Bioresearch Communications
Volume 06, Issue 01, January 2020
Journal Homepage: www.bioresearchcommunications.com

Original Article

High-Risk Non-synonymous SNPs of Human Bcl-2 Gene Alters Structural Stability and Small Molecule Binding

A. S. M. Zisanur Rahman[1,3], Arittra Bhattacharjee[1], Abdullah All Jaber[1], Maqsud Hossain[1,2], Kazi Nadim Hasan[1], Sohidul Islam[1] and Zaied Ahmed Bhuyan[1]*

[1]Department of Biochemistry & Microbiology, North South University, Bashundhara, Dhaka- 1229, Bangladesh. [2]NSU Genome Research Institute, North South University, Bashundhara, Dhaka-1229, Bangladesh. [3]Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada

ABSTRACT: B-cell lymphoma 2 (Bcl-2) gene, which encodes Bcl-2 protein, is vital for apoptotic programmed cell death. In this study, computational approaches were performed to reveal the effect of non-synonymous single nucleotide polymorphisms (nsSNPs) in Bcl-2 gene. A total of 79 nsSNPs were studied. Amongst all, 11 nsSNPs (M166T, G141E, R129C, S105P, F104S, R98L, L97P, H94P, V93A, G27S, V15L) were located at highly conserved region or functionally important domains with a high probability of being deleterious for Bcl-2 structure or role. Except for H94P, rest of the high-risk nsSNPs displayed decreased stability of BcL-2 structure. However, F104S attracted significant attention because it exhibited considerably reduced binding affinity than wild type Bcl-2 protein against HA 14-1 antagonist. This extensive computational experiment will assist as a valuable data source for upcoming population-based studies.

Keywords: Bcl-2 gene, apoptosis, non-synonymous SNPs, HA 14-1

Article History
Received: 10 September 2019
Accepted: 03 December 2019

Scan the QR code to see the online version or,visit- www.bioresearchcommunications.com

Corresponding author
Zaied Ahmed Bhuyan
Email: zaied.bhuyan@northsouth.edu
Phone: +880-2-55668200 Ext-6256

Citation: Rahman, ASMZ., Bhattacharjee, A., Jaber, AA., Hossain, M., Hasan, KN., Islam, S. and Bhuyan, ZA. High-Risk Non-synonymous SNPs of Human Bcl-2 Gene Alters Structural Stability and Small Molecule Binding. 2020. ***Biores Comm.*** 6(1), 791-800

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most common genetic variation in humans and characterized by a change at a single nucleotide base in a particular DNA sequence. SNPs can act as an important biological marker in identifying genes and colligating them with diseases[1]. Even though most of the SNPs do not show any phenotypic effect, SNPs occurring in a gene, especially in the protein-coding region or its regulatory region can play potential role in instigating various diseases [2]. Therefore, SNPs that are located in the protein-coding region of a gene and cause a change in the amino acid, are called non-synonymous SNPs (nsSNPs), which can lead to the alteration in the protein stability or solubility and consequently structure and function of the protein[3, 4].

Bcl-2 (B-cell lymphoma-2) gene which encodes a protein that helps control the apoptosis process, was initially identified to be connected with human follicular lymphoma of B cell and described as an oncogene [5, 6, 7]. However, Bcl-2 was later discovered to have anti-apoptotic activity [8]. Bcl-2 gene, located on the long arm of chromosome 18, have the cytogenetic location 18q21.33 [5]. Bcl-2 gene contains six exons and two promoters (P1 and P2), each having different functions [9]. Although Bcl-2 is located on chromosome 18, the association of Bcl-2 translocation with many leukemia and lymphomas has been established where overexpression of Bcl-2 protein suppresses apoptosis [7]. Bcl-2, a family of protein-containing pro-apoptotic as well as anti-apoptotic members, are central regulators of the caspase activation and play a central role in programmed cell death by governing the

mitochondrial and endoplasmic reticulum membrane integrity [10]. Analysis of truncated Bcl-2 revealed that the amino-terminal of Bcl-2 protein is essential for the formation of heterodimerization with pro-apoptotic Bax protein as well as for conferring anti-apoptotic activity [11]. Hydrophobic residues substitution of BH4 domain located in the amino terminal of Bcl-2 causes the Bcl-2 to lose its ability to form heterodimer with Bax. One particular mutation in the BH4 domain, V15E, was especially informative because it results in the complete loss of ability of Bcl-2 to form a heterodimer with Bax and thus inhibit the pro-apoptotic function of Bax to induce apoptosis [11].

Bcl-2 SNPs can carry out the prospective indication for treatment outcomes in various cancer patients [12, 13]. However, the lack of experimental data has prompted to predict their functional and/or structural effects using various bioinformatics tools. Most of the currently available tools predict either the functional or structural level. Thus, a comprehensive assessment approach was taken for collecting the data from tools that predict the structural and functional effects of SNPs. These tools produce data based on various algorithms, data, and resources [14, 15, 16].

Here, in this study, we analyzed the impact of nsSNPs in Bcl-2 gene using bioinformatics tools to predict their putative function and/or structural changes. SNPs having functional significance were selected and prioritized with the significance of their function, location and the minor allele frequency (MAF). 10 SNPs were identified at conserved regions or functionally important domains with a high probability of being deleterious for Bcl-2 structure or role. This extensive computational experiment will serve as a valuable data source for upcoming population-based studies.

MATERIALS AND METHODS

SNP Mining

The SNPs for the Bcl-2 gene were obtained from dbSNP (https://www.ncbi.nlm.nih.gov/SNP/) and the Ensembl Genome Browser (http://www.ensembl.org/index.html). The Ensembl transcript ID ENST00000398117.1 of the Bcl-2 gene was used to retrieve the variant descriptions, the validation status, allele frequency, and the corresponding genomic coordinates. Using the variant descriptions, filtration nsSNPs were performed. The protein sequence of the Bcl-2 was obtained from UniProt (http://www.uniprot.org/) (UniProt ID P10415).

Assessment of Functional Effect of nsSNPs

Functional effects of nsSNPs were predicted with Sorting Intolerant from Tolerant (SIFT) (http://sift.jcvi.org/), a multistep algorithm, considers the conservation profile for each amino acid residues to predict the consequence of missense mutation on protein function [17]. SIFT score less than 0.05 were considered as harmful mutations. Polymorphism Phenotyping v2 (PolyPhen 2) (http://genetics.bwh.-harvard.edu/pph2/) was applied to predict whether a substitution in amino acid will have an impact in the stability and the function of the protein in question-based on the provided sequence. It calculated and provided the score by aligning the query sequence with homologous proteins [18]. Protein Variation Effect Analyzer (PROVEAN) (http://provean.jcvi.org/seq_submit.php) tool predicted the effect of amino acid substitution on the function of the protein by using pairwise sequence alignment. [19].

Protein Analysis through Evolutionary Relationships (PANTHER) is an online tool that measured the influence of an nsSNP on the function of the protein, using a method called PANTHER-PSEP (position-specific evolutionary preservation) [20]. Multivariate Analysis of Protein Polymorphism (MAPP) (http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html) calculated the impact of an amino acid substitution by considering variation in physiochemical properties including polarity, charge, volume, free energy, hydropathy in helical and sheet conformation to assess the impact of missense mutants on protein function [21]. Subsequently, Screening for Non-Acceptable Polymorphisms(SNAP) (https://www.rostlab.org/services/SNAP/) calculated the impact of nsSNPs on the protein function using neural network-based method[22]. nsSNPAnalyzer 2.0 (http://snpanalyzer.uthsc.edu/) predicted the phenotypic effect that can be resulted from nsSNPs using a method called Random Forest [23].

Prediction of Disease-Associated Non-synonymous SNP (nsSNP)

Additional algorithms were applied to predict disease association by using SNPs&GO, PhD-SNP, and PMut (http://mmb.pcb.ub.es/PMut/) [24, 25].

Identifying SNPs located in evolutionarily conserved regions in the gene

Evolutionary conservation of amino acid residues in Bcl-2 was determined using the ConSurf web server (http://consurf.tau.ac.il/2016/). ConSurf determined the level of conservation by comparing with the homologues for each amino acid residue and based on that, identifies the functionally significant and evolutionary conserved regions in proteins [26]. In ConSurf, Bcl-2 homologues were aligned and position-specific conservation scores (CS) were calculated using an empirical Bayesian algorithm where CS of 1 to 4 are considered variable, 5 to 6 are considered intermediate and 7 to 9 are considered conserved. nsSNPs having high conservation scores were considered high-risk nsSNPs and selected for further analysis.

Prediction of Protein-Protein Interactions

Interaction of Bcl-2 protein was predicted by the STRING database version 10.5. The interactions comprise physical and functional relations derived from prior data, genomic context, high-throughput studies and co-expression [27].

Detection of nsSNPs Location in Protein Structure

Mutation3D (http://mutation3d.org/index.shtml), a tool based on the complete-linkage clustering method, predicted the functional impact of amino acid substitution and visualize the cluster of amino acid substitution on the 3D structure of the protein. Mutation3D clustering has the ability to differentiate between the functional and non-functional mutations. It performs bootstrapping approach to estimate the statistical significance for each cluster in terms of P-value as the percentile rank[28].

Prediction of post-translational modification sites

UbPred (www.ubpred.org) was applied for predicting putative ubiquitylation sites. Sumoylation sites prediction was carried out by pSumo-CD (http://www.jci-bioinfo.cn/pSumo-CD)[29], JASSA (http://www.jassa.fr/index.php?m=doc)[30], and GPS-SUMO (http://sumosp.biocuckoo.org/)[31] programs. Putative phosphorylation sites prediction were performed by GPS 2.1 (http://gps.biocuckoo.org/)[32] and NetPhos 3.1 (http://www.cbs.dtu.dk/services/NetPhos/)[33] programs Methylation sites were predicted using PSSMe (http://bioinfo.ncu.edu.cn/PSSMe.aspx)[34].

Surface accessibility of Native and Mutant Bcl-2 Protein

NetSurfP webserver was utilized to predict the surface accessibility of native and mutant Bcl-2 protein. Protein sequence in FASTA format is the input option for NetSurfP and based on the provided sequence, it predicted the secondary structure and surface accessibility of the protein. It also provided Z-score which determined the reliability of the predicted result [35].

Protein Stability Analysis

The nsSNP induced change in the stability of Bcl-2 protein was predicted using I-mutant 2.0 and MUpro (http://mupro.proteomics.ics.uci.edu/) web servers [24]. I-Mutant calculated Gibbs free energy value (ΔG) for the wild-type protein and subtracted it from that of the mutant protein (ΔΔG = ΔG mutant – ΔG wild-type) to estimate the total free energy change value (ΔΔG). Negative ΔΔG value indicated a decrease in protein stability, and a positive ΔΔG value indicated an increase in protein stability [24, 36]. The pH was set to 7.4 and the temperature was set to 37ºC for all submissions.

MUpro estimated the changes in protein stability due to a single site mutation by the support vector machine (SVM) method. Here, negative score (< 0) for a variant indicated the decrease in stability caused by the variant and a positive score (> 0) for a variant indicated the increase in stability [35].

Prediction of structural effect of point mutation

Project HOPE (http://www.cmbi.ru.nl/hope/home) was used to analyze the effect of the point mutation in the native Bcl-2 by providing and visualizing the 3D structure(s) of the mutant protein(s) [37]. The energy minimization for 3D structures was performed using YASARA (Yet Another Scientific Artificial Reality Application) Server [38]. Tm-Align https://zhanglab.ccmb.med.umich.edu/TM-align/) was used to calculate Tm-scores and root mean square deviation (RMSD) [39].

Binding Sites Prediction and Molecular Docking Simulation

To find whether nsSNPs of Bcl-2 gene are present on any ligand or protein binding region, binding site prediction was executed through FTsite server (http://ftsite.bu.edu/) which uses energy-based method and has a high accuracy rate for the prediction of ligand binding site [40].Here, the native or mutant structures were uploaded as PDB file format.

The comparative molecular docking analysis of native and F104S mutant was carried out for confirmation of binding site alteration using Autodock vina [40,41]. Semiflexible docking protocols were applied for this step where the ligand was kept flexible but the receptor was held rigid. Before docking simulation, a grid box was used around the ligand-binding pocket as previously identified using FTsite [42]. Docking studies were performed using HA14-1 ((PubChem CID: 3549) as ligand. HA14-1 is a Bcl-2 inhibitor [43]. Receptors and ligand were analyzed and prepared for docking analysis using AutoDock Tools integrated into the MGL tools (http://mgltools.scripps.edu). The grid box size was set at X=30, Y=33 and Z=28 axis and spacing between the grid points was set to 0.375 Å. The receptors and ligand were converted in PDBQT format for molecular docking simulation via AutoDock Vina. Based on the lower free energy with lower root mean square score, the best output files were selected and visualized by PyMOL Molecular Visualization Software [44].

RESULTS AND DISCUSSION

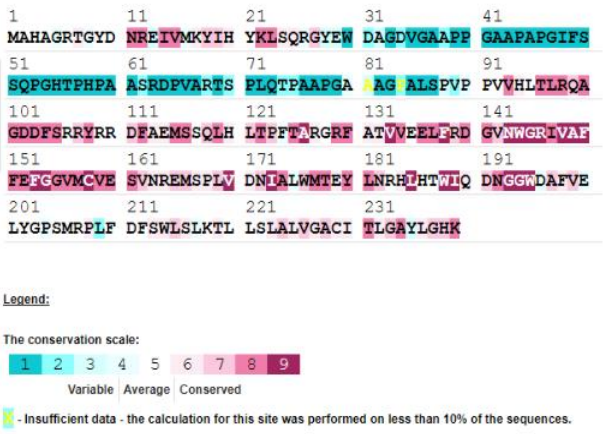
SNP Mining

According to the dbSNP database, the Bcl-2 gene contains 82 nsSNPs in the coding region. Among them, 79 nsSNPs introduce single amino acid changes (missense mutations) into the Bcl-2 gene.

Assessment of Functional Impacts and Disease Association of Non-synonymous SNP

To determine whether a given missense mutation affected Bcl-2 function and its association with disease, selected nsSNPs were subjected to a variety of

in silico SNP prediction algorithms. The results that are provided in **Table 1**, identified 11 nsSNPs with a high probability of being deleterious to Bcl-2 structure and/or function as predicted by at least five different prediction algorithms.



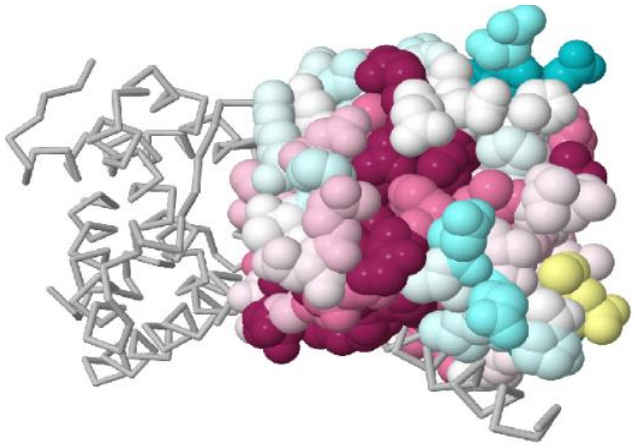


Figure 1. Amino acids of Bcl-2 protein were ranked on a conservation scale of 1–9 and are highlighted as follows: blue residues (1–4) are variable, white residues (5) are average, and purple residues (6–9) are conserved. ConSurf analysis depicts the M166, G141, R129, S105, F104, R98, L97, V93, G27, and V15 mutations occurred in a highly conserved area of Bcl-2 protein.

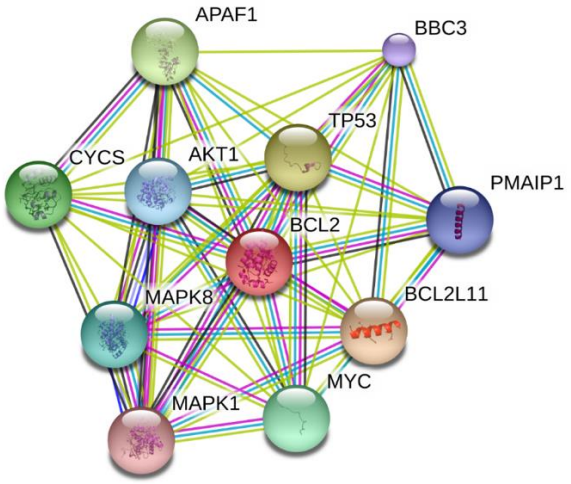


Figure 2. Protein-protein interaction network of Bcl-2.

Identifying SNPs located in Evolutionary Conserved Regions in the Gene

Amino acids located in enzymatic sites or essential for protein-protein interactions usually are more conservative than other residues. Thus, the consequence of nsSNPs positioned at highly conserved area tend to be more deleterious than those located at non-conversed sites. To further investigate the possible impacts of the high-risk nsSNPs in **Table 1**, the ConSurf web server was employed to calculate the level of evolutionary conservation at all amino acid sites in the Bcl-2 protein. ConSurf analysis revealed that residues M166, G141, R129, S105, F104, R98, L97, V93, G27, and V15 are highly conserved with conservation Score ranging from 7 to 9 (**Figure 1**).

Protein-Protein Interactions Network of Bcl-2

Bcl-2 interacts with MAPK8 (Mitogen-activated protein kinase 8), AKT1 (V-akt murine thymoma viral oncogene homolog 1), PMAIP1 (Phorbol-12-myristate-13-acetate-induced protein 1), CYCS(Cytochrome c), BBC3 (BCL2 binding component 3), TP53 (tumor protein p53), BCL2L11

Table 1. Functionally significant and disease-associated Bcl-2 nsSNPs predicted by at least five SNP prediction algorithms

nsSNP ID	Mutation	Mutation Type
rs201085318	M166T	Missense
rs755252289	G141E	Missense
rs777784952	R129C	Missense
rs763718170	S105P	Missense
rs751038951	F104S	Missense
rs755908754	R98L	Missense
rs528042823	L97P	Missense
rs551395951	H94P	Missense
rs565014924	V93A	Missense
rs779372254	G27S	Missense
rs745851862	V15L	Missense

Table 2. Putative phosphorylation sites predicted using NetPhosP and GPS 2.1. Sites that were predicted using both tools are considered as phosphorylated sites in this study

Position	Amino Acid
7	Thr
24	Ser
28	Tyr
51	Ser
56	Thr
62	Ser
70	Ser
74	Thr
87	Ser
96	Thr
105	Ser

116	Ser
125	Thr
161	Ser
180	Tyr
213	Ser
216	Ser
222	Ser

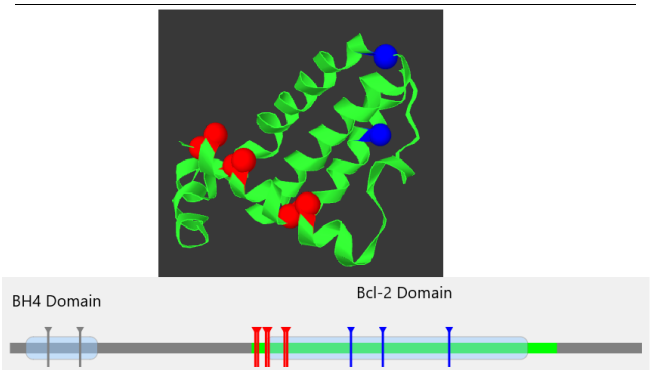


Figure 3. Structural model of native Bcl-2 showing 9 high-risk nsSNPs in the Bcl-2 domain and 2 in the BH4 domain.

(Bcl-2-like 11), MAPK1 (Mitogen-activated protein kinase 1) proteins, APAF1 (Apoptotic peptidase activating factor 1), MYC (V-myc myelocytomatosis viral oncogene homolog) (**Figure 2**).

Detection of nsSNPs Location in Protein Structure

Mutation 3D predicted that 9 nsSNPs are located in the Bcl-2 domain and 2 nsSNPs are located in the BH4 domain of the Bcl-2 protein (**Figure 3**). It also identified 3 clusters of amino acid substitutions (indicated in red color in **Figure 3**). These 11snSNPs are considered the higher risk for Bcl-2 protein.

Prediction of Post-translational Modification Sites

The putative ubiquitylation site prediction was implemented through UbPred and BDM-PUB programs. No change in ubiquitination site was predicted by any program. Putative sumoylation sites were predicted using the pSumo-CD, JASSA, PCI-SUMO and GPS-SUMO programs. GPS-SUMO determined both SUMO-interaction Motifs (SIMs) and sumoylation sites. One SUMO-interacting motif (SIM)was predicted by GPS-SUMO. The SIM is located at position 92-96 with a sequence VVHLT. Putative phosphorylation sites were predicted using GPS program and NetPhos 3.1. Sites that were predicted to be phosphorylated by both tools were considered in this study. The result of putative phosphorylation sites considered in this study is given in **Table 2**. Putative methylation sites were predicted using PSSMe which predicted two lysine residues for methylation sites- K17 and K218 with support vector machine (SVM) probability 0.5593 and 0.77136, respectively.

Surface Accessibility of Native and Mutant Bcl-2 Protein

In order to compare the biophysical property of native and mutant amino acids, the solvent accessibility was calculated using NetSurfP. A huge drift in the Z-score, which provides reliable predictions for buried and exposed amino acids, was observed for mutant models F104S, L97P, V93A and12 V15L (**Table 3)**.

Protein Stability Analysis

I-Mutant 2.0, was used to predict protein stability changes upon single base mutation. The I-Mutant 2.0 predictions are expressed in 2 possibilities- increase or decrease the stability of protein after mutation, with reliability index at pH 7.0 and temperature 25 °C. It estimated the free energy change value (ΔΔG) and a negative ΔΔG corresponds to a decrease in protein stability, whereas a positive ΔΔG corresponds to an increase in protein stability. To add another layer of confirmation for mutated protein stability, further

Table 3. Surface accessibility of native and mutant Bcl-2 protein

SNPs	AA	AA position	RSA	ASA	Z-Score	Class assignment
rs201085318	M	166	0.269	53.787	0.244	E
	T		0.255	35.410	0.219	E
rs755252289	G	141	0.372	29.284	-1.296	E
	E		0.399	69.918	-1.355	E
rs777784952	R	129	0.188	43.144	1.528	B
	C		0.112	15.697	1.385	B
rs763718170	S	105	0.175	20.475	0.405	B
	P		0.188	26.748	0.346	B
rs751038951	F	104	0.030	5.921	0.560	B
	S		0.031	3.680	0.210	B
rs755908754	R	98	0.246	56.288	1.529	E
	L		0.245	44.878	1.491	E
rs528042823	L	97	0.013	2.417	1.867	B
	P		0.019	2.625	1.574	B
rs551395951	H	94	0.095	17.299	0.992	B
	P		0.054	7.620	0.756	B

rs565014924	V	93	0.033	5.041	-0.389	B
	A		0.025	2.788	0.300	B
rs779372254	G	27	0.476	37.501	-1.938	E
	S		0.538	63.077	-1.788	E
rs745851862	V	15	0.022	3.320	1.103	B
	L		0.024	4.321	0.908	B

Keys: AA-Amino acid, **RSA**-relative surface accessibility, **ASA**-absolute surface accessibility, **E**- Exposed, **B**- Buried. Amino acids highlighted in bold were the mutant amino acids.

analysis of these 11 nsSNPs was conducted using MUpro. Results for both methods are expressed as two possibilities increase or decrease the stability of protein after mutation with a confidence score. For the 11 nsSNPs, all except H94P predicted to have decreased in the stability of the protein (**Table 4**) by both tools.

Prediction of Structural Effects of Point Mutation

Project HOPE predicted that in M166T, R129C, F104S, R98L, L97P, H94P, V93A mutations; the mutant residue is smaller than the native residue. This may cause a possible loss of external interactions and can cause an empty space in the core of the protein. M166T, G141E, F104S, R98L, H94P mutant residues are less hydrophobic than native residues which can cause loss of hydrophobic interactions with other molecules on the surface of the protein. In case of G141E, S105P, G27S, and V15L mutations, the mutant residue is larger than the native residue.

Structure Analysis

To broaden the structural analysis, the calculated Tm-score and root mean square deviation (RMSD) value for each nsSNP model were accomplished using Tm-Align web tool. Tm-score provided a mean for evaluating the topological similarity between wild-type and mutant models, whereas RMSD values provided the distance between the α-carbon backbones of wild-type and mutant models. High RMSD indicates a greater deviation between wild-type and mutant protein structures. The Tm-score and RMSD for each nsSNP model are listed in **Table 5**. The maximum RMSD was 2.27 for G27S, and the minimum RMSD value was 0.40 for M166T. The energy value of native models was found to be -217669.3 kJ/mol after minimization, whereas mutant models exhibited total energy value ranging from -219296 kJ/mol to -108234.2 kJ/mol (**Table 6**).

Table 4. Results of stability prediction for the 11 nsSNPs by I-Mutant 2.0 and MUpro

Mutation	Stability						
	MUpro				*I-Mutant*		
	Support Vector Machine(SVM)	*Confidence*	*Neural Network*	*Confidence*	*Stability*	*T(°C), pH*	*DDG Value Prediction*
M166T	Decrease	-1	Decrease	-0.99721803	Decrease	25, 7	-1.75
G141E	Decrease	-0.6438476	Decrease	-0.99619385	Decrease	25, 7	-0.82
R129C	Decrease	-0.9244694	Decrease	-0.93083892	Decrease	25, 7	-0.5
S105P	Decrease	-0.0384512	Decrease	-0.7102215	Decrease	25, 7	-1.05
F104S	Decrease	-1	Decrease	-0.99882333	Decrease	25, 7	-1.88
R98L	Decrease	0.30227434	Decrease	-0.66915414	Decrease	25, 7	-0.63
L97P	Decrease	-1	Decrease	-0.96213578	Decrease	25, 7	-1.47
H94P	Increase	0.22234996	Decrease	-0.60127869	Increase	25, 7	0.57
V93A	Decrease	-0.9265469	Decrease	-0.90813655	Decrease	25, 7	-3.15
G27S	Decrease	-0.2581561	Decrease	-0.74161313	Decrease	25, 7	-1.26
V15L	Decrease	-1	Decrease	-0.99983933	Decrease	25, 7	-1.05

Table 5. RMSD (A°) and TM-score for the 11 high-risk nsSNPs of Bcl-2

nsSNPs	Mutation	RMSD (A°)	TM-score
rs201085318	M166T	0.40	0.99473
rs755252289	G141E	1.23	1.00
rs777784952	R129C	1.28	0.993
rs763718170	S105P	0.70	0.87
rs751038951	F104S	2.18	0.78

rs755908754	R98L	1.11	0.79
rs528042823	L97P	1.69	0.81
rs551395951	H94P	0.81	0.86
rs565014924	V93A	0.93	0.77
rs779372254	G27S	2.27	0.76
rs745851862	V15L	1.59	0.79

Table 6. The total energy of native and mutant structures before and after energy minimization

Amino acid change	Total energy (kJ/mol)			
	Before minimization	*Z-score*	*After minimization*	*Z-score*
Native	2619187.4	-1.54	-217669.3	0.72
M166T	22059.4	-1.12	-108234.2	0.79
G141E	2619122.7	-1.55	-219296.0	0.76
R129C	2619924.2	-1.57	-216164.1	0.67
S105P	202195388.3	-1.58	-215440.7	0.57
F104S	2620065.4	-1.59	-216819.5	0.63
R98L	2622601.6	-1.57	-215736.1	0.70
L97P	3584753.8	-1.61	-215732.5	0.60
H94P	2620727.3	-1.57	-216361.2	0.68
V93A	2619443.2	-1.57	-219187.7	0.80
G27S	2619443.2	-1.55	-217818.2	0.68
V15L	2620899.7	-1.55	-218583.5	0.70

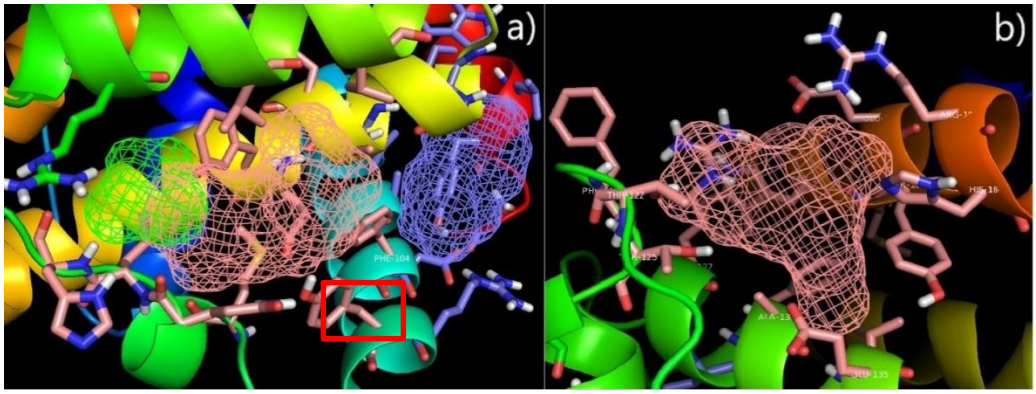


Figure 4. The figure illustrates ligand binding sites. Here Pink mesh, green mesh and purple mesh are the representation of the binding site 1, 2 and 3, respectively. **a)** Representation of binding site 1 in native Bcl-2 protein showing F104 (Phe104 in the red box) residue at site 1 **b)** Binding site changes due to point mutation at F104.

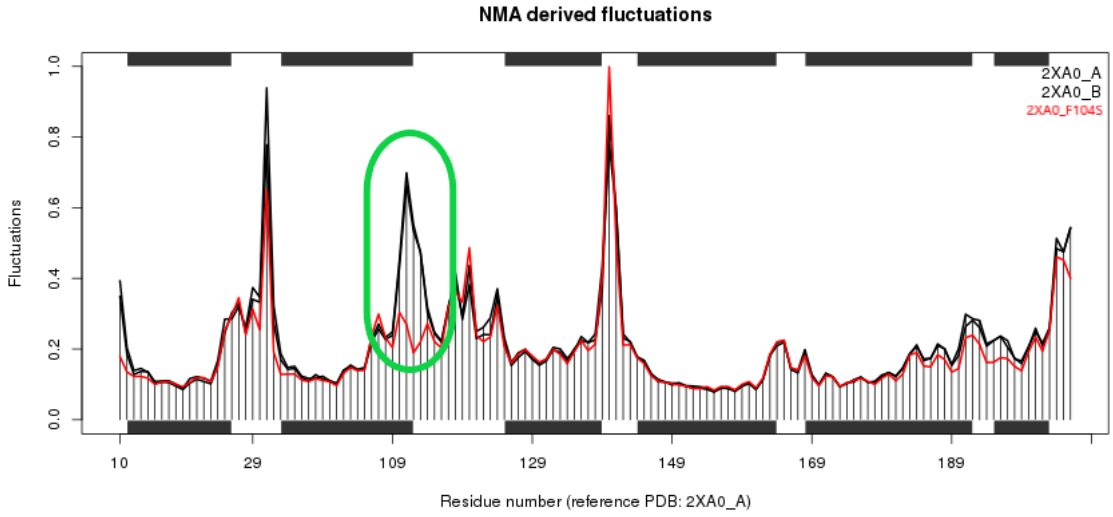


Figure. 5 Residual fluctuation analysis of native Bcl-2 protein and F104S mutant using crystal structure B 2XA0 from Protein Data bank (https://www.rcsb.org/). NMA of residual fluctuation shows that F104S results in the structural alteration as seen in the residual fluctuation differences indicated in the green circle.

Binding Sites Prediction and Analysis

FTsite was employed to identify ligand binding sites for Bcl-2 native protein. Ligand binding site prediction showed that binding site 1 and 3 involves F104 residue which is consistent with literature published [45]. Here, first and third ligand binding sites of F104S mutants alter the ligand binding and function of Bcl-2 protein (**Figure. 4**). For further understanding of structural alteration due to the F104S mutation, Normal Mode Analysis (NMA) of the native Bcl-2 and the F104S mutant structures was performed (**Figure. 5)**. Since F104 residue is involved in alternations of protein and ligand-binding regions, only F104S mutant was selected to predict how this F104S alters the protein structure and binding site. NMA analysis showed that F104S mutation caused differences in residual fluctuation between the native Bcl-2 and F104S mutant which proves structural alteration resulted from the F104S mutation.

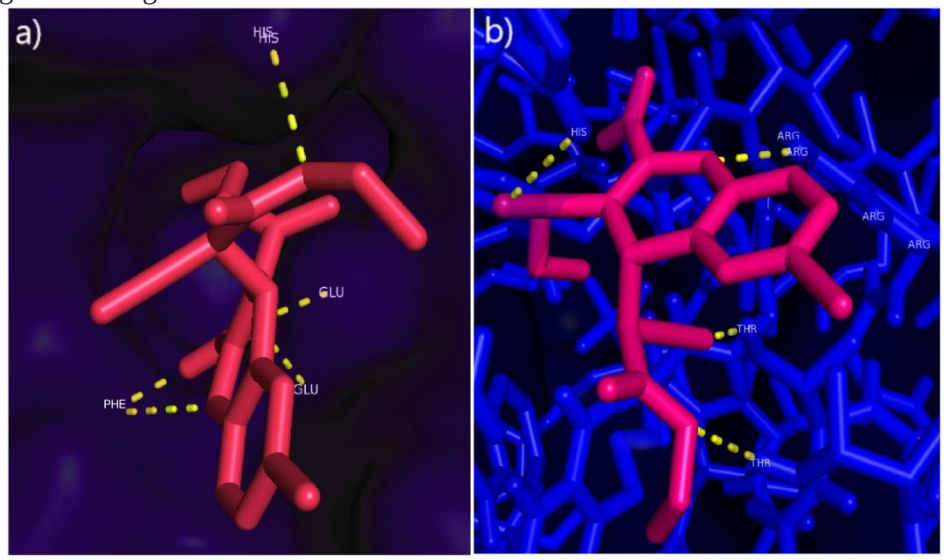


Figure 6. Comparative docking analysis of Bcl-2 binding site 1 using HA 14-1 as ligand. **a)** HA-14-1 bound to native Bcl-2 with 5 H-bonds involving F104, His120 and Glu136 **b)** HA 14-1 bound to F104S mutant with 4H-bonds involving Thr122, Thr125, Arg129 and His184 residues. Yellow dashes indicate hydrogen bonds formed between the Bcl-2 and the ligand molecule

To confirm the binding site alterations are due to F104S mutation, comparative molecular docking analysis was performed using HA 14-1 as ligand and native and F104S mutant Bcl-2 as receptors. HA 14-1 was allowed to bind to the binding site using the parameters mentioned in method section 2.12. Docking results showed that native Bcl-2 binds to the ligand at binding site 1 involving F104 with 5 H-bonds. On the other hand, F104S mutant binds to the ligand in a different binding site with 4 H-bonds (**Figure. 6**). Moreover, the binding affinity of native Bcl-2 and HA 14-1 was -8.4 kcal/mol, whereas the binding affinity of F104S mutant and HA 14-1 was -5.6 kcal/mol, indicating a reduced interaction between receptor and ligand. The detailed molecular interactions of ligand (HA 14-1) with native and F104S mutant are shown in **Figure. 6.**

CONCLUSION

In silico analysis of Bcl-2 nsSNPs demonstrated that multiple nsSNPs located in the Bcl-2 gene may be deleterious to the structure/ function and thus considered as high-risk nsSNPs. Amongst the studied nsSNPs, most of these high-risk nsSNPs are located at evolutionarily conserved regions or functional domains that are significant for the normal function of Bcl-2. This may result in altered protein-protein or protein-ligand interactions. Among the analyzed nsSNPs, 10 of the high-risk nsSNPs decrease the stability of Bcl-2 structure. One particular mutation, F104S, completely alters the binding site and results in reduced interactions between the Bcl-2 protein and its antagonist. Therefore, in the field of personalized cancer medicine, better drug designing is critical for individuals with F104S (rs751038951). In conclusion, we recommend that these nsSNPs should be taken as high-risk biomarkers for cancer prognosis in population-based studies.

ACKNOWLEDGEMENTS

The authors express their gratitude to Dr. Chaman Ara Keya, Associate Professor, Department of Biochemistry and Microbiology, North South University, for her inspiration and cooperation during this study.

REFERENCES

1. Ramensky, V., Bork, P. and Sunyaev, S. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 30(17):3894-3900.

2. Radivojac, P., Vacic, V., Haynes, C., Cocklin, R.R., Mohan, A., Heyen, J.W., Goebl, M.G. and Iakoucheva, L.M. 2010.

Identification, analysis, and prediction of protein ubiquitination sites. *Proteins*. 78(2): 365–380.
3. Doniger, S.W., Kim, H.S., Swain, D., Corcuera, D., Williams, M., Yang, S.P. and Fay, J.C. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* 4(8): e1000183.
4. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29(1): 308–311.
5. Tsujimoto, Y., Gorham, J., Cossman, J., Jaffe, E. and Croce, C.M. 1985. The t(14; 18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. *Science*. 229(4720):1390-3.
6. Cleary, M.L. and Sklar, J. 1985. Nucleotide sequence of a t(14; 18) chromosomal breakpoint in follicular lymphoma and demonstration of a breakpoint-cluster region near a transcriptionally active locus on chromosome 18. *Proc Natl Acad Sci U S A.* 82(21):7439-43.
7. Bakhshi, A., Jensen, J.P., Goldman, P., Wright, J.J., McBride, O.W., Epstein, A.L. and Korsmeyer, S.J. 1985. Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around Jhon chromosome 14 and near a transcriptional unit on 18. *Cell*. 41(3):899-906.
8. Vaux, D.L., Cory, S. and Adams, J.M. 1988. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature*. 335(6189):440-2.
9. Seto, M., Jaeger, U., Hockett, R.D., Graninger, W., Bennett, S., Goldman, P. and Korsmeyer, S.J. 1988. Alternative promoters and exons, somatic mutation and deregulation of the Bcl-2-Ig fusion gene in lymphoma. *EMBO J.* 7(1):123-31.
10. Germain, M. and Shore, G. C. 2003. Cellular distribution of Bcl-2 family proteins. *Sci STKE* 2003(173):pe10.
11. Hardwick, J.M. and Soane, L. 2013. Multiple functions of BCL-2 family proteins. *Cold Spring Harb Perspect Biol.* 5(2). pii: a008722.
12. Moon, J.H., Sohn, S.K., Lee, M.H., Jang, J.H., Kim, K., Jung, C.W. and Kim, D.H. 2010. BCL2 gene polymorphism could predict the treatment outcomes in acute myeloid leukemia patients. *Leuk Res.* 34(2):166-72.
13. Park, Y.H., Sohn, S.K., Kim, J.G., Lee, M.H., Song, H.S., Kim, M.K., Jung, J.S., Lee, J.J., Kim, H.J. and Kim, D.H. 2009. Interaction between BCL2 and interleukin-10 gene polymorphisms alter outcomes of diffuse large B-cell lymphoma following rituximab plus CHOP chemotherapy. *Clin Cancer Res.* 15(6):2107-15.
14. Bhatti, P., Church, D.M., Rutter, J.L., Struewing, J.P. and Sigurdson, A.J. 2006. Candidate single nucleotide polymorphism selection using publicly available tools: A guide for epidemiologists. *Am J Epidemiol.* 164(8):794-804.
15. Johnson, A. D. 2009. Single-Nucleotide Polymorphism Bioinformatics: A Comprehensive Review of Resources. *Circ Cardiovasc Genet.* 2(5):530-536.
16. Li, L. and Wei, D. 2015. Bioinformatics tools for discovery and functional analysis of single nucleotide polymorphisms. *Adv Exp Med Biol.* 827:287-310.
17. Kumar, P., Henikoff, S. and Ng, P.C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 4:1073–1082
18. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. 2010. A method and server for predicting damaging missense mutations. *Nat. Methods.* 7(4):248.
19. Choi, Y. and Chan, A.P. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 31(16):2745-2747.
20. Thomas, P.D., Kejariwal, A., Campbell, M.J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S. and Vandergriff, J.A. 2003. PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* 31(1):334-341.
21. Stone, E.A. and Sidow, A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* 15(7):978-986.
22. Bromberg, Y., Yachdav, G. and Rost, B. 2008. SNAP predicts effect of mutations on protein function. *Bioinformatics.* 24(20):2397-2398.
23. Yoo, J., Lee, Y., Kim, Y., Rha, S.Y. and Kim, Y. 2008. SNPAnalyzer 2.0: A web-based integrated workbench for linkage disequilibrium analysis and association analysis. *BMC Bioinformatics.* 9(1):290.
24. Capriotti, E., Fariselli, P. and Casadio, R. 2005. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 33(Web Server issue):W306-10.
25. Cheng, J., Randall, A. and Baldi, P. 2006. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins.* 62(4):1125-1132.
26. Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E. and Ben-Tal, N. 2003. ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics.* 19(1):163-164.
27. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P. and Kuhn, M. 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43(Database issue):D447-52.
28. 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. and Yu, H. 2016. mutation3D: Cancer Gene Prediction Through Atomic Clustering of Coding Variants in the Structural Proteome. *Hum Mutat.* 37(5):447-456.
29. Jia, J., Zhang, L., Liu, Z., Xiao, X. and Chou, K.C. 2016. pSumo-CD: Predicting sumoylation sites in proteins with covariance discriminant algorithm by incorporating sequence-coupled effects into general PseAAC. *Bioinformatics. 32*(20):3133-3141.
30. Beauclair, G., Bridier-Nahmias, A., Zagury, J.F., Saïb, A. and Zamborlini, A. 2015. JASSA: A comprehensive tool for prediction of SUMOylation sites and SIMs. *Bioinformatics.* 31(21):3483-3491.
31. Zhao, Q., Xie, Y., Zheng, Y., Jiang, S., Liu, W., Mu, W., Liu, Z., Zhao, Y., Xue, Y. and Ren, J. 2014. GPS-SUMO: A tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res.* 42(Web Server issue):W325-30.
32. Xue, Y., Liu, Z., Cao, J., Ma, Q., Gao, X., Wang, Q., Jin, C., Zhou, Y., Wen, L. and Ren, J. 2011. GPS 2.1: Enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Eng Des Sel.* 24(3):255-260.
33. Blom, N., Gammeltoft, S. and Brunak, S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 294(5):1351-1362.
34. Wen, P.P., Shi, S.P., Xu, H.D., Wang, L.N. and Qiu, J.D. 2016. Accurate in silico prediction of species-specific methylation sites based on information gain feature optimization. *Bioinformatics.* 32(20):3107-3115.
35. Cheng, J., Randall, A. and Baldi, P.2006. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function and Genetics*. Mar 1;62(4):1125-32.
36. Mavroconstanti, T., Johansson, S., Winge, I., Knappskog, P.M. and Haavik, J. 2013. Functional Properties of Rare Missense Variants of Human CDH13 Found in Adult Attention Deficit/Hyperactivity Disorder (ADHD) Patients. *PLoS One.* 8(8), p.e71445.
37. Venselaar, H., te Beek, T.A., Kuipers, R.K., Hekkelman, M.L. and Vriend, G. 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics.* 11(1):548.
38. Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. and Karplus, K. 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology

modeling: Four approaches that performed well in CASP8. *Proteins: Structure, Function, and Bioinformatics.* 77 Suppl 9:114-22.
39. Zhang, Y. and Skolnick, J. 2005. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33(7):2302-2309.
40. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem. 19(14):1639-1662.
41. Trott, O. and Olson, A.J. 2010. AutoDock Vina: inproving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem.* 31(2):455-461.
42. Ngan, C.H., Hall, D.R., Zerbe, B., Grove, L.E., Kozakov, D. and Vajda, S. 2012. FtSite: High accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics.* 28(2):286-287.
43. Arisan, E.D., Kutuk, O., Tezil, T., Bodur, C., Telci, D. and Basaga, H. 2010. Small inhibitor of Bcl-2, HA14-1, selectively enhanced the apoptotic effect of cisplatin by modulating Bcl-2 family members in MDA-MB-231 breast cancer cells. *Breast Cancer Res Treat.* 119(2):271.
44. Lil, M. A., and Danielson, M. L. 2011. Computer-aided drug design platform using PyMOL. *J Comput Aided Mol Des.* 25(1):13-19.
45. Iyer, D., Vartak, S.V., Mishra, A., Goldsmith, G., Kumar, S., Srivastava, M., Hegde, M., Gopalakrishnan, V., Glenn, M., Velusamy, M. and Choudhary, B. 2016. Identification of a novel BCL2-specific inhibitor that binds predominantly to the BH1 domain. *FEBS J.* 283(18):3408-3437.