A Simulation Analysis and Screening of Deleterious Non-Synonymous Single Nucleotide Polymorphisms (SNPs) in Human CDKN1A Gene

G. M. Shazzad Hossain Prince[1,*], Trayee Dhar[1]

[1]Department of Microbiology, Noakhali Science and Technology University, Noakhali, 3814, Bangladesh.

* Corresponding Author: shazzad.prince27@gmail.com

Abstract

CDKN1A also known as *p21^{CIP1} /p21^{WAF1}*, a cyclin dependent kinase 1, interacts with proliferating cell nuclear antigen (PCNA) resulting in cell cycle inhibition in human. Non-synonymous single nucleotide polymorphisms (nsSNPs), which reside in the coding region of a gene, might distort the normal function of the corresponding protein. In silico analysis in this study followed many different algorithms. Following the final screening of 118 nsSNPs from dbSNP (NCBI), 12 missense SNPs (R19C (C→T), G23D (A→G), V25G (G→T), V25L (C→G), Q29P (A→C→G), F51L (C→T), E56K (A→G), T57I (C→T), G61R (C→G), G61D (A→G), Y151C (A→G) and R156W (C→G→T) were predicted to have deleterious effect by all the algorithms. Of them, R19C, G23D, F51L, Y151C and R156W occurred at the highly conserved site. G23D, F51L variants also occurred at the CDI domain. Homology structures of the protein predicted decrease of energy in mutant models. GV-GD scores predicted only two variants as neutral (V25L, F51L).

Keywords CDKN1A, p21^{cip1}, nsSNPs, In silico, Cyclin dependent kinase 1

1. Introduction

CDKN1A or *p21^{cip1} /p21^{WAF1}* negatively regulates cell cycle at G1 check point by binding to proliferating cell nuclear antigen (PCNA) which allows cells to repair damaged DNA. Cyclin dependent kinase inhibitor 1 or *p21^{cip1}* inactivates PCNA. This promotes functions of *p53* (Yates et al., 2015). Dysfunction at G1 checkpoint may give rise to mutation and inappropriate cell proliferation which in turn might induce cancer progression (Bahl et al., 2000; Facher et al., 1997). CDKN1A or *p21^{cip1}* also displays anti-apoptotic roles while inhibiting stress activated protein kinase (SAPK), apoptosis signal-regulating kinase 1 (ASK1), and inhibiting Fas-mediated apoptosis. (Yates et al., 2015). CDKN1A inhibits functional subunits of cyclinA-cdk2, cyclin E-cdk2, cyclin D-cdk4, and cyclin D-cdk6 and is unregulated by *TP53* pathway (Ravitz and Wenner, 1997). It plays central role in cellular growth arrest, terminal differentiation, and apoptosis. Direct regulation of p21 expression occurs by *p53*, and if *p21^{cip1} /p21^{WAF1}* is cleaved by Caspase-3-mediated pathway, apoptosis of cancer cell follows (Ralhan et al., 2000). Mutations in *TP53* is found in association with chronic lymphocytic leukemia (CLL). Dysfunction in the interaction between *TP53* and CDKN1A can cause structural abnormality of *TP53* (Tracy et al., 2017).

Some common sequence variants, such as polymorphism on codon 31 in p21, was found associated with the procession of breast cancer. Mutation in CDKN1A, commonly known as or p21^{cip1} /p21^{WAF1}, was hardly responsible for various types of cancers, despite being a principle downstream regulator of TP53 (Powell et al., 2002). But two common polymorphisms, C→A transversion at codon 31 (Ser→Arg) and C→T transition in the 3'UTR of exon 3, 20bp downstream from strop codon) in CDKN1A or p21^{cip1} /p21^{WAF1} has been suggested to induce development of some types of cancer i.e. esophageal cancer. (Bahl et al., 2000; Facher et al., 1997; Powell et al., 2002; Tracy et al., 2017).

Another polymorphism A→G at codon 149 (Asp→Gly) was observed in esophageal squamous cell carcinomas (ESCCs) in Indian patients and was predicted to interfere with *p53* pathway as well as esophageal tumorigenesis (Bahl et al., 2000; Powell et al., 2002). Polymorphism in CDKN1A was also reported as a risk factor to Alzheimer's disease (AD) (Yates et al., 2015). In silico methods with great accuracy was also followed to screen disease associated nsSNPs in many studies using computational tool such as- PolyPhen 2.0, SIFT, PANTHER, I-mutant 3.0, PhD-SNP, SNP&GO, Pmut, and Mutpred etc. (Ali et al., 2017; Kamaraj and Purohit, 2013; Mathe et al., 2006; Stojiljkovic et al., 2016). Transactivation activity of 1514 missense substitution was analyzed by a Align GV-GD algorithm which scores substitution from C0 (neutral) to C65 (deleterious) (Ali et al., 2017; Mathe et al., 2006). Conservation analysis, based on the evolutionary information, of a protein unfolds its amino acid positions along the chain, which is essential to retain its structural integrity and function (Ashkenazy et al., 2016). Many different computer based algorithms to predict missense variants should be optimized as well as sequence alignment to find the best result (Hicks et al., 2011). Application of many computational analysis was employed in this study to find deleterious nsSNPs of CDKN1A as well as their functional effect on the corresponding protein.

2. Methods

2.1. Collecting SNPs and Protein's Sequence from the databases

Human CDKN1A gene SNPs were collected from the National Center for Biological Information (NCBI) online dbSNP database (https://www.ncbi.nlm.nih.gov/snp). From this server, the non-synonymous SNP (nsSNP) information (Chromosome coordinates and alleles) was retrieved for the computational analysis in this study (Sherry et al., 2001). Protein IDs from the analysis of nsSNPs in SIFT server (http://sift.jcvi.org/) were used to retrieve FASTA sequence of the protein encoded by Human CDKN1A gene from the Ensemble database (https://ensembl.org/). Ensemble is a growing database which has been working for aggregating, processing, integrating and redistributing genomic datasets (Zerbino et al., 2017).

2.2. Deleterious SNPs predicted by SIFT and Polyphen-2

Sorting intolerant from the tolerant (SIFT) server was used to predict the damaging nsSNPs (Kumar et al., 2009). It receives a query sequence and employs multiple alignment to predict either tolerated or damaging substitution in each position of the query sequence (Ng and Henikoff, 2003). nsSNPs with a normalized probability score less than 0.05 is predicted to be harmful and a score greater than 0.05 is considered as tolerated (Ng and Henikoff, 2006). Polyphen-2 server available at (http://genetics.bwh.harvard.edu/pph2/) was utilized, using the data gathered from SIFT result of the nsSNPs, to find either deleterious or neutral amino acid substitutions. By naive Bayesian classifier, it identifies the potential amino acid substitution and the effect of mutation (Adzhubei et al., 2010). It classifies SNPs as "Probably damaging" (0.85 to 1.00), "Possibly damaging' (0.15 to 0.84) and "Benign" (<0.14) (Ali et al., 2017).

2.3. Disease related amino acid substitution by Predict-SNP, MAPP, PhD-SNP, and SNAP

Predict SNP, a consensus classifiers for prediction of disease related amino acid mutation, available at (https://loschmidt.chemi.muni.cz/predictsnp1/) with integrated tools- MAPP, PhD-SNP 2.06 and SNAP 1.1.30, was used to determine the deleterious or neutral nsSNPs (Bendl et al., 2014). Multivariate Analysis of Protein Polymorphism (MAPP) defines deleterious nsSNPs based on the physicochemical variation present in a column of protein sequence akignment (Stone and Sidow, 2005). Predictor of Human Deleterious Single Nucleotide Polymorphism (PhD-SNP) is a Support Vector Machine (SVM) which uses evolutionary information to sort out SNPs related to Mendelian and complex diseases from the neutral ones (Capriotti et al., 2006). SNAP screens out non-acceptable polymorphisms using neural network. It determines 80% of the deleterious substitution at 77% accuracy and 76% of the neutral substitutions at 80% accuracy (Bromberg and Rost, 2007).

2.4. Analysis of protein's stability change upon amino acid substitution

2.4.1. I-Mutant 3.0 server

A Support Vector Machine (SVM) based predictor I-Mutant 3.0, available at (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi, was used in this study to calculate the change in the stability of protein resulting from amino acid substitution (Capriotti et al., 2005). It computes the DDG (kcal/mol) value and RI value (reliability index) of a submitted mutation of a protein.

2.4.2. MUpro server

MUpro server uses both the Support Vector Machine (SVM) and neural network to predict the change in the stability of the protein upon single site mutation. It predicts, with an 84% accuracy, amino acid substitutions responsible for either decreasing or increasing the stability of the related protein via cross validation methods on a large dataset of single amino acid mutations. No tertiary structure of the protein is required for the prediction. (Cheng et al., 2006). MUpro web server available at (http://mupro.proteomics.ics.uci.edu/) was used in this study and a confidence score <0 was predicted to cause decrease in the stability and a score >0 is predicted to cause increase in the stability.

2.5. Structural conformation and conservation analysis by ConSurf sever

Highly conserved functional regions of the protein coded by CDKN1A gene was identified by ConSurf tool (http://consurf.tau.ac.il/). It constructs a phylogenetic tree, following multiple sequence alignment of a query sequence, to run a high-throughput identification process of the functional regions of a protein based on evolutionary data (Ashkenazy et al., 2016).

2.6. Prediction of secondary structure by PSIPRED

Secondary structure of CDKN1A was predicted using PSIPRED server available at (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan et al., 2013). It is based on a two stage neural network with the implementation of position specific scoring matrices constructed from PSI-BLAST to predict the available secondary structures of a protein. (Jones, 1999).

2.7. Homology Modelling

3D structures of the protein encoded by CDKN1A was constructed using four different homology modelling tools as no crystal structure with appropriate length of this protein was available in protein data bank.

2.7.1. Homology modelling by Muster server

MUlti-score ThreadER (MUSTER) algorithm (https://zhanglab.ccmb.med.umich.edu/MUSTER/) , with the implementation of MODELLER v8.2 to construct full length protein model following sequence-template alignment, was used in this study. It uses information such as- sequence profiles, secondary structures, structure fragment profiles, solvent accessibility, dihedral torsion angles, hydrophobic scoring matrix of different sequences to construct the model (Wu and Zhang, 2008). It provides several template based models with a scoring function expressed as Z value to determine the best model.

2.7.2. Homology modelling by Phyre-2 server

Phyre-2 tool (http://www.sbg.bio.ic.ac.uk/phyre2/), based on Hidden Markov method, was used to predict the homology based three dimensional structure of the query amino acid sequence. It combines four steps to build a model: (1) collecting homologous sequence, (2) screening of fold library, (3) loop modelling, (4) multiple template modelling by ab-initio folding simulation-Poing and (4) Side chain placement (Kelley et al., 2015).

2.7.3. Homology modelling by RaptorX server

RaptorX server available at (http://raptorx.uchicago.edu/StructurePrediction/predict/), using RaptorX-Boost and RaptorX-MSA to construct three dimensional structure of a protein, predicted the 3D model of the protein coded by

CDKN1A. It combines a nonlinear scoring function and a probabilistic-consistency algorithm for predicting the model structure (Källberg et al., 2012).

2.7.4. Homology modelling by Swiss-Model server

Swiss-Model workplace available at (https://swissmodel.expasy.org/), a web based tool for the homology modelling of protein, was used to avail the three dimensional structure of CDKN1A (Arnold et al., 2006). The accuracy of the constructed model is calculated by CAMEO system. Swiss-model is an automated tool based on evolutionary information which searches for the best sequence-template alignment from its high-throughput template library (SMTL) to build the model (Biasini et al., 2014).

2.8. Model Refinement, energy minimization and mutation

High resolution structure refinement of the protein model at the atomic level was carried out by ModRefiner tool (Xu and Zhang, 2011) and energy minimization executed for the improvement of the quality of models with the GROMOS 96 forcefield (van Gunsteren, 1996) implementation of DeepView v.4.1.0 (swiss pdb viewer) (Guex and Peitsch, 1997). Execution of energy minimization was done in vacao without any reaction field. Swiss PdbViewer v.4.1.0 was also implemented to build mutated models by browsing a rotamer library of the protein model. Energy minimization of mutant models was brought about to increase the quality.

2.9. Visualization of different models by BIOVIA Discovery Studio Visualizer 2016

Discovery Studio Visualizer (*Dassault Systèmes*, 2016) tool was used to display the 3D structures of the constructed models following structure refinement and energy minimization.

2.10. Validation of models

2.10.1. Ramachandran plot analysis

Ramachandran plot analysis of the protein models, available at ((http://www-cryst.bioc.cam.ac.uk/rampage) determined the energetically allowed sites of amino acid residues in a protein structure through the calculation of backbone dihedral angles of ψ against φ of amino acid residues (Lovell et al., 2003).

2.10.2. QMEAN6 Z-value

The degree of nativeness of the predicted models are evaluated by the QMEAN scoring (https://swissmodel.expasy.org/qmean/). The major five geometrical features of the protein is predicted by this composite scoring function (Benkert et al., 2008).

2.11. Screening of most frequent deleterious substitution

IMB SPSS Statistics for Windows v.20 (*IBM Corp.*, 2011) was used to find the most frequent variants having deleterious effect based on the prediction of six evaluation methods- Predict SNP, MAPP, PhD SNP, SNAP, I-Mutant 3.0 and Mupro, following the initial screening by SIFT and Polyphen-2, which predicted 53 nsSNPs as being harmful out of 118.

2.12. Visualization of selected mutations using Mutation 3D server

Mutation 3D server (http://mutation3d.org/), an interactive tool to visualize amino acid substitutions along with their position on functional domains, was employed in this study. The algorithm used here is an approach based on 3D clustering method and was used to predict driver genes in cancer (Meyer et al., 2016).

2.13. Prediction of biophysical activity by Align-GVGD

Align-GVGD predicted the biophysical characteristics of the amino acid substitutions (http://agvgd.hci.utah.edu/). Grantham Variation (GV) and Grantham Deviation (GD) scores (0 to >200) and graded classifiers (C0 to C65) were

calculated to predict either deleterious or neutral variants (Mathe et al., 2006). An extension of Grantham difference to multiple sequence alignments and true simultaneous multiple comparisons, Align-GVGD, predicts missense substitutions (Tavtigian et al., 2006).

2.14. Harmful effects of selected variants by MutPred2 server

MutPred2 (http://mutpred2.mutdb.org/) web server proposed the reasons behind the disease related amino acid substitution at the molecular level (Pejaver et al., 2017). The result generated by this algorithm predicts a general probability score to define a substitution as deleterious or disease related and includes top five molecular features with a P-value. Scores are interpreted in three categories- actionable hypotheses (general probability score, $g > 0.5$ and property score, $p < 0.05$), confident hypotheses ($g > 0.75$ and $p < 0.05$), and very confident hypotheses ($g > 0.75$ and $p < 0.01$) (Li et al., 2009).

3. Results

NCBI dbSNP database contains information on SNPs of various genes. Mining this database on December 11, 2017, revealed 1766 SNPs of CDKN1A gene, of which 1718 were from *Homo sapiens*. Of the 118 CDKN1A nonsynonymous SNPs found in humans, 113 were missense mutation and the rest 5 were nonsense mutation. FASTA format of the protein encoded by CDKN1A gene, retrieved from Ensemble database, is 164 amino acid long.

3.1. Prediction of functional context of nsSNPs

Functional features of the 118 CDKN1A nsSNPs analyzed by SIFT and Polyphen-2 screened out 65 nsSNPs as because they were predicted as "Tolerated" by SIFT and "Benign" by Polyphen-2 server. All 5 nonsense mutations were found to be either tolerated or benign by these servers.

Table 1 represents the final scores of 53 nsSNPs generated by the sequence homology based software SIFT and Polyphen-2. In SIFT, 23 nsSNPs out of 53 had a tolerance index ≤0.05 and predicted as "Damaging". While analyzing with Polyphen-2, two score were calculated-HumDiv and HumVar. Polyphen-2 gives score ranging from 0 (neutral) to 1 (damaging). HumDiv predicted 34 nsSNPs as "Probably Damaging (High Confidence)" and 12 as "Possibly Damaging (Low confidence)" out of 53 nsSNPs. HumVar reported 28 nsSNPs as "Probably Damaging (High Confidence)" and 10 as "Possibly Damaging (Low confidence)" out of 53 nsSNPs.

SNP ID	Protein Id	Allele	Variant	Polyphen-2				SIFT	
				HumDiv	Value	HumVar	Value	SIFT Prediction	Score
rs148679597	ENSP00000244741	A/G	C117Y	Pro. D[a]	0.995	Pro. D	0.976	Damaging	0.05
rs4986867	ENSP00000244741	A/C/T	F63L	Pos. D	0.749	Pos. D[b]	0.492	Tolerated	0.31
rs12721594	ENST00000244741	A/G	R48Q	Pro. D	1	Pro. D	0.961	Damaging	0.03
rs45548832	ENSP00000244741	A/G/T	R67H	Pro. D	0.999	Pro. D	0.927	Tolerated	0.09
	ENST00000244741	A/G/T	R67L	Pro. D	0.968	Pos. D	0.843	Tolerated	0.3
rs114149607	ENSP00000244741	C/T	T80M	B[c]	0.112	B	0.002	Damaging	0.04
rs151314631	ENSP00000244741	C/G	V25L	Pro. D	0.999	Pro. D	0.997	Tolerated	0.22
rs199801759	ENSP00000244741	A/G	Y151C	Pro. D	1	Pro. D	0.997	Damaging	0.04

rs368128681	ENSP00000244741	A/G	G23D	Pro. D	0.998	Pro. D	0.983	Damaging	0
rs369108178	ENSP00000244741	C/T	P134L	B	0.046	B	0.03	Damaging	0.04
rs370291979	ENSP00000244741	A/G	T57A	Pro. D	0.995	Pro. D	0.953	Tolerated	0.56
rs371219528	ENSP00000244741	C/T	S2P	Pos. D	0.838	B	0.446	Tolerated	0.16
rs372390764	ENSP00000244741	C/T	R67C	Pro. D	0.999	Pro. D	0.927	Damaging	0.04
rs372882329	ENSP00000244741	C/T	L113P	Pro. D	0.999	Pro. D	0.987	Tolerated	0.29
rs373450720	ENSP00000244741	C/T	R32C	Pro. D	0.999	Pro. D	0.928	Damaging	0.01
rs374154006	ENSP00000244741	C/G/T	R156W	Pro. D	1	Pro. D	0.998	Damaging	0
rs374965936	ENSP00000244741	C/T	R83W	B	0.001	B	0.001	Damaging	0.02
rs375050346	ENSP00000244741	A/C/G	R20H	Pos. D	0.894	Pos. D	0.621	Tolerated	0.22
rs376481017	ENSP00000244741	A/C/G	D33N	Pro. D	1	Pro. D	0.994	Tolerated	0.38
	ENSP00000244741	A/C/G	D33H	Pro. D	1	Pro. D	0.996	Tolerated	0.2
rs541505866	ENSP00000244741	C/G	D52E	Pro. D	0.998	Pro. D	0.976	Damaging	0.01
rs559151286	ENSP00000244741	C/T	A39V	Pos. D	0.947	Pos. D	0.518	Tolerated	0.25
rs569995916	ENSP00000244741	A/G	G104R	Pos. D	0.952	Pos. D	0.621	Tolerated	0.48
rs746709171	ENSP00000244741	A/C/T	A36E	Pos. D	0.810	Pos. D	0.49	Tolerated	1
	ENSP00000244741	A/C/T	A36V	Pro. D	0.974	Pos. D	0.673	Tolerated	0.26
rs748758330	ENSP00000244741	C/T	R69W	B	0.004	B	0.005	Damaging	0.04
rs748842488	ENSP00000244741	C/T	F51L	Pro. D	0.999	Pro. D	0.998	Damaging	0.01
rs750249677	ENSP00000244741	C/T	R155C	Pro. D	1	Pro. D	0.981	Damaging	0
rs752126483	ENSP00000244741	A/C	D112E	Pos. D	0.884	B	0.433	Tolerated	1
rs752557277	ENSP00000244741	C/T	R9C	B	0.004	B	0.002	Damaging	0.02
rs752792828	ENSP00000244741	C/T	Y77H	Pos. D	0.61	B	0.28	Damaging	0.03
rs753291170	ENSP00000244741	A/G	E56K	Pro. D	0.999	Pro. D	0.926	Tolerated	0.18
rs753529000	ENSP00000244741	C/T	R122C	Pro. D	0.978	B	0.353	Tolerated	0.24
rs756383281	ENSP00000244741	A/C/G	Q29P	Pro. D	0.958	Pos. D	0.802	Damaging	0.01
	ENSP00000244741	A/C/G	Q29R	Pos. D	0.762	Pos. D	0.478	Tolerated	0.08
rs758877354	ENSP00000244741	C/G	G61R	Pro. D	0.999	Pro. D	0.988	Damaging	0
rs760466052	ENSP00000244741	C/T	S98L	Pos. D	0.845	B	0.421	Tolerated	0.3
rs762978904	ENSP00000244741	C/T	R46C	B	0.015	B	0.006	Damaging	0.02

rs763967941	ENSP00000244741	A/G	E28K	Pro. D	1	Pro. D	0.999	Tolerated	0.1
rs765248879	ENSP00000244741	G/T	V25G	Pro. D	1	Pro. D	0.999	Tolerated	0.1
rs765503766	ENSP00000244741	C/G	D33E	Pro. D	0.998	Pro. D	0.968	Tolerated	0.77
rs767974221	ENSP00000244741	G/T	L115R	Pro. D	0.999	Pro. D	0.983	Tolerated	0.52
rs769049147	ENSP00000244741	A/G	G70D	Pro. D	0.986	Pos. D	0.684	Tolerated	0.43
rs769229553	ENSP00000244741	C/T	R19C	Pro. D	1	Pro. D	1	Damaging	0
rs770464093	ENSP00000244741	A/G	G23S	Pro. D	0.999	Pro. D	0.979	Damaging	0.01
rs771753909	ENSP00000244741	C/G	L71V	Pro. D	0.939	B	0.433	Tolerated	0.28
rs772419969	ENSP00000244741	C/T	T57I	Pro. D	1	Pro. D	0.99	Tolerated	0.22
rs778199042	ENSP00000244741	A/G	G61D	Pro. D	0.997	Pro. D	0.977	Tolerated	0.06
rs779063779	ENSP00000244741	A/G	R156Q	Pro. D	1	Pro. D	0.991	Damaging	0.02
rs779250316	ENSP00000244741	A/G	R19H	Pro. D	1	Pro. D	0.999	Damaging	0
rs779717089	ENSP00000244741	C/T	R86W	B	0.035	B	0.005	Damaging	0.01
rs780909704	ENSP00000244741	A/G	A64T	Pos. D	0.898	B	0.207	Tolerated	0.51
rs935843132	ENSP00000244741	A/C	K75Q	Pos. D	0.917	B	0.446	Tolerated	0.55

Table 1. Screening of Deleterious Single Nucleotide Polymorphism (SNP) Predicted by SIFT and Polyphen-2. [a] Probably damaging (more confident); [b] Possibly damaging (less confident); [c] Benign.

3.2. Prediction of deleterious and disease related amino acid substitutions

Predict-SNP is a consensus classifiers for the prediction of disease related mutations. Analysis of 53 of nsSNPs by this server also predicted MAPP, PhD-SNP and SNAP scores of nsSNPs. All results are summarized in

Table *2*. Out of this 53 nsSNPs, 31, 50, 19 and 29 were predicted as "Deleterious" by Predict-SNP, MAPP, PhD-SNP and SNAP respectively.

Variant	Predict-SNP Prediction	Predict-SNP Expected Accuracy	MAPP Prediction	MAPP Expected Accuracy	PhD-SNP Prediction	PhD-SNP Expected Accuracy	SNAP Prediction	SNAP Expected Accuracy
S2P	D[a]	0.549464	D	0.856822	N	0.832104	N[b]	0.768234
R9C	N	0.738345	D	0.409295	N	0.718714	N	0.611679
R19H	D	0.755662	D	0.765367	N	0.552027	D	0.885196
R19C	D	0.869084	D	0.913793	D	0.607981	D	0.869565
R20H	N	0.602564	N	0.742731	N	0.832104	D	0.555519
G23S	D	0.755662	D	0.571214	N	0.552027	D	0.720388

G23D	D	0.869084	D	0.877061	D	0.732603	D	0.885196
V25G	D	0.869084	D	0.81934	D	0.773385	D	0.720388
V25L	D	0.718713	D	0.771193	D	0.577953	D	0.555519
E28K	D	0.755662	D	0.841829	N	0.508246	D	0.622082
Q29R	D	0.755662	D	0.877061	N	0.508246	D	0.555519
Q29P	D	0.869084	D	0.877061	D	0.773385	D	0.805103
R32C	D	0.654946	D	0.783358	N	0.782877	D	0.720388
D33E	N	0.631518	D	0.81934	N	0.782877	N	0.665242
D33H	D	0.755662	D	0.774944	N	0.68184	D	0.720388
D33N	D	0.718713	D	0.765367	N	0.782877	D	0.622082
A36V	N	0.631518	D	0.783358	N	0.782877	N	0.611679
A36E	N	0.736888	D	0.587706	N	0.782877	N	0.708784
A39V	N	0.628313	D	0.877061	N	0.782877	N	0.611679
R46C	N	0.625291	D	0.81934	N	0.660879	N	0.611679
R48Q	D	0.605483	D	0.765367	N	0.58231	D	0.805103
F51L	D	0.869084	D	0.76087	D	0.732603	D	0.848454
D52E	D	0.869084	D	0.81934	D	0.577953	D	0.848454
E56K	D	0.869084	D	0.774944	D	0.67621	D	0.805103
T57I	D	0.869084	D	0.877061	D	0.732603	D	0.555519
T57A	D	0.605483	D	0.877061	D	0.607981	N	0.500147
G61R	D	0.869084	D	0.877061	D	0.607981	D	0.848454
G61D	D	0.869084	D	0.877061	D	0.577953	D	0.720388
F63L	D	0.505959	D	0.76087	D	0.607981	N	0.583594
A64T	N	0.736888	D	0.587706	N	0.832104	N	0.583594
R67H	D	0.654946	D	0.573143	N	0.718714	D	0.622082
R67C	D	0.718713	D	0.783358	N	0.660879	D	0.720388
R67L	N	0.712788	D	0.858321	N	0.552027	N	0.500147
R69W	N	0.738345	D	0.461769	N	0.552027	N	0.500147
G70D	N	0.631518	D	0.656672	D	0.588855	N	0.583594
L71V	N	0.736888	D	0.571214	N	0.782877	N	0.665242
K75Q	N	0.738345	D	0.409295	N	0.718714	N	0.665242
Y77H	D	0.789035	D	0.766117	N	0.446708	D	0.885196
T80M	N	0.631518	N	0.65022	N	0.782877	N	0.500147
R83W	N	0.625291	N	0.642291	N	0.58231	N	0.500147
R86W	N	0.75204	D	0.573143	N	0.508246	N	0.583594
S98L	N	0.602564	D	0.913793	N	0.68184	N	0.500147
G104R	N	0.733469	D	0.913793	N	0.718714	N	0.611679
D112E	N	0.738345	D	0.508996	N	0.832104	N	0.768234
L113P	D	0.505959	D	0.877061	D	0.67621	N	0.665242
L115R	N	0.625291	D	0.877061	D	0.607981	N	0.554397
C117Y	D	0.869084	D	0.771193	D	0.577953	D	0.720388

R122C	N	0.602564	D	0.656672	N	0.660879	D	0.555519
P134L	N	0.602564	D	0.856822	N	0.832104	N	0.583594
Y151C	D	0.869084	D	0.656672	D	0.67621	D	0.720388
R155C	D	0.755662	D	0.856822	N	0.552027	D	0.805103
R156W	D	0.869084	D	0.771193	D	0.588855	D	0.848454
R156Q	D	0.755662	D	0.765367	N	0.660879	D	0.848454

Table 2. Prediction of Disease Related Mutations by Predict-SNP, MAPP, PhD-SNP, and SNAP. [a] Deleterious effect; [b] Neutral effect.

3.3. I-mutant 3.0 and Mupro analysis of protein's stability change upon amino acid substitution

Prediction of the changes in protein's stability upon amino acid substitution performed by I-Mutant 3.0 server and Mupro Server predicted 46 and 41 amino acid substitution, out of 53, respectively, to bring decrease in the stability of the cyclin dependent kinase 1 protein encoded by CDKN1A gene. All Scores from these servers are represented in

Table 3. Stability change in I-Mutant 3.0 was calculated at 25 ºC and pH 7 and expressed as DDG value=DG (new protein)- DG (wild type) in kcal/mol.

SNP ID	Protein Id	Variant	Sign of DDG	DDG value prediction kcal/mol	RI	MUpro Prediction	Confidence score
rs148679597	ENSP00000244741	C117Y	D[a]	-0.34	0	I[b]	0.467551
rs4986867	ENSP00000244741	F63L	D	-0.99	7	I	0.69283285
rs12721594	ENST00000244741	R48Q	D	-0.79	8	D	-0.96731
rs45548832	ENSP00000244741	R67H	D	-1.19	9	D	-1
	ENST00000244741	R67L	D	-0.43	9	D	-0.49746
rs114149607	ENSP00000244741	T80M	D	-0.38	4	I	0.075972
rs151314631	ENSP00000244741	V25L	D	-1	7	D	-0.27178
rs199801759	ENSP00000244741	Y151C	D	-1.33	0	D	-1
rs368128681	ENSP00000244741	G23D	D	-0.93	6	D	-0.43437301
rs369108178	ENSP00000244741	P134L	I	-0.12	2	I	0.491005
rs370291979	ENSP00000244741	T57A	D	-0.93	7	D	-0.68358

rs371219528	ENSP00000244741	S2P	I	-0.15	3	I	0.048219
rs372390764	ENSP00000244741	R67C	D	-0.93	7	D	-1
rs372882329	ENSP00000244741	L113P	D	-1.58	3	D	-0.76282
rs373450720	ENSP00000244741	R32C	D	-0.79	5	D	-1
rs374154006	ENSP00000244741	R156W	D	-0.61	5	D	-0.31493523
rs374965936	ENSP00000244741	R83W	D	-0.26	3	I	0.46858235
rs375050346	ENSP00000244741	R20H	D	-1.42	9	D	-0.86875024
rs376481017	ENSP00000244741	D33N	D	-0.73	0	D	-0.88042
	ENSP00000244741	D33H	D	-0.52	3	D	-1
rs541505866	ENSP00000244741	D52E	I	0.14	6	D	-1
rs559151286	ENSP00000244741	A39V	D	-0.25	4	I	0.959321
rs569995916	ENSP00000244741	G104R	I	-0.27	1	I	0.66691507
rs746709171	ENSP00000244741	A36E	D	-0.34	6	D	-0.19889
	ENSP00000244741	A36V	D	-0.15	4	D	-0.1802
rs748758330	ENSP00000244741	R69W	D	-0.3	7	I	0.050719
rs748842488	ENSP00000244741	F51L	D	-0.89	3	D	-0.68682
rs750249677	ENSP00000244741	R155C	D	-1.23	4	D	-0.73539
rs752126483	ENSP00000244741	D112E	I	0.3	8	D	-0.58613
rs752557277	ENSP00000244741	R9C	D	-0.8	3	D	-0.80301
rs752792828	ENSP00000244741	Y77H	D	-1.08	7	D	-1
rs753291170	ENSP00000244741	E56K	D	-0.82	9	D	-0.25532

rs753529000	ENSP00000244741	R122C	D	-0.85	2	D	-0.04479
rs756383281	ENSP00000244741	Q29P	D	-0.65	4	D	-0.70387
	ENSP00000244741	Q29R	D	-0.31	1	D	-0.28217
rs758877354	ENSP00000244741	G61R	D	-0.4	3	D	-0.75934
rs760466052	ENSP00000244741	S98L	I	0.34	1	I	1
rs762978904	ENSP00000244741	R46C	D	-0.88	5	D	-0.93791
rs763967941	ENSP00000244741	E28K	D	-1.22	9	I	0.377192
rs765248879	ENSP00000244741	V25G	D	-2.43	9	D	-1
rs765503766	ENSP00000244741	D33E	I	-0.01	6	D	-0.59711968
rs767974221	ENSP00000244741	L115R	D	-1.72	7	D	-1
rs769049147	ENSP00000244741	G70D	D	-0.86	5	D	-0.21741
rs769229553	ENSP00000244741	R19C	D	-1.06	5	D	-0.34018
rs770464093	ENSP00000244741	G23S	D	-1.26	8	D	-1
rs771753909	ENSP00000244741	L71V	D	-1.55	8	D	-0.63308
rs772419969	ENSP00000244741	T57I	D	-0.21	0	D	-0.85972
rs778199042	ENSP00000244741	G61D	D	-1.03	7	D	-0.50922
rs779063779	ENSP00000244741	R156Q	D	-1.27	9	D	-0.74046
rs779250316	ENSP00000244741	R19H	D	-1.42	9	I	0.31757804
rs779717089	ENSP00000244741	R86W	D	-0.3	2	D	-0.49632
rs780909704	ENSP00000244741	A64T	D	-0.67	6	D	-0.07652
rs935843132	ENSP00000244741	K75Q	D	-0.27	1	D	-0.20877166

Table 3. Characterization of the Effect of nsSNPs on Protein Stability by I-Mutant 3.0 and MUpro. [a] Decrease the stability of protein structure; [b] Increase the stability of protein structure.

3.4. Screening of most frequent deleterious amino acid substitution

Following the screening of harmful nsSNPs by SIFT and Polyphen-2 server, the rest 53 nsSNPs, predicted to be deleterious, were further analyzed. A distribution of harmful substitution predicted by Predict SNP, MAPP, PhD-SNP, SNAP, I-Mutant 3.0 and Mupro is shown in **Figure 1**. Disease related mutations exposed by Predict SNP, MAPP, PhD-SNP and SNAP predicted 58.5%, 93.3%, 35.8%, and 54.7% nsSNPs, respectively, as deleterious. Analysis by I-Mutant 3.0 and Mupro suggested that 86.8% and 77.4% nsSNPs, respectively, were affected by amino acid substitution and resulted in decrease in stability. Frequency of deleterious nsSNPs in various servers was calculated by IBM SPSS v20 and Microsoft office 2013.

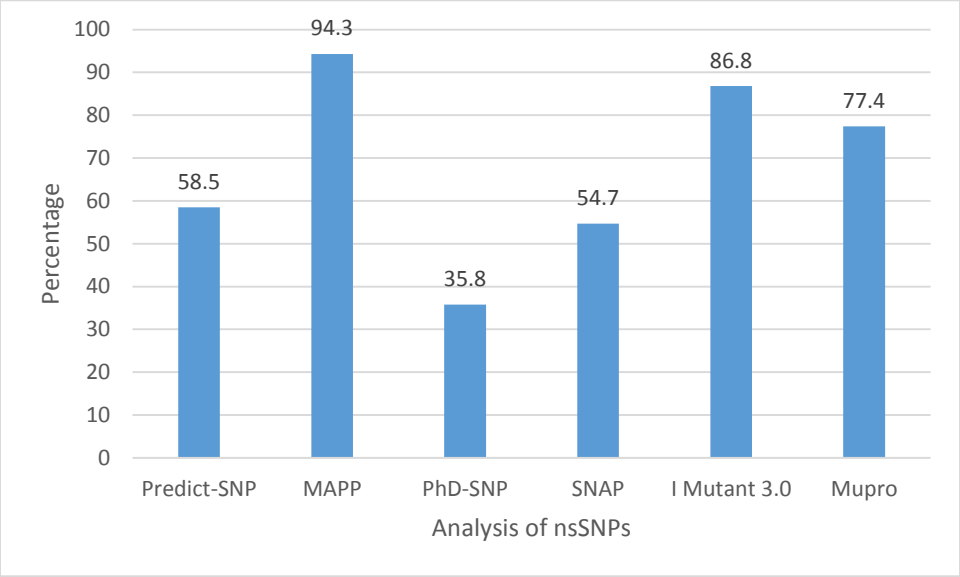

Figure 1. Overview of the analysis of deleterious and stability decreased cyclin dependent kinase inhibitor 1 due to nsSNPs by various in silico tools

Prediction and distribution of deleterious nsSNPs by all six methods- Predict-SNP, MAPP, PhD-SNP, SNAP, I-Mutant 3.0 and Mupro, illustrated in **Figure 2**, revealed the most frequent 12 amino acid substitution predicted to be either "Deleterious" or "Stability Decreasing". Amino acid substitution- R19C, G23D, V25G, V25L, Q29P, F51L, E56K, T57I, G61R, G61D, Y151C and R156W were predicted as most harmful by all the above methods.

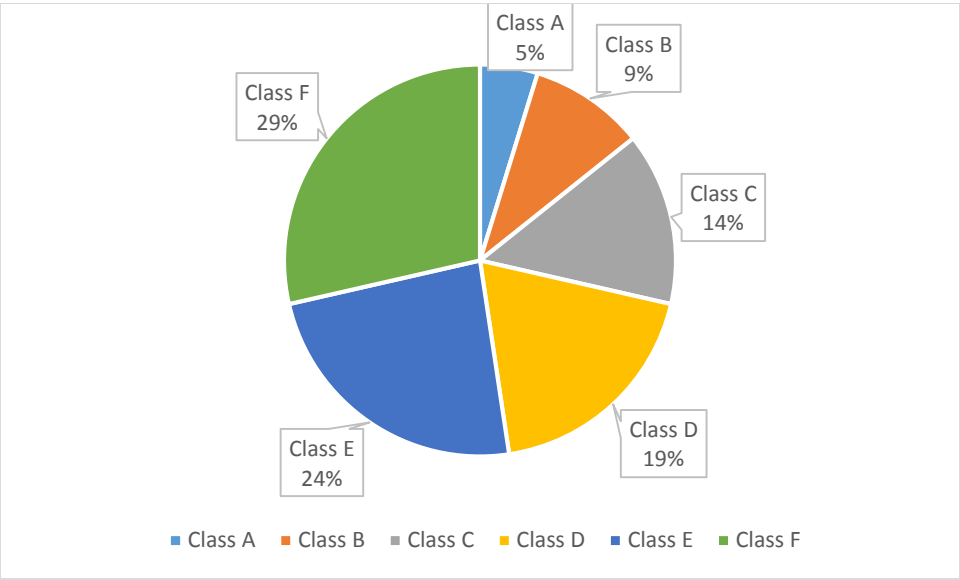

Figure 2. Prediction and distribution of deleterious nsSNPs by Predict-SNP, MAPP, PhD-SNP, SNAP, I-Mutant 3.0 and Mupro methods; Class A, predicted damaging by at least one method, T80M, R83W, S98L, G104R and P134L; Class B, predicted damaging by at least two methods, D33E,A39V, R69W, D112E; Class C, predicted damaging by at least three methods, R9C, R20H, A36V, A36E, R46C, A64T, R67L, L71V, R86W; Class D, predicted damaging by at least four methods, S2P, R19H, E28K, F63L, G70D, K75Q, L115R, R122C; Class E, predicted damaging by at least five methods, G23S, Q29R, R32C, D33H, D33N, R48Q, D52E, T57A, R67H, R67C, Y77H, L113P, C117Y, R155C, R156Q; Class F, predicted damaging by all six methods, R19C, G23D, V25G, V25L, Q29P, F51L, E56K, T57I, G61R, G61D, Y151C, R156W.

3.5. Structural conformation and conservation analysis by ConSurf sever

A colorimetric conservation score was produced as result by the ConSurf server (**Figure 3**). Highly conserved functional region of the protein was revealed by ConSurf tool. It was found that R19C, G23D, F51L, Y151C and R156W have a conservation score of 9; V25G, V25L, Q29P, T57I, G61R and G61D have a conservation score of 8; E56K have a conservation score of 7.

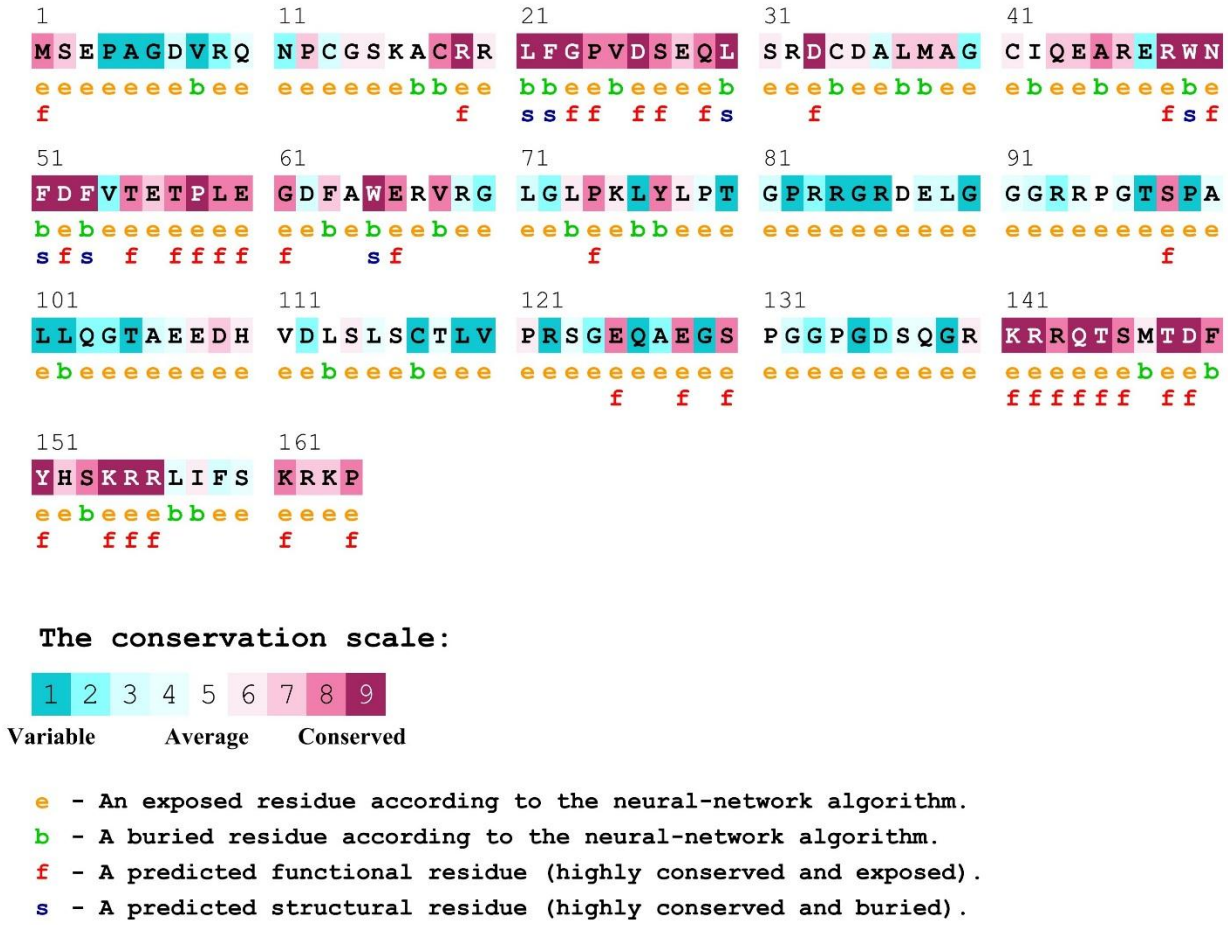

Figure 3. Prediction of evolutionary conserved amino acid residues by ConSurf server. Conservation score is represented as the color coding bars.

3.6. Prediction of secondary structure by PSIPRED

Distribution of alpha helix and beta sheet and coil is exposed by PSIPRED (**Figure 4**). Among the secondary structures, the highest in percentage was coils (82.93%) followed by alpha helix (14.02%) and beta sheet (3.05%).

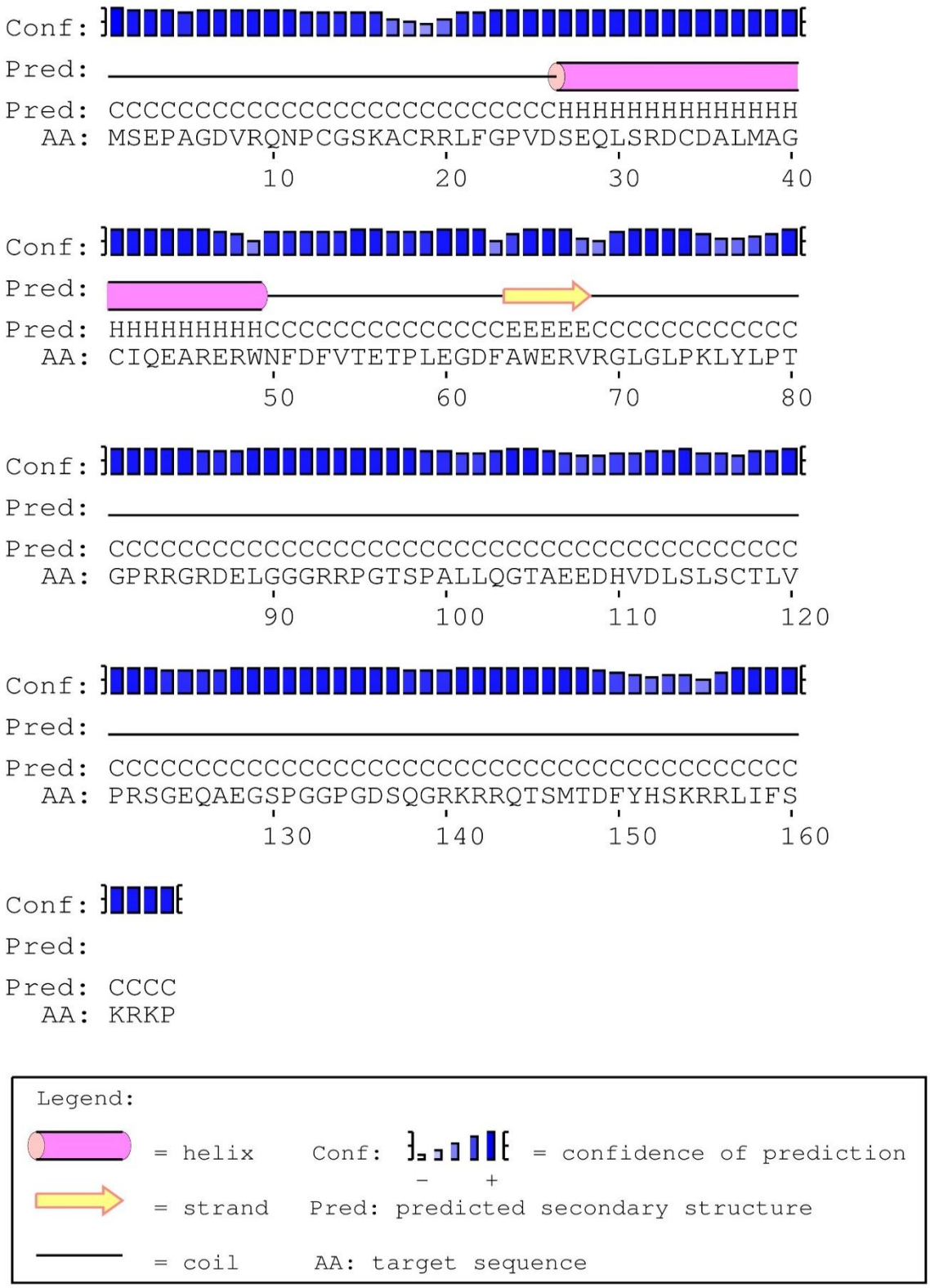

Figure 4. Prediction of secondary structure of by PSIPRED server

3.7. Homology Modelling

Structure refined and energy minimized homology models of CDKN1A by MUSTER, Phyre-2, RaptorX and Swiss model serves are illustrated using computational program Discovery Studio Modeling Environment 2016 (**Figure 5**). These models were checked for validation through Ramachandran plot analysis and QMEAN Z scores.

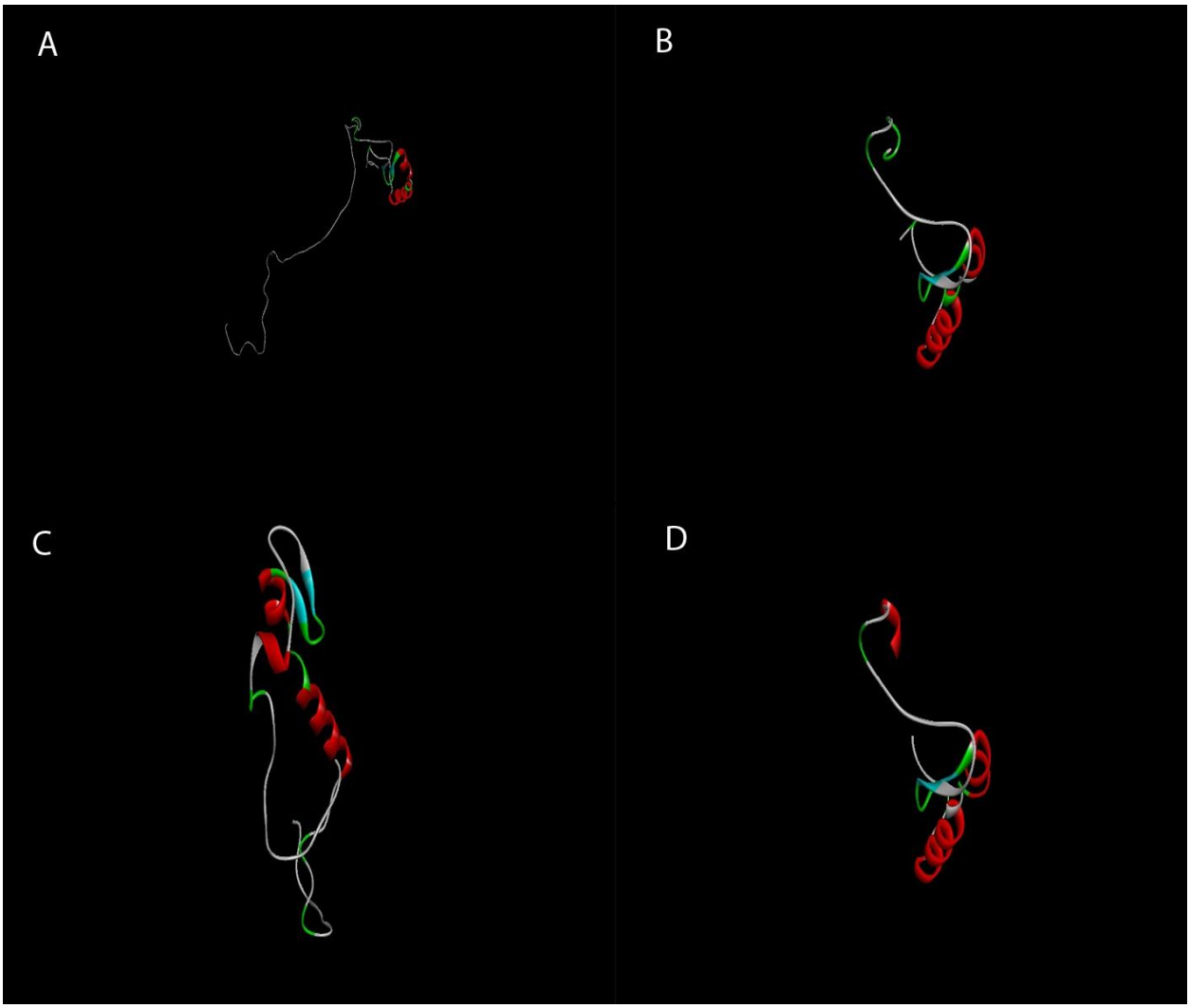

Figure 5. Homology models from different servers; A. CDKN1A_Muster, B. CDKN1A_Phyre2, C. CDKN1A_RaptorX, D. CDKN1A_Swiss model

3.7.1. Ramachandran plot analysis

Ramachandran plot is an x-y plot of phi/psi dihedral angles between NC-alpha and Calpha-C bonds to evaluate a protein's backbone conformation. Ramachandran plots of structure refined and energy minimized homology models are illustrated in **Figure 6**.

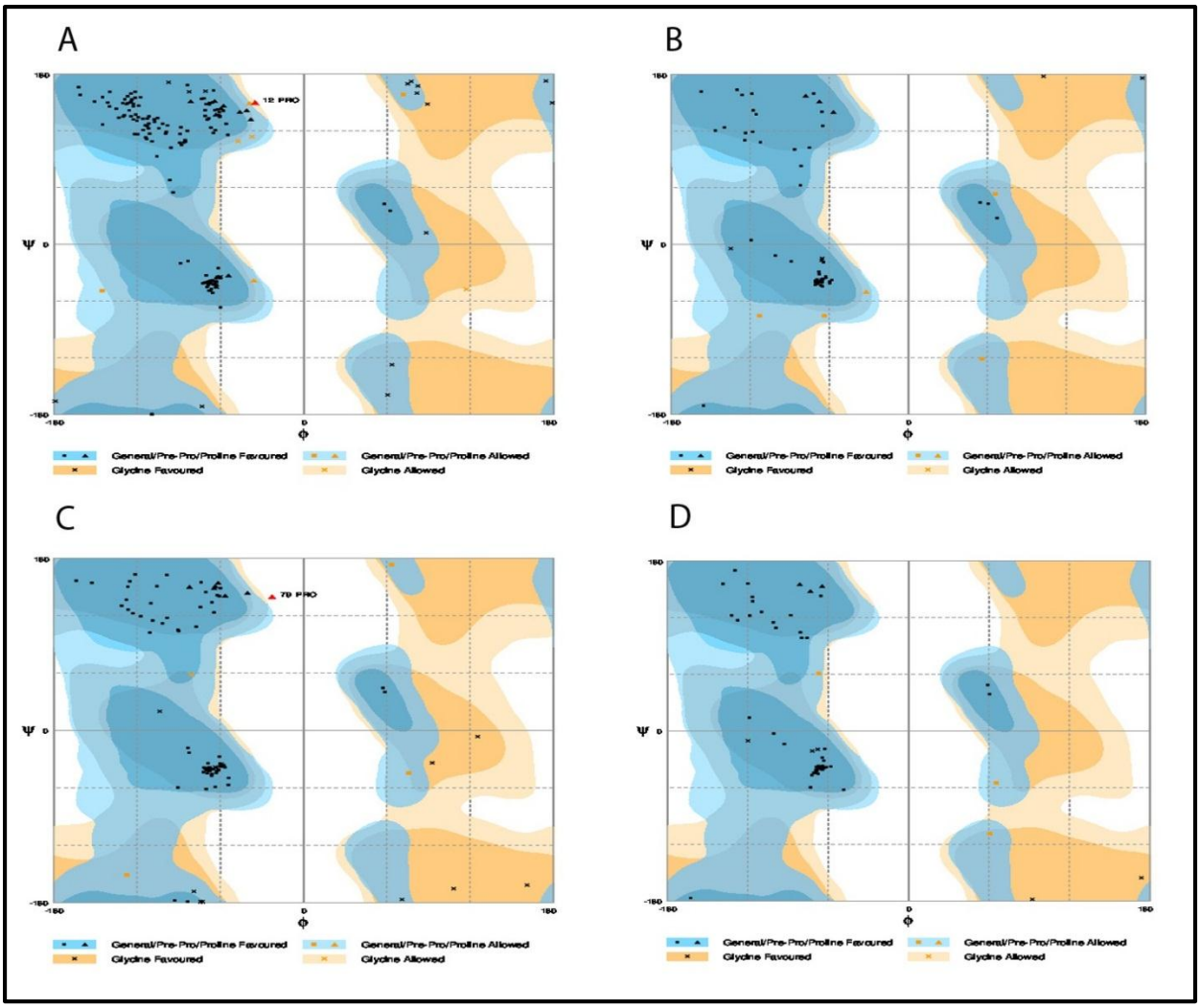

Figure 6. Ramachandran plots of different models; A. CDKN1A_Muster, B. CDKN1A_Phyre2, C. CDKN1A_RaptorX, D. CDKN1A_Swiss model

Table 4 represents the number of residues of different models (structure refined and energy minimized) in favored, allowed and outlier regions generated by the Ramachandran plot. Model generated by MUSTER had highest number of residues (95.1%) in favored region followed by Swiss model (95%), RaptorX (94.5%) and Phyre-2 (92.3%) server.

Model	**Residues in favored region**		**Residues in allowed region**		**Residues in outlier region**	
	No of residues	% of residues	No of residues	% of residues	No of residues	% of residues
CDKN1A_muster	154	95.1	7	4.3	1	0.6
CDKN1A_phyre 2	60	92.3	5	7.7	0	0
CDKN1A_raptor X	86	94.5	4	4.4	1	1.1
CDKN1A_swiss model	57	95.0	3	5.0	0	0

Table 4. Ramachandran plot analysis of different homology models of CDKN1A

3.7.2. QMEAN6 Z-value

Table 5 shows the Z scores obtained from QMEAN Server (Swiss Model) using structure refined and energy minimized models. Overall, the total QMEAN score (Z value) of the models were better for RaptorX (-1.68), followed by Swiss model (-2.01), Phyre-2 (-2.12) and MUSTER (-2.30).

Models	All atom pair wise energy	C beta interaction energy	Solvation energy	Torsion angle energy	Secondary Structure agreement	ACC Agreement	Total QMEAN score
CDKN1A_muster	-1.71	-0.61	-1.77	-1.48	-1.50	-0.18	-2.30
CDKN1A_phyre 2	0.62	-1.22	-1.55	-0.54	-1.83	-1.13	-2.12
CDKN1A_raptor X	-0.29	-1.14	-1.06	-0.94	-1.19	0.20	-1.68
CDKN1A_swiss model	0.53	0.05	-1.35	-1.02	-1.57	-1.36	-2.01

Table 5. Z scores obtained from QMEAN Server using energy minimized models

3.7.3. Z scores of MUSTER generated models

Ten corresponding template was constructed by MUSTER, but only one template was considered as good (Z score >7.5). The rest had a Z score <7.5, thus was considered as bad (

Table 6).

Rank	Template	Align length	Coverage	Z-score	Seq_id	Type
1	1jsuC	69	0.420	9.345	0.391	Good
2	3jbmA	153	0.932	5.296	0.124	Bad
3	2mxxA	155	0.945	4.759	0.116	Bad
4	2muxA	137	0.835	4.705	0.095	Bad
5	5n5xA2	162	0.987	4.623	0.111	Bad
6	4l8nA2	160	0.975	4.564	0.106	Bad
7	4fmyA	162	0.987	4.426	0.086	Bad
8	1su3B3	147	0.896	4.424	0.095	Bad
9	4uisA1	140	0.853	4.331	0.093	Bad
10	3jb8A	161	0.981	4.318	0.112	Bad

Table 6. Z scores of different models generated by MUSTER

3.8. Visualization of selected mutations using Mutation 3D server

Mutation 3D server illustrated selected harmful substitutions (**Figure 7**). Only one domain, CDI (PF02234), of the 164 amino acids in long cyclin dependent kinase inhibitor 1 protein (encoded by CDKN1A gene), was revealed. Among 12 mutations, 3 mutations were uncovered (R19C, Y151C, R156W), the rest 9 mutations (G23D, V25G, V25L, Q29P, F51L, E56K, T57I, G61R, and G61D) were in the domain region, thus considered as higher risk mutation for the cyclin dependent kinase inhibitor 1 protein.

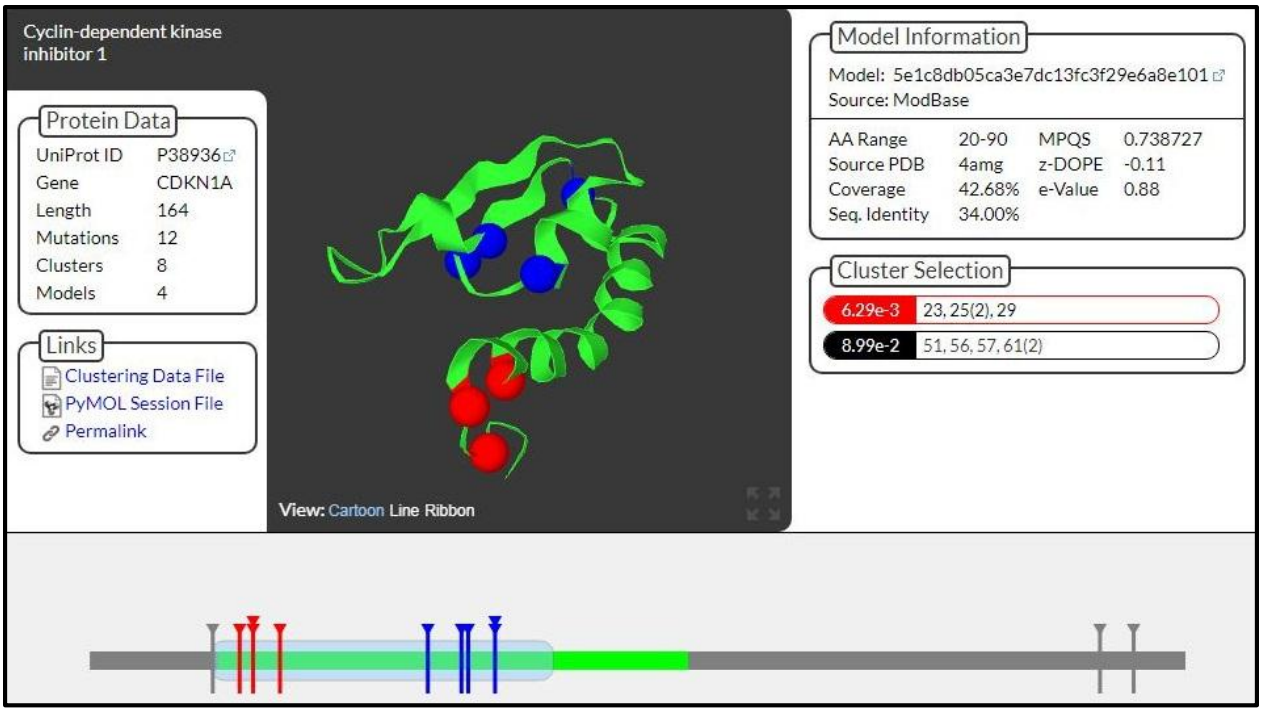

Figure 7. 3D structure of cyclin dependent kinase inhibitor 1 protein, generated by mutation 3D server, and representation of selected mutations on protein's domain

3.9. Comparison of total energy and electrostatic constraints between energy minimized native and mutated models

Total energy and electrostatic constraint (expresses as kJ/mol), after energy minimization by GROMOS 96 implementation of Swiss-PdbViewer, of the native (Swiss Model) structure and the mutant modeled structures are shown in **Table 7**. Out of 12 amino acid substitutions, G23D, V25L, F51L, E56K, T57I, G61R, G61D were found to have decrease in both total energy and electrostatic constraint in comparison with the native structure.

Amino acid variants	Total energy after minimization (kJ/mol)	Electrostatic constraint (kJ/mol)
Native	-2257.007	-1918.52
R19C	-2232.437	-1700.97
G23D	-2502.866	-2030.48
V25G	-2411.375	-1883.49
V25L	-2473.934	-1956.50
Q29P	-2278.592	-1798.33
F51L	-2449.725	-1954.00
E56K	-2491.817	-1951.96
T57I	-2489.148	-1963.93
G61R	-2813.512	-2290.11
G61D	-2542.411	-2025.01

Table 7. Total energy and electrostatic constraint after energy minimization in native and mutated models

3.10. Biophysical analysis of selected amino acid substitutions

Functional effect of missense amino acid substitution, analyzed by Align GVGD, is represented in

Table 8. All 12 nsSNPs existed at strongly conserved residues (GV=0). Nine variants, R19C, G23D, V25G, Q29P, T57I, G61R, G61D, Y151C, R156W, were predicted to be in the class C65. Thus, they were inferred as most likely to interfere with the function. One variant, E56K, was predicted to be in class C55. Thus, it was defined as interfering with function. The additional variants, F51L and V25L, were predicted as less likely to interfere with the function.

Amino acid variants	**A-GVDV**		
	GV	**GD**	**Prediction class**
R19C	0.00	179.53	Class C65
G23D	0.00	93.77	Class C65
V25G	0.00	108.79	Class C65
V25L	0.00	31.78	Class C25
Q29P	0.00	75.14	Class C65
F51L	0.00	21.82	Class C15
E56K	0.00	56.87	Class C55
T57I	0.00	89.28	Class C65
G61R	0.00	125.13	Class C65
G61D	0.00	93.77	Class C65
R156W	0.00	101.29	Class C65
Y151C	0.00	193.72	Class C65

Table 8. Analysis of the functional effect upon nonsynonymous mutation by Align GVGD

Classifiers according to A-GVGD, ordered from most likely to interfere with function to least likely:

GD>=65+Tan(10)x(GV^2.5) => Class C65 <=> most likely

GD>=55+Tan(10)x(GV^2.0) => Class C55

GD>=45+Tan(15)x(GV^1.7) => Class C45

GD>=35+Tan(50)x(GV^1.1) => Class C35

GD>=25+Tan(55)x(GV^0.95) => Class C25

GD>=15+Tan(75)x(GV^0.6) => Class C15

Else (GD<15+Tan(75)x(GV^0.6)) => Class C0 <=> less likely

3.11. Harmful effects of selected variants by MutPred2 server

MutPred2 tool takes the physio-chemical properties in consideration to predict the degree of tolerance for nsSNPs. The summarized results from MutPred server for the selected 12 amino acid substitution is represented in **Table 9**. Gain of intrinsic disorder (P= 0.0065) was predicted at V25G; Amino acid substitution at F51L and E56K was predicted to have a loss of proteolytic cleavage at D52 (P=0.0095); at the substitution G61D, altered metal binding (P=0.0055) was predicted. Amino acid substitution at Y151C was predicted to have loss of phosphorylation at Y151 (P=0.005) along with altered disordered interface (P=0.0019). Loss of acetylation at K161 (P=0.0029) and gain of strand (P=0.0076) were predicted at R156W.

Amino acid variants	**Probability of deleterious amino acid substitution**	**Top 5 features**
R19C	0.787	Altered disordered interface(P=0.01)
G23D	0.667	Altered disordered interface (P=0.03) Gain of intrinsic disorder(P=0.05) Gain of proteolytic cleavage at D26(P=0.04)
V25G	0.711	Gain of intrinsic disorder (P= 0.0065)

		Altered disordered interface (P=0.03) Gain of B-factor (P=0.02)
V25L	0.352	N/A
Q29P	0.627	N/A
F51L	0.642	Altered metal binding (P=0.03) Loss of proteolytic cleavage at D52 (P=0.0095)
E56K	0.527	Loss of loop (P=0.02) Altered metal binding (P=0.01) Loss of proteolytic cleavage at D52 (P=0.0095)
T57I	0.473	N/A
G61R	0.640	Altered metal binding (P=0.03)
G61D	0.690	Altered metal binding (P=0.0055)
Y151C	0.625	Loss of phosphorylation at Y151(P=0.005) Altered disordered interface(P=0.0019) Loss of acetylation at K154(P=0.03) Loss of proteolytic cleavage at D149(P=0.01)
R156W	0.619	Altered disordered interface(P=0.01) Loss of acetylation at K161(P=0.0029) Loss of phosphorylation at Y151(P=0.01) Gain of strand(P=0.0076) Loss of proteolytic cleavage at R156(P=0.03) Altered metal binding (P=0.05)

Table 9. Prediction of the effect of nsSNPs by MutPred server

4. Discussion

Non synonymous single nucleotide polymorphism in genes pose a profound impact on the functionality of a protein. The determination of the phenotypic features attributed to nsSNPs by in silico tools may bring the susceptible variants interfering with the natural function. CDKN1A gene, located chromosome 6, codes for the protein $p21^{Cip1}$, also known as cyclin dependent kinase inhibitor 1 protein. This protein inhibits CDK2 and halts cell cycle progression. It acts as a genome guardian by triggering apoptosis upon DNA damage or stress (Georgakilas et al., 2017); this function might become the opposite if mutation in CDKN1A arises (Liu and Kwiatkowski, 2015).

In this in silico study, following the final screening using the SIFT and Polyphen-2 tool, 53 missense substitutions, out of 118 nsSNPs, was predicted to be harmful. In a study to evaluate the effect of missense variants, SIFT, Polyphen-2 and SNAP tool gave better result with higher accuracy than the other to identify deleterious nsSNPs (Thusberg and Vihinen, 2009). Of these 53 nsSNPs of this analytical study, SIFT predicted 43.39% (n=23) substitutions as damaging where Polyphen-2 predicted 64.15% (34) and 52.83% (28) nsSNPs as deleterious according to HumDiv and HumVar scores, respectively. Non synonymous SNPs associated with disease were highest according to MAPP (50), followed by Predict SNP (31), SNAP (29) and PhD SNP (19). Changes in the stability upon mutation was highest according to I-Mutant 3.0 tool (46) and MUpro predicted 41 nsSNPs, out of 53 as responsible to decrease the stability of the protein. I-Mutant 3.0 was found to be the most reliable predictor to identify the changes in stability upon mutation (Khan and Vihinen, 2010). Among these 53 substitutions- R19C (C→T, rs769229553), G23D (A→G, rs368128681), V25G (G→T, rs765248879), V25L (C→G, rs151314631), Q29P (A→C→G, rs756383281), F51L (C→T, rs748842488), E56K (A→G, rs753291170), T57I (C→T, rs772419969), G61R (C→G, rs758877354), G61D (A→G, rs778199042), Y151C (A→G, rs199801759) and R156W (C→G→T, rs374154006) were predicted as both deleterious and responsible for decreasing stability of the protein by all above methods. Thus, further analysis on their effect on the functionality was executed. A similar study based on the dbSNP was conducted to find the deleterious SNPs of

BARD1 gene (Alshatwi et al., 2012). Homology model generated by the Swiss-model was taken as native model, considering its overall validation score. Homology modelling has been one of the most promising method to build reliable structures (Bordoli et al., 2008). Following energy minimization, mutant model R19C and V25G presented an increase (less favorable) in the total energy count (kcal/mol) comparing to the native model. Total scores for Y151C and R156W could not be computed due to the limitation in the amino acid length of the model structure (62 amino acid). Only three (R19C, Y151C and R156W) out of these twelve substitutions was uncovered and occurred out of the CDI (PF02234) domain predicted by Mutation 3D algorithm. Domains are the functionally active site of a protein structure and mutations on this site may have a tremendous effect on its activity (Yang et al., 2015). Biophysical analysis by Align-GVGD proposed only two substitutions (V25L and F51L) as less confidently interfering with the function of the protein. More importantly, among these 12 highly damaging mutations- R19C, G23D, F51L, Y151C and R156W occurred at the most conserved site, based on evolutionary history, on the amino acid sequence of CDKN1A, which are predicted to be the positions interfering mostly with the function of the protein. Some molecular features of some selected substitutions revealed by MutPred2 application predicted them as being actionable or confident hypotheses that they might be deleterious or associated with disease. MutPred2 general probability of deleterious mutation score for R19C, G23D, V25G, F51L, E56K, G61R, G61D, Y151C, and R156W was 0.787, 0.667, 0.711, 0.642, 0.527, 0.640, 0.690, 0.625 and 0.619 respectively. The effect of missense mutation at the molecular level was also predicted by in silico analysis in Chinese patients having congenital dyserythropoietic anaemia (CDA) (Wang et al., 2018)

5. Conclusion

A simulation based study employed in this study to detect the deleterious nsSNPs using many computational program has revealed, 12 amino acid substitutions (R19C, G23D, V25G, V25L, Q29P, F51L, E56K, T57I, G61R, G61D, Y151C and R156W) of CDKN1A which might be responsible for the functional discrepancies comparing to the novel wild CDKN1A gene having a major role in the regulation of cell cycle progression at G1 phase. Majority of the variants identified in this study are located on domain site and conserved region of cyclin dependent kinase inhibitor 1. In situ evaluation of this gene, and its protein cyclin dependent kinase 1 in this study, which predicted some novel deleterious non-synonymous single nucleotide polymorphisms along with their functional sites and molecular features might come in help for further in vitro study.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICTS OF INTEREST

This is to certify that, from the authors, there is no conflict of interests regarding the publication of this paper.

REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249. https://doi.org/10.1038/nmeth0410-248

Ali, S.K., Sneha, P., Priyadharshini Christy, J., Zayed, H., George Priya Doss, C., 2017. Molecular dynamics-based analyses of the structural instability and secondary structure of the fibrinogen gamma chain protein with the D356V mutation. J. Biomol. Struct. Dyn. 35, 2714–2724.

Alshatwi, A.A., Hasan, T.N., Syed, N.A., Shafi, G., Grace, B.L., 2012. Identification of Functional SNPs in BARD1 Gene and In Silico Analysis of Damaging SNPs: Based on Data Procured from dbSNP Database. PLOS ONE 7, e43939. https://doi.org/10.1371/journal.pone.0043939

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinforma. Oxf. Engl. 22, 195–201. https://doi.org/10.1093/bioinformatics/bti770

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350. https://doi.org/10.1093/nar/gkw408

Bahl, R., Arora, S., Nath, N., Mathur, M., Shukla, N.K., Ralhan, R., 2000. Novel polymorphism in p21waf1/cip1 cyclin dependent kinase inhibitor gene: association with human esophageal cancer. Oncogene 19, 323.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J., Damborsky, J., 2014. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. PLoS Comput. Biol. 10. https://doi.org/10.1371/journal.pcbi.1003440

Benkert, P., Tosatto, S.C.E., Schomburg, D., 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins 71, 261–277. https://doi.org/10.1002/prot.21715

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252-258. https://doi.org/10.1093/nar/gku340

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., Schwede, T., 2008. Protein structure homology modeling using SWISS-MODEL workspace. Nat. Protoc. 4, 1. https://doi.org/10.1038/nprot.2008.197

Bromberg, Y., Rost, B., 2007. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. 35, 3823–3835. https://doi.org/10.1093/nar/gkm238

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., Jones, D.T., 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 41, W349–W357. https://doi.org/10.1093/nar/gkt381

Capriotti, E., Calabrese, R., Casadio, R., 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinforma. Oxf. Engl. 22, 2729–2734. https://doi.org/10.1093/bioinformatics/btl423

Capriotti, E., Fariselli, P., Casadio, R., 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33, W306-310. https://doi.org/10.1093/nar/gki375

Cheng, J., Randall, A., Baldi, P., 2006. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62, 1125–1132. https://doi.org/10.1002/prot.20810

Dassault Systèmes, 2016. , Discovery Studio Modeling Environment. Dassault Systèmes, San Diego.

Facher, E.A., Becich, M.J., Deka, A., Law, J.C., 1997. Association between human cancer and two polymorphisms occurring together in the p21Waf1/Cip1 cyclin-dependent kinase inhibitor gene. Cancer 79, 2424–2429.

Georgakilas, A.G., Martin, O.A., Bonner, W.M., 2017. p21: A Two-Faced Genome Guardian. Trends Mol. Med. 23, 310–319. https://doi.org/10.1016/j.molmed.2017.02.001

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723. https://doi.org/10.1002/elps.1150181505

Hicks, S., Wheeler, D.A., Plon, S.E., Kimmel, M., 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Hum. Mutat. 32, 661–668.

IBM Corp., 2011. , IBM SPSS Statistics for Windows. IBM Corp., Armonk, NY.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202. https://doi.org/10.1006/jmbi.1999.3091

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., Xu, J., 2012. Template-based protein structure modeling using the RaptorX web server. Nat. Protoc. 7, 1511–1522. https://doi.org/10.1038/nprot.2012.085

Kamaraj, B., Purohit, R., 2013. In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. BioMed Res. Int. 2013.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845. https://doi.org/10.1038/nprot.2015.053

Khan, S., Vihinen, M., 2010. Performance of protein stability predictors. Hum. Mutat. 31, 675–684. https://doi.org/10.1002/humu.21242

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081. https://doi.org/10.1038/nprot.2009.86

Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D.N., Mooney, S.D., Radivojac, P., 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinforma. Oxf. Engl. 25, 2744–2750. https://doi.org/10.1093/bioinformatics/btp528

Liu, Y., Kwiatkowski, D.J., 2015. Combined CDKN1A/TP53 mutation in bladder cancer is a therapeutic target. Mol. Cancer Ther. 14, 174–182. https://doi.org/10.1158/1535-7163.MCT-14-0622-T

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins Struct. Funct. Bioinforma. 50, 437–450. https://doi.org/10.1002/prot.10286

Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P., Tavtigian, S.V., 2006. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic Acids Res. 34, 1317–1325. https://doi.org/10.1093/nar/gkj518

Meyer, M.J., Lapcevic, R., Romero, A.E., Yoon, M., Das, J., Beltrán, J.F., Mort, M., Stenson, P.D., Cooper, D.N., Paccanaro, A., Yu, H., 2016. mutation3D: Cancer Gene Prediction Through Atomic Clustering of Coding Variants in the Structural Proteome. Hum. Mutat. 37, 447–456. https://doi.org/10.1002/humu.22963

Ng, P.C., Henikoff, S., 2006. Predicting the effects of amino acid substitutions on protein function. Annu. Rev. Genomics Hum. Genet. 7, 61–80. https://doi.org/10.1146/annurev.genom.7.080505.115630

Ng, P.C., Henikoff, S., 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31, 3812–3814.

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.-J., Mort, M., Cooper, D.N., Sebat, J., Iakoucheva, L.M., Mooney, S.D., Radivojac, P., 2017. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv 134981. https://doi.org/10.1101/134981

Powell, B.L., van Staveren, I.L., Roosken, P., Grieu, F., Berns, E.M., Iacopetta, B., 2002. Associations between common polymorphisms in TP53 and p21WAF1/Cip1 and phenotypic features of breast cancer. Carcinogenesis 23, 311–315.

Ralhan, R., Agarwal, S., Mathur, M., Wasylyk, B., Srivastava, A., 2000. Association between polymorphism in p21Waf1/Cip1 cyclin-dependent kinase inhibitor gene and human oral cancer. Clin. Cancer Res. 6, 2440–2447.

Ravitz, M.J., Wenner, C.E., 1997. Cyclin-dependent kinase regulation during G1 phase and cell cycle regulation by TGF-beta. Adv. Cancer Res. 71, 165–207.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311.

Stojiljkovic, M., Klaassen, K., Djordjevic, M., Sarajlija, A., Brasil, S., Kecman, B., Grkovic, S., Kostic, J., Rodriguez-Pombo, P., Desviat, L.R., 2016. Molecular and phenotypic characteristics of seven novel mutations causing branched-chain organic acidurias. Clin. Genet. 90, 252–257.

Stone, E.A., Sidow, A., 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res. 15, 978–986. https://doi.org/10.1101/gr.3804205

Tavtigian, S.V., Deffenbaugh, A.M., Yin, L., Judkins, T., Scholl, T., Samollow, P.B., de Silva, D., Zharkikh, A., Thomas, A., 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J. Med. Genet. 43, 295–305. https://doi.org/10.1136/jmg.2005.033878

Thusberg, J., Vihinen, M., 2009. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. Hum. Mutat. 30, 703–714. https://doi.org/10.1002/humu.20938

Tracy, I., Tapper, W., Parker, A., Gardiner, A., Sadullah, S., Pratt, G., Copplestone, A., Oscier, D., Best, O.G., 2017. Type C Tp53-cdkn1a pathway dysfunction occurs independently of cdkn1a gene polymorphisms in chronic lymphocytic leukaemia and is associated with tp53 abnormalities. Br. J. Haematol. 178, 824–826.

van Gunsteren, W.F., 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Biomos ; Zürich.

Wang, Y., Ru, Y., Liu, G., Dong, S., Li, Y., Zhu, X., Zhang, F., Chang, Y.-Z., Nie, G., 2018. Identification of CDAN1, C15ORF41 and SEC23B mutations in Chinese patients affected by congenital dyserythropoietic anemia. Gene 640, 73–78. https://doi.org/10.1016/j.gene.2017.10.027

Wu, S., Zhang, Y., 2008. MUSTER: Improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins 72, 547–556. https://doi.org/10.1002/prot.21945

Xu, D., Zhang, Y., 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J. 101, 2525–2534. https://doi.org/10.1016/j.bpj.2011.10.024

Yang, F., Petsalaki, E., Rolland, T., Hill, D.E., Vidal, M., Roth, F.P., 2015. Protein Domain-Level Landscape of Cancer-Type-Specific Somatic Mutations. PLOS Comput. Biol. 11, e1004147. https://doi.org/10.1371/journal.pcbi.1004147

Yates, S.C., Zafar, A., Rabai, E.M., Foxall, J.B., Nagy, S., Morrison, K.E., Clarke, C., Esiri, M.M., Christie, S., Smith, A.D., 2015. The effects of two polymorphisms on p21cip1 function and their association with Alzheimer's disease in a population of European descent. PloS One 10, e0114050.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, J.K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A., Flicek, P., 2017. Ensembl 2018. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx1098