Deleterious missense single nucleotide polymorphism alters the structural conformation of human anti-apoptotic Bcl-2 protein

P. P. Anand* and Y. Shibu Vardhanan

Biochemistry and Toxicology Division, Department of Zoology, University of Calicut, Malappuram 673 635, India

Identification and understanding the mechanism behind tumorigenesis is crucial for developing new strategies in cancer treatments. In this *in silico* study, we used the 14 different tools to evaluate the convergent deleterious missense Bcl-2 SNPs. Out of 37 missense single nucleotide polymorphism (SNPs), five are deleterious. Molecular modelling and structural and functional evaluation of the mutant proteins were carried out to understand their clinical significance. All deleterious missense SNPs alter the structural stability of the protein. Also H94P deleterious missense SNPs alter the ligand-binding ability of Bcl-2. The results indicate that the mutant antiapoptotic Bcl-2 protein may contribute to tumorigenesis in different ways, depending upon the mutation location.

Keywords: Apoptosis, anti-apoptotic protein Bcl-2, deleterious single nucleotide polymorphism, *in silico* evaluation, protein modelling.

THE Bcl-2 family of proteins play a vital regulatory role in different cell-death pathways, and the deregulation of these proteins leads to chemoresistance and tumorigenesis[1–5]. The Bcl-2 family contains various proteins with a single BH domain, which has a wide range of cellular functions[6–8]. Anti-apoptotic proteins such as Bc-2 and Bcl-xl are found to be over expressed in different types of cancers and inhibit apoptosis by binding to Bax and Bak; they can also inhibit autophagy[8,9]. Bcl-2 and Bcl-xl are expressed in the membranes of several cellular organelles such as mitochondria, endoplasmic reticulum and nucleus[10]. The anti-apoptotic protein prevents the release of cytochrome *c* and apoptosis inducing factor (AIF) into the cytoplasm[10,11].

The missense single nucleotide polymorphism (SNPs) are associated with single amino acid alteration in the coding regions of a gene. These mutations may be deleterious or neutral; the deleterious mutations affect the structural and functional aspects of the protein[12]. The

strength of the anti-apoptotic activity of Bcl-2 protein is well known[13]. However, the single nucleotide variation in Bcl-2 protein is not yet thoroughly studied. Several studies on Bcl-2 functionality have revealed the effect of point mutations. The mutations in coding regions could lead to several functional defects with an impact on tumorigenesis[14]. Computational strategies are intended to predict the effects of the amino acid residue changes in proteins widely used to assess whether the changes are deleterious or not[15]. Numerous tools are currently used to indicate the potential and the structural and functional impact caused by amino acid changes[16]. The present study evaluates the effects of deleterious missense SNPs in Bcl-2 and predicts their structural deformations.

Methodology

Datasets

The missense SNPs and protein sequence (FASTA format) of Bcl-2 were retrieved from NCBI-dbSNP database during April 2018 (NCBI accession no.: AAH27258.1 and UniProt Id: P10415)[17]. Screening of human-validated missense SNPs was done according to Anand[12] and Pires *et al.*[18].

Selection of deleterious missense SNPs

Fourteen tools under different categories were used for predicting the deleterious missense Bcl-2 protein SNPs. These categories are as follows[12,18]: (i) sequence homology and evolutionary conservation – SIFT[19], Provean[20], Panther[21], Align-GD, and MAPP; (ii) supervised learning methods – SNAP2[22], MutPred, SuSpect[18], PhD-SNP and nsSNP Analyzer; (iii) protein sequence and structure-based methods – Polyphen-2[23], and (iv) consensus-based method – Predict SNP[22,24] and PON-P2[22,24]. Among these 14 prediction tools, more than 12 tools showing convergent deleterious missense SNPs were used for further analysis.

*For correspondence. (e-mail: anandpp633@gmail.com)
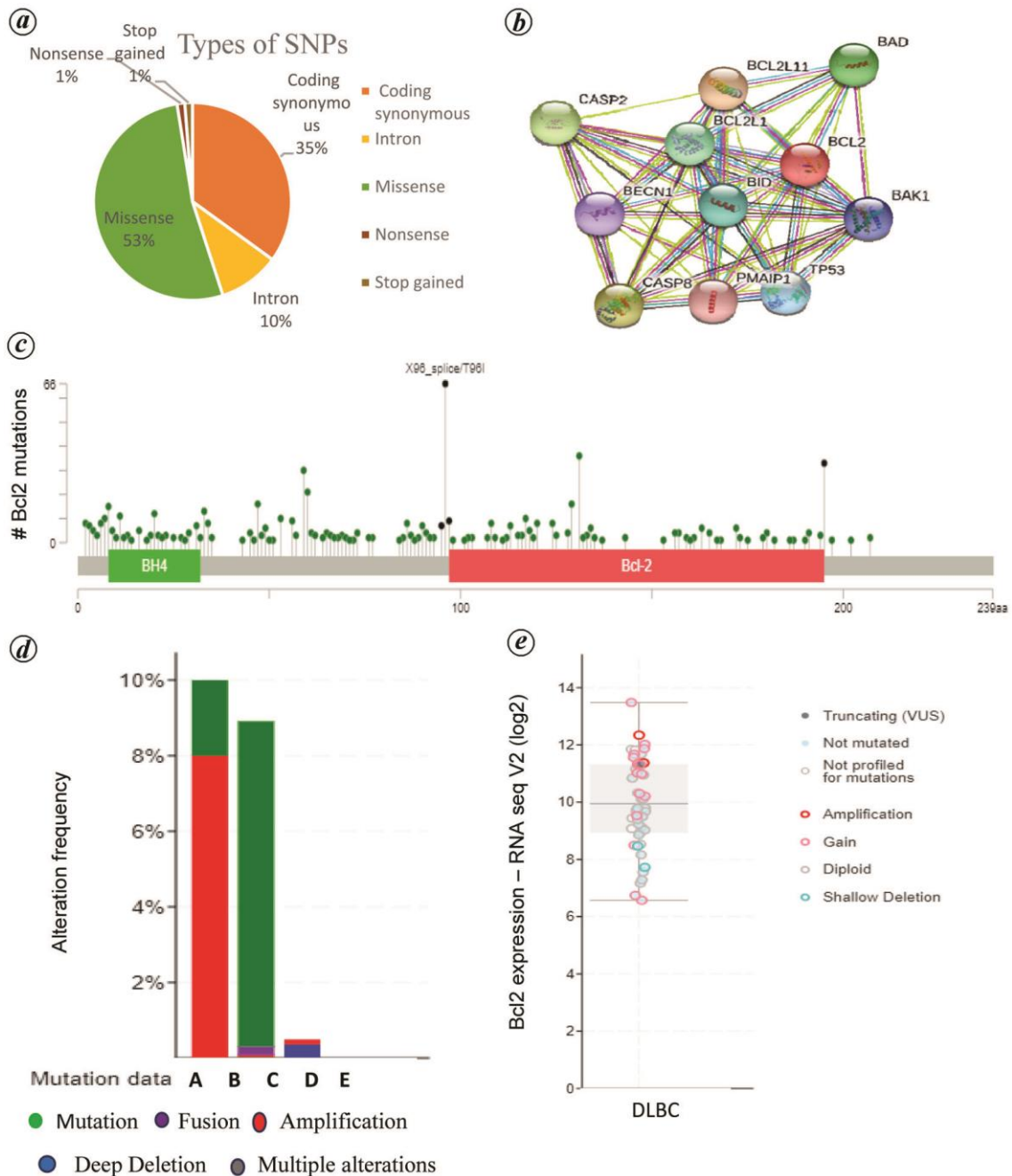
Figure 1. *a*, Distribution frequency of SNPs within the BCL-2 gene. It can be concluded that the majority of SNPs are observed in missense (53%) followed by coding synonymous (35%), intron (10%), nonsense, and stop gained (1%). *b*, Bcl-2 interacting proteins (generated by STRING tool). *c*, Bcl-2 mutation rate – TCGA database. *d*, Bcl-2 alteration frequency – TCGA database: A, Non-Hodgkin lymphoma; B, Mature B-cell neoplasms, C, B-lymphoblastic leukaemia/lymphoma, D, Leukaemia, E, Mature T and NK neoplasms. *e*, Bcl-2 expression and mutation rate – TCGA database.

Mutations in Bcl-2 protein

Using TCGA datasets[25,26] helps determine the mutation and amplification frequency of BCL-2 in various cancers. Bcl-2 interacting proteins were analysed using the STRING server[27], and estimating the evolutionary conservation of each amino acid in the Bcl-2 protein was done using ConSurf server[28,29].

Molecular modelling

The structural modelling of native Bcl-2 and convergent deleterious missense Bcl-2 SNPs was separately done using the MUSTER server[30]. The modelling was based on the previous sequence profile–profile alignment (PPA) method. The PROCHECK server was used for evaluating the best model[31]. In the Ramachandran plot[32], for a good
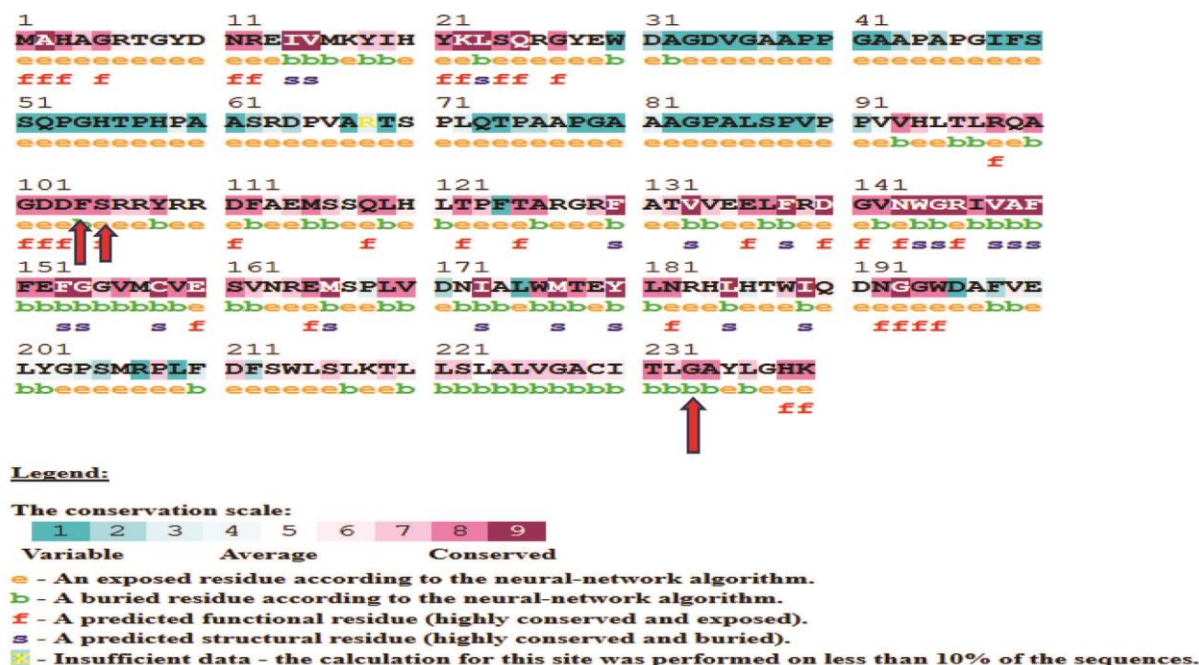
Figure 2. Colour grade indicating the degree of conservation status of amino acid residues. The grade number (colour) increases with evolutionary conservation (1 is highly variable and 9 is highly conserved site). Three convergent deleterious (F104S, S105P, and G233D) predicted SNPs are present in the highly conserved region represented as red arrows.

protein structure, it is expected that there should be more than 90% of the residues in the core of favoured regions of the protein; and the protein model visualized in PyMol tool and validated in PDBsum server[33,34].

Evaluation of deleterious missense SNPs – structural impact analysis

Project HOPE server analyses the structural effects of point mutation in proteins[35]. ExPASy ProtParam server divulges the physico-chemical characteristic feature of the proteins, such as isoelectric point (pI), molecular weight (Mw), extinction coefficient (EC), instability index (II), aliphatic index (AI), and hydropathicity (GRAVY)[36]. NOMAD-Ref Gromacs server is used for energy minimization of modelled proteins for *in vivo* simulation[37]. Solvent accessible surface area (ASA) of native and deleterious Bcl-2 proteins was evaluated using VADAR server[18]. I-Mutant server was used for estimating the stability of deleterious proteins[32,38], and the post-translation modification (PTM) of Bcl-2 protein was estimated using ModPred server[39]. FT-Site server was used to predict the ligand-binding site of the Bcl-2 protein[2,18,40].

Results

Distribution frequency of Bcl-2 SNPs

Human-validated Bcl-2 contained a total of 75 SNPs. Among these 37 are missense, 28 coding synonymous,

eight introns, one stope-gained, and one non-sense variant. The missense SNPs were validated based on the 1000 genome project, frequency or genotype, and refSNP cluster. They were used for initial screening and predicting the convergent deleterious missense SNPs (Figure 1 *a*).

Mutations in Bcl-2 protein

The Bcl-2 protein interacted with several other proteins such as BAD, BAK1, BID, TP53 (Figure 1 *b*). So its alteration is directly or indirectly correlated with the protein–protein interaction mechanisms. Mutation in the Bcl-2 protein leads to different types of cancers, such as non-Hodgkin lymphoma, mature B-cell neoplasms, B-lymphocytic leukaemia/lymphoma, leukaemia, and mature T and NK neoplasms. Among the five groups, only the first three neoplasms are mainly caused by Bcl-2 gene alteration. The highest gene mutational alteration frequency was observed in non-Hodgkin lymphoma and mature B-cell Neoplasms (Figure 1 *c–e*).

Convergent deleterious missense SNPs of Bcl-2

Among the 37 missense SNPs, only five (G233D, L97P, H94P, F104S and S105P) were found to be more deleterious compared to the others, and these variants were used for further analysis (Table 1; detailed prediction

Table 1. Results of Bcl-2 convergent deleterious predicted missense SNPs analysed by 14 prediction tools. N, Neutral; D, Deleterious; PB, Probably benign; PSD, Possible damaging; PD, Probably damaging; U, Unknown; B, Benign; P, Pathogenic (the complete prediction details also available as Supplementary Tables 1–16)

| Variant ID	Amino acid variation	SIFT	PROVEAN	PANTHER	Align GV-GD	MAPP	SNAP-2	Mut-Pred-2	SusPect	PMut	PhD-SNP	nsSNPAnalyzer	Polyphen-2	Predict SNP	PON-P2
rs1800477	A43T	N	N	PSD	PSD	N	D	U	N	N	N	U	PSD	N	U
rs61733415	F124L	N	N	PB	N	N	D	U	N	N	N	U	N	N	U
rs138465313	G237C	D	N	PSD	D	N	D	D	N	N	N	U	PD	D	P
rs148811059	G203S	D	D	PD	PSD	N	D	U	N	N	N	D	B	N	U
rs201827119	A82V	D	N	PSD	PSD	D	D	U	N	N	N	U	B	N	U
rs368344129	G47S	N	N	PSD	PSD	U	N	U	N	N	N	U	B	N	U
rs369277721	P46L	N	N	PB	D	D	D	U	N	N	N	U	B	N	U
rs369294037	R207Q	N	N	PSD	N	N	D	U	N	N	N	N	B	N	U
rs373889054	L95V	N	N	PSD	N	D	N	U	N	D	N	U	B	N	P
rs375594294	G83E	N	N	PSD	D	D	D	U	N	N	N	U	B	N	U
rs376149674	G233D	D	D	PD	D	D	D	D	PSD	D	D	U	PD	D	P
rs528042823	L97P	D	D	PD	D	D	D	D	D	D	D	D	PD	D	P
rs540701354	D34Y	D	N	PSD	D	D	D	U	N	N	N	U	PSD	D	U
rs546449806	A32G	N	N	PSD	PSD	U	D	U	N	N	N	U	B	N	U
rs551395951	H94P	D	D	PD	D	D	D	D	PSD	D	D	D	PD	D	P
rs557714365	H3R	D	N	PD	N	U	D	U	N	N	N	U	B	N	P
rs559397524	Q99R	N	N	PSD	N	D	N	U	N	N	N	U	B	N	U
rs565014924	V93A	D	N	PSD	PSD	D	D	U	N	D	N	N	PSD	D	U
rs565741123	K17N	N	N	PSD	D	N	N	U	N	N	N	U	B	N	U
rs574148249	L221F	N	N	PSD	N	N	D	D	N	N	N	D	B	N	U
rs746401408	R68G	N	N	PB	D	U	D	U	N	N	N	U	B	N	U
rs747504890	G128R	N	N	PSD	D	D	D	U	N	N	N	U	B	N	U
rs747696230	S62Y	D	N	PB	D	U	D	U	N	N	N	U	B	N	U
rs748122615	R207W	D	D	PSD	D	D	D	U	N	N	N	D	B	D	U
rs749649089	V66I	N	N	PB	N	U	N	U	N	N	N	U	B	N	U
rs751038951	F104S	D	D	PD	D	D	D	D	D	D	D	D	PD	D	P
rs758123306	H20Q	N	N	PD	N	N	D	U	PSD	D	N	U	PD	N	U
rs762635201	S105F	D	D	PSD	D	D	D	U	PSD	D	N	U	PD	D	P
rs763718170	S105P	D	N	PSD	D	D	D	D	PSD	D	D	U	PD	D	P
rs766508625	D31N	N	N	PSD	N	N	N	U	N	N	N	U	PSD	N	U
rs770293857	S87R	N	N	PSD	D	D	N	U	N	N	N	U	B	N	U
rs772554403	A61T	N	N	PB	PSD	U	N	U	N	N	N	U	B	N	U
rs775404824	G8E	D	N	PD	D	D	D	U	N	N	N	U	PD	D	P
rs775686567	K239R	D	N	PSD	N	N	D	U	N	D	N	U	PD	N	U
rs777784952	R129S	N	N	PD	D	N	N	U	N	N	N	D	PSD	N	P
rs780606244	A67T	N	N	PB	PSD	U	N	U	N	N	N	U	B	N	U
rs781718615	P88R	N	N	PD	D	D	N	U	N	N	N	U	PD	N	P

reports are available in the Supplementary Tables 1–16). Among the five convergent deleterious missense SNPs, three (F104S, S105P and G233D) occurred in the evolutionary highly conserved region of the Bcl-2 protein (Figure 2; Supplementary Figure 17).

Molecular modelling

The complete Bcl-2 protein was modelled using the MUSTER tool. For the homology modelling PDB id:1LXL (A) was used as template; the native and mutant variants were visualized using PyMol tool (Figure 3 a; Supplementary Figures 8–13).

Model validation

The modelled proteins were validated using the PDBsum server. The Ramachandran plot of the native and all variant mutant proteins showed that 90% of residues occurred in the core and additional allowed regions. For a good protein structure, it is expected that there should be 90% of the residues in the core or favoured regions of the protein. So, the native and all mutant proteins were structurally stable. In the native protein, 84.2% of amino acid residues were observed in the core or the most favoured regions, and 10.7% were observed in additional allowed regions. In the mutant protein model, more than 80% of the amino acid residues were present in the core region of
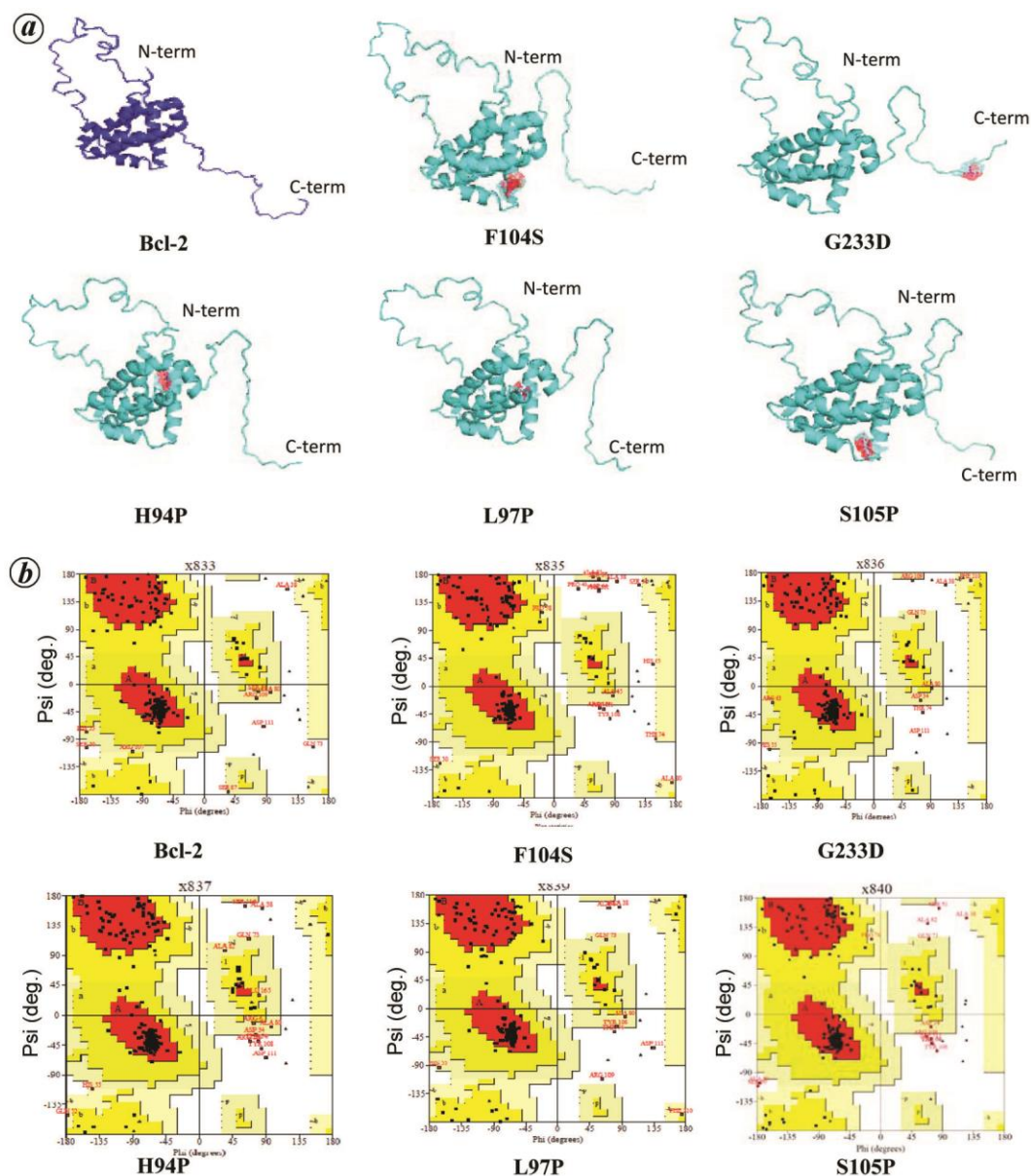
Figure 3. *a*, Ribbon structure of native and mutant Bcl-2 protein, red colour mesh indicates the mutant site (visualized in PyMol tool). *b*, Ramachandran plot of native and mutant Bcl-2 proteins, generated in PDBsum and PROCHECK server. (More details are provided in the Supplementary Figures 1–13).

the Ramachandran plot, except for L97P and S105P variants (Table 2 and Figure 3 *b*; Supplementary Figures 1–6).

Structural effect of convergent deleterious missense SNPs

The promotif structural documentation of mutant proteins showed an entirely different structural conformational arrangement compared to native Bcl-2 (Table 2). The deleterious mutations almost shared similar physicochemical characteristic features compared to the native Bcl-2 protein. The only variation was in the hydrophobi-city nature. The site-specific mutations alter the hydrophobicity nature of proteins; they will alter the normal interactions between other molecules and proteins (Table 2). TM-align score is used to understand the topological similarities and structural difference between the native and mutant protein models, the TM score between 0.7 to 0.8 indicates the all mutant protein showed insignificant structural deviations from the native protein (Table 2). The total ASA of L97P (17972.4 Å) was greater than the native ASA (1.7889.1 Å). The other variants had a lower total ASA than the native model. In the case of percentage of side ASA, hydrophobicity of F104S (44.71) was more than that of native model (43.88), whereas

Table 2. Results of Ramachandran plot analysis (generated by PDBsum). Promotif documentation generated in PROCHECK. Physico-chemical properties of native and mutant proteins were generated in ExPASy ProtoParam server. pI, Isoelectric point; MW, Molecular weight; II, Instability index; AI, Aliphatic index; EC, Extinction coefficient; GRAVY, Hydrophobicity. TM-Align score (0.0 < TM Score < 0.30 random structural similarity and 0.5 < TM score < 1.00 in about the same fold)

| | Ramachandran plot | | Promotif documentation | | | | Physico-chemical properties | | | | | | TM-align |
| | Most favoured regions (%) | Additional allowed regions (%) | Beta turns | Gamma turns | Helix | Helix–helix interactions | pI | Mw (Da) | II | AI | EC | GRAVY | TM score |
Model													
Native	84.2	10.7	25	2	9	20	6.75	26265.89	51.63	78.03	45045	–0.136	*******
G233D	84.3	10.7	24	3	9	21	6.75	26205.79	53.56	78.03	45045	–0.151	0.73747
L97P	79.1	15.8	22	3	9	21	6.49	26323.93	51.99	78.03	45045	–0.149	0.73574
H94P	80.6	12.2	20	4	10	22	6.70	26225.87	52.44	78.03	45045	–0.129	0.72465
F104S	83.8	9.1	18	5	9	21	6.75	26249.85	50.51	76.40	45054	–0.159	0.75052
S105P	77.6	16.3	22	6	9	21	6.75	26275.93	51.32	78.03	45054	–0.139	0.70017

Table 3. Energy values of wild-type protein and variant models after minimization by NOMAD Ref-Gromacs server. RMSD values (generated by TM-Align) assess the backbone movement, while native and mutant Bcl-2 protein stabilities analysed by the I-Mutant server

| | | Solvent accessible surface area (ASA) | | | | | | I-Mutant server | | |
Protein	Nucleotide variation	Total ASA (Å)	Exposed nonpolar ASA (Å)	Exposed polar ASA (Å)	Exposed charged ASA (Å)	% side ASA hydrophobic	Total energy after EM (kJ/mol)	RMSD (Å)	I-mutant score	RI	DDG value (kcal/mol)
Native	Nil	17889.1	11351.2	3306.7	3231.1	43.88	–11931.557	0.00	Nil	Nil	Nil
G233D	G–A	17712.0	11124.8	3343.5	3243.7	42.92	–12166.502	2.94	Decrease	6	–0.97
L97P	T–C	17972.4	11301.1	3263.8	3407.5	42.65	–11788.940	2.66	Decrease	4	–1.75
H94P	A–C	17866.0	11097.2	3376.0	3392.8	41.84	–11997.131	2.68	Increase	7	0.33
F104S	T–C	17524.4	11383.4	3192.4	2948.6	44.71	–11535.550	3.31	Decrease	9	–1.98
S105P	T–C	16819.0	10581.8	3033.5	3203.6	42.93	–11742.408	2.72	Increase	5	–0.44

the remaining four SNPs had lower values. This deviation from the native model indicates structural changes in the modelled variants and also loss of interactions (Table 3).

The five mutant proteins were chosen to study the deleterious effects to mimic the *in vivo* folding conditions and parameters. Energy minimization of native and all mutant proteins was done using the Nomad-Ref Gromacs server and conjugant gradient force fields. Total energy after energy minimization (EM) of native Bcl-2 protein was –11931.557 kJ/mol. G233D (–12166.502 kJ/mol) and H94P (–11997.131 kJ/mol) mutant proteins showed the highest energy minimization compared to native protein, while the others showed small variations compared to native protein, i.e. L97P (–11788.940 kJ/mol), S105P (–11742.408 kJ/mol), and F104S (–11535.550 kJ/mol). These energy variations alter the protein functions and stability (Table 3). The root mean square deviation (RMSD) value assesses the backbone movement between native and mutant proteins. Among the five mutants, F104S was found to have the highest RMSD value of 3.31 Å. RMSD value of remaining variants was below 3 Å, i.e. G233D (2.94 Å), S105P (2.72 Å), H94P (2.68 Å), and L97P (2.66 Å). These structural changes can affect the normal functions of the protein (Table 3). The I-Mutant server predicts the stability of the mutant protein. Among the five missense SNPs, the stability of

two mutants increased (H94P and S105P), while that of the remaining variants decreased (Table 3).

Project HOPE server was used to study the structural effect of mutations. The mutant residue, variant F104S, is smaller than the wild type; it will cause space in the protein's core, and wild type residue is more hydrophobic than the mutant residue. In both the PDB file and the PISA assembly (PISA database contains protein assemblies that are highly likely to be biologically relevant), this residue was involved in a multimer contact. The mutation is located within a stretch of residues annotated in UniProt as a special motif: BH3. The differences in amino acid properties will disturb the motif and probably affect its function.

In G233D, the mutant residue was larger than the wild-type residue, leading to bumps. The residue is located in a region annotated in the UniProt database as a transmembrane domain. The wild-type residue had neutral charge, and the mutant residue was negative as the wild-type residue was more hydrophobic than the mutant residue. As a result, this can cause repulsion with other residues in the protein or ligand and lipid membrane. The wild-type residue is glycine; this flexible residue might be necessary for the protein function. Since glycine is the only flexible amino acid that can make torsion angles, mutation into another residue will force the local backbone into an incorrect confirmation and disturb the local

structure. H94P variant is located within the stretch residue annotated in UniProt as a unique motif: BH3. The difference in amino acid properties will disturb the motif and probably affect its function. The mutant residue is smaller than the wild-type residue; this will cause space in the protein's core region. The hydrophobicity of the wild-type and mutant residues differs; thus the mutation will cause loss of hydrogen bonds in the core of the protein and as a result, disturb the correct folding.

In the L97P variant, the mutation is located within a stretch of residue annotated in UniProt as a special motif: BH3. Mutant residue proline disrupts an α-helix (proline disrupts an α-helix when not located at one of the first three positions of that helix), and this can have severe effects on the structure of the protein. The mutant residue is smaller than the wild-type residue, and this mutation also results in empty space in the core of the protein. In the variant S105P, the mutation is also located within a stretch of residues annotated in UniProt as a special motif: BH3. The mutant residue proline disturbs the α-helix, which can severely affect the protein structure. The mutant residue is more significant than the wild-type residue that might lead to bumps, and hydrophobicity of the wild type and mutant residues differs; thus, there is repulsion with other residues in the protein or ligands. The residue is located on the protein; its mutation can disturb interactions with other molecules or other parts of the protein.

Post-translational modification

In Bcl-2, among the five mutants, only two were present in the post-translation modification site (H94P and F104S). These sites act as proteolytic cleavage, and show a low confidence level. In F10S, site-specific mutation does not alter the proteolytic cleavage property, but in H94P the proteolytic cleavage activity is enhanced. The high expression level of post-translational modification changes the biological property of Bcl-2.

Ligand binding site of Bcl-2 protein

The FT-site sever accurately identifies the binding site. This method is based on experimental evidence that the ligand site also binds small molecules of various shapes and polarities. The FT-site server depicted three ligand sites in Bcl-2 with corresponding amino acids in these sites. Only the HIS × 94 position was present in the third ligand-binding site. So, the mutation H94P may be more deleterious, because it can affect the ligand-binding affinity of the Bcl-2 protein (Table 4 and Figure 4 *a* and *b*).

Discussion

Identification of SNPs in the coding region of a gene that may have implications in inherited human diseases is the fundamental research objective in medical genetics[12,41]. Several studies have attempted to predict the functional effects of SNPs, namely whether they are deleterious or neutral[15]. This study focused on the functional effects of convergent deleterious missense SNPs of Bcl-2 protein. In 37 missense SNPs, five showed the damaging and deleterious effects on Bcl-2. Among these three convergent deleterious missense SNPs (F104S, G233D, and S105P) occurred in the evolutionary highly conserved region of the genome. So, this deleterious mutation is evolutionarily highly conserved, and it will be inherited from one generation to the next[28,29,42]. Models for the mutated proteins depict varying energies compared to the native protein structure. The varying energy indicates damage to protein structure leading to abnormal protein function[37]. Variants H94P and S105P were found to increase the stability of the protein structure; the other three missense SNPs (F104S, G233D, and L97P) were found to have decreased protein stability. The Ramachandran plot[32] analysis of native Bcl-2 and variants protein 90% of residue occurred in core and additional allowed regions. TM-score clearly indicated that the mutant protein shows structural deviation compared to the native protein that will alter the normal protein function and may cause deleterious effects[37]. In FT-site server analysis, the H94P mutant was present in the ligand-binding site of Bcl-2 protein[43]; so, this mutation will affect the ligand-binding affinity[37,44].

The all convergent deleterious mutant residue differs in size and charge compared to the native residue. This will alter the normal interaction and stability of the protein. G233D is located in a region annotated in the UniProt database as a transmembrane domain. This mutation can cause repulsion with other residues in the protein or ligand and lipid membrane. The native glycine amino acid gives flexibility to the protein, which might be necessary for protein function. This mutation may lead to incorrect conformation and disturb the local structure. In

Table 4. FT-site, ligand-binding sites of Bcl-2 protein model and amino acids occurring in the respective sites. The predicted variant H94P was present in the ligand-binding sites. So the variant affects the ligand-binding affinity of the protein

FT-site server		
Binding site 1	Binding site 2	Binding site 3
THR × 7	ARG × 12	GLU × 13
ASP × 10	VAL × 15	MET × 16
PRO × 88	MET × 16	LYS × 17
VAL × 89	TRP × 30	HIS × 20
PRO × 90	ASP × 34	TYR × 21
PRO × 91	VAL × 35	TRP × 30
VAL × 92	GLY × 35	ASP × 31
VAL × 93	ASP × 171	HIS × 94
TRP × 195	ALA × 174	
ASP × 196	LEU × 175	
VAL × 199	THR × 178	
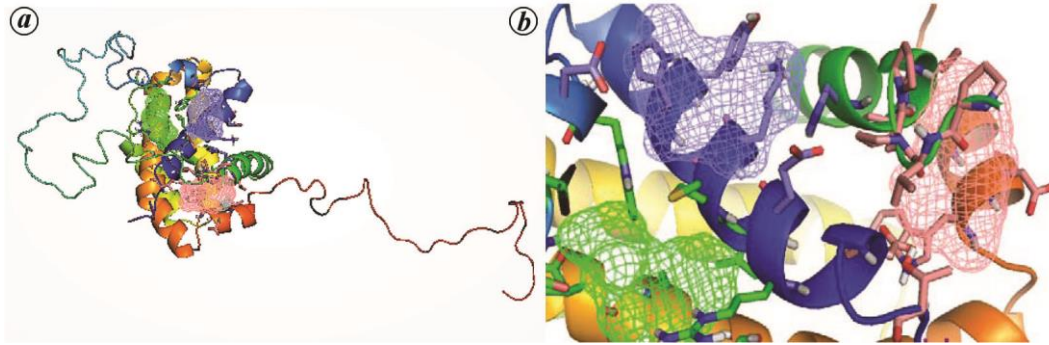
Figure 4. *a*, Ligand-binding site of Bcl-2 protein generated by FT-site server depicted in blue, green and pink mesh-like structure. *b*, Enlarged view of the ligand-binding site.

the case of L97P and S105P, presence of proline alter the alpha-helix structural conformation of the Bcl-2 and it will cause severe effects of protein structure and function. F104S residue was found to be involved in a multimer contact, and S105P mutation located on the protein surface; mutation of this residue can disturb interactions with other molecules or other parts of the protein[4,5]. Except for G233D variants, other missense mutations (S105P, F104S, L97P and H94P) were located within a residue stretch annotated in UniProt as a special motif: BH3. In the Bcl-2 protein, one of the molecular functions is BH3 domain binding. This domain is a potent death domain and has an essential role in protein–protein interactions and cell death[7,43,45]. Bax and Bak contain unique domains (BH3 domains) necessary for binding to Bcl-2 and Bcl-XL respectively[46–48]. Disruption of BH3 binding is sufficient to disrupt the formation of the full protein–protein heterodimers[8]. Biochemical and genetic studies indicate that BH3-only proteins function as triggers of cell death[7,8,49,48]. A new approach to enhancing the efficacy of endocrine treatment in estrogen receptor- positive (ER+) breast cancer is by negating the anti-apoptotic properties of BCL-2 using BH3 mimetics[50,51].

The Bcl-2 anti-apoptotic protein is best known for its ability to inhibit apoptosis[8,10,52]. Various studies however, have suggested that it may undergo a transformation from protector to killer under some conditions[4,14,53]. For example, proteolytic removal of N-terminal sequences by caspase-mediated cleavage reverses the phenotype of Bcl-2 protein (ref. 54). Interestingly, mutations in translocated BCL-2 alleles have been identified in human lymphomas that ablate the aspartic acid residue required for caspase cleavage[55], suggesting that some tumours may evolve strategies for avoiding Bcl-2 phenotype reversal[53]. Besides, the orphan nuclear receptor Nur77 can be induced to translocate from nucleus to cytosol, binding Bcl-2 and effecting a conformational change in it that probably mimics what happens during caspase cleavage[9], exposing the customarily buried BH3 domain of Bcl-2 and causing it to function as a proapoptotic protein[7,53,56]. From the

above analysis all deleterious mutants alter the structural conformation of the protein and also its biological properties, mainly protein–protein interactions or interactions with other molecules. Such mutations may lead to several types of lymphomas and may also be linked with tumour heterogeneity. Additional *in vivo* studies are needed to address the expression pattern of the deleterious missense SNPs.

Conclusion

The validated human missense Bcl-2 SNPs were initially analysed using 14 different categories of computational tools. More than 12 tools showed convergent deleterious effects that can be used for further analysis. Among the 37 missense SNPs, five showed convergent deleterious effects. Among these, three SNPs were present in the evolutionary highly conserved region of the Bcl-2 protein. This mutant is inherited from one generation to the next[14]. Mutation in H94P affects the ligand-binding affinity of the protein, controlling this expression by masking the ligand-binding association as therapeutical practices[29,52]. Mutations (H94P) affect the distinct protein interactions and may lead to tumour heterogeneity. The structural effects of mutations other than G233D, disrupt the binding of BH3 (ref. 45). This cancer mutation affects specific Bcl-2 activities[14]. So, the mutation functionally affects both oncogenes and tumour suppressors[7]. This observation strongly suggests that the mutated Bcl-2 will contribute to tumorigenesis in different ways, depending on where the mutation is located[49]. In summary, the mutation of Bcl-2 affects homo- and heterodimerization. It is important for tumour heterogeneity with great implications for clinical outcomes[6,56]. It also supports the recent discovery that protein–protein interactions within cancer gene mutations may play a critical role in tumorigenesis[13,44]. A systematic *in silico* analysis of deleterious missense SNPs of Bcl-2 and the potential structural and functional impacts of variants have not been analysed yet.

The data available from this study may help in further *in vitro* and population genetics studies to determine the impact of these variations on the Bcl-2 protein structure and function. This will lead to advances in rational drug design based on Bcl-2 specific SNPs[56].

Conflict of interests: The authors declare that they have no competing interests.

1. Green, D. R. and Reed, J. C., Mitochondria and apoptosis. *Science*, 1998, **281**, 1309–1312.
2. Favaloro, B. *et al.*, Role of apoptosis in disease. *Aging*, 2012, **4**(5), 330–349.
3. Adams, J. M. and Cory, S., The Bcl-2 protein family: arbiters of cell survival. *Science*, 1998, **281**, 1322–1326.
4. Adams, J. M. and Cory, S., The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene*, 2007, **26**, 1324–1337.
5. Andreeff, M. *et al.*, Expression of Bcl-2 related genes in normal and AML progenitors: changes induced by chemotherapy and retinoic acid. *Leukaemia*, 1999, **13**, 1881–1892.
6. Hanada, M. *et al.*, Structure–function analysis of Bcl-2 protein. Identification of conserved domains important for homodimerization with Bcl-2 and heterodimerization with Bax. *J. Biol. Chem.*, 1995, **270**, 11962–11969.
7. Huang, D. C. and Strasser, A., BH3-only proteins essential initiator of apoptotic cell death. *Cell*, 2000, **103**, 839–842.
8. Maiuri, M. C. *et al.*, BH3-only proteins and BH3 mimetics induce autophagy by competitively disrupting the interaction between Beclin 1 and Bcl-2/Bcl-XL. *Autophagy*, 2007, **3**(4), 374–376.
9. Yang, S. E. *et al.*, Down-modulation of Bcl-Xl, release of cytochrome *c* and sequential activation of caspases during honokiol induced apoptosis in human squamous lung cancer CH27 cells. *Biochem. Pharamcol.*, 2002, **63**(9), 1641–1651.
10. Martinou, J. C. and Youle, R. J., Mitochondria in apoptosis: Bcl-2 family members and mitochondrial dynamics. *Dev. Cell*, 2011, **21**(1), 92–101.
11. Gross, A., McDonnell, J. M. and Korsmeyer, S. J., Bcl-2 family members and the mitochondria in apoptosis. *Genes Dev.*, 1999, **13**(15), 1899–1911.
12. Anand, P. P., Computational modelling of human sarcomeric telethonin protein and predicting the functional effect of missense single nucleotide polymorphism. *Curr. Sci.*, 2019, **117**(4), 638–648; doi:10.18520/cs/v117/i4/638-648.
13. Zhang, Y. *et al.*, An antiapoptotic Bcl-2 family protein index predicts the response of leukemia cells to the pan-Bcl-2 inhibitor S1. *BJC*, 2013, **108**, 1870–1878.
14. Garcia Saez, A. J., The secrets of the Bcl-2 family. *Cell Death Differ.*, 2012, **19**, 1733–1740.
15. Zhang, Z. *et al.*, Analyzing effects of naturally occurring missense mutations. *Comput. Math. Methods Med.*, 2012, 805827; doi: 10.1155/2012/805827.
16. Rodrigues, C. *et al.*, Performance of *in silico* tools for the evaluation of UGT1A1 missense variants. *Hum. Mutat.*, 2015, **36**, 1215–1125.
17. Sherry, S. T. *et al.*, dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, 2001, **29**, 308–311.
18. Pires, A. S. *et al.*, *In silico* analyses of deleterious missense SNPs of human apolipoprotein E3. *Sci. Rep.*, 2017, **7**, 2509.
19. Kumar, P., Henikoff, S. and Ng, P. C., Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protoc.*, 2009, **4**, 1073–1081.
20. Choi, Y. *et al.*, Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE*, 2012, **7**, e46688; doi:10.1371/journal.pone.0046688.
21. Mi, H. *et al.*, The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids*, 2005, **33**,284–288.
22. Bendl, J. *et al.*, Predict SNP: robust and accurate consensus classifier for prediction of diseases-related mutations. *PLoS Comput. Biol.*, 2014, **10**, e1003440.
23. Adzhubei, I. *et al.*, A method and server for predicting damaging missense mutations. *Nature Methods*, 2010, **7**, 248–249.
24. Niroula, A., Urolagin, S. and Vihinen, M., PON-P2: prediction method for fast and reliable identification of harmful variants. *PLoS ONE*, 2015, **10**, e0117380.
25. Cerami *et al.*, The cBIo cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, 2012, **2**, 401–404.
26. Gao, J. *et al.*, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, 2013, **6**.
27. Szklarczk, D. *et al.*, The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.*, 2017, **45**, D362–D368.
28. Askenazy, H. *et al.*, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, 2010, **38**, W529–W533; doi:10.1093/nar/gkq399.
29. Celniker, G. *et al.*, ConSurf: using evolutionary data to raise testable hypothesis about protein function. *Isr. J. Chem.*, 2013, **53**, 199–206.
30. Sitao, W. and Zhang, Y., MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. *Proteins*, 2008, **72**(2), 547–556.
31. Laskowaski, R. A. *et al.*, PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 1993, **26**, 283–291.
32. Gopalakrishnan, K. *et al.*, Ramachandran plot on the web (2.0). *Protein Pept. Lett.*, 2007, **14**, 669–671.
33. Arnold, K. *et al.*, The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 2006, **22**, 195–201.
34. Laskowaski, R. A., Chistyakov, V. V. and Thornton, J. M., PDB-sum more: new summaries and analyses of the known 3D structure of proteins and nucleic acids. *Nucleic Acids Res.*, 2005, **33**, 266–268.
35. Venselaar, H., Beek, T. A. H. T., Kuipers, R. K. P., Hekkelman, M. L. and Vriend, G., Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformat.*, 2010, **11**(548); doi: 10.1186/1471-2105-1-548.
36. Anand, P. P. and Shibu Vardhanan, Y., Computational modelling of wet adhesive mussel foot proteins (Bivalvia): insights into the evolutionary convolution in diverse perspectives. *Sci. Rep.*, 2020, **10**, 2612; doi:https://doi.org/10.1038/s41598-020-59169-y.
37. Lindahl, E. *et al.*, NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all atom normal mode analysis. *Nucleic Acids Res.*, 2006, **34**, 52–60.
38. Gopalakrishnan, C. *et al.*, Computational modelling of complete HOXb13 protein for predicting the functional effect of SNPs and associated role in hereditary prostate cancer. *Sci. Rep.*, 2017, **7**(43830); doi:10.1038/srep43830.
39. Pejaver, V., *et al.*, The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci.*, 2014, **23**(8), 1077–1093.
40. Kozakov, D. *et al.*, The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nature Protoc.*, 2015, **10**(5), 733–755.
41. Wang, Z. and Moult, J., Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death. *Science*, 2001, **292**, 727–730.
42. Kall, L., Krogh, A. and Sonnhammer, E. LL., Advantages of combined transmembrane topology and signal peptide prediction – Phobius

web server. *Nucleic Acids Res.*, 2007, **35**; doi:10.1093/nar/gkm256.

43. Day, C. L. *et al.*, Solution structure of pro-survival Mcl-1 and characterization of its binding by proapoptotic BH3-only ligands. *J. Biol. Chem.*, 2005, **280**, 4738–4744.

44. Proto, W. F., Franco, O. L. and Alencar, S. A., Computational analysis and prediction of guanylin deleterious SNPs. *Peptides*, 2015, **69**, 92–102.

45. Frenzel, A. *et al*., Suppression of B-cell lymphomagenesis by the BH3-only protein Bmf and bad. *Blood*, 2010, **15**, 995–1005.

46. Chittenden, T. *et al.*, A conserved domain in Bak, distinct from BH1 and BH2, mediates cell death and protein binding functions. *EMBO J.*, 1995, **14**, 5589–5596.

47. Zha, H. *et al.*, Proapoptotic protein Bax heterodimerizes with Bcl-2 and homodimerizes with Bax via a novel domain (BH3) distinct from BH1 and BH2. *J. Biol. Chem.*, 1996, **271**, 7440–7444.

48. Willis, S. *et al*., The Bcl-2 regulated apoptotic pathway. *J. Cell Sci.*, 2003, **116**, 4053–4056; doi:10.1242/jcs.00754.

49. Letai A. *et al.*, Distinct BH3 domains either sensitize or activate mitochondrial apoptosis, serving as prototype cancer therapeutics. *Cancer Cell*, 2002, **2**, 183–192.

50. Martin, L. A. and Dowsett, M., BCL-2: a new therapeutic target in estrogen receptor-positive breast cancer? *Cancer Cell*, 2013, **24**(1), 7–9; doi:10.1016/j.ccr.2013.06.006.

51. Diaz, J. L. *et al*., A common binding site mediates heterodimerization and homodimerization of Bcl-2 family members. *J. Biol. Chem.*, 1997, **272**(17), 11350–11355.

52. Van Delft, M. F. and Huang, D. C., How the Bcl-2 family of proteins interact to regulate apoptosis. *Cell Res.*, 2006, **16**(2), 203–213.

53. Lin, B. *et al.*, Conversion of Bcl-2 from protector to killer by interaction with nuclear orphan receptor TR3/NGFI-B/Nur77. *Cell*, 2004, **116**, 527–540.

54. Chen, Z. X. and Prevaiz, S., Involvement of cytochrome *c* oxidase subunits Va and Vb in the regulation of cancer cell metabolism by Bcl-2. *Cell Death Differ.*, 2010, **17**, 408–420.

55. Tanaka, S. *et al.*, Frequent somatic mutations in translocated Bcl-2 gene of non-Hodgkin's lymphoma patients. *Blood*, 1992, **79**, 229–237.

56. Reed, J. C., Bcl-2-family proteins and hematologic malignancies: history and future prospectus. *Blood*, 2008, **111**(7), 3322–3330.