figure 9a. The *C310R* mutant in the transmembrane domain also deviated from the WT structure, with the mutant type displaying lower fluctuation than the WT, as shown in Figure 9b. The average RMSD values for the WT, *W87S*, and *W112R* are 1.88 Å, 2.13 Å, and 2.38 Å, respectively. Furthermore, the WT and *C310R* mutants have average RMSD values of 6.18 Å and 4.89 Å, respectively.

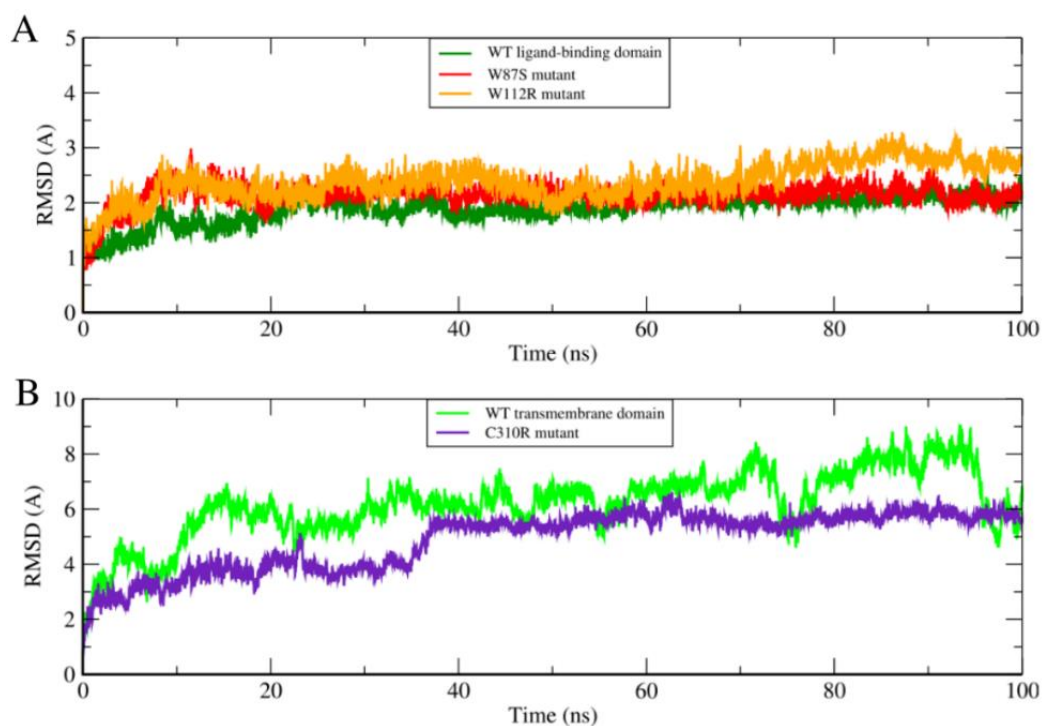


Figure 9 Analysis of RMSD values of the domain backbone of WT and mutant structures, ligand-binding domain (A), and transmembrane domain (B) at 100 ns simulation. The ordinate is RMSD (Å), and the abscissa is time (ns).

The RMSF analysis of each residue illustrates the effect of mutations on their dynamics. The study of RMSF values revealed significant differences in fluctuation between the WT and mutant structures in the N-terminal region of the *W87S* mutant and the N-terminal region as well as positions 90-99 in the *W112R* mutant in the ligand-binding domain after 100 ns of molecular dynamic simulation (Figure 10a). Similarly, in the transmembrane domain, the analysis of RMSF values indicated significant differences in fluctuation between the WT and mutant structures in positions 38-57, 77-92, and the C-terminal region of the *C310R* mutant (Figure 10b). The RMSF plots show that residues in positions 1-22 and 65-100 of the *W112R* mutant in the ligand-binding domain, as well as residues in positions 39-53 and 77-99 of the *C310R* mutant in the transmembrane domain, exhibit a relatively flexible region compared to other residues. Additionally, the highest residual fluctuation is observed at positions 1 (8.01 Å) and 2 (5.96 Å) in the *W87S* mutant, positions 1 (6.90 Å) and 2 (4.84 Å) in the *W112R* mutant, and positions 85 (7.95 Å) and 86 (7.71 Å) in the *C310R* mutant, when compared to their respective WT structures.

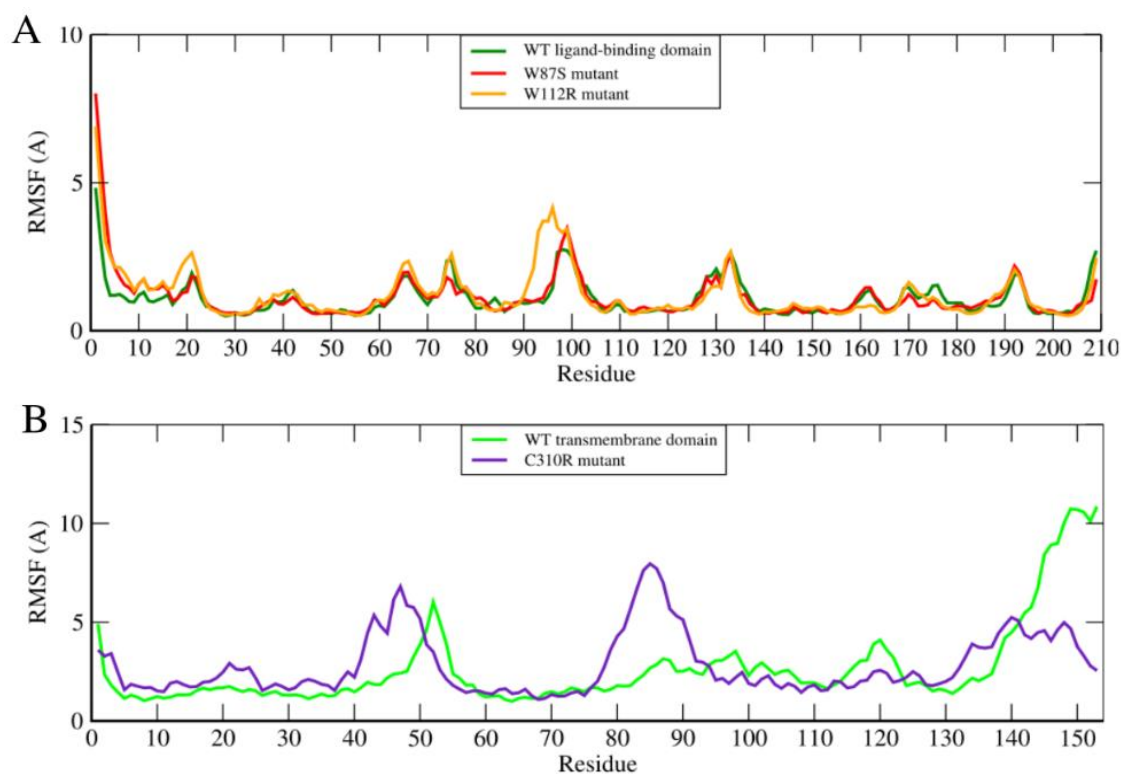


Figure 10 Analysis of RMSF values of the domain backbone of WT and mutant structures, ligand-binding domain (A), and transmembrane domain (B) over the entire simulation. The ordinate is RMSF (Å), and the abscissa is residue.

Based on the R_g analysis of the WT and mutant structures, it is evident that the *W112R* mutant exhibits a higher average R_g value (19.83 Å) in the ligand-binding domain compared to the WT (19.48 Å) and *W87S* mutant (19.41 Å), as depicted in Figure 11a. The ligand-binding domain's WT and *W87S* mutant structures display similar average R_g values. Conversely, in the transmembrane domain, the *C310R* mutant demonstrates a significantly lower average R_g value (16.31 Å) than its WT structure (17.03 Å), as illustrated in Figure 11b. This suggests a potential decrease in the flexibility of the *C310R* mutant, and interestingly, the *C310R* mutant appears to deviate from its R_g value after 30 ns.

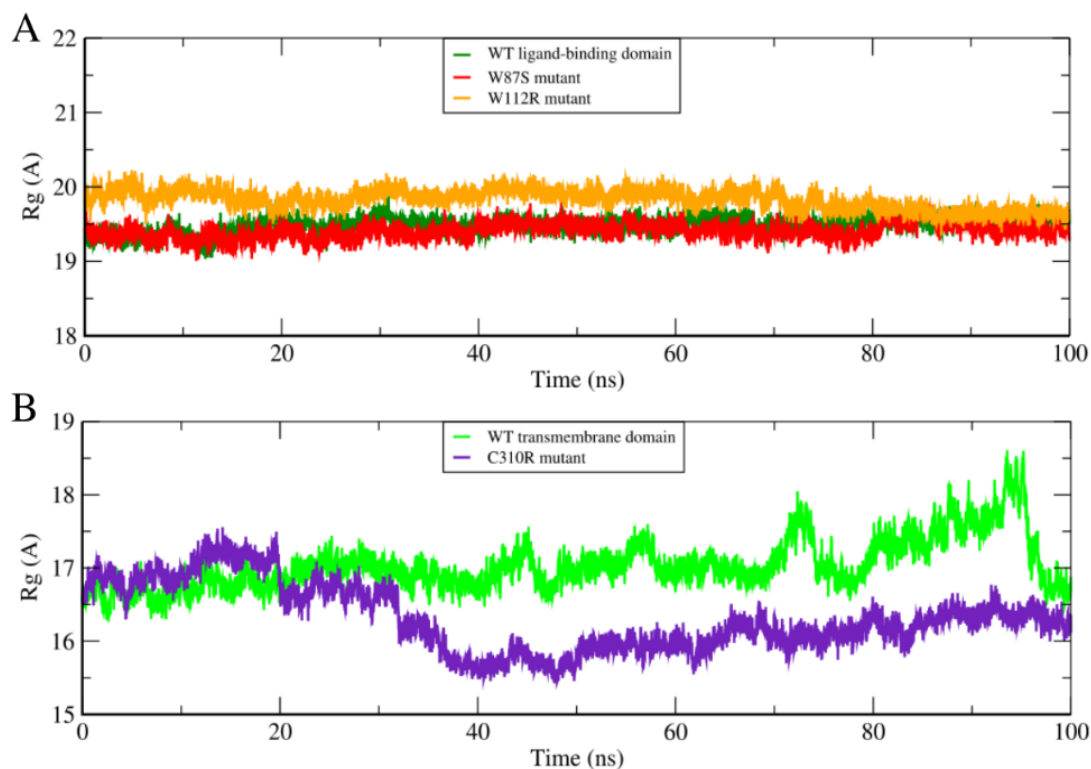


Figure 11 Analysis of R_g of the domain backbone of WT and mutant structures, ligand-binding domain (A), and transmembrane domain (B) over the entire simulation. The ordinate is R_g (Å), and the abscissa is time (ns).

Furthermore, the SASA analysis reveals that in the ligand-binding domain, the *W112R* mutant exhibits a higher average SASA value (11663.50 Å²) than the WT (11468.27 Å²). In comparison, the *W87S* mutant displays a lower average SASA value (11370.25 Å²) than the WT, as shown in Figure 12a. In the transmembrane domain, the *C310R* mutant exhibits a lower average SASA value (8796.62 Å²) than its WT (8936.70 Å²), as depicted in Figure 12b. Since a higher SASA value indicates protein expansion, it can be inferred that in the ligand-binding domain, the WT and *W87S* mutant are more stable than the *W112R* mutant, and the *W87S* mutant is more stable than the WT. The *C310R* mutant is more stable than its WT structure in the transmembrane domain.

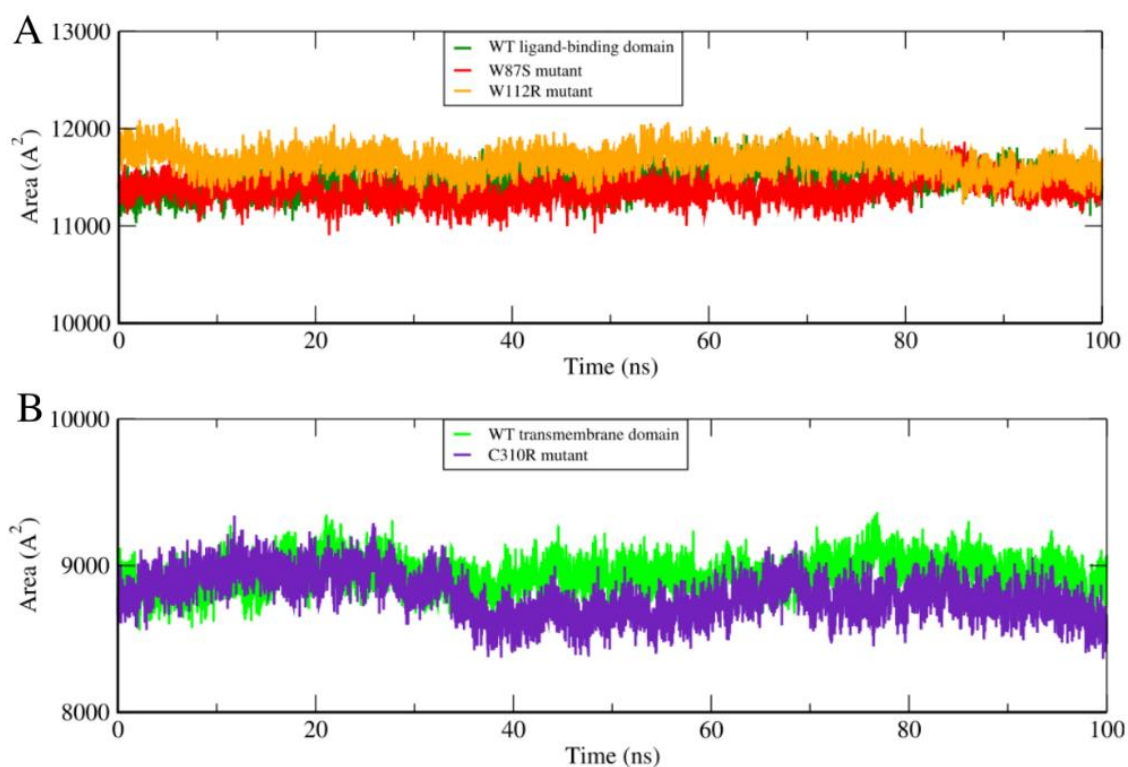


Figure 12 Analysis of SASA of the domain backbone of WT and mutant structures, ligand-binding domain (A), and transmembrane domain (B) over the entire simulation. The ordinate is SASA (\AA^2), and the abscissa is time (ns).

To further investigate the structural changes resulting from substitutions in the *GABRA6* gene in each mutant type, the number of different secondary structures in the mutated types was calculated and compared with the wild types, as presented in Supplementary Figures 4 to 10. Additionally, the contribution of different secondary structures in the protein structure during the simulation is summarized in Table 6. Moreover, to understand the secondary structural changes of the mutant types, the DSSP parameter was calculated during the simulation, as shown in Figure 13 and Figure 14.

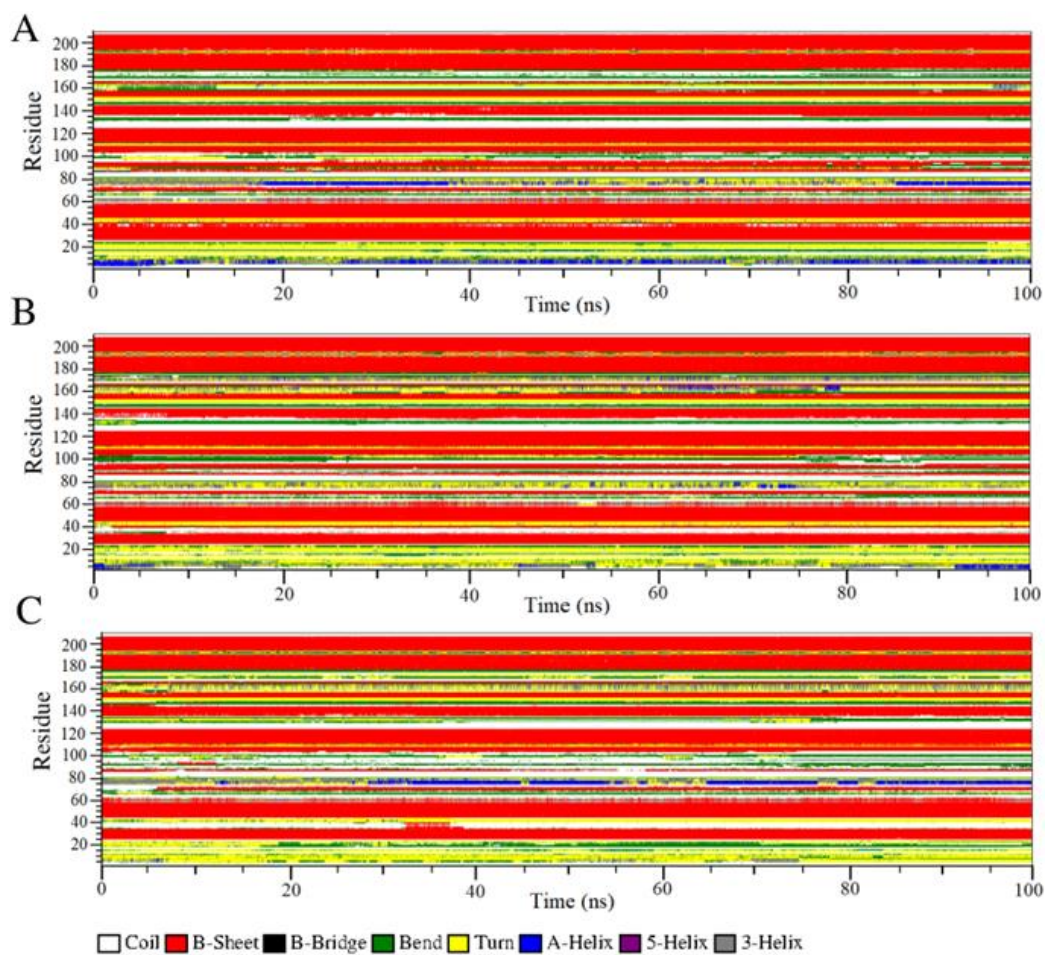


Figure 13 Variation in secondary structure elements for WT (A), *W87S* (B), and *W112R* (C) structures with respect to time, in the ligand-binding domain.

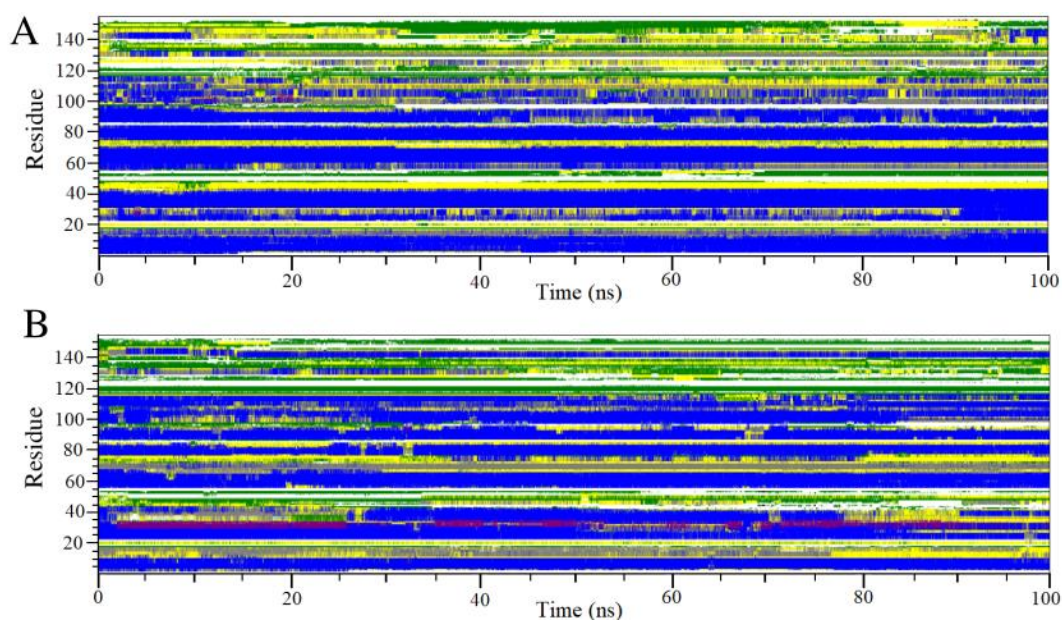


Figure 14 Variation in secondary structure elements for WT (A) and *C310R* (B) structures with respect to time in the transmembrane domain.

Table 6 Contribution of different secondary structures in protein structure during the simulation.

Protein	Structure	Coil	β -sheet	β -bridge	Bend	Turn	α -helix	5-helix	3_{10} -helix
WT ligand-binding domain	0.63	0.22	0.48	0.00	0.13	0.12	0.02	-	0.02
W87S mutant	0.62	0.23	0.46	0.01	0.12	0.14	0.01	-	0.03
W112R mutant	0.60	0.26	0.45	0.00	0.13	0.13	0.02	-	0.02
WT transmembrane domain	0.63	0.15	0.00	0.00	0.13	0.21	0.42	0.00	0.08
C310R mutant	0.58	0.17	0.00	0.00	0.15	0.17	0.41	0.02	0.08

(Structure = α -helix + β -sheet + β -bridge + turn).

The average coil percentage in the ligand-binding domain was 22.09%, 22.86%, and 25.85% for the WT, W87S, and W112R mutant structures, respectively (Figure S4a). In the transmembrane domain, the average coil percentage for the WT and C310R mutant structures was found to be 15.49% and 17.28%, respectively (Figure S4b). Secondary structure analysis in the ligand-binding domain reveals a significant difference in fluctuation between the WT and W112R mutant structures in the coil region residues. Similarly, in the transmembrane domain, a notable difference in fluctuation in the coil region can be observed between the WT and C310R mutant structures after approximately 50 ns of molecular dynamics simulation.

In the ligand-binding domain, the average percentage of β -sheet was found to be 48.32%, 46%, and 44.54% for the WT, W87S, and W112R mutant structures, respectively (Figure S5a). However, the average β -sheet percentage in the transmembrane domain was practically 0 (Figure S5b). The secondary structure analysis in the ligand-binding domain demonstrates a considerable difference in fluctuations between the WT and mutant structures, particularly in the residues of the β -sheet region.

The average β -bridge percentages in the ligand-binding domain were determined to be 0.35% for the WT structure, 0.83% for the W87S mutant structure, and 0.28% for the W112R mutant structure (Figure S6a). In the transmembrane domain, the β -bridge percentages were found to be 0.03% for the WT structure and 0.04% for the C310R mutant structure (Figure S6b).

The average percentages of bend were similar among the WT, W87S mutant, and W112R mutant structures in the ligand-binding domain, with values of 12.69%, 12.11%, and 12.65%, respectively (Figure S7a). In the transmembrane domain, the average bend percentages were 12.85% for the WT structure and 14.75% for the C310R mutant structure (Figure S7b). Notably, the C310R mutant structure exhibited a significant increase in fluctuation in the bend region residues during the first 30 ns compared to the WT structure in the ligand-binding domain.

Regarding the average turn percentages, the ligand-binding domain showed values of 12.10% for the WT structure, 14.06% for the W87S mutant structure, and 13.43% for the W112R mutant structure (Figure S8a). In the transmembrane domain, the turn percentages were determined to be 20.67% for the WT structure and 16.60% for the C310R mutant structure (Figure S8b). Notably, there were significant differences in fluctuation between the WT and three mutant structures in both domains within the turn region residues.

Regarding average α -helix percentages, the ligand-binding domain exhibited values of 2% for the WT structure, 1.18% for the W87S mutant structure, and 1.62% for the W112R mutant structure (Figure S9a). In the transmembrane domain, the average α -helix percentages were 42.47% for the

WT structure and 41.37% for the *C310R* mutant structure (Figure S9b). Notably, there was a considerable difference in fluctuation in the α -helix region residues between the WT and *C310R* mutant structures throughout the molecular dynamics simulation times in the transmembrane domain.

No 5-helix secondary structure was observed in the ligand-binding domain. However, in the transmembrane domain, the average percentages of 5-helix were 0.21% for the WT structure and 1.95% for the *C310R* mutant structure.

Lastly, the average 3_{10} -helix percentages in the ligand-binding domain were 2.41% for the WT structure, 2.93% for the *W87S* mutant structure, and 1.60% for the *W112R* mutant structure (Figure S10a). In the transmembrane domain, the average 3_{10} -helix percentages were 8.24% for the WT structure and 7.98% for the *C310R* mutant structure (Figure S10b). Notably, a decrease in fluctuation was observed in the *W112R* mutant structure compared to the WT structure in the ligand-binding domain. Conversely, an increase in fluctuation was visible in the transmembrane domain of the *C310R* mutant structure compared to the WT structure.

4. Discussion

This study aimed to identify the most detrimental missense nsSNPs associated with the human *GABRA6* gene, which encodes the alpha-6 subunit of the GABA_A receptor, and to investigate their molecular, functional, and structural effects on the protein. It is widely observed that evolutionarily conserved regions tend to contain disease-related SNPs. According to the neutral theory of evolution proposed by Kimura in 1983, amino acids that differ between species or SNPs that do not occur in coding regions of a protein are usually not influenced by natural selection or are under low selection pressure [67]. This suggests that such amino acid changes are generally tolerable and have minimal impact on protein function. Consequently, these amino acids are considered less mutable and more stable. However, nsSNPs occurring in coding regions of human genes can have phenotypic effects and may undergo natural selection. Amino acid substitution patterns can provide insights into selection patterns in protein-coding genes. An exciting example illustrating this is the distribution pattern of disease-related and non-pathogenic mutations in human genes. For instance, studies utilizing disease-related mutation data and phylogenetic analysis of multiple species have demonstrated that disease-associated substitutions (DAS) occur more frequently in evolutionarily conserved regions compared to regions subject to variation. Conversely, silent and polymorphic mutations exhibit the opposite trends, being randomly distributed. These patterns support the notion that protein-conserved regions are subjected to evolutionary pressure as these amino acids are crucial for the normal functioning of a particular gene. Conversely, due to their random distribution, silent mutations have minimal impact on organisms and may not be influenced by natural selection [68]. Therefore, this study has selected non-synonymous SNPs in coding regions for analysis.

Experimental investigation through wet laboratory techniques can be time-consuming, expensive, and resource-intensive. However, prior utilization of bioinformatics tools can determine SNPs' molecular, functional, and structural effects on proteins, thereby reducing the preclinical stage's time frame and economic costs before further *in vivo* and *in vitro* analysis. To predict missense nsSNPs, a combination of tools include SIFT for predicting functional impacts based on sequence homology, PolyPhen-2 for assessing structural and functional consequences through

machine-learning classification, PROVEAN for aligning protein sequences to predict functional effects, Condel for combining multiple predictions into a consensus score, SNPs & GO for utilizing gene ontology terms in prediction, PMut for evaluating amino acid substitution effects, SNAP2 for considering sequence, structural, and evolutionary information, MutPred2 for associating non-synonymous coding SNPs with diseases, and SNPeffect4.0 for machine-learning-based prediction of functional impacts were employed. I-Mutant3.0 and MUpro tools were also used for protein stability analysis, ConSurf for evolutionary conservation analysis, and SOPMA and PSIPRED for protein secondary structure analysis. Project HOPE was used to predict structural changes between mutant and wild residues, such as size, charge, and hydrophobicity value. For 3D structure modeling, the Pfam server was employed for domain region identification, the I-TASSER server for 3D structure model generation, and the PROCHECK and ERRAT servers for structural validation. Ligand binding site prediction was conducted using the FTSite and COACH servers. Finally, molecular dynamic simulations were performed for each WT and mutant structure using GROMACS to investigate conformational changes in physiological environments, with RMSD, RMSF, Rg, SASA, and DSSP values calculated.

4.1 Missense nsSNPs Prediction

In this study, 3 most deleterious missense nsSNPs in the human *GABRA6* gene coding region (Table 1), after analyzing by SIFT, PolyPhen-2, PROVEAN, and Condel tools were identified with high confidence scores, which are high-risk missense nsSNPs.

These selected missense nsSNPs were further analyzed using five tools to study the effects of mutations on the protein. Each computational method has different results due to predictions based on various databases and algorithms. SNPs & GO and PMut were used to predict disease-related missense nsSNPs; SNPs & GO's result consists of three different algorithms; SNPs & GO, PhD-SNP, and PANTHER. All 3 missense nsSNPs in the *GABRA6* gene (*W87S*, *W112R*, and *C310R*, Table 2) were considered deleterious in all 4 algorithms. SNAP2, MutPred2, and SNPeffect4.0 tools were used to predict the molecular and phenotypic consequences of selected missense nsSNPs; SNPeffect tool predicts SNPs consequence by TANGO (predict aggregation-prone regions), WALTZ (predict amyloid forming regions), and LIMBO (predict hsp70 chaperone binding sites). Results of SNAP2 and MutPred2 of 3 selected missense nsSNPs in the *GABRA6* gene showed that all 3 missense nsSNPs have damaging effects on protein structure and function. TANGO analysis showed that the *C310R* variant increases the aggregation tendency of the protein, and WALTZ analysis showed that this variant decreases the amyloid propensity of the protein (Table 3).

4.2 Protein Stability Analysis

Protein stability is critical for biomolecules' function, activity, and regulation. Missense mutations often impair the affected polypeptide propensity for the functional conformation folding and decrease the functional conformation stability; that leads to an increase in the proportion of mutant polypeptides that exist in a non-functional conformation, and the non-functional conformation is more prone to aggregation or degradation than the functional conformation. Any decreased stability and incorrect folding are the significant consequences of deleterious missense variants [69]. The effects of selected missense nsSNPs on protein stability were estimated by I-Mutant 3.0 and MUpro tools by calculating Gibbs free energy change ($\Delta\Delta G$) of the mutant and wild-type proteins.

The $\Delta\Delta G$ values represent the change in free energy of folding between mutant and wild-type proteins, indicating protein stability changes upon mutation. ΔG measures the energy difference between folded and unfolded states. A positive $\Delta\Delta G$ (>0) signifies increased stability in the mutant protein compared to the wild-type, suggesting favorable folding. Conversely, a negative $\Delta\Delta G$ (<0) indicates decreased stability, implying less favorable folding. This parameter offers insights into how mutations affect protein stability: positive values imply stabilizing mutations, while negative values indicate destabilizing ones, aiding in understanding the functional impact of genetic variations on protein structure and function [70]. Selected missense nsSNPs in the *GABRA6* gene showed decreasing protein stability in the mutant types by both tools, except the *C310R* variant, which showed increasing protein stability in the mutant type by I-Mutant3.0 analysis. *W87S* and *W112R* variants in the *GABRA6* gene have $\Delta\Delta G$ values < -1 kcal/mol in both tools, which are expected to alter the function and structure of the protein.

4.3 Evolutionary Conservation Analysis and Ligand Binding Sites Prediction

Protein conformational changes are essential for the function of proteins. In recent years, some important investigations have proposed the structure of complexes that exhibit conformational changes upon binding and further understanding these proteins' functions. As more experimental work is done to characterize the dynamics of protein interactions, it has become increasingly clear that proteins can exist in different conformational states. The flexibility within the protein regions allows it to adopt a new conformation and bind ligands with different structures. The proteins' ability to adopt several structures allows functional diversity without relying on the sequence diversity evolution, which can significantly facilitate the potential to rapidly evolve new functions and new structures [71]. Many properties of proteins derive from two opposing factors: flexibility and conformation rigidity. Protein molecules need flexibility and rigidity in order to function. Flexibility is required for the regulation and function, and rigidity is essential for maintaining globular structure. In some proteins, there are flexible local sites that are highly conserved. This indicates the relevance of flexibility to the overall stability of proteins. So, flexibility and rigidity in protein conformation must be exquisitely balanced [72].

ConSurf was used to identify selected missense nsSNPs position in evolutionarily conserved regions; FTSite and COACH algorithms were used to predict whether these selected deleterious missense nsSNPs are present in protein binding regions. ConSurf results for the human *GABRA6* gene showed that two residues (*W87* and *W112*) located in highly conserved regions (with scores 9) and predicted to have structural impacts on the protein, and another residue (*C310*) predicted as buried residue (with score 8). Further analysis using FTSite and COACH algorithms revealed no significant deleteriousness in selected missense nsSNPs at binding sites in its two protein domains. All the missense nsSNPs that were predicted as highly conserved by the ConSurf server could be used in pathogenic SNPs screening because neutral nsSNPs are more frequent in variable regions. In contrast, the deleterious nsSNPs are more common in the conserved areas [73].

4.4 Protein Secondary Structure Analysis

According to secondary structure predicting tools of proteins used in this study, 74.61% of amino acids sites of *GABRA6* protein are located in alpha helices and random coils (Figure 4), which corresponds to the evidence from the literature that sites of pathogenic and polymorphic mutations

are commonly found to be in alpha helices and random coils regions, and not frequently in beta-strand areas [74]. Out of 3 amino acid residues corresponding to *GABRA6* selected missense nsSNPs, SOPMA results indicated one residue (C310) located in alpha helices, and other two residues (W87 and W112) in random coils; and PSIPRED results indicated W87 residue located in strands, W112 residue in coils, and C310 residue in helices (Figure S1).

4.5 Project HOPE Analysis

Mutations in or near specific amino acids that contribute to the conformation of functional spatial are very likely to cause pathology. Missense mutations can cause amino acids to be substituted, resulting in size, charge, and hydrophobicity changes, affecting protein folding and interactions. According to the project HOPE analysis, mutation-related changes lead to loss of interaction or structural disorder, especially in the transmembrane domain. Furthermore, introducing or losing hydrophobicity or charge could cause misfolding, loss of interactions, or repulsion. As Ser and Arg residues in *W87S* and *W112R* mutants are more minor than Trp residue in wild-type forms, it may lead to loss of interactions in protein structure. Also, Arg residue in the *C310R* mutant is more significant than the Cys residue in the wild-type form, which may lead to bumps in protein structure. The charge of Trp and Cys residues in *W112R* and *C310R* mutants were neutral, then turned to Arg residues with a positive charge. Whenever the mutations introduce any charges, it might cause the repulsion of ligands or other residues with the same charge. In the *W87S* mutant, there is no charge change. The hydrophobicity in these three missense nsSNPs decreased compared to the wild-type forms, which might cause hydrophobic interactions to be lost in the core of the protein or on the surface (Table S4).

4.6 3D Structure Modeling

Pfam server was utilized for domain region identification in *GABRA6* protein; then, selected missense nsSNPs positions were located in different domains (Table 5). After generating 3D structure models by I-TASSER server and quality validation of predicted models by PROCHECK and ERRAT servers, PyMol software was used to visualize the 3D structural changes in the wild and mutant protein structures; the differences in the structures between the wild type and mutant type which visualized by PyMol software and 2D schematic structures from project HOPE proved the pathogenic impact of these deleterious missense nsSNPs on the protein (Figures 5 to 7).

4.7 Substitution Effects on Structure Via Molecular Dynamics Simulations

We conducted molecular dynamics simulations to investigate the potential deleterious effects of selected mutations on the protein. Specifically, we performed 100 ns simulations to examine these mutations' impact on the protein's structural dynamics. Molecular dynamics simulations probe the structural dynamics and stability of both wild-type and mutant proteins, shedding light on how mutations impact protein conformation, stability, and function. These simulations provide intricate atomic-level details, facilitating comprehension of disease mechanisms.

In the case of the *W87S* mutant, the RMSD value was consistently higher during most of the 100 ns simulation, particularly in the first 50 ns. This suggests that this mutant is likely to destabilize the protein structure compared to the WT. Although the RMSF value remained relatively constant in

most regions, it was higher in certain areas, indicating a potential loss of stability in the protein structure. Additionally, the Rg value was slightly lower throughout the simulation, which may contribute to an increase in the strength of the protein structure. Furthermore, the SASA value was lower than the WT and *W112R* mutant structures, indicating that the *W87S* mutant exhibits the highest level of stability among these variants.

For the *W112R* mutant, the RMSD value consistently remained higher throughout the entire simulation, especially in the first 20 ns and the last 30 ns. This suggests that this mutant is likely to destabilize the protein structure compared to the WT. Similarly, the RMSF value was higher in specific regions, indicating a potential loss of stability in the protein structure. Moreover, the Rg value was higher throughout the 100 ns simulation, suggesting a possible decrease in the strength of the protein structure. The SASA value of the *W112R* mutant was the highest among the WT and mutant structures, indicating that this mutant is likely to be the least stable among them.

In the case of the *C310R* mutant, the RMSD value consistently remained much lower throughout the simulation, suggesting that this mutant is more likely to stabilize the protein structure than the WT. The RMSF value showed higher and lower variations than the WT, indicating the possibility of both stability loss and gain in the protein structure. The Rg value was lower throughout the simulation, particularly after 30 ns, suggesting a potential increase in the stability of the protein structure. Additionally, the SASA value was lower than the WT structure, indicating that this mutant may cause a loss of stability in the protein structure.

We conducted secondary structure analyses to understand the disruption in secondary structures during the simulation. We observed significant differences in coil, β -sheet, bend, turn, α -helix, and 3_{10} -helix regions for the mutant types compared to the WT structures, as shown in Figures 13 and 14. The results revealed that the total number of secondary structures involved in forming secondary structures is higher in the WT compared to all mutant types.

Specifically, compared to the WT, the *W87S* and *W112R* mutants exhibited a decrease of 2.32% and 3.78% in β -sheets and turns, respectively, and an increase of 1.96% and 1.33% in β -sheets and turns, respectively. The *W112R* mutant also showed a reduction of 3.76% in coils. On the other hand, the *C310R* mutant displayed an increase of 1.79% in coils, 1.9% in bends, and 1.74% in 5-helices, along with a decrease of 4.07% in turns, compared to the WT. These alterations in secondary structures in the mutant types can influence the tertiary structure of the protein and potentially lead to changes in its three-dimensional structure or instability.

Furthermore, the analysis revealed that in the *W87S* and *W112R* mutants, the amino acids removed from the secondary structure (Try) reduced β -sheets and converted to turn and coil structures. Conversely, the *C310R* mutant had no significant preference for conversion to any specific structure, as turns were reduced and converted to 5-helices, coils, and bends. Overall, the turn structure exhibited the most significant changes in the mutated types among the secondary structures.

In conclusion, our molecular dynamics simulations suggest that the selected mutations can potentially affect the stability and secondary structure of the protein. The *W87S* and *W112R* mutants are likely to destabilize the protein structure, while the *C310R* mutant may contribute to its stabilization. These findings provide valuable insights into the impact of these mutations on the protein's structural dynamics and can aid in further investigations and understanding of their functional implications.

Although the research's simulation yielded insightful outcomes, certain constraints need acknowledgment. Firstly, the analysis was confined by a limited scope of reported missense nsSNPs, lacking verification through laboratory experiments. Additionally, the prediction of missense nsSNPs relied solely on publicly available data, hindering access to comprehensive clinical or genetic information. The study's focus on the *GABRA6* gene's coding region restricted the generation of 3D protein structures using I-TASSER, potentially overlooking crucial amino acid positions. Furthermore, the simulation's duration may have been insufficient for capturing comprehensive system behavior, impacting the depth of the insights obtained. Lastly, the reliance on the parameters of the Amber ff99SB protein force field introduces potential biases in the results, suggesting the need for caution in interpretation. Given these limitations, complementing the study with wet laboratory experiments is advisable to validate and strengthen the findings.

5. Conclusion

This study aimed to investigate the functional and structural effects of missense nsSNPs in the human *GABRA6* gene using advanced in-silico prediction tools. While experimental methods are more reliable for discerning deleterious nsSNPs, conducting repeated experiments on all pathogenic mutations is time-consuming. The approaches employed in this study provide evidence of the diverse effects of mutations, which can aid in characterizing their pathogenicity. Notably, three high-risk missense mutations in the *GABRA6* gene were identified, exhibiting the most detrimental impacts on protein structure and function. However, it's essential to recognize the limitations of our study, particularly the lack of experimental validation and the uncertain significance of these variants in clinical and genomic databases. These missense nsSNPs may have the potential to be considered as diagnostic markers or drug targets for epilepsy and other neuropsychiatric disorders. Consequently, future experimental designs can incorporate this in-silico data to analyze the biological context. To explore avenues for mitigating the effects of these three high-risk mutations in the mutated protein, it is suggested that research, such as targeted mutagenesis in vitro or animal models, be conducted to modify the functional or structural consequences of the mutations.

Author Contributions

Tahere Mohammadpour collected, analyzed, and interpreted the data and also drafted the manuscript. Reza Mohammadzadeh designed and revised the work.

Competing Interests

The authors have declared that no competing interests exist.

Data Availability Statement

All data generated or analyzed during this study are included in this article and in its additional files.

Additional Materials

The following additional materials are uploaded at the page of this paper.

1. Figure S1: PSIPRED analysis of *GABRA6* gene residues. The black boxes indicate the most deleterious missense nsSNPs.
2. Figure S2: The first predicted 3D structure model of wild-type (A), the result of PROCHECK server (B), and the result of ERRAT server (C) for the ligand-binding domain.
3. Figure S3: The first predicted 3D structure model of wild-type (A), the result of PROCHECK server (B), and the result of ERRAT server (C) for the transmembrane domain.
4. Figure S4: coil secondary structure comparison between WT and mutant structures.
5. Figure S5: β -sheet secondary structure comparison between WT and mutant structures.
6. Figure S6: β -bridge secondary structure comparison between WT and mutant structures.
7. Figure S7: bend secondary structure comparison between WT and mutant structures.
8. Figure S8: turn secondary structure comparison between WT and mutant structures.
9. Figure S9: α -helix secondary structure comparison between WT and mutant structures.
10. Figure S10: 310-helix secondary structure comparison between WT and mutant structures.
11. Table S1: All the missense nsSNPs in the *GABRA6* gene.
12. Table S2: Prediction of deleterious missense nsSNPs by four tools (SIFT, PolyPhen-2, PROVEAN, and CONDEL) in the *GABRA6* gene.
13. Table S3: Detailed output of MutPred2 tool for the three most deleterious nsSNPs in the *GABRA6* gene.
14. Table S4: Prediction of structural effects of the three most deleterious nsSNPs in the *GABRA6* gene by project HOPE.
15. Table S5: Predicted ligand-binding sites by FTSite.
16. Table S6: Predicted ligand-binding sites by COACH.

References

1. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al. ILAE official report: A practical clinical definition of epilepsy. *Epilepsia*. 2014; 55: 475-482.
2. Myers CT, Mefford HC. Advancing epilepsy genetics in the genomic era. *Genome Med*. 2015; 7: 91.
3. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: Position paper of the ILAE commission for classification and terminology. *Epilepsia*. 2017; 58: 512-521.
4. Wang J, Lin ZJ, Liu L, Xu HQ, Shi YW, Yi YH, et al. Epilepsy-associated genes. *Seizure*. 2017; 44: 11-20.
5. Brejc K, Van Dijk WJ, Klaassen RV, Schuurmans M, Van Der Oost J, Smit AB, et al. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*. 2001; 411: 269-276.
6. Macdonald RL, Olsen RW. GABA_A receptor channels. *Annu Rev Neurosci*. 1994; 17: 569-602.
7. Brooks-Kayal AR, Russek SJ. Regulation of GABA_A receptor gene expression and epilepsy. *Jasper's basic mechanisms of the epilepsies*. 4th ed. Bethesda, MD: National Center for Biotechnology Information; 2012.
8. Fritschy JM. Epilepsy, E/I balance and GABA_A receptor plasticity. *Front Mol Neurosci*. 2008; 1: 201. doi: 10.3389/neuro.02.005.2008.

9. Dibbens LM, Harkin LA, Richards M, Hodgson BL, Clarke AL, Petrou S, et al. The role of neuronal GABAA receptor subunit mutations in idiopathic generalized epilepsies. *Neurosci Lett.* 2009; 453: 162-165.
10. Hernandez CC, Gurba KN, Hu N, Macdonald RL. The GABRA6 mutation, R46W, associated with childhood absence epilepsy, alters $\alpha 6\beta 2\gamma 2$ and $\alpha 6\beta 2\delta$ GABA_A receptor channel gating and expression. *J Physiol.* 2011; 589: 5857-5878.
11. Syvänen AC. Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nat Rev Genet.* 2001; 2: 930-942.
12. Mah JT, Chia KS. A gentle introduction to SNP analysis: Resources and tools. *J Bioinform Comput Biol.* 2007; 5: 1123-1138.
13. George Priya Doss C, Rajasekaran R, Sudandiradoss C, Ramanathan K, Purohit R, Sethumadhavan R. A novel computational and structural analysis of nsSNPs in CFTR gene. *Genomic Med.* 2008; 2: 23-32.
14. Fielden MR, Matthews JB, Fertuck KC, Halgren RG, Zacharewski TR. In silico approaches to mechanistic and predictive toxicology: An introduction to bioinformatics for toxicologists. *Crit Rev Toxicol.* 2002; 32: 67-112.
15. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: Methods for virtual ligand screening and profiling. *Br J Pharmacol.* 2007; 152: 9-20.
16. 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.
17. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* 2001; 29: 308-311.
18. The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021; 49: D480-D489.
19. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 2012; 40: W452-W457.
20. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc.* 2016; 11: 1-9.
21. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: Comprehensive and non-redundant UniProt reference clusters. *Bioinformatics.* 2007; 23: 1282-1288.
22. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using polyphen-2. *Curr Protoc Hum Genet.* 2013; 76: 7.20.1-7.20.41.
23. Altschul SF, Madden TL, Schäffer AA, Zhang JH, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25: 3389-3402.
24. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012; 7: e46688.
25. 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.
26. Clifford RJ, Edmonson MN, Nguyen C, Buetow KH. Large-scale analysis of non-synonymous coding region single nucleotide polymorphisms. *Bioinformatics.* 2004; 20: 1006-1014.
27. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* 2005; 15: 978-986.

28. Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* 2007; 8: R232.
29. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet.* 2011; 88: 440-449.
30. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs & GO: A web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics.* 2013; 14: S6.
31. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics.* 2006; 22: 2729-2734.
32. Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs. mendelian disease: Evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci.* 2004; 101: 15398-15403.
33. 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.
34. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. *BMC Genomics.* 2015; 16: S1.
35. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, et al. The human gene mutation database (HGMD®): Optimizing its use in a clinical diagnostic or research setting. *Hum Genet.* 2020; 139: 1197-1207.
36. Mottaz A, David FP, Veuthey AL, Yip YL. Easy retrieval of single amino-acid polymorphisms and phenotype information using Swiss Var. *Bioinformatics.* 2010; 26: 851-852.
37. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun.* 2020; 11: 5918.
38. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol.* 2004; 22: 1302-1306.
39. Maurer-Stroh S, Debulpaep M, Kuemmerer N, De La Paz ML, Martins IC, Reumers J, et al. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods.* 2010; 7: 237-242.
40. Van Durme J, Maurer-Stroh S, Gallardo R, Wilkinson H, Rousseau F, Schymkowitz J. Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput Biol.* 2009; 5: e1000475.
41. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: An online force field. *Nucleic Acids Res.* 2005; 33: W382-W388.
42. De Baets G, Van Durme J, Reumers J, Maurer-Stroh S, Vanhee P, Dopazo J, et al. SNPeff 4.0: On-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res.* 2012; 40: D935-D939.
43. Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. *Bioinformatics.* 2005; 21: ii54-ii58.
44. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins.* 2006; 62: 1125-1132.

45. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 2016; 44: W344-W350.
46. Geourjon C, Deleage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics.* 1995; 11: 681-684.
47. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000; 16: 404-405.
48. 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.
49. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021; 49: D412-D419.
50. Wu S, Zhang Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 2007; 35: 3375-3382.
51. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000; 28: 235-242.
52. Zhang Y, Kolinski A, Skolnick J. TOUCHSTONE II: A new approach to ab initio protein structure prediction. *Biophys J.* 2003; 85: 1145-1164.
53. Zhang Y, Skolnick J. SPICKER: A clustering approach to identify near-native protein folds. *J Comput Chem.* 2004; 25: 865-871.
54. Yang J, Roy A, Zhang Y. BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* 2012; 41: D1096-D1103.
55. Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010; 5: 725-738.
56. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. *Nat Methods.* 2015; 12: 7-8.
57. Yang J, Zhang Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* 2015; 43: W174-W181.
58. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993; 26: 283-291.
59. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* 1993; 2: 1511-1519.
60. DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsl Protein Crystallogr.* 2002; 40: 82-92.
61. Ngan CH, Hall DR, Zerbe B, Grove LE, Kozakov D, Vajda S. FTSite: High accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics.* 2012; 28: 286-287.
62. Roy A, Yang J, Zhang Y. COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. *Protein Sci.* 2012; 40: W471-W477.
63. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comput Biol.* 2009; 5: e1000585.
64. Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc Natl Acad Sci.* 2008; 105: 129-134.

65. Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*. 2013; 29: 2588-2595.
66. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1: 19-25.
67. Kimura M. The neutral theory of molecular evolution. Cambridge, UK: Cambridge University Press; 1983.
68. Shastri BS. SNPs in disease gene mapping, medicinal drug development and evolution. *J Hum Genet*. 2007; 52: 871-880.
69. Bross P, Corydon TJ, Andresen BS, Jørgensen MM, Bolund L, Gregersen N. Protein misfolding and degradation in genetic diseases. *Hum Mutat*. 1999; 14: 186-198.
70. Ferrer-Costa C, Orozco M, de la Cruz X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *J Mol Biol*. 2002; 315: 771-786.
71. Goh CS, Milburn D, Gerstein M. Conformational changes associated with protein-protein interactions. *Curr Opin Struct Biol*. 2004; 14: 104-109.
72. Vihinen M. Relationship of protein flexibility to thermostability. *Protein Eng Des Sel*. 1987; 1: 477-480.
73. Bond LM, Peters JP, Becker NA, Kahn JD, Maher LJ. Gene repression by minimal lac loops in vivo. *Nucleic Acids Res*. 2010; 38: 8072-8082.
74. Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr Opin Struct Biol*. 2015; 32: 18-24.