

In Silico Analysis of nsSNPs of Human KRAS Gene and Protein Modeling Using Bioinformatic Tools

Duoduo Xu, Qiqi Shao, Chen Zhou, Arif Mahmood, and Jizhou Zhang*

Cite This: *ACS Omega* 2023, 8, 13362–13370

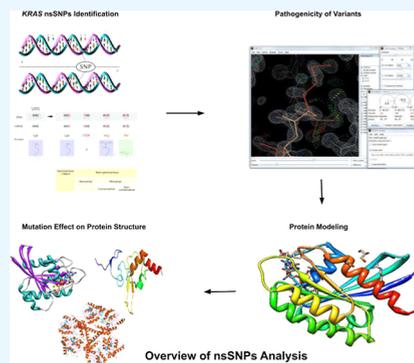
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The *KRAS* gene belongs to the RAS family and codes for 188 amino acid residues of KRAS protein, with a molecular mass of 21.6 kD. Non-synonymous single-nucleotide polymorphisms (nsSNPs) have been identified within the coding region in which some are associated with different diseases. However, structural changes are not well defined yet. In this study, we first categorized SNPs in the *KRAS* coding area and then used computational methods to determine their impact on the protein structure and stability. In addition, the three-dimensional model of KRAS was taken from the Protein Data Bank for structural modeling. Furthermore, genomic data were extracted from a variety of sources, including the 1000 Genome Project, dbSNPs, and ENSEMBLE, and assessed through *in silico* methods. Based on various tools used in this study, 10 out of 48 missense SNPs with rSIDs were found deleterious. The substitution of alanine for proline at position 146 pushed several residues toward the center of the protein. Arginine instead of leucine has a minor effect on protein structure and stability. In addition, the substitution of proline for leucine at the 34th position disrupted the structure and led to a bigger size than the wild-type protein, hence interrupting the protein interaction. Using the well-intended computational approach and applying several bioinformatic tools, we characterized and identified most damaging nsSNPs and further explored the structural dynamics and stability of KRAS protein.



INTRODUCTION

Ras is a gene family that contains *NRAS* (neuroblastoma-RAS), *HRAS* (Harvey-RAS), and *KRAS* (Kirsten-RAS). *NRAS*, *HRAS*, and *KRAS* genes have been identified in tumor cell lines in the 1980s, and researchers have been studying the structure and biology of Ras.¹ Ras proteins are tiny GTPases that act as main regulators of many signaling pathways that are involved in a wide range of cellular functions. *HRAS* (Harvey), *KRAS* (Kristen), and *NRAS* (neuroblastoma) are the three members of the Ras gene family.^{2–4} The Ras gene human family includes *Kras* (*v-Ki-ras2* Kirsten rat sarcoma oncogene) and has a unique position on chromosome 12p12.1. *KRAS* is recognized as a very common gene in the RAS superfamily and acts as a group of tiny GTP-binding proteins, known as RAS-like GTPases. In mammalian genomes, more than 150 RAS-like genes have been discovered.⁵

KRAS1 and *KRAS2* are the two copies of the *KRAS* gene found in the human genome. The *KRAS* gene had a molecular mass of 21.6 kD, and the KRAS protein contained 188 amino acid residues and plays an important role in intracellular signal transduction.⁶ Until it binds to GTP, the KRAS protein is inactive. When GTP binds to the KRAS protein, it experiences conformational changes that affect two domains of the protein, causing it to become activated. Switch 1 (amino acids 30–38) and switch 2 (amino acids 59–67) are two main domains that create an effector loop, determining the specificity of this GTPase's binding to its effector molecules.⁷ *KRAS* is the most

often mutated isoform, accounting for a total of 86% of RAS mutations. *KRAS-4B* is the most common isoform in human malignancies, associated with pancreatic cancer (90%), colon cancer (30 to 40%), and lung cancer [15 to 20%, mainly non-small-cell lung cancer (NSCLC)]. *KRAS* is also associated with different types of cancers such as cancers of the biliary tract, cervical cancer, endometrial cancer, bladder cancer, myeloid leukemia, liver cancer, and breast cancer.⁸ Moreover, haplotype analysis of two *KRAS* SNPs rs712 and rs7973450 revealed that the TG haplotype was associated with the positive lymph node status in LSCC patients.⁹

Single-nucleotide polymorphisms (SNPs) are the most common genetic variation, accounting for about 90% of human genetic variation, and some loci have been shown to be related to gene phenotypes and tumor susceptibility. SNPs are genetic markers existing in every 200–300 base pairs of the human genome.¹⁰ In the exon region of the human genome, there are approximately 0.5 million SNPs.¹¹ The protein structure, stability, and function may be affected by

Received: February 7, 2023

Accepted: March 23, 2023

Published: April 3, 2023



substitution of amino acids in the coding region of genes. Nonsynonymous SNPs (nsSNPs) with a high risk of causing mutations or changing protein function are known as high-risk nonsynonymous SNPs.^{12,13} Indeed, several studies have found that most nsSNPs are responsible for around half of the variations that are associated with hereditary genetic diseases. nsSNPs in cancer-causing genes have attracted a lot of attention in the current years.^{14,15} Multiple nsSNPs have been recognized in numerous studies that have the possibility to cause autoimmune diseases, infections, and inflammatory illness progression.^{16,17} Genes which are related to immunity are extremely polymorphic, and many nsSNPs in these genes have yet to be identified. The bioactivity of SNPs influenced the sensitivity of drug reactions among the signaling pathways of regularly used immunosuppressants, such as glucocorticoids, mycophenolate mofetil, azathioprine, tacrolimus, cyclophosphamide, as well as methotrexate.

In this study, we used different *in silico* tools to screen the nsSNPs for identification of the most deleterious nsSNPs in human KRAS protein and to check their impact on the structure and function of KRAS protein, which provide a better understanding of dynamic properties and potential therapeutic targets/options.

MATERIALS AND METHODS

Data Collection from SNP Database. To identify functional SNPs of KRAS, we searched different databases including dbSNP, ENSEMBLE, SNP500 cancer, GeneCards, and UniPort. Other sources of information in GenBank, PubMed, LocusLink, and the Human Genome Project data was also retrieved. Such tools are an excellent source of genomic variation data, and it is updated on a regular basis after new entries are submitted.¹⁸ ENSEMBLE (www.ensembl.org/) was used to obtain the nucleotide and protein sequence of KRAS.

Data Analysis of Deleterious nsSNPs. The SIFT (short intolerance from tolerance) algorithm was used to determine whether nsSNPs have a tolerant or deleterious effect on protein function. SIFTs arranged SNPs into intolerant or tolerant groups based on homologous alignment. Normalization probability of definite amino acids less than a chosen threshold value (0.05) was likely to be intolerant; however, tolerance indexes more than >0.05 were considered acceptable.¹⁹

nsSNP Consequence by the Structural Homology-Based Method. The structure of a protein determines/regulates its function, and any changes to the structure can cause the protein's function to be disrupted. The effect of nsSNPs on the protein structure is required to determine the impact of mutations on its activity. We used Polymorphism Phenotyping version 29 (PolyPhen2) to predict the harmful effects of nsSNPs on the protein structure and function. For categorization, this tool employed a naive Bayesian method with a score range of 0–1. The mutations are classified as beginning, possibly damaging, or probably destructive based on their score. The scores that are closest to 1 are regarded to be probably damaging, with a significant impact on the protein structure.²⁰

Functional nsSNP Characterization. SNP&GO, PhD-SNP, nsSNP analyzer, PROVEAN, and P-MUT were used to characterize functional nsSNPs. The support vector machine approach was used by the predictor of human deleterious single-nucleotide polymorphism (PhD SNP) to classify and

explain the non-synonymous SNP effect on protein. PhD SNP analyses classified genes into two categories: deleterious and neutral.²¹

PROVEAN was used to determine pathogenic SNPs. PROVEAN BLASTED the query sequence against NCBI and categorized the mutation as harmful or neutral based on the cutoff value. PROVEAN's threshold value is -2.5 , and values higher than this are classified as deleterious.²²

SNP&GO also used the SVM algorithm. PMUT uses neural networking to analyze different types of sequence information and classify mutants based on the obtained information. The mutated protein sequence in FASTA format was submitted to the PMUT, and the result might be disease or neutral based on the sequence's probability score.²³

nsSNP Prediction on the Molecular Phenotype of Protein. The SNP effect was used to predict the nsSNPs on the coding region of the KRAS protein. It is not only calculated the conservation score but also predicted the landscape of protein homeostasis.²⁴ The SNP effect uses a variety of methods, including TANGO to predict mutant aggregation propensity, WALTZ to predict amyloid propensity, and LIMBO to predict chaperon binding. The TANGO score of wild and mutant amino acids was determined using different TANGO scores (dTANGO), which were used to determine aggregation propensity. The dWALTZ score was used to predict the amyloid-forming region of a protein. The chaperone binding propensity mutant is predicted using the dLIMBO score. The crystal structure of KRAS protein was retrieved from Protein Data Bank (PDB). For modeling of the mutated structure, FoldX (<http://foldxsuite.crg.eu/>) was used.

Impact of Nonsynonymous Mutations on Protein Stability. Protein stability is important for maintaining its structure and function. The I-MUTANT 3.0 program was used to anticipate how a mutation will affect the stability of the KRAS protein. The authenticity of the I-MUTANT result was verified by comparing it to the MUPRO result. Both systems utilized the same algorithm to determine whether a mutation would increase or reduce stability.²⁵ The SRide server was used to predict the stabilizing residues of the normal and mutant proteins.

Prediction of 3D Structures of Mutant Models. Protein functions, such as binding affinity in the presence and absence of mutation, and other key features are better explained by the protein's three-dimensional (3D) structure. The experimental methods used to obtain the 3D structures of proteins are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. However, these procedures are both costly and time-consuming.²⁶ Because the computational methods are rapid and cost-effective, they are utilized as an alternative to experimental methods to predict the protein structure.²⁷ Mutant models were predicted using I-TASSER Swiss models.

In Silico Site-Directed Mutagenesis and RMSD Calculation. In the program TRITON interfaced with MODELLER,²⁸ the 3D structure of mutant protein was built based on homologous modeling. PROCHECK²⁹ predicted stereo chemical quality of each model, while ERRAT checked the environment profile. NOMAD-Ref and CHIMERA were used to reduce the potential energy of mutant structures and calculate the RMSD of native and mutant structures.

Statistical Analysis. Continuous normally distributed data are expressed as the means \pm SDs. All statistical calculations were carried out using SPSS statistical software. A *t*-test was used to compare data between two groups. Data were analyzed

Table 1. Screening and Identification of Pathogenic nsSNPs Using Different Bioinformatic Tools

variant ID	allele change	amino acid change	SIFT	PolyPhen2	SNP&GO	PANTHER	PHD SNP	PROVEAN	P-MUT
rs4362222	T/A/C/G	R161S	0	0.97	disease	neutral	disease	deleterious -5.303	disease 0.71
rs539423712	A/G	V160A	0.01	0.998	neutral	disease	neutral	deleterious -3.536	disease 0.56
rs387907206	T/C	K147E	0	0.967	disease	disease	disease	deleterious -3.625	disease 0.63
rs121913527	C/A/G/T	A146P	0	1	disease	disease	disease	deleterious -4.516	disease 0.53
rs1565884227	C/T	A134T	0	0.946	disease	neutral	neutral	deleterious -3.307	disease 0.61
rs1463850736	C/A/G/T	A130P	0.04	0.978	disease	neutral	disease	deleterious -3.253	disease 0.69
rs730880471	C/T	D119N	0	0.995	disease	neutral	disease	deleterious -4.566	neutral 0.47
rs770248150	T/A/G	K117N	0.01	0.994	neutral	neutral	disease	deleterious -4.558	neutral 0.47
rs780974222	C/G	G75A	0.03	0.988	neutral	neutral	neutral	deleterious -5.938	disease 0.54
rs387907205	A/C/G	Y71D	0	0.996	disease	disease	disease	deleterious -9.374	disease 0.70
rs1555194026	C/A	S65I	0.01	0.935	neutral	neutral	disease	deleterious -5.318	neutral 0.32
rs727503108	C/A	G60V	0	1	neutral	disease	neutral	deleterious -8.427	neutral
rs104894359	C/G/T	G60R	0	1	disease	disease	neutral	neutral	disease 0.71
rs104894359	C/G/T	G60S	0	1	disease	disease	disease	deleterious -5.617	disease 0.71
rs104886029	G/A	A59V	0.03	0.972	disease	neutral	disease	deleterious -3.706	disease 0.71
rs104894364	G/A	T58I	0	1	disease	disease	disease	deleterious -5.623	disease 0.71
rs727503109	T/C	I36M	0	0.976	neutral	neutral	disease	neutral -2.345	disease 0.70
rs104894366	G/A/C	P34L	0	1	disease	disease	disease	deleterious -8.449	disease 0.71
rs104894366	G/A/C	P34R	0	1	neutral	disease	neutral	deleterious -7.598	disease 0.53
rs794727277	T/A	N26Y	0	0.967	neutral	neutral	disease	deleterious -5.622	disease 0.65
rs730880472	A/C	L23R	0.05	0.994	disease	disease	disease	deleterious -4.925	disease 0.58
rs121913538	C/A/G	L19F	0.01	0.999	disease	neutral	disease	deleterious -3.373	disease 0.54
rs121913236	G/C/T	Q22R	0.01	0.982	neutral	neutral	neutral	deleterious -3.324	neutral 0.49
rs776785730	A/C/G	S17R	0	0.999	disease	disease	disease	deleterious -4.344	disease 0.70
rs1555195579	C/A	G15V	0	0.999	disease	disease	disease	deleterious -7.434	disease 0.71
rs104894365	C/T	V14I	0	0.959	neutral	neutral	disease	neutral -0.819	disease 0.71
rs121913535	C/A/G/T	G13C	0.02	0.997	neutral	neutral	disease	deleterious -7.619	disease 0.57
rs104894361	T/A/C/G	K5N	0.01	0.989	neutral	neutral	disease	deleterious -3.603	disease 0.64

via analysis of variance with the Tukey–Kramer multiple comparisons test. *P* values <0.05 were considered significant.

RESULTS

Nonsynonymous (nsSNP) Analysis Showed 28 out of 335 SNPs as Detrimental. To evaluate the tolerance index (TI) score of a mutant protein, 335 nsSNPs were subjected to the sorting intolerant from tolerant (SIFT) algorithm. By matching the homologous sequence, SIFT determines the TI score in the range of 0–1 to elevate the conserved amino acid. SIFT evaluated 28 SNPs as detrimental out of 335, with TI scores ranging from 0 to 0.04. Out of the 28 SNPs, around 17 occupy with the lowest TI score of 0, indicating that they have a significantly deleterious effect on the protein structure. Five SNPs have a TI of 0.01, one has a TI score of 0.02, three have a TI score of 0.03, and only one has a TI score of 0.04. These digits indicate that SNPs are of great importance. Table 1 shows the SIFT TI score. Distribution of SNPs in different regions are shown in Figure 1.

Functional Modification of Coding nsSNP Prediction Showed 21 of the 335 nsSNPs as Potentially Harmful. The effect of nsSNPs on the protein structure and function was predicted using the PolyPhen2 online server. The PolyPhen2 server looked up the sequence and 3D structure of the protein queries and compared them to what were already known. The likelihood score predicts the effect of mutation on the structure and function of protein by aligning both the structure and similarity observed in them. Only 21 of the 335 nsSNPs uploaded to the server had a probability score greater than 0.98, indicating that they are likely to be harmful. As shown in Table 1, A146P, G60R, G60V, G60S, E34L, T58I, and P34R

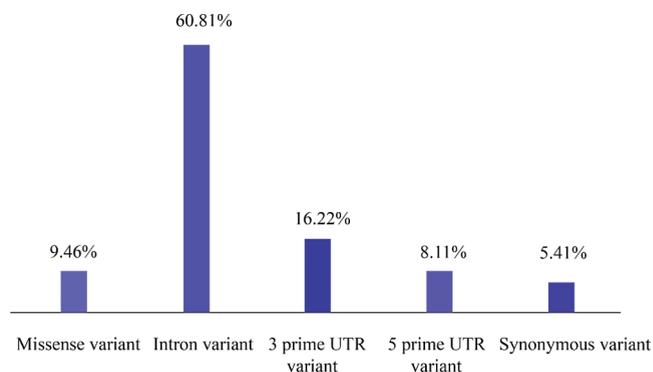


Figure 1. Different regions showing distributions of SNPs of the human KRAS gene.

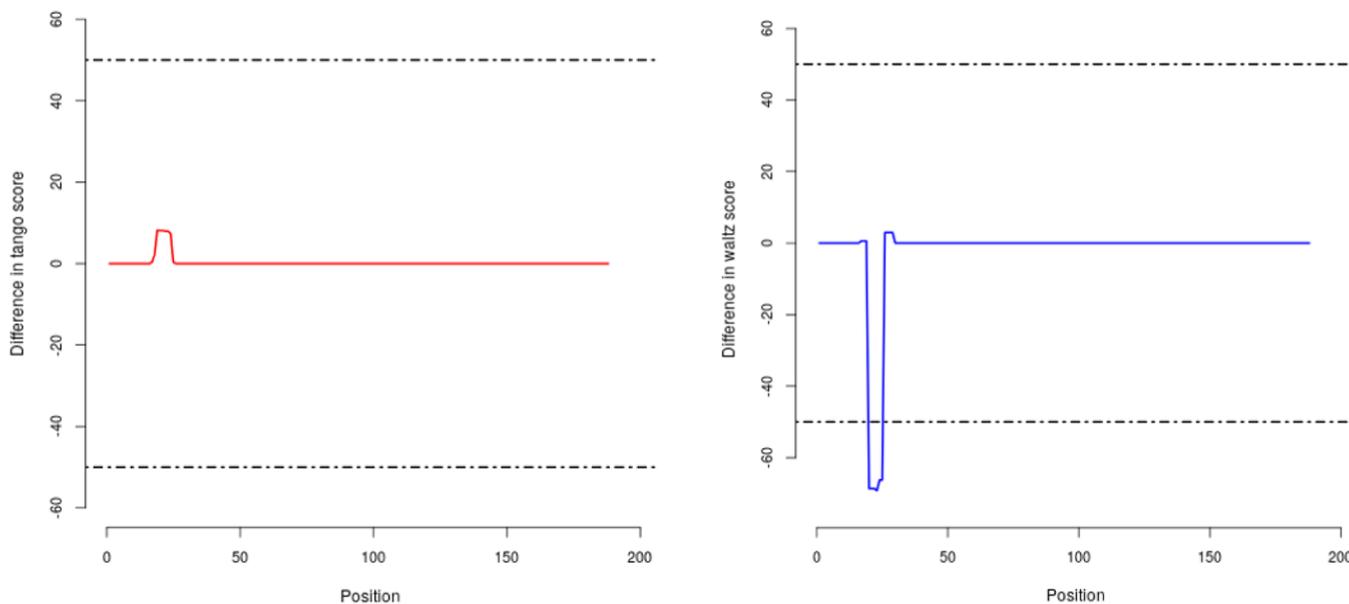
have the highest score of 1, while others have scores in the range of 0.98–0.999.

Twelve nsSNPs were discovered to be mutual in the results of both servers when the SIFT and PolyPhen2 results were combined. Even though the two servers utilized different methodologies, such as SIFT's conclusion being based on structural details and PolyPhen2 being a structure-based tool, the research revealed a good correlation by two ways (Table 1).

Phenotypic Influence of Mutation by the SNP Effect. The SNP effect was exploited to investigate the phenotypic impact of KRAS variants. The server assessed the tendency of chaperon binding, aggregation, and amyloid propensity, but the solution revealed that a majority of these characteristics were unaffected by these variants. Those variations that can

Table 2. Analysis of nsSNPs in Human KRAS Protein through the SNP Effect

variant ID	dTANGO score	aggregation tendency	dWALTZ	amyloid propensity	dLIMBO	chaperone binding tendency
rs387907206	1.11	not effected	-0.10	not effected	0.03	not effected
rs121913527	0.00	not effected	-0.00	not effected	-0.43	not effected
rs387907205	1.58	not effected	0.11	not effected	-27.93	not effected
rs104894359	0.00	not effected	-39.44	not effected	0.00	not effected
rs104894364	0.00	not effected	0.02	not effected	-3.74	not effected
rs104894366	0.00	not effected	10.68	not effected	-0.05	not effected
rs730880472	-49.71	not effected	-393.90	decrease	8.08	not effected
rs121913538	50.70	increase	-1.02	not effected	0.00	not effected
rs776785730	-28.26	not effected	10.36	not effected	0.03	not effected
rs1555195579	40.34	not effected	0.11	not effected	0.00	not effected

**Figure 2.** Difference between different variants. WALTZ amyloid propensity (right panel) and TANGO aggregation difference (left panel).**Table 3. DDG (Change in Free Energy Variation upon Mutation) Was Applied in Which DDG < 0 = Decrease in the Stability, While the DDG Value > 0 = Increase in the Stability of Protein**

variant id	position	wild type	mutant	DDG-value (kcal/mol)	I-MUTANT	MUPRO
rs387907206	147	K	E	-0.38	decrease	decrease
rs121913527	146	A	P	-0.24	increase	decrease
rs387907205	71	Y	D	-1.24	decrease	decrease
rs104894359	60	G	S	-1.13	decrease	decrease
rs104894364	58	T	I	0.19	increase	increase
rs104894366	34	P	L	-0.46	decrease	decrease
rs730880472	23	L	R	-1.78	decrease	decrease
rs121913538	19	L	F	-1.02	decrease	decrease
rs776785730	17	S	R	0.04	increase	decrease
rs1555195579	15	G	V	-0.35	decrease	increase

change the properties and structure of the KRAS protein are probably not these characters. The burden of permanent β -strand-driven aggregation of protein and amyloid generation on living organisms is enormous. Because of these characteristics, not only the protein loses its function but it also places abnormal pressure on the cell, as a considerable amount of energy is used in expressing the gene and degrading the defective protein. L23R-only (rs730880472) showed a decrease in amyloid propensity with a dWALTZ score of -393.90, while the L19F (rs121913538) showed an increase in

mass tendency with a dTANGO score of 50.70, as indicated in Table 2 and Figure 2.

Effect of Mutation on Protein Stability. Using the I-MUTANT server, the stability of KRAS was related to that of a mutant structure. The results of I-MUTANT are constructed on the ProtTherm database, which contains the most extensive collection of experimental thermodynamics data on free energy differences in protein stability due to mutations.

For DDG (Δ Gibbs free energy) stability prediction, models with mutations such as K147E, Y71D, A146P, G60S, E34L, T58I, L32R, S17R, L19F, and G15V were uploaded to the I-

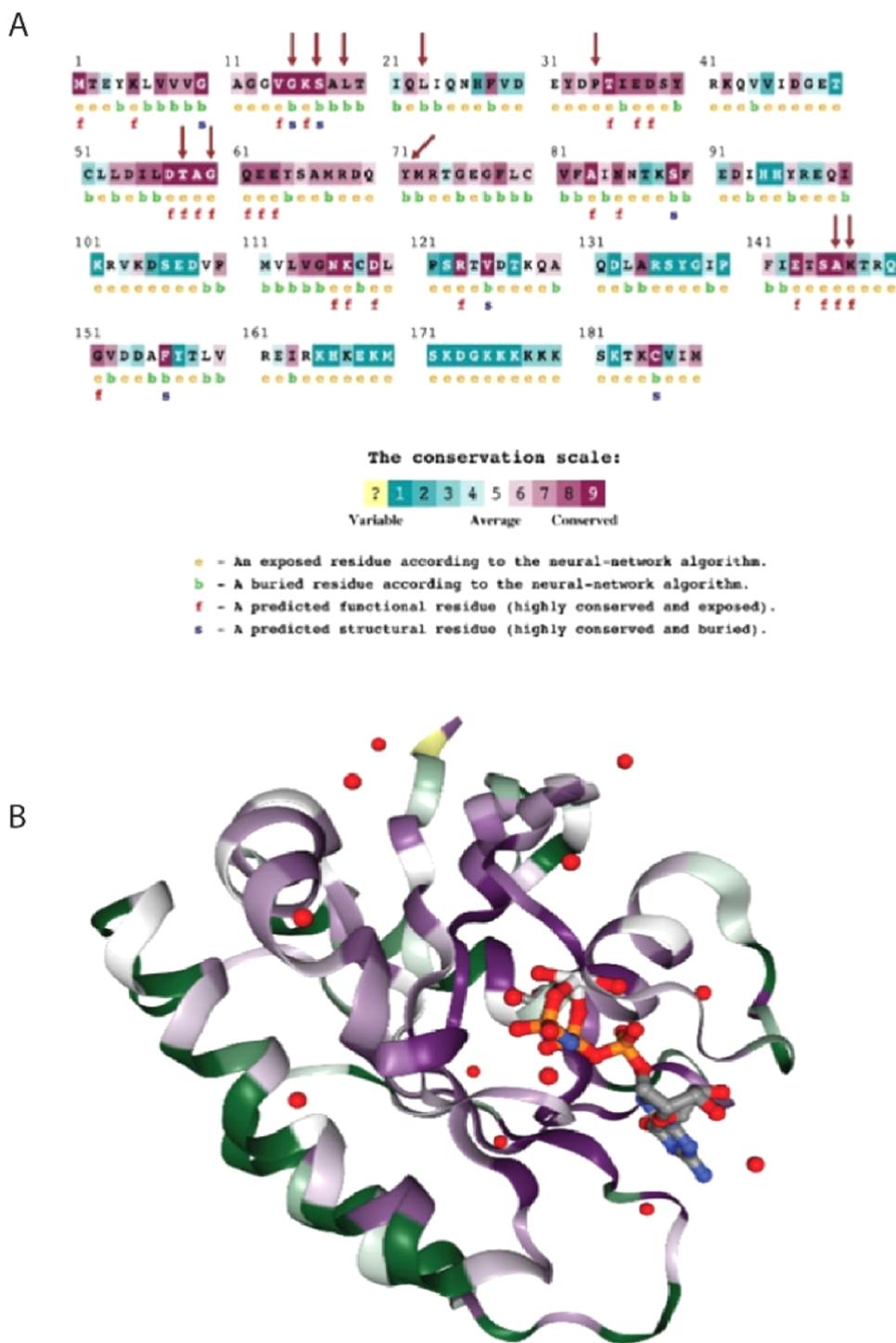


Figure 3. Structure and residues of KRAS protein. (A) Detection of evolutionary conservation of residues in human KRAS protein predicted by the ConSurf server. (B) Score from 1 to 9 defines the conserved region. With the increase in value, the residues were more conserved. Red dots indicate positively charged amino acids.

MUTANT server. Except for A146P, S17R, and T58I, which have been linked to increased structural stability, all the mutations demonstrate a decrease in protein stability. The lowest DDG value (-1.78 kcal/mol) was due to the mutation in L23R. DDG values for all other mutations varied from -0.24 to -1.78 kcal/mol, implying lower protein stability due to DDG values being less than 0. Table 3 displays the results. The mutant was also sent to MUPRO to ensure that the results were valid. Except for the A146P and S17R mutations, the

results were nearly identical. As indicated in Table 3, I-MUTANT showed an increase in stability, while MUPRO showed a decrease in stability, as shown in Table 3.

Analysis of Conservation by Using the ConSurf Server. The degree of protection of KRAS protein deposits was found and analyzed. All ten harmful mutations were found in highly conserved areas in our findings (7-8-9). P34L was anticipated to be an exposure mutation, while mutants like K147E, G60S, A146P, and T58I were projected to be

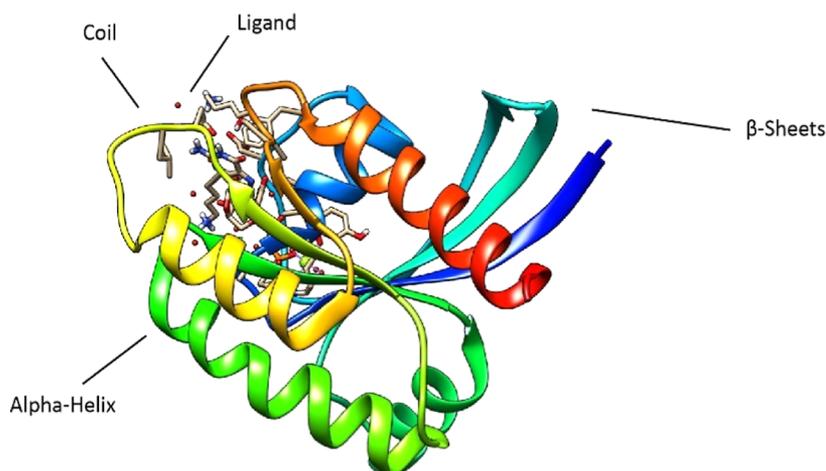


Figure 4. Three-dimensional crystal structure of the KRAS protein representing the helix, β sheets, and strand with its unique ligands.

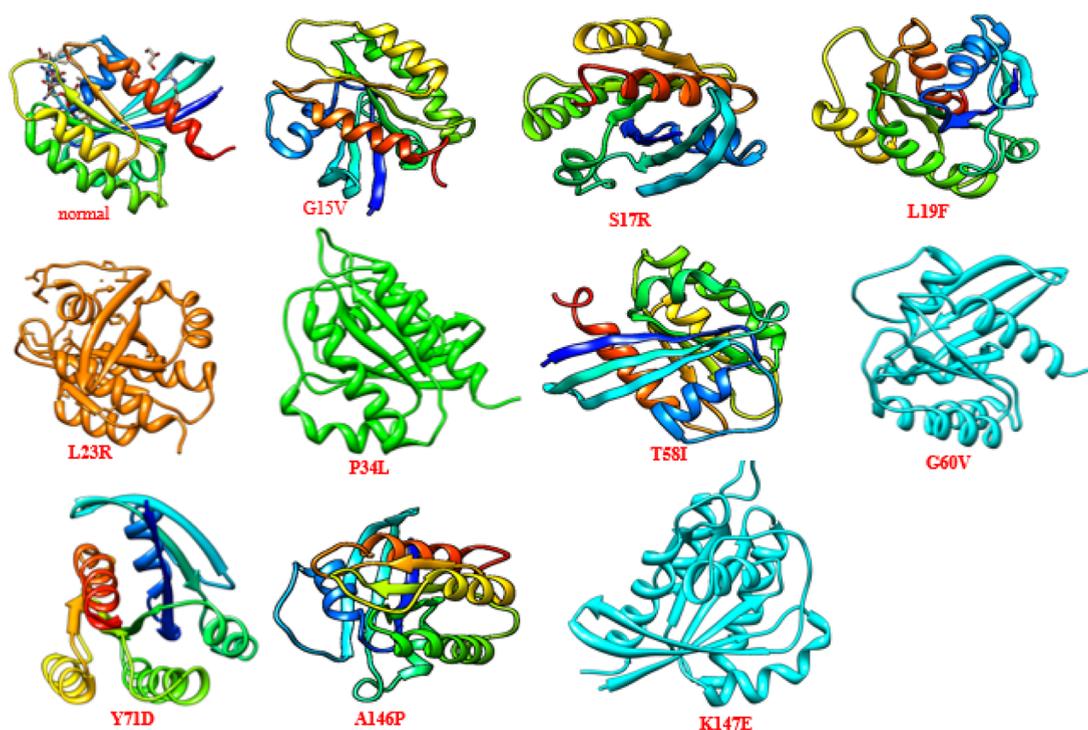


Figure 5. 3D structures of native and mutant models of human KRAS protein. Changes in the protein structure can be observed in each model.

functional and exposure mutations. As illustrated in Figure 3A, the mutations G15V and S17R were expected to be buried structural amino acids, while three mutations identified, L23R, L19F, and Y71D, were anticipated to be buried residues in the structure of protein. Figure 3B shows a colored 3D ConSurf model of the KRAS structure of protein.

Wild-type 3D Model of KRAS Protein. The sequence determines a protein's structure, and the structure determines a protein's unique function; hence, the 3D model of a protein is important to determine various attributes such as affinities of the protein ligand, domain, binding pocket, motif, and interface protein sites. The complete wild-type 3D model of KRAS protein with PDB ID (4OBE) with a resolution score of 2.85 was downloaded from the PDB, as illustrated in Figure 4. The wild-type model consists of an α helix, β sheets, and loops with ligands bound to distinct locations.

Models of Mutants by Swiss Models. The Swiss model server was used to estimate the structure of mutants through the homology modeling method after retrieving the native full structure of HRAS from the PDB. PyMOL and CHIMERA software were used to generate the mutated model, which included rs104894228 (G13R), rs750680771 (D38H), rs730880460 (G60V), rs730880460 (G60D), rs727504747 (A59L), rs1564789552 (Y64H), rs917210997 (G115R), rs1204223913 (P110L), rs369106578 (R123G), rs730880464 (R123P), and rs1564789700 (I46T) (Figure 5). Mutations cause structural changes in the HRAS protein. GROMOS 96 was used to minimize energy, which reduced the amount of energy and force that acts on each atom in a group of atoms to obtain the most thermodynamically stable structure of KRAS. The energy value of the final and stable KRAS model was $-18\,756$ kJ/mol, compared to $-108\,915$ kJ/mol before the energy minimization process. The mutant

residue is smaller in size than the wild-type, and the mutation at this site can add charge at this position, changing the amino acid at position 147 from lysine (K) to glutamic acid (E), which can interrupt interactions with other molecules. The alteration of amino acid alanine (A) mutates into proline (P) at position 146, causing the wild-type residue to sink to the center of the protein (Figure 6). The wild-type group is smaller in size as compared to the mutated one, which is possibly not able to adjust in the allowed space and hence affects the protein structure, and the change of tyrosine to aspartic acid at position 71 may result in a loss of external contact. These mutations cause a lack of hydrophobic interaction on the protein's surface. At position 60, a wild-type residue glycine is mutated into serine (Figure 6), due to which torsion angles are certainly out and disturb the local structure because the mutant residue is not fitted in the correct site. It will be in a difficult position to form correct hydrogen bonds since the mutant residue at position T58I is larger than the wild type.

With the substitution of leucine for proline at position 34, the wild-type size becomes smaller than the mutant type; therefore, the protein interaction is interrupted. At position 23, leucine is converted to an arginine (Figure 6). Hence, the wild-type residue is tiny, and the mutant residue could not fit in the center of the protein, causing hydrophobic contact to be disrupted. At position 19, leucine changes into phenylalanine, so the mutant residue will be buried in the protein's core, and due to its larger size, it will most likely not fit. The tiny wild-type serine residue is replaced by a larger mutant-type arginine in mutation of S17R, and such mutant residues altered charge when they come into contact with wild-type serine residues, causing protein folding issues. The protein will lose hydrophobic characteristics because of this mutation. Due to the fact that wild-type amino acid (glycine) is far more elastic to all residues, replacing glycine at position 15 with valine could disrupt the protein structure due to unique torsion angles. The mutant residue will not fit in the protein center due to its larger size.

DISCUSSION

The KRAS gene belongs to the RAS family that has 188 amino acid residues and a molecular mass of 21.6 kD, which plays an important role in intracellular signal transduction, while non-synonymous SNPs cause amino acid variation in a chain of proteins, which has an impact on the structure and function of the protein. nsSNPs are responsible for the majority of genetic diseases. It is difficult to distinguish between normal and harmful SNPs, as well as to identify the amino acid that plays a significant role in development of disease.³⁰ By combining several algorithms and database information, *in silico* analysis can help to discriminate between neutral and deleterious SNPs. On the basis of the structure and phylogenetic information, a mutated amino acid is evaluated, and quite precise results were gained.³¹ Non-synonymous SNPs in the coding area can modify the amino acid sequence of a protein, affecting its function and increasing disease susceptibility.¹⁹

Using computational approaches, the SNP databases were evaluated in this work to explore SNPs that could possibly be deleterious for KRAS. There were 335 hits while searching for nsSNPs in KRAS. rsIDs of mutants were uploaded to SIFT and PolyPhen2 servers to investigate the function impacts of nsSNPs. SIFT identified 28 nsSNPs as non-tolerable, while PolyPhen2 identified 21 nsSNPs as potentially or certainly

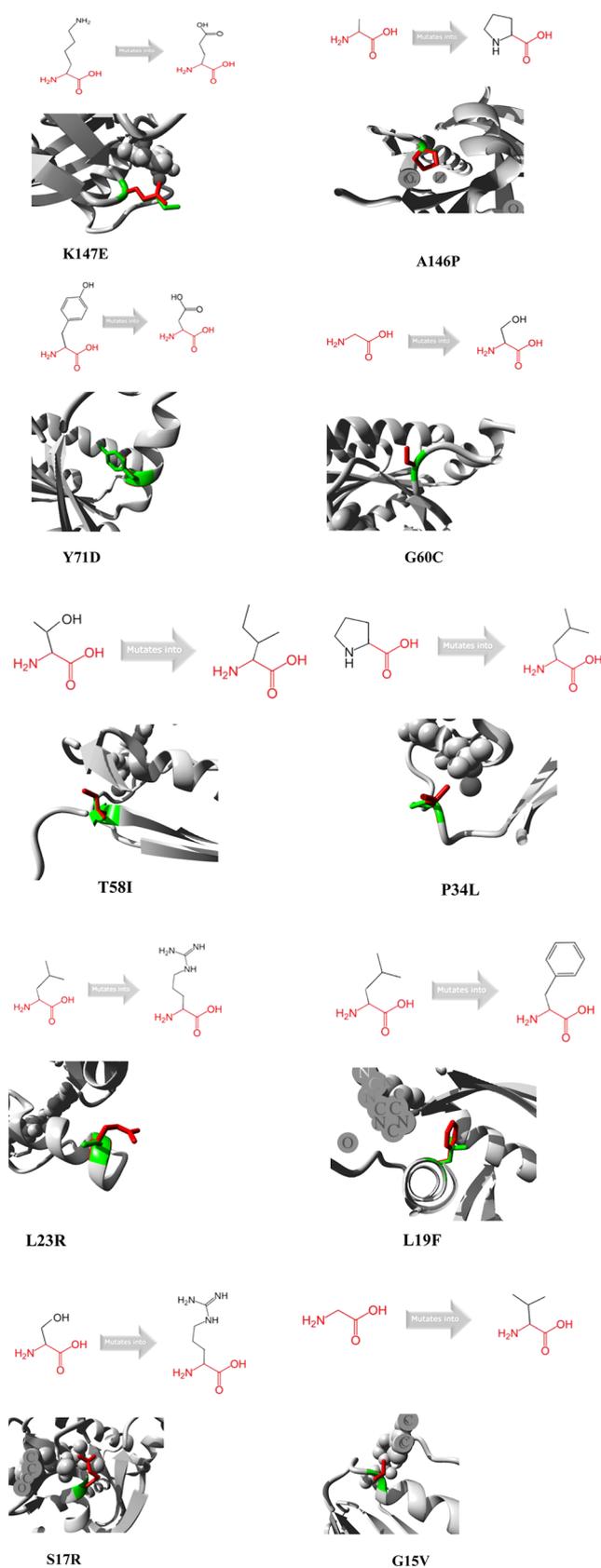


Figure 6. Mutagenesis of different residue substitutions at different positions in KRAS protein by applying the Project HOPE server. The substitution of alanine to proline at the 146th position led the residue to bend toward the central core. The substitution of tyrosine to aspartic acid at the 71st position may result in the loss of external

Figure 6. continued

contact and interaction. The alteration of glycine to serine at the 60th position possibly affects hydrophobic interaction of the protein surface. With the substitution of a leucine for a proline at position 34, the wild-type size becomes smaller than the mutant type; therefore, the protein interaction is interrupted. At position 23, the leucine is converted to an arginine, the wild-type residue is tiny, and the mutant residue could not fit in the center of the protein, causing hydrophobic contact.

harmful. The modifications in results could be due to the server's usage of a different algorithm.

Polyphen 2 and SNAP were claimed to have superior performance in finding functional nsSNPs than other Insilco Technologies.³² The above-mentioned server can be utilized to discriminate between casual and non-casual associations between nsSNPs and the phenotype of interest in this scenario. The amyloid and chaperone binding SNP effect was exploited to test the influence of SNPs on aggregation of protein propensity. The links between nsSNPs and their placement in the protein structure have been explored by several groups.³³ Therefore, the complete tertiary structure was downloaded from the Protein Data Bank with its PDB ID (4OBE), and its mutant models were modeled by using the Swiss model server. The mutant and wild-type models were subjected to CHIMERA for their energy minimization and RMSD calculations. The aberration between the two structures is measured by their RMSD values, which can alter stability and functional movement.³⁴

ConSurf identified all 10 residues P34, K147, G60, A146, P581, G15, S17, L23, L19, and Y71 as extremely conserved with a score of 9.³⁵ By Project HOPE server, K147 is required for binding of protein, but substitution into E147 alters the domain structure and reduces the protein's binding capacity. Each mutation disrupted the protein's architecture and function in the same way G60, A146, P581, G15, S17, L23, L19 and Y71 are protein residues with high ConSurf conservation scores in the range of 7–9. HOPE, on the other hand, revealed a functionally and structurally significant residue.

The presence of disease-associated nsSNPs is comforting in today's genomic assessment. As a result, we used an *in silico* method to detect nsSNPs associated with disease in the KRAS gene, as well as a literature review to find the nsSNP associations with other different cancer types that had not been previously documented. We identified that only 11 of the 335 non-synonymous changes in the coding area were categorized as harmful using sequence and through-sequence structure-based software. The results of structural analysis showed that these nsSNPs can alter the structure and function of KRAS and also change their location of ligand-binding and stability of protein and up-regulate KRAS functions. As a result, this research could be useful in determining the role of KRAS nsSNPs in the development of various cancers, as well as preventing the impacts of changes in KRAS gene activity. The novelty of this study is that we finished computational screening and analysis of deleterious nsSNPs in human KRAS protein through bioinformatics approaches. However, this study is limited to the computational approach, while experimental validation of this study is needed for better understanding of protein functions. The advantage of this study was to screen the nsSNPs for the identification of the

most deleterious nsSNPs in the human KRAS gene and to check their impact on the structure and function of KRAS protein.

CONCLUSIONS

In conclusion, we carried out comprehensive analysis of SNPs of the KRAS gene and characterized them according to their pathogenicity and different properties by applying several computational tools. From total SNPs, 10 SNPs out of 48 missense variants were found to be highly pathogenic and/or deleterious, while the remaining SNPs were likely neutral. Structural analysis of these pathogenic variants revealed a high degree of disrupted protein structures and stability, which suggests that these variants affect protein function by abolishing the wild/normal 3D structure of KRAS protein.

ASSOCIATED CONTENT

Data Availability Statement

The data used to support this study are available from the corresponding author upon reasonable request.

AUTHOR INFORMATION

Corresponding Author

Jizhou Zhang – Oncology Department, Wenzhou Hospital of Traditional Chinese Medicine Affiliated to Zhejiang Chinese Medicine University, Wenzhou 325000, China;
Email: jizhouzhangjet@163.com

Authors

Duoduo Xu – Oncology Department, Wenzhou Hospital of Traditional Chinese Medicine Affiliated to Zhejiang Chinese Medicine University, Wenzhou 325000, China

Qiqi Shao – Department of Nursing, Central Health Center of Zeya Town, Wenzhou 325000, China

Chen Zhou – Ultrasonography Department, Wenzhou Hospital of Traditional Chinese Medicine Affiliated to Zhejiang Chinese Medicine University, Wenzhou 325099, China

Arif Mahmood – Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha 410078 Hunan, China;
orcid.org/0000-0002-0309-1230

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acsomega.3c00804>

Author Contributions

D.X. and Q.S. contributed equally. D.X. performed the experiments and wrote the manuscript. Q.S. collected and analyzed the data. C.Z. did some of the experiments. A.M. wrote and revised the manuscript. J.Z. supervised, wrote the manuscript, and funded the study.

Funding

This paper was funded by the Natural Science Foundation of Zhejiang Province (no. LY20H290001) and Wenzhou Public Welfare Science and Technology Program (Y20180211).

Notes

The authors declare no competing financial interest.

REFERENCES

- (1) Cox, A. D.; Der, C. J. Ras history: The saga continues. *Small GTPases* 2010, 1, 2–27.

- (2) Loyo, M.; Li, R. J.; Bettegowda, C.; Pickering, C. R.; Frederick, M. J.; Myers, J. N.; Agrawal, N. Lessons learned from next-generation sequencing in head and neck cancer. *Head Neck* **2013**, *35*, 454–463.
- (3) Suh, Y.; Amelio, I.; Guerrero Urbano, T.; Tavassoli, M. Clinical update on cancer: molecular oncology of head and neck cancer. *Cell Death Dis.* **2014**, *5*, No. e1018.
- (4) Takács, T.; Kudlik, G.; Kurilla, A.; Szeder, B.; Buday, L.; Vas, V. The effects of mutant Ras proteins on the cell signalome. *Cancer Metastasis Rev.* **2020**, *39*, 1051–1065.
- (5) Jančík, S.; Drábek, J.; Radzich, D.; Hajdúch, M. Clinical relevance of KRAS in human cancers. *J. Biomed. Biotechnol.* **2010**, *2010*, 1–13.
- (6) Timar, J.; Kashofer, K. Molecular epidemiology and diagnostics of KRAS mutations in human cancer. *Cancer Metastasis Rev.* **2020**, *39*, 1029–1038.
- (7) Corso, G.; Velho, S.; Paredes, J.; Pedrazzani, C.; Martins, D.; Milanezi, F.; Pascale, V.; Vindigni, C.; Pinheiro, H.; Leite, M.; et al. Oncogenic mutations in gastric cancer with microsatellite instability. *Eur. J. Cancer* **2011**, *47*, 443–451.
- (8) Lee, J.-E.; Choi, J. H.; Lee, J. H.; Lee, M. G. Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2005**, *573*, 195–204.
- (9) Rajasekaran, R.; Georgepriyadoss, C.; Sudandiradoss, C.; Ramanathan, K.; Rituraj, P.; Rao, S. Computational and structural investigation of deleterious functional SNPs in breast cancer BRCA2 gene. *Chin. J. Biotechnol.* **2008**, *24*, 851–856.
- (10) 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.
- (11) Chitrala, K. N.; Yeguvapalli, S. Computational screening and molecular dynamic simulation of breast cancer associated deleterious non-synonymous single nucleotide polymorphisms in TP53 gene. *PLoS One* **2014**, *9*, No. e104242.
- (12) Jia, M.; Yang, B.; Li, Z.; Shen, H.; Song, X.; Gu, W. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2 gene. *PLoS One* **2014**, *9*, No. e104311.
- (13) Ramensky, V.; Bork, P.; Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **2002**, *30*, 3894–3900.
- (14) Radivojac, P.; Vacic, V.; Haynes, C.; Cocklin, R. R.; Mohan, A.; Heyen, J. W.; Goebel, M. G.; Iakoucheva, L. M. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 365–380.
- (15) Doniger, S. W.; Kim, H. S.; Swain, D.; Corcuera, D.; Williams, M.; Yang, S.-P.; Fay, J. C. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **2008**, *4*, No. e1000183.
- (16) Ng, P. C.; Henikoff, S. Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* **2006**, *7*, 61–80.
- (17) Sherry, S. T.; Ward, M.-H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311.
- (18) Akhtar, M.; Jamal, T.; Jamal, H.; Din, J. U.; Jamal, M.; Arif, M.; Arshad, M.; Jalil, F. Identification of most damaging nsSNPs in human CCR6 gene: In silico analyses. *Int. J. Immunogenet.* **2019**, *46*, 459–471.
- (19) Adzhubei, I.; Jordan, D. M.; Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, *76*, 7.20.1–7.20.41.
- (20) 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.
- (21) Choi, Y.; Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747.
- (22) 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.
- (23) Calabrese, R.; Capriotti, E.; Casadio, R. Predicting the Insurgence of Human Genetic Diseases Due to Single Point Protein Mutation using Machine Learning Approach. In *Communications to SIMAI Congress*, 2006.
- (24) Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **2005**, *33*, W299–W302.
- (25) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **2010**, *38*, W529–W533.
- (26) Yates, C. M.; Sternberg, M. J. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein–protein interactions. *J. Mol. Biol.* **2013**, *425*, 3949–3963.
- (27) Prokop, M.; Adam, J.; Kříž, Z.; Wimmerová, M.; Koča, J. TRITON: a graphical tool for ligand-binding protein engineering. *Bioinformatics* **2008**, *24*, 1955–1956.
- (28) Abdulazeez, S.; Sultana, S.; Almandil, N. B.; Almohazey, D.; Bency, B. J.; Borgio, J. F. The rs61742690 (S783N) single nucleotide polymorphism is a suitable target for disrupting BCL11A-mediated foetal-to-adult globin switching. *PLoS One* **2019**, *14*, No. e0212492.
- (29) Johnson, M. M.; Houck, J.; Chen, C. Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response. *Cancer Epidemiol., Biomarkers Prev.* **2005**, *14*, 1326–1329.
- (30) Shah, A. A.; Amjad, M.; Hassan, J. U.; Ullah, A.; Mahmood, A.; Deng, H.; Ali, Y.; Gul, F.; Xia, K. Molecular Insights into the Role of Pathogenic nsSNPs in GRIN2B Gene Provoking Neurodevelopmental Disorders. *Genes* **2022**, *13*, 1332.
- (31) Karimianpour, N.; Mousavi, S. P.; Ziaei, A.; Akbari, M. T.; Pourmand, G. R.; Abedi, A.; Ahmadi, A.; Afshin, A. H. Mutations of RAS gene family in specimens of bladder cancer. *Urol. J.* **2008**, *5*, 237–242.
- (32) Yue, P.; Moul, J. Identification and analysis of deleterious human SNPs. *J. Mol. Biol.* **2006**, *356*, 1263–1274.
- (33) Varfolomeev, S. D.; Uporov, I. V.; Fedorov, E. V. Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases. *Biochemistry* **2002**, *67*, 1099–1108.
- (34) Essadssi, S.; Krami, A. M.; Elkhatabi, L.; Elkarhat, Z.; Amalou, G.; Abdelghaffar, H.; Rouba, H.; Barakat, A. Computational analysis of nsSNPs of ADA gene in severe combined immunodeficiency using molecular modeling and dynamics simulation. *J. Immunol. Res.* **2019**, *2019*, 1–14.
- (35) Essadssi, S.; Krami, A. M.; Elkhatabi, L.; Elkarhat, Z.; Amalou, G.; Abdelghaffar, H.; Rouba, H.; Barakat, A. Computational Analysis of nsSNPs of ADA Gene in Severe Combined Immunodeficiency Using Molecular Modeling and Dynamics Simulation. *J. Immunol. Res.* **2019**, *2019*, 5902391.