Gene 508 (2012) 188–196

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

In silico analysis of Single Nucleotide Polymorphisms (SNPs) in human *BRAF* gene

Muhammad Ramzan Manwar Hussain [a,b,*,1], Noor Ahmad Shaik [b,1], Jumana Yousuf Al-Aama [b], Hani Z. Asfour [b], Fatima Subhani Khan [d], Tariq Ahmad Masoodi [e], Muhammad Akhtar Khan [c], Nazia Sultana Shaik [f]

[a] *Institute of Molecular Sciences and Bioinformatics Lahore, Pakistan*
[b] *Princess Al-Jawhara Center of Excellence in Research of Hereditary Diseases, Department of Genetic Medicine, King AbdulAziz University, Saudi Arabia*
[c] *Center of Excellence in Environment Studies, King AbdulAziz University, Saudi Arabia*
[d] *Department of Bioinformatics, International Islamic University Islamabad, Pakistan*
[e] *College of Applied Medical Sciences, King Saud University, Saudi Arabia*
[f] *Department of Biotechnology, KL University, Vijayawada, Andhra Pradesh, India*

ARTICLE INFO

Article history:
Accepted 9 July 2012
Available online 21 July 2012

Keywords:
BRAF gene
Single Nucleotide Polymorphism
Pathogenic variants
Genomics
In silico analysis

ABSTRACT

BRAF gene mutations are frequently seen in both inherited and somatic diseases. However, the harmful mutations for *BRAF* gene have not been predicted in silico. Owing to the importance of *BRAF* gene in cell division, differentiation and secretion processes, the functional analysis was carried out to explore the possible association between genetic mutations and phenotypic variations. Genomic analysis of BRAF was initiated with SIFT followed by PolyPhen and SNPs&GO servers to retrieve the 85 deleterious non-synonymous SNPs (nsSNPs) from dbSNP. A total of 5 mutations i.e. c.406T>G (S136A), c.1446G>T (R462I), c.1556 A>G (K499E), c.1860 T>A (V600E) and c.2352 C>T (P764L) that are found to exert benign effects on the *BRAF* protein structure and function were chosen for further analysis. Protein structural analysis with these amino acid variants was performed by using I-Mutant, FOLD-X, HOPE, NetSurfP, Swiss PDB viewer, Chimera and NOMAD-Ref servers to check their solvent accessibility, molecular dynamics and energy minimization calculations. Our in silico analysis suggested that S136A and P764L variants of *BRAF* could directly or indirectly destabilize the amino acid interactions and hydrogen bond networks thus explain the functional deviations of protein to some extent. Screening for BRAF, S136A and P764Lvariants may be useful for disease molecular diagnosis and also to design the molecular inhibitors of BRAF pathways.

1. Introduction

Human B-type Raf kinase (*BRAF*) belongs to the raf/mil family of serine/threonine protein kinases. It regulates the MAP kinase/ERKs signaling, a critical cascade for cell division, differentiation and secretion processes (Garnett and Marais, 2004). The corresponding *BRAF* gene is located on chromosome 7q34 and it is expressed in various tissues including brain, placenta, and testis. Due to its reported involvement in both inherited and somatic human diseases, the *BRAF* gene is increasingly being focused in medical genetics. For example, germline *BRAF* gene mutations can cause cardiofaciocutaneous (CFC), Noonan or LEOPARD syndromes (Sarkozy et al., 2009), whereas somatic mutations increases the risk of lung, colon, thyroid and skin cancers (El-Osta et al., 2011; Fedorenko et al., 2011; Balch et al., 2001; Jemal et al., 2001).

More than 100 *BRAF* gene mutations have been described, with varied frequency of >80% in melanomas to as little as 0–18% in other tumors such as 5–15% in thyroid tumors, 1–3% in lung cancers and 5% in colorectal cancer (Ikenoue et al., 2004; Pratilas and Solit, 2007). Most of the *BRAF*-activating non-synonymous mutations of the kinase domain are clustered in exons 11 and 15 regions (Davies et al., 2002). The T1799A transversion in exon 15, that converts valine to glutamic acid at codon 600 (formerly designated as V599E) accounts for more than 80% of all the *BRAF* gene mutations. Other potential BRAF mutants documented in human diseases are SER-301; ILE-462; SER-463; GLU-464; VAL-464; ALA-466; GLU-466; VAL-466; ALA-469; GLU-469; SER-581; LYS-586; LEU-595; ARG-596; ARG-597; VAL-597; GLU-600 AND ASP-600 (Davies et al., 2002; Greenman et al., 2007; Hingorani et al., 2003; Rajagopalan et al., 2002; Sjoeblom et al., 2006). Mechanistically, mutated BRAF exert most of its oncogenic effects through the activation of the MAPK pathway (Pratilas and Solit, 2007).

Abbreviations: SIFT, Sorting Intolerant from Tolerant; Polyphen, Phenotype Polymorphism; CFC, Cardio-Facio-Cutaneous; SNP, Single Nucleotide Polymorphism; nsSNP, Nonsynonymous Polymorphism; OMIM, Online Mendelian Inheritance in Man; ANOLEA, Atomic Non-Local Environment Assessment; NLE, Non-Local Environment; A, Alanine; R, Arginine; N, Asparginine; D, Aspartic; E, Glutamic acid; Q, Glutamine; G, Glycine; H, Histidine; I, Isoleucine; L, Leucine; K, Lysine; F, Phenylalanine; P, Proline; S, Serine; T, Threonine; W, Tryptophan; V, Valine.

* Corresponding author at: Princess Al-Jawhara Center of Excellence in Research of Hereditary Diseases, Department of Genetic Medicine, King AbdulAziz University, Saudi Arabia. Tel.: +966 6402000x21053–52608.
E-mail address: ramzan.manwar@yahoo.com (M.R.M. Hussain).

[1] Both the authors contributed equally.

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.07.014

Sr. #	ENSP #	Amino acid substitution	SIFT Prediction	SIFT Score
1.	ENSP00000288602	R444W	DAMAGING	0
2.	ENSP00000288602	R462I	DAMAGING	0
3.	ENSP00000288602	I463S	DAMAGING	0
4.	ENSP00000288602	G464E	DAMAGING	0
5.	ENSP00000288602	G464V	DAMAGING	0
6.	ENSP00000288602	G464R	DAMAGING	0
7.	ENSP00000288602	G466A	DAMAGING	0
8.	ENSP00000288602	G466E	DAMAGING	0
9.	ENSP00000288602	G466V	DAMAGING	0
10.	ENSP00000288602	G466R	DAMAGING	0
11.	ENSP00000288602	G469A	DAMAGING	0
12.	ENSP00000288602	G469E	DAMAGING	0
13.	ENSP00000288602	G469V	DAMAGING	0
14.	ENSP00000288602	G469R	DAMAGING	0
15.	ENSP00000288602	G469S	DAMAGING	0
16.	ENSP00000288602	V471F	DAMAGING	0
17.	ENSP00000288602	L485F	DAMAGING	0
18.	ENSP00000288602	K499E	DAMAGING	0
19.	ENSP00000288602	E501K	DAMAGING	0.01
20.	ENSP00000288602	E501G	DAMAGING	0
21.	ENSP00000288602	G534R	DAMAGING	0
22.	ENSP00000288602	I572F	DAMAGING	0
23.	ENSP00000288602	N581S	DAMAGING	0.04
24.	ENSP00000288602	N581D	DAMAGING	0.05
25.	ENSP00000288602	E586K	DAMAGING	0
26.	ENSP00000288602	D587A	TOLERATED	0.07
27.	ENSP00000288602	D587E	DAMAGING	0.02
28.	ENSP00000288602	I592K	DAMAGING	0
29.	ENSP00000288602	I592M	DAMAGING	0
30.	ENSP00000288602	I592V	DAMAGING	0
31.	ENSP00000288602	D594E	DAMAGING	0
32.	ENSP00000288602	D594G	DAMAGING	0
33.	ENSP00000288602	D594V	DAMAGING	0
34.	ENSP00000288602	F595S	DAMAGING	0
35.	ENSP00000288602	F595L	DAMAGING	0
36.	ENSP00000288602	G596R	DAMAGING	0
37.	ENSP00000288602	L597E	DAMAGING	0
38.	ENSP00000288602	L597R	DAMAGING	0
39.	ENSP00000288602	L597S	DAMAGING	0
40.	ENSP00000288602	L597V	DAMAGING	0
41.	ENSP00000288602	T599I	DAMAGING	0
42.	ENSP00000288602	V600A	TOLERATED	0.32
43.	ENSP00000288602	V600E	DAMAGING	0
44.	ENSP00000288602	V600K	DAMAGING	0
45.	ENSP00000288602	V600R	DAMAGING	0
46.	ENSP00000288602	V600D	DAMAGING	0
47.	ENSP00000288602	V600L	DAMAGING	0.01
48.	ENSP00000288602	V600M	DAMAGING	0
49.	ENSP00000288602	K601E	DAMAGING	0
50.	ENSP00000288602	K601N	DAMAGING	0
51.	ENSP00000288602	R603T	DAMAGING	0
52.	ENSP00000288602	R603V	DAMAGING	0
53.	ENSP00000288602	R603W	DAMAGING	0
54.	ENSP00000288602	R603Y	DAMAGING	0
55.	ENSP00000288602	R603A	DAMAGING	0
56.	ENSP00000288602	R603C	DAMAGING	0
57.	ENSP00000288602	R603D	DAMAGING	0
58.	ENSP00000288602	R603E	DAMAGING	0
59.	ENSP00000288602	R603F	DAMAGING	0.04
60.	ENSP00000288602	R603G	DAMAGING	0
61.	ENSP00000288602	R603H	DAMAGING	0
62.	ENSP00000288602	R603I	DAMAGING	0
63.	ENSP00000288602	R603K	DAMAGING	0
64.	ENSP00000288602	R603L	DAMAGING	0.03
65.	ENSP00000288602	R603M	TOLERATED	1
66.	ENSP00000288602	R603N	DAMAGING	0
67.	ENSP00000288602	R603P	DAMAGING	0
68.	ENSP00000288602	R603Q	DAMAGING	0
69.	ENSP00000288602	R603S	DAMAGING	0
70.	ENSP00000288602	R603	DAMAGING	0
71.	ENSP00000288602	S605F	DAMAGING	0
72.	ENSP00000288602	S605N	DAMAGING	0.03
73.	ENSP00000288602	D638E	DAMAGING	0
74.	ENSP00000288602	L692S	DAMAGING	0
75.	ENSP00000288602	S136A	TOLERATED	1
76.	ENSP00000288602	P764L	DAMAGING	0.01
77.	ENSP00000288602	I572F	DAMAGING	0
78.	ENSP00000288602	L384M	TOLERATED	0.1
79.	ENSP00000288602	Q93H	DAMAGING	0
80.	ENSP00000288602	Q356E	TOLERATED	0.85
81.	ENSP00000288602	A246P	DAMAGING	0
82.	ENSP00000288602	Q257R	DAMAGING	0.04
83.	ENSP00000288602	K499E	DAMAGING	0
84.	ENSP00000288602	E501G	DAMAGING	0
85.	ENSP00000288602	N581D	DAMAGING	0.05

Fig. 1. SIFT prediction for the 85 nsSNPs of *BRAF*.

Basing on the importance of *BRAF* gene in diverse range of human diseases its functional genomics based on mutation analysis is conceived to give key leads in disease diagnosis and therapy (Pratilas and Solit, 2007). But, the harmful mutations for *BRAF* gene have not been predicted to date in silico. Therefore, frequently reported genetic variants such as c.467 N>G (S136A), c.1446 G>T (R462I), c.1556 A>G (K499E), c.1860 T>A (V600E) and c.2352 C>T (P764L) were examined in this study. To explore the possible associations between genetic mutation and phenotypic variation different genomics tools were used for prioritization of high-risk non-synonymous mutations in coding regions that are likely to have an effect on the structure and function of *BRAF*. Our in silico analysis suggests that the presence of deleterious mutations (S136A and P764L) in *BRAF* gene can tremendously affect the structure and functions of protein, in the form of mis-folding and intermolecular interactions thus they may play an important role in disease susceptibility.

2. Materials and methods

The data on human *BRAF* gene was collected from Online Mendelian Inheritance in Man (OMIM) and Entrez Gene on National Center for Biological Information (NCBI) web site. The SNPs information (Protein accession number and SNP ID) of the *BRAF* gene was retrieved from the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/), and SWISSProt databases (http://expasy.org/).

2.1. SIFT

SIFT (version 2) predicts the tolerated and deleterious SNPs and identifies the impact of amino acid substitution on protein function and phenotype alterations. According to the previously reported data, SIFT (http://blocks.fhcrc.org/sift/SIFT.html) discriminates between functionally neutral and deleterious polymorphisms during mutagenesis studies in humans (Ng and Henikoff, 2001). SIFT performs analysis on the basis of different algorithms and interprets the homologous sequences using the Swiss-Prot (version 51.3) and TrEMBL (version 34.3). The threshold intolerance score for SNPs is 0.05 or less (Fig. 1).

2.2. PolyPhen

Polyphen (version 2) (http://genetics.bwh.harvard.edu/pph2/) predicts the influence of amino acid substitution on the structure and function of proteins by using the specific empirical rules. Protein sequence, database ID/accession number, amino acid position and amino acid variant details are the input options for Polyphen (Ramensky et al., 2002). The tool estimates the position-specific independent count (PSIC) score for every variant and calculates the score difference between variants.

2.3. SNPs&GO

Single Nucleotide Polymorphism Database (SNPs) & Gene Ontology (GO) are support vector machine (SVM) based accurate methods used to predict the disease related mutations from protein sequences with a scoring accuracy of 82% and Matthews correlation coefficient of 0.63. For SNPs&GO, FASTA sequence of whole protein is considered to be an input option and output will be the prediction results based on the discrimination among disease related and neutral variations of protein sequence. The probability score higher than 0.5 reveals the disease related effect of mutation on the parent protein function (Calabrese et al., 2009; Thomas et al., 2003).

2.4. I-Mutant and FOLD-X

I-Mutant (version 2.0) (http://folding.uib.es/i-mutant/i-mutant2.0.html) is a neural network based tool for the routine analysis of protein stability and alterations by taking into account the single-site mutations. The FASTA sequence of protein retrieved from UniProt is used as an input to predict the mutational effect on protein stability. I-Mutant also provides the scores for free energy alterations, calculated with the FOLD-X energy based web server. FOLD-X is a computer algorithm for quantitative estimation of interactions facilitating the stability of proteins. The FOLD-X tool provides the comparison between wild type and mutant models in the form of van der Waals clashes, which greatly influence the energy decomposition. Sometimes the mutations cause strain in the original native structure and sometimes reduce (Abagyan and Totrov, 1994; Schymkowitz et al., 2005).

Table 2
I-TASSER results carrying C-score, TM-score and RMSD regarding selected secondary structure (native protein Model 1).

Model #	C-score	Exp. TM-Score	Exp. RMSD	No. of decoys	Cluster
Model 1	−2.70	0.40±0.14	15.1±3.5	191	0.0305
Model 2	−3.22			114	0.0182
Model 3	−3.76			66	0.0105
Model 4	−3.86			60	0.0096
Model 5	−4.00			52	0.0083

2.5. SAAP and dbSNPs

Single Amino Acid Polymorphisms (SAAP) (http://www.bioinf.org.uk/saap/) (Cavallo and Martin, 2005) and Single Nucleotide Polymorphism Database (dbSNPs) (http://www.bioinf.org.uk/saap/db/) were used to recognize the single point mutation within the encoded protein of *BRAF*. Damaging effect of these mutations was predicted by SIFT and PolyPhen. Energy minimization was done by SWISS-PDB viewer and Atomic Non-Local Environment Assessment (ANOLEA), with the harmonization between the results of these two programs. ANOLEA performs energy calculations by examining the "Non-Local Environment" (NLE) for each heavy atom in the residues of protein sequence (Melo and Feytmans, 1998). The energy base calculations are conducted at the atomic levels in protein structures and the output energy profile gives the energy value for every amino acid. High energy zones (HEZs) are linked with errors or latent interacting zones of protein.

2.6. MUSTER

MUSTER (v1.0) "http://zhanglab.ccmb.med.umich.edu/MUSTER/" is a valuable threading tool for protein structure prediction. Muster provides the Z-score and complete full length models by using MODELLER v8.2. If the calculated Z-score is greater than 7.5, the corresponding template is considered good otherwise designated as bad. Sequence-derived profiles, secondary structures, structured-derived profiles, solvent accessibility, backbone torsion angles and hydrophobic scoring matrix are the six different sources used by MUSTER (Wu and Zhang, 2008).

2.7. I-TASSER

I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) creates the full length protein models by excising continuous fragments from threading alignments and further reconstructs them by using replica-exchanged Monte Carlo simulations. During the simulation, decoys (low temperature replicas) are clustered by SPICKER and the full length

Table 1
Alteration in protein stability due to mutations.

Mutations	Position	WT	NEW	pH	Temp (°C)	Stability	RI	DDG (kcal/mol)
Mutation 1: 136 S→A	136	S	A	7.0	25	Decrease	8	−1.29
Mutation 2: 462 R→I	462	R	I	7.0	25	Decrease	1	−0.41
Mutation 3: 499 K→E	499	K	E	7.0	25	Increase	2	0.35
Mutation 4: 600 V→E	600	V	E	7.0	25	Decrease	6	−1.51
Mutation 5: 764 P→L	764	P	L	7.0	25	Decrease	2	−0.62

Where, "WT" is the amino acid in native protein, "NEW" is mutant amino acid, "RI" is the reliability index and DDG is the stability (DDG<0: Decrease Stability, DDG>0: Increase Stability).

Table 3
Surface accessibility of wild type and mutant variants.

Sr #	Amino acid	Class assignment	Position	Relative surface accessibility (RSA)	Absolute surface accessibility	Z-fit score for RSA prediction
1	S	Exposed	136	0.353	41.348	−1.586
	A	Exposed	136	0.378	41.700	−1.338
2	R	Exposed	462	0.334	76.555	1.019
	I	Exposed	462	0.354	65.397	1.144
3	K	Buried	499	0.217	44.555	1.489
	E	Buried	499	0.209	36.582	1.420
4	V	Exposed	600	0.313	48.077	−0.393
	E	Buried	600	0.264	46.191	−0.429
5	P	Exposed	764	0.361	51.183	−1.604
	L	Exposed	764	0.367	67.161	−2.098

Values associated with wild type strain are colored green while those linked with mutant variants are colored red.

atomic models are generated by selecting the top five cluster centroids. The number of decoys at the unit space of SPICKER cluster, defines the cluster density. The quality of predicted structure is estimated by I-TASSER in the form of confidence score (C-score). The C-score lies in the range of −5 to 2, with higher values depicting the high confidence for predicted model. The C-score has a correlation with the TM-score and RMSD. If the native structure is known, the TM-score and RMSD are used to measure the accuracy of predicted structure else, these are used to predict the quality of the modeling prediction by calculating the distance between two predicted models (by considering the one model as a native on the basis of C-score). The TM-score is a recently proposed scale to solve the local error problem of RMSD. A TM-score >0.5 highlights a model of correct topology and a TM-score <0.17 indicates a random similarity (Roy et al., 2010; Zhang, 2008; Zhang and Skolnick, 2004).

2.8. Chimera

University of California, San Francisco (UCSF) Chimera (http://www.cgl.ucsf.edu/chimera) is a high-quality extensible program for interactive conception and analysis of molecular assemblies and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles (Pettersen et al., 2004).

2.9. NetSurfP server ver. 1.1

NetSurfP server (http://www.cbs.dtu.dk/services/NetSurfP/) predicts the surface accessibility and secondary structure of amino acids, when they are in a sequence. The reliability of this prediction method is in the form of Z-score. The Z-score highlights the surface prediction reliability, but not associated with secondary structure (Petersen et al., 2009). The Z-score is related to the surface prediction, and not the secondary structure.

2.10. Project HOPE

Project Have yOur Protein Explained (HOPE; http://www.cmbi.ru.nl/hope/home) is an automatic mutant analysis server to study the insight structural features of native protein and the variant models. HOPE provides the 3D structural visualization of mutated proteins, and gives the results by using UniProt and DAS prediction servers. Input method of Project HOPE carries the protein sequence and selection of Mutant variants. HOPE server predicts the output in the form of structural variation between mutant and wild type residues (Venselaar et al., 2010).

3. Results

3.1. Retrieval of SNPs

The *BRAF* gene investigated in this work reported to have a total of 106 coding SNPs, of which 85 were nsSNPs (non-sense, missense and frame shift) and 21 were synonymous SNPs. Only non-synonymous coding SNPs were chosen for further analysis.

3.2. Prediction of tolerated and deleterious SNPs

The SIFT tool estimates the consequence of an amino acid substitution on protein function by considering the sequence homology and physical properties of amino acids.

SIFT scores were classified as damaging (0.00–0.05), potentially damaging (0.051–0.10), borderline (0.101–0.20), or tolerant (0.201–1.00). Of the 85 screened nsSNPs (Fig. 1), S136A showed the tolerance score of 1.00, whereas R462I, K499E and V600E showed the score of 0.00 and remaining P764L showed 0.01 score (Ng and Henikoff, 2001).

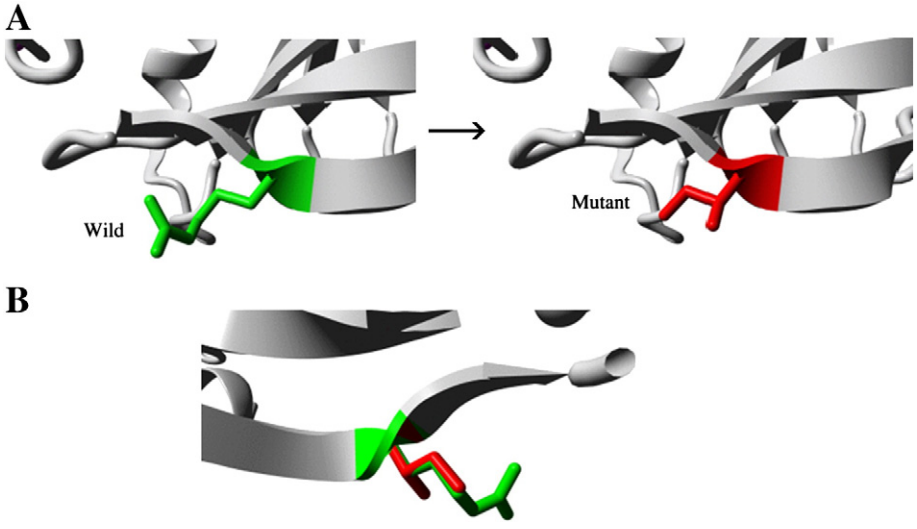

Fig. 2. Close-up view of the mutation at 462 position in *BRAF*. A) Differentiation between the size and stereochemical orientation of wild type and mutant residue. The main protein core is shown in gray color while the wild type and mutated residues are shown in green and red color respectively. B) The color illustration of superimposed structure.

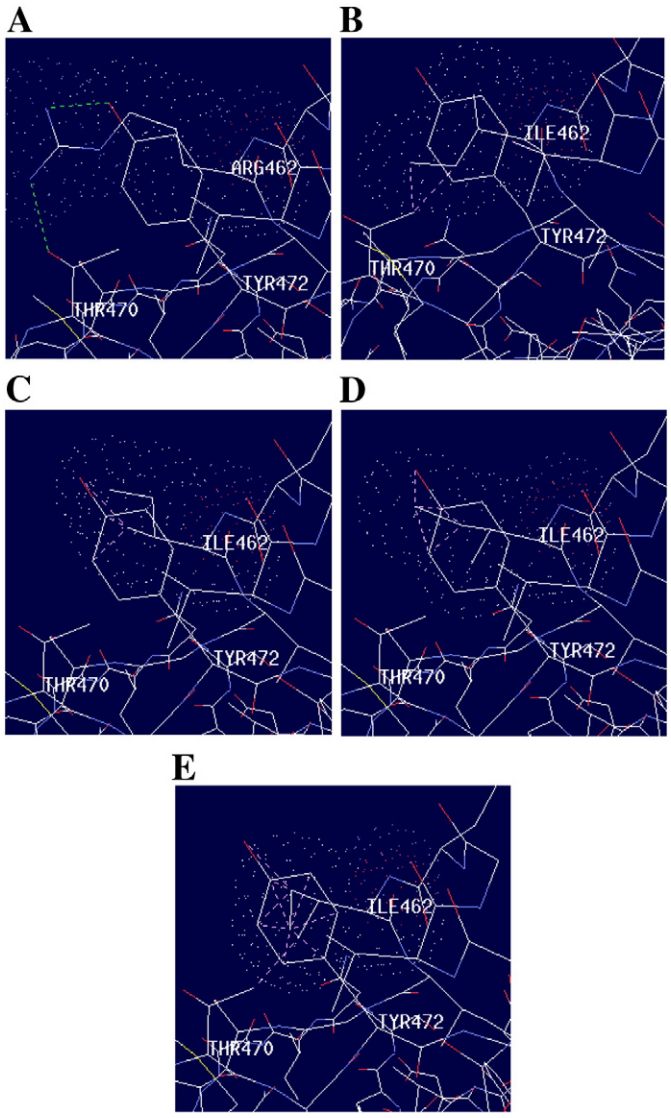

Fig. 3. Molecular orientation, interactions and distortion as a result of mutation at 462 (Arg→Ile), A) green dotted/discontinuous lines highlight the H-bonding interaction between N–H of Arg462 and O–H of Tyr472 and Thr470, B) pink lines indicate the distortion between the mutated residue (Ile462) and the Tyr472, C) van der Waals clashes between the delocalized electron density of Tyr472 and Ile462, D) conformational change in Ile462, and the electrostatic repulsion between phene ring of Tyr472 and aliphatic domain of Ile462, E) high distortion within and outside the cyclic environment of Tyr472. All the interactions and distortions are operated within the domain of 2.35 and 3.2 Å in each model. The dotted spheres around the wild type and mutant residues indicate the van der Waals radii.

3.3. Prediction of protein structural and functional modifications

An amino acid substitution carries the potential impact on protein structure and function. The PolyPhen server characterizes protein structural modifications by using the SIFT results as input. The prediction variant with accession number P15056, mutant position 136, amino acid substitution of S (AA1)→A (AA2) was predicted to be benign with a PSIC score of 0.000 (sensitivity: 1.00; specificity: 0.00 in human division). PolyPhen scores were assigned probably damaging (2.00 or more), possibly damaging (1.40–1.90), potentially damaging (1.20–1.50), benign (0.00–0.90). Out of the total 85 nsSNPs screened by SIFT and PolyPhen, 5 nsSNPs were found to be benign with the PSIC score ranging from 0.00 to 0.2 (Ramensky et al., 2002).

3.4. Prediction of disease related mutations by SNPs&GO

SNPs&GO is trained and tested with cross-validation procedures in which similar proteins are placed in same dataset to calculate the LGO score derived from the GO data base. The disease probability score was predicted to be higher for R462I (0.868), K499E (0.547) and V600E (0.799) as compared to S136A (0.226) and P764L (0.373) mutants. The S136A and P764L were predicted neutral with reliability index (RI) values of 5 and 3, respectively. But the RI values for R462I, K499E and V600E mutations were found to be 7, 1 and 6 respectively.

3.5. Prediction of protein structural stability

I-Mutant is a neural network based routine tool used in the analysis of protein stability alterations by considering the single-site mutation. I-Mutant also provides the scores for free energy alterations, calculated with the FOLD-X energy based web server. By assimilating the FOLD-X estimations with those of I-Mutant, the 93% precision can be achieved. The five mutations (S→A, R→I, K→E, V→E and P→L) in *BRAF* gene have been selected on the basis of prediction scores of PolyPhen, the additional mutations were chosen from published reports. These variants were given to I-Mutant web server to predict the DDG stability and reliability index (RI) upon mutation (Table 1).

3.6. In silico solvent accessibility of amino acid residue in protein structure, and MD simulation of native and mutant residues

Amino acid alterations due to point mutations can drastically modify the protein stability; hence the modeling of protein structural information is necessary for absolute understanding of its functionality. The deleterious nsSNPs retrieved from dbSNP, and SAAPdb (http://www.bioinf.org.uk/saap/db/), usually maintain the SNPs information of both dbSNP and HGVBASE (Human Genome Variation database) that displays the data onto the translated regions of the gene to establish the nature and location of a mutation in protein. The *BRAF* protein secondary structure was taken from I-TASSER (Table 2) and mutations were achieved at the consequent positions by using Swiss-Prot. The NOMAD-Ref Gromacs server and Deep View-The Swiss-Pdb viewer were used to collect the energy minimization by selecting the force field for the native and mutant residues.

The 3D analysis of wild type and mutant protein structures was performed by using project HOPE (http://www.cmbi.ru.nl/hope/home) and Deep View-The Swiss-Pdb viewer. Each amino acid in wild type and mutant structures carries specific properties like solvent accessibility, charge density, hydrophobicity, rigidity, molecular surfaces and electrostatic potential values. Native and mutant residues sometimes differ due to carrying such specific properties and can disrupt the structural and functional features of the original protein. For example, the c.406T>G variant (rs138086018) leads to Ser→Ala conversion at position 136, the mutant residue is smaller, hydrophobic, not conserved and contains high positive charge density as compared to wild type. The mutant residue showed the high relative surface accessibility (RSA), absolute surface accessibility and Z-fit score for RSA (Table 3).

The c.1446G>T mutation (rs180177032), which leads to Arg→Ile conversion at 462 position is reported in ExPASy as VAR_018613. For this variant, mutated residue is found to be small (Fig. 2), neutral and more hydrophobic than the wild type. The Swiss Pdb-Deep viewer and HOPE server based 3D visualization and follow-up analysis of wild and mutated residues in protein molecule (Fig. 3) highlighted the hydrogen bonding interaction of Arg with nearby Tyr472 and Thr470 (Fig. 3A). The mutated Ile462 induces different conformational changes and distort the protein stability due to clashes with electronic density of Tyr472 and Thr470 (Figs. 3B–E).

The c.1556 A>G variant (rs180177037) that leads to Lys→Glu conversion at 499 position, is reported on ExPASy as VAR_026116 and also

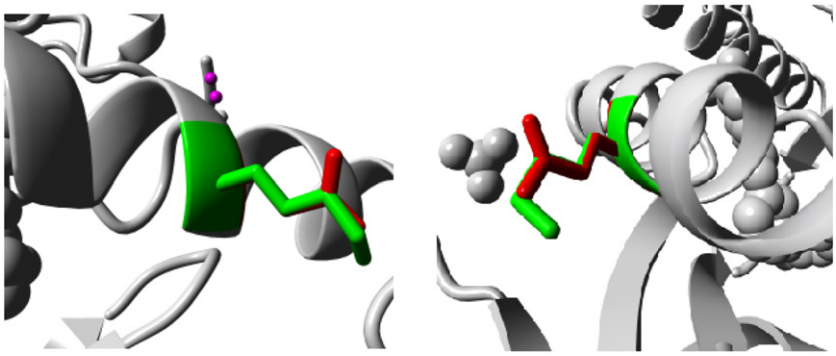

Fig. 4. Deep view of superimposed structure of wild and mutant residue at 499 position.

located within domain, annotated in UniProt. In this vital mutation the positively charged residue is replaced by negatively charged small residue (Fig. 4). So, this mutation induces the opposite charge with high charge density and can possibly disturb the electrostatic interactions with other molecules. The wild-type residue is not conserved at this position but the mutated residue is located in highly conserved region. The current study will estimate the effects of mutational interactions made by mutated moiety, the structural conformations of the residue, and the H-bonding within the domain of 2.35 and 3.2 Å. The introduction of opposite charge can disturb the electrostatic interaction of amino acid residues in the vicinity of position 499 (Fig. 5). The carboxylic acid anion (delocalization of charge density between two oxygen atoms) on Glu at 499 might interact with the neighboring residues (Arg444,Gln496, Leu495, Val502, Gly503 and Tyr519) and distorts the original conformation of residues within the protein.

The c.1860 T>A variant (rs121913378) leads to Val→Glu conversion at 600 position is different from other mutations because the hydrophobic neutral amino acid is transformed into negatively charged acidic species. In this mutation the exposed Val large residue (Fig. 6) is converted into buried Glu small moiety (Table 3). The conformational changes in mutated Glu600 is observed by using Swiss-Pdb viewer and graphically assembled in Fig. 7. In some of structures, Glu600 showed the H-bonding with the nearby Ser602 while in others it exhibited the van der Waals clashes with Lys601 and Ser614. This mutation can introduce the high solvent accessibility, ionic potential and high rigidity in the mutated model.

The c.2352 C>T variant (rs139420557) that leads to Pro→Leu conversion at 764 positions is a rarely observed variant due to the transformation of cyclic amino acid into chain residue. Prolines are known for their rigidity and therefore create special backbone conformation, which might be disturbed at this position due to Leu. In this mutation the small size secondary amine structure is mutated into large size primary aliphatic amine which might not fit into the position and can disturb the local structure of protein. The current mutation can affect the structure and function of native structure due to carrying the less charge density, aliphatic domain and less rigidity.

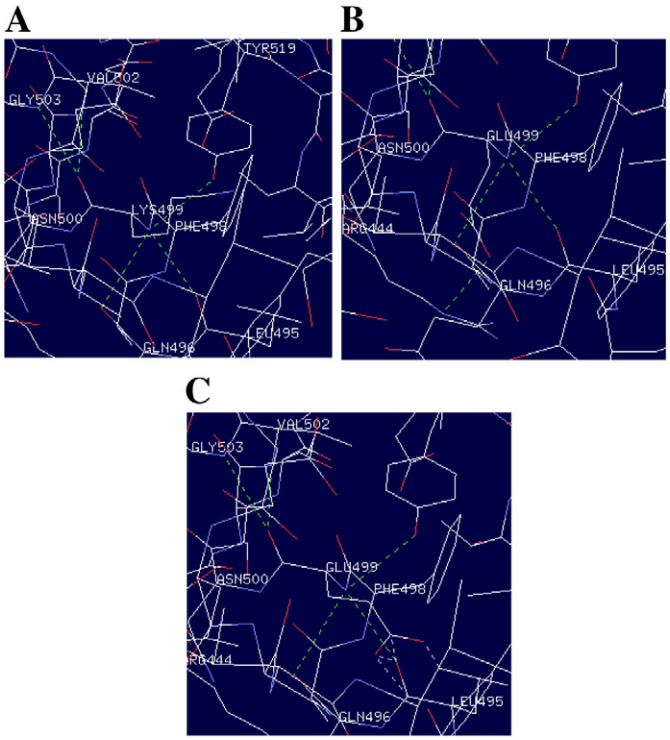

Fig. 5. H-bonding interactions of wild and mutated residue with the vicinal moieties. A) Structural discrepancy between Lys499 (native) and Glu499 (mutant) residues on the basis of stereochemical interactions, operated at 2.35 and 3.2 Å range. Lys499 interacts with the neighboring residues (Gln496, Leu495, Val502, Gly503 and Tyr519), as visualized by the 3-D portray. B) Conversion of wild type into mutant variant (Glu499) induces the H-bonding interaction (green discontinuous/dotted line) with Arg444. C) van der Waals clashes with Leu495 as a result of change in stereochemical orientation of Glu499 (indicated with the dotted pink line).

4. Discussion

The SNPs are hypothesized to play an important role in several human diseases. Approximately, more than 4 million unique human SNPs have now been reported by a number of SNP databases such as

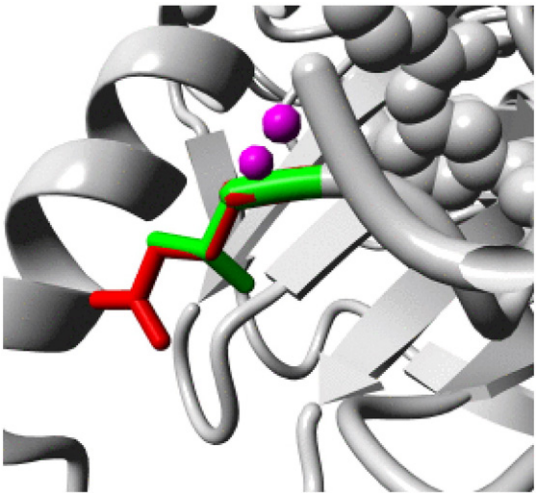

Fig. 6. Superimposed structure of (wild type) Val and (mutant) Glu residues at 600 position.

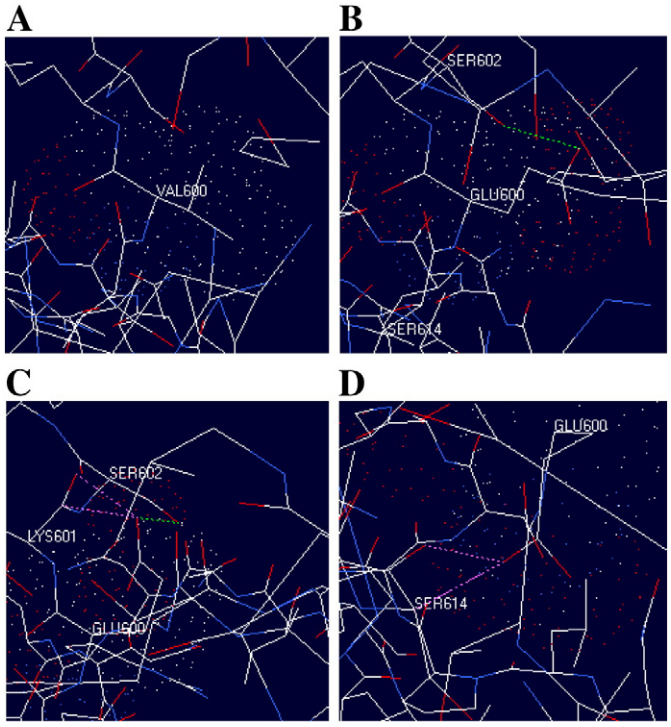

Fig. 7. Three dimensional analysis of wild type and mutated residue at 600 position. A) Deep 3D analysis of Val600 is indicating the lack of H-bonding and van der Waals clashes with the nearby residues in 2.35 and 3.2 Å range, and the dotted spheres showed the van der Waals radii of Val600. B) The H-bonding interaction between the Glu600 and Ser602 in the form of dotted green lines. C) One discontinuous (dotted) green line highlights the H-bonding between Glu600 and Ser602, while two dotted pink lines portrays the conformational clashes of Glu600 with Lys601, respectively. D) Two distortions are observed with Ser at 614 position.

human genome variation database, HGVBase (Smigielski et al., 2000) and the National Center for Biotechnology Information database, dbSNP (Rajasekaran et al., 2007). Approximately, 2% of the all known single nucleotide variants associated with genetic diseases are non-synonymous SNPs occurring in coding regions and contributes to the functional diversity of the encoded proteins in the human population (Fredman et al., 2002). Nonetheless, to date the complete mechanisms by which a nucleotide variant may result in a phenotypic change are for the most part unknown. But, in silico analysis using powerful software tools can facilitate predicting the phenotypic effect of non-synonymous coding SNPs on the physico-chemical properties of the concerned proteins. Such information is critical for genotype–phenotype correlations and also to understand disease biology.

Given the fact that nsSNPs in critical cellular genes such as *BRAF* modify the normal programs of cell proliferation, differentiation and death, it is believed to play an important role in disease predisposition also. Therefore, an effort was made to identify SNPs that can modify the structure, function and expression of the *BRAF* gene. As described in the methodology section, a computational approach was undertaken to study systematic analysis of SNPs to predict the benign mutations by using the SIFT, PolyPhen and SNP&GO servers. According to the SIFT results, most of the SNPs can cause possible damaging effect. But the PolyPhen results predicted only 5 mutations S136A, R462I, K499E, V600E and P764L can produce the benign effects. Out of which, S136A and P764L mutations are reported in dbSNP, these variants showed higher benign tendency over the remaining ones. R462I and V600E are earlier reported for their role in colorectal cancers, sarcomas, colorectal adenocarcinoma, metastatic melanoma, and ovarian serous carcinomas (Davies et al., 2002; Greenman et al., 2007; Hingorani et al., 2003; Rajagopalan et al., 2002; Sjoeblom et al., 2006). The K499E variant was previously reported in "Cardiofaciocutaneous syndrome" (Rodriguez-Viciana et al., 2006; Schulz et al., 2008).

Although different mutations in kinase domain of *BRAF* protein are reported, the native primary structure (with complete 766 amino acid sequence) is not yet available on either PDB or Swissprot databases. Native *BRAF* protein structure is essential for the in silico testing of structural and functional impacts of activating mutations, which are increasingly being identified in several human diseases. The 3D structures for most of the proteins are analyzed and deposited in PDB server. But, not all mutant protein structures are analyzed and deposited in the webservers. Therefore, in the present study, we predicted the secondary structure of native *BRAF* protein using I-TASSER and analyzed the effect of highly damaging nsSNPs on the structural and functional aspects of protein. The TM-score (>0.5) for our predicted secondary structure highlights correct topology of the model. Additionally, the C-score reflects the high confidence for the quality of secondary structure and has a correlation with the TM-score and RMSD. These values (TM-score and RMSD) corroborate the quality of our predicted structure, but we have assessed the Ramachandran plot (Fig. 8) to corroborate quality of our *BRAF* secondary structure (Lovell et al., 2002).

Molecular dynamics (MD) simulation was performed to study the explicit solvent behavior of native and mutant structures to examine the difference in dynamics and stability of native and fetal mutations. We compared RMSD value and total energy values (kcal/mol) of native structure and mutated modeled structure for *BRAF* gene variants. The total energy of wild type structure prior to energy minimization was +26,826 kJ/mol, while it is −9620.395 kJ/mol after energy minimization (Table 4). Our results indicated that S136A variant has higher surface accessibility score due to which the core of native protein structure can lose the hydrogen bonding and disrupts the proper folding of *BRAF* protein complex. With regards to Ile462 mutation, molecular and structural variations between Arg and Ile can lead to different conformational changes thus it mediate the distortion and structural destability due to clashes between the electronic density of nearby Tyr472 and Thr470 amino acid residues in *BRAF* protein (Figs. 3B–E). The Lys→Glu variant at 499 introduces the opposite charge with high charge density and possibly disturbs the electrostatic interactions with other molecules. The Val→Glu variant at 600 position is also significant because the hydrophobic neutral amino acid is transformed into negatively charged acidic species. This mutant residue is bigger than the wild type and cannot fit within the available space. The free carboxylic acid group on Glu can disturb the ionic interaction of nearby residues and impart the vital vulnerable effects. The Pro→Leu variant at 764 position is a rare type due to the transformation of cyclic amino acid into chain residue form. Due to carrying the less rigid behavior, the mutant residue can possibly disturb the original core structure of native protein. The two identified *BRAF* variants i.e. S136A and P764L could de-regulate the ras-MAPK signaling cascade, thus alter the cell division, differentiation and secretion processes resulting in human pathologies. Hence, the S136A and P764L variants constitute a unique resource of genetic markers that may considerably increase the power of *BRAF* gene mutation-screening in disease epidemiological studies.

In conclusion, the current study reports the in silico analysis of five genetic variants in *BRAF* gene (including two vital mutations i.e. S136A and P764L), which can directly or indirectly influence the intermolecular and intramolecular interactions of amino acid residues, and can culminate into disease risks in body. By analyzing the conformational changes and interactions of amino acid residues within *BRAF* protein, we have identified significant structural changes that can explain the activity deviations, caused by several mutations. Screening for *BRAF*, c.467 N>G and c.2352 C>T variants may be useful for molecular diagnosis and development of vital molecular inhibitors of *BRAF* pathways. This study demonstrates the significance of different bioinformatics tools to figure out the phenotypic changes and protein function, associated with the structure–function relationship of *BRAF*.

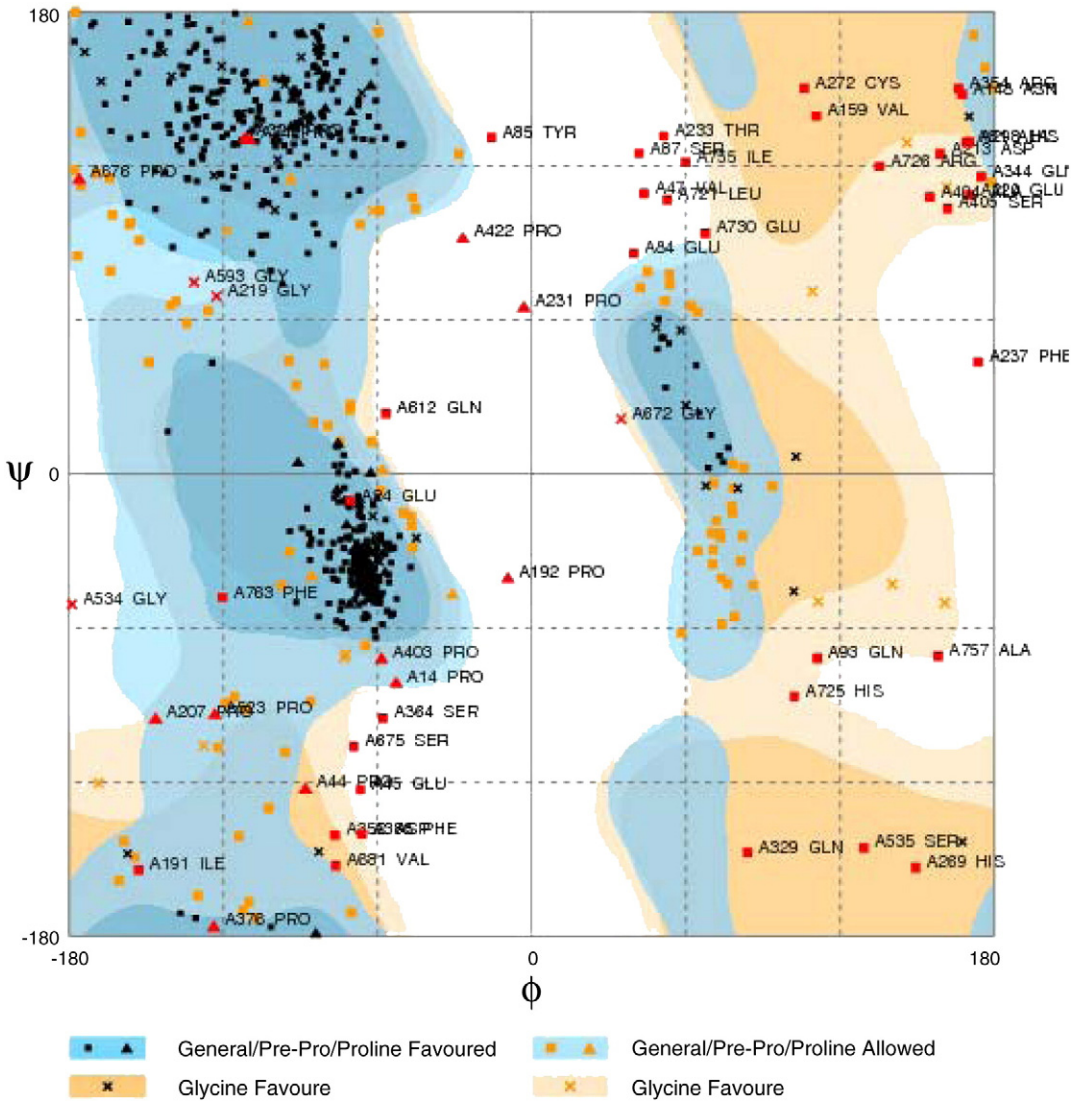

Fig. 8. Ramachandran plot of *BRAF* secondary structure.

Table 4
Total energy of native and mutant structures after energy minimization.

Amino acid variants	Total energy after energy minimization (kJ/mol)	Total electrostatic potential
Native	−9620.395	−18,057.46
S→A	−17,594.932	−19,141.68
R→I	−15,654.336	−18,593.48
K→E	−19,974.053	−19,306.34
V→E	−17,597.764	−19,153.93
P→L	−22,657.219	−19,488.01

The total energy of native structure before energy minimization was 26,826 kJ/mol.

References

Abagyan, R., Totrov, M., 1994. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J. Mol. Biol. 235, 983–1002.

Balch, C.M., et al., 2001. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. J. Clin. Oncol. 19, 3635–3648.

Calabrese, R., et al., 2009. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum. Mutat. 30, 1237–1244.

Cavallo, A., Martin, A.C., 2005. Mapping SNPs to protein sequence and structure data. Bioinformatics 8, 1443–1450.

Davies, H., et al., 2002. Mutations of the BRAF gene in human cancer. Nature 417, 949–954.

El-Osta, H., et al., 2011. BRAF mutations in advanced cancers: clinical characteristics and outcomes. PLoS One 6 (10), e25806.

Fredman, D., et al., 2002. HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. Nucleic Acids Res. 30, 387–391.

Garnett, M.J., Marais, R., 2004. Guilty as charged: B-RAF is a human oncogene. Cancer Cell 6, 313–319.

Greenman, C., et al., 2007. Patterns of somatic mutation in human cancer genomes. Nature 446, 153–158.

Hingorani, S.R., et al., 2003. Suppression of BRAF(V599E) in human melanoma abrogates transformation. Cancer Res. 63, 5198–5202.

Ikenoue, T., et al., 2004. Rapid detection of mutations in the BRAF gene using real-time polymerase chain reaction and melting curve analysis. Cancer Genet. Cytogenet. 149, 68–71.

Jemal, A., Devesa, S.S., Hartge, P., Tucker, M.A., 2001. Recent trends in cutaneous melanoma incidence among whites in the United States. J. Natl. Cancer Inst. (Bethesda) 93, 678–683.

Lovell, S.C., et al., 2002. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Protein Struct. Funct. Genet. 50, 437–450.

Melo, F., Feytmans, E., 1998. Assessing protein structures with a non-local atomic interaction energy. J. Mol. Biol. 277, 1141–1152.

Ng, P.C., Henikoff, S., 2001. Predicting deleterious amino acid substitutions. Genome Res. 11, 863–874.

Petersen, B., et al., 2009. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct. Biol. 9, 51 http://dx.doi.org/10.1186/1472-6807-9-51.

Pettersen, E.F., et al., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Pratilas, C.A., Solit, D.B., 2007. Therapeutic strategies for targeting BRAF in human cancer. Rev. Recent Clin. Trials 2 (2), 121–134.

Rajagopalan, H., et al., 2002. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. Nature 418, 934-934.

Rajasekaran, R., et al., 2007. Identification and in silico analysis of functional SNPs of the BRCA1 gene. Genomics 90, 447–452.

Ramensky, V., Bork, P., Sunyaev, S., 2002. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 30, 3894–3900.
Rodriguez-Viciana, P., et al., 2006. Germline mutations in genes within the MAPK pathway cause cardio-facio-cutaneous syndrome. Science 311, 1287–1290.
Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738.
Sarkozy, A., et al., 2009. Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. Hum. Mutat. 30, 695–702.
Schulz, A.L., et al., 2008. Mutation and phenotypic spectrum in patients with cardio-facio-cutaneous and Costello syndrome. Clin. Genet. 73, 62–70.
Schymkowitz, J., et al., 2005. The FoldX web server: an online force field. Nucleic Acids Res. 33, W382–W388.
Sjoeblom, T., et al., 2006. The consensus coding sequences of human breast and colorectal cancers. Science 314, 268–274.
Smigielski, E.M., et al., 2000. dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. 28, 352–355.
Thomas, P.D., et al., 2003. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. 31, 334–341.
Venselaar, H., Te Beek, T.A., Kuipers, R.K., Hekkelman, M.L., Vriend, G., 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics 11, 548.
Wu, S., Zhang, Y., 2008. MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins 72, 547–556.
Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 40.
Zhang, Skolnick, 2004. Scoring function for automated assessment of protein structure template quality. Proteins 57, 702–710.