Biochemistry and Biophysics Reports 42 (2025) 101957

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

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

In-silico analysis of nsSNPs in *BCL-2* family proteins: Implications for colorectal cancer pathogenesis and therapeutics

Amanda Shen-Yee Kong [a], Yong Chiang Tan [b], Hin-Yee Thew [c], Kok-Song Lai [d], Swee-Hua Erin Lim [d], Sathiya Maran [c,**], Hwei-San Loh [a,*]

[a] School of Biosciences, University of Nottingham Malaysia, Jalan Broga, 43500, Semenyih, Selangor Darul Ehsan, Malaysia
[b] International Medical University, 57000, Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia
[c] School of Pharmacy, Monash University Malaysia, Bandar Sunway, 47500, Subang Jaya, Selangor Darul Ehsan, Malaysia
[d] Health Sciences Division, Abu Dhabi Women's College, Higher Colleges of Technology, Abu Dhabi, 41012, United Arab Emirates

ARTICLE INFO

Keywords:
Colorectal cancer
BCL-2 family proteins
nsSNPs
Molecular docking
Diagnostic markers
Therapeutic interventions

ABSTRACT

Colorectal cancer (CRC) is a multifaceted disease characterized by abnormal cell proliferation in the colon and rectum. The *BCL-2* family proteins are implicated in CRC pathogenesis, yet the impacts of genetic variations within these proteins remains elusive. This *in-silico* study employs diverse sequence- and structure-based bioinformatics tools to identify potentially pathogenic nonsynonymous single nucleotide polymorphisms (nsSNPs) in *BCL-2* family proteins. Leveraging computational tools including SIFT, PolyPhen-2, SNPs&GO, PhD-SNP, PANTHER, and Condel, 94 nsSNPs were predicted as deleterious, damaging, and disease-associated by at least five tools. Stability analysis with I-Mutant2.0, MutPred, and PredictSNP further identified 31 nsSNPs that reduce protein stability. Conservation analysis highlighted highly functional, exposed variants (rs960653284, rs758817904, rs1466732626, rs569276903, rs746711568, rs764437421, rs779690846, and rs2038330314) and structural, buried variants (rs376149674, rs1375767408, rs1582066443, rs367558446, rs367558446, rs1319541919, and rs1370070128). To explore the functional effects of these mutations, molecular docking and molecular dynamics simulations were conducted. G233D (rs376149674) and R12G (rs960653284) mutations in the *BCL2* protein exhibited the greatest differences in docking scores with D-α-Tocopherol and Tocotrienol, suggesting enhanced protein-ligand interactions. The simulations revealed that D-α-Tocopherol and Tocotrienol (strong binders) contributed to greater stability of *BCL-2* family proteins, while Fluorouracil, though weaker, still demonstrated selective binding stability. This work represents the first comprehensive computational analysis of functional nsSNPs in *BCL-2* family proteins, providing insights into their roles in CRC pathogenesis. While these findings demand experimental validation, they hold great promise for guiding future large-scale population studies, facilitating drug repurposing efforts, and advancing the development of targeted diagnostic and therapeutic modalities for CRC.

1. Introduction

Colorectal cancer (CRC) is a leading cause of cancer-related deaths worldwide [1]. In Malaysia, CRC is the third most common cancer, with an age-standardized incidence rate of 19.6 per 100,000 [2]. The incidence of CRC is particularly high among Chinese Malaysians, who have an age-standardized incidence rate of 27.35 per 100,000 [3]. The molecular mechanisms of CRC are not fully understood, but research suggests that a number of factors contribute to the development of the disease, including genetic mutations, environmental factors, and lifestyle choices. Despite the advances that have been made in early diagnosis and clinical management of CRC, there is still a lack of universally applicable biomarkers that can be used to identify individuals at risk of invasion, metastasis, and drug resistance [4].

Apoptosis is a tightly regulated process of programmed cell death that is disrupted in tumorigenesis [5]. This disruption is often driven by the overexpression of anti-apoptotic *BCL-2* family proteins, which inhibit the intrinsic apoptotic pathway. The *BCL-2* family proteins are

* Corresponding author.
** Corresponding author.
E-mail addresses: sathiya.Maran@monash.edu (S. Maran), sandy.loh@nottingham.edu.my (H.-S. Loh).

https://doi.org/10.1016/j.bbrep.2025.101957
Received 9 December 2024; Received in revised form 12 February 2025; Accepted 14 February 2025

categorized as either pro-apoptotic or anti-apoptotic, depending on their role in modulating apoptotic [6] (Fig. 1). Pro-apoptotic proteins such as *BAX* and *BAK* promote apoptosis by activating caspases, a family of proteases responsible for cleaving cellular proteins. In contrast, anti-apoptotic proteins, such as *BCL2* and *BCL-xL* inhibit apoptosis by blocking caspase activation or promoting cell survival. In our previous review, we explored the involvement of *BCL-2* family proteins in CRC pathogenesis, and highlighting the relationship between BH3-only pro-apoptotic proteins and caspase-independent cell death, offering insights into alternative cell death mechanisms [7]. In this study, we focus on *BCL-2* family proteins to further elucidate their role in tumorigenesis and to identify promising therapeutic targets.

Single nucleotide polymorphisms (SNPs) are the most prevalent form of genetic variation in the human genome, accounting for approximately 90 % of all genetic variations [8]. Among these, non-synonymous SNPs (nsSNPs) represents a critical subset that induces amino acid substitutions, potentially altering protein structure and function. While not all nsSNPs are deleterious, identifying those with pathogenic potential is crutial. Recent studies have highlighted the role of pathogenic nsSNPs in several oncogenes, underscoring the utility of *in-silico* approaches for analyzing large datasets to identify damaging nsSNPs [9–11]. Such variants can profoundly impact protein functionality, potentially driving oncogenesis.

Despite the established significance of *BCL-2* family proteins in apoptosis and cancer, their pathogenic nsSNPs in CRC remains largely unexplored. The sheer volume of nsSNPs makes experimental characterization of each variant impractical, necessitating computational tools to prioritize potentially deleterious nsSNPs for further investigation [12]. These tools enable the initial screening of nsSNPs that may influence disease susceptibility and drug metabolism, providing a foundation for subsequent laboratory validations.

In this study, we employed *in-silico* analysis to identify pathogenic nsSNPs in *BCL-2* family proteins and explored their effects on protein three-dimensional structure and function. Additionally, we evaluated the relationship between these nsSNPs and three CRC treatment modalities — Fluorouracil, D-α-Tocopherol, and Tocotrienol — by analyzing binding energies and assessing their potential impact on therapeutic efficacy. Fluorouracil remains the gold standard in CRC chemotherapy, with a well-established role in disrupting DNA synthesis and inducing apoptosis [13]. In contrast, D-α-Tocopherol, and Tocotrienol, often marketed together as Tocotrienol-rich fractions (TRFs) containing both Tocopherols and Tocotrienols, are gaining recognition for their superior anticancer potential, particularly in CRC [14]. This dual approach enabled us to assess the potential for these compounds to enhance existing therapies by targeting apoptosis regulation through *BCL-2* family proteins. To the best of our knowledge, this is the first comprehensive analysis of pathogenic nsSNPs in *BCL-2* family proteins, shedding light on their implications for protein stability, conservation, and function in the context of CRC. A flow chart depicting the tools used for the analysis is provided in Fig. 2.

2. Materials and methods

2.1. Retrieving nsSNPs for BCL-2 family proteins from databases

nsSNPs of *BCL2* (OMIM: 151430), *BCL2L2 (BCL-W)* (OMIM: 601931), *BCL2L10 (BCL-B)* (OMIM: 606910), *BID* (OMIM: 601997), *BCL2L11 (BIM)* (OMIM: 603827), *PMAIP1 (NOXA)* (OMIM: 604959), *BBC3 (PUMA)* (OMIM: 605854), *BAK1 (BAK)* (OMIM: 600516), and *BAX* (OMIM: 600040) were extracted from the National Center for Biological Information (NCBI) dbSNP database [https://www.ncbi.nlm.nih.gov/snp/] (version GRCh37.p13) on August 2022. Relevant information regarding the nsSNPs such as SNP IDs, allele changes, positions, protein accession numbers, residue changes, and global minor allele frequencies (MAFs), were obtained from this database [9]. Amino acid sequences of these genes were acquired from the NCBI protein database [https://www.ncbi.nlm.nih.gov/protein/] (accessed on June 15, 2022).

2.2. Identifying pathogenic nsSNPs

The functional impact of nsSNPs was predicted using six bioinformatics tools: Condel (Consensus Deleteriousness score) [http://bbglab.irbbarcelona.org/fannsdb/query/condel] (accessed on August 20, 2022), SIFT (Sorting Intolerant From Tolerant) [https://sift.bii.a-star.

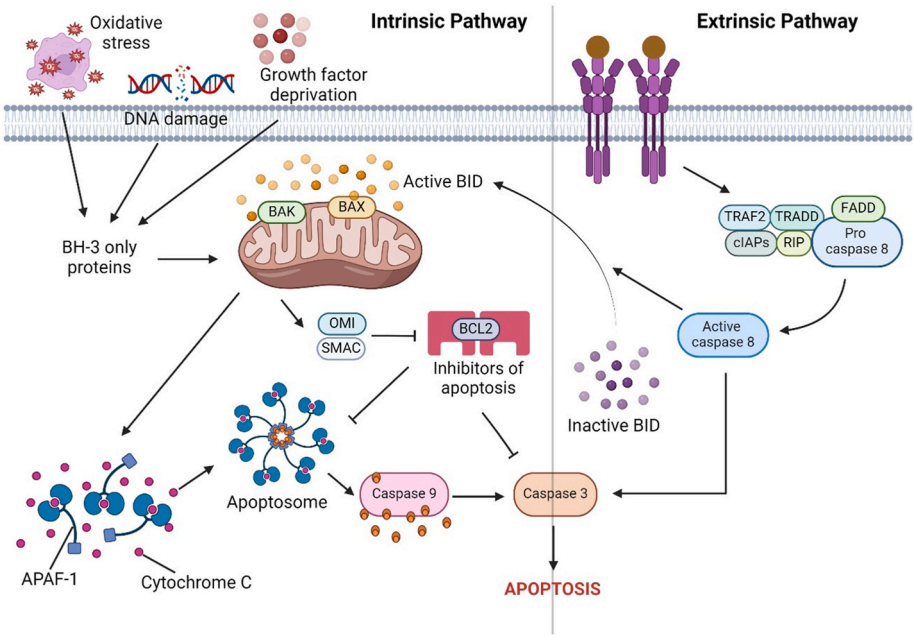

Fig. 1. A schematic representation of the apoptosis signaling pathway. Created with BioRender.com (accessed on September 17, 2024).
APAF-1, Apoptotic protease-activating factor-1; cIAPs, Cellular inhibitor of apoptosis proteins; FADD, Fas-associated death domain; OMI, High-temperature requirement; RIP, Receptor-interacting protein; SMAC; Second mitochondrial activator of caspases; TRADD, TNF receptor-associated death domain; TRAF2, TNF receptor-associated factor 2.

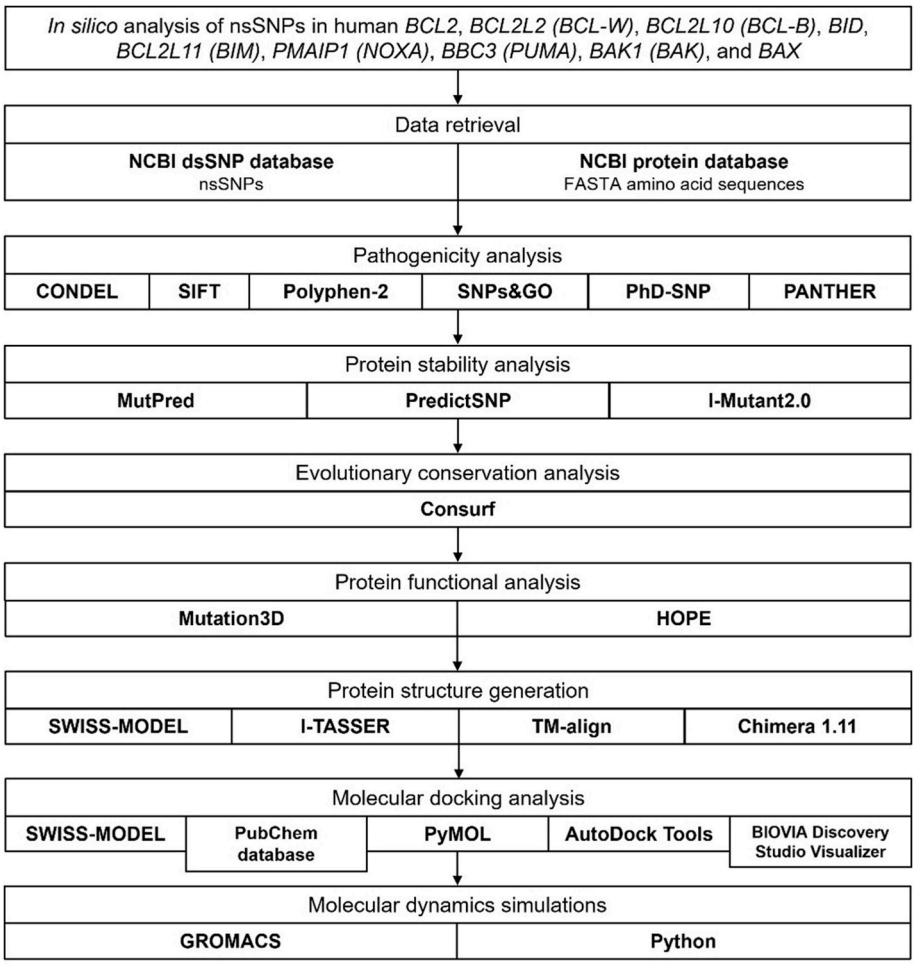

Fig. 2. Diagrammatic representation of methodology employed in this study.

edu.sg/www/Extended_SIFT_chr_coords_submit.html] (accessed on August 20, 2022), PolyPhen-2 (Polymorphism Phenotyping v2) [http://genetics.bwh.harvard.edu/pph2/] (accessed on August 19, 2022), and SNPs&GO (Single Nucleotide Polymorphisms and Gene Ontology) integrated with PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms) and PANTHER (Protein Analysis Through Evolutionary Relationships) [https://snps.biofold.org/snps-and-go/snps-and-go.html] (accessed on August 22, 2022) [15]. The pathogenic nsSNPs that were predicted to be damaging, deleterious, or disease-associated by at least five bioinformatics tools were selected for downstream analysis [9].

2.3. Protein stability assessment

The stability of nsSNPs was evaluated using three prediction tools: MutPred2 [http://mutpred.mutdb.org/#qform] (accessed on October 26, 2022), PredictSNP [https://loschmidt.chemi.muni.cz/predictsnp1/] (accessed on October 27, 2022), and I-Mutant2.0 [https://folding.biofold.org/cgi-bin/i-mutant2.0.cgi] (accessed on October 27, 2022).

MutPred predicts the probability of disease association and molecular mechanisms associated with amino acid substitutions that potentially affect the phenotype [16]. PredictSNP integrates consensus predictions from six methods, including SIFT, PhD-SNP, PolyPhen-1, PolyPhen-2, Semi-HMM-based nucleic acid parser, and multivariate analysis of protein polymorphism. This provides more accurate and reliable predictions than those of individual tools [17]. PredictSNP determines the impact of amino acid substitutions by utilizing experimental annotations from UniProt and Protein Mutant databases.

I-Mutant2.0 employs a support vector machine (SVM)-based methodology and experimental thermodynamic data from the ProTherm database to predict alterations in protein stability [18]. The degree of stability is reflected by the corresponding free energy change values (DDG), with positive values indicating increased stability. The tool also provides a reliability index (RI) from 0 to 10 to ensure reliable predictions at default settings of 25 °C and pH of 7.

2.4. Identifying functionally important regions of proteins through evolutionary conservation

The ConSurf tool [https://consurf.tau.ac.il/consurf_index.php] (accessed on October 31, 2022) was utilized to predict the evolutionary conservation and sequence homology preservation of residue position within proteins. This web-based tool employs an empirical Bayesian algorithm and phylogenetic relationships to calculate a colorimetric conservation score (1–9) for each amino acid position [9]. Residues are classified as variable (1–4), intermediately conserved (5–6), or highly conserved (7–9). Additionally, the tool determines whether a residue position is exposed on the protein surface or buried within the core. Highly conserved residues that are exposed are typically functional, whereas those that are buried are typically structural.

2.5. Protein functional analysis

Protein functional analysis utilized Mutation3D [http://www.mutation3d.org/] (accessed on November 11, 2022) and HOPE tools [https://www3.cmbi.umcn.nl/hope/input/] (accessed on November 08, 2022).

Mutation3D predicts protein function by analyzing spatial arrangements of amino acid substitutions in protein models and identifying functional hotspots from somatic mutations across cancer sequencing studies [19]. HOPE provides insights into the structural consequences of mutations and interprets disease-related phenotypes in human proteins by integrating data from various web services and databases, generating comprehensive reports, figures, and animations [20]. Both tools were employed to distinguish disease mutations from common variants, cluster mutational data, and assess the relative locations of amino acids within proteins.

2.6. Computational generation and comparison of protein structures

The 3D structures of wild-type and pathogenic nsSNPs in *BCL-2* family proteins were modeled using I-TASSER (Iterative Threading Assembly Refinement) (https://zhanggroup.org/I-TASSER/) (accessed on March 15, 2023). I-TASSER utilizes a hierarchical approach for protein structure prediction [21]. The mutant model with the highest confidence score (C-score), indicating the best compatibility, was selected for superimposition over the native structure. Similarities between the modeled wild-type and mutant protein structures were assessed using Template Modeling-align (TM-align) (https://zhanggroup.org/TM-align/) (accessed on April 6, 2023) and computed for template modeling-score (TM-score) and root-mean-square deviation (RMSD) values. TM-score ranges from 0 to 1, with 1 indicating a perfect match between both structures [22]. TM-score values between 0.0 and 0.30 indicate random structural similarity, while values between 0.50 and 1.00 indicate the same fold. Lower TM-score and higher RMSD values indicate greater structural deviation of mutant models from wild-type. Model validation was performed using ERRAT and PROCHECK Ramachandran plots from SAVES server (https://saves.mbi.ucla.edu/). The resulting structures were viewed using Chimera 1.11.

2.7. Molecular docking

Molecular docking was conducted to evaluate the binding affinities between ligands and the *BCL-2* family proteins, including *BCL2*, *BCL2L11 (BIM)*, *BAK1 (BAK)*, *BAX* [10]. The 3D structures of the ligands — Fluorouracil, D-α-Tocopherol, and Tocotrienol — were retrieved from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) (accessed on February 07, 2023) and converted into the required format using PyMOL (https://pymol.org/2/) (accessed on February 10, 2023) [23]. Protein-ligand interactions were prepared and analyzed using AutoDock Tools (version 1.5.7) (https://vina.scripps.edu/), with water molecules removed and polar hydrogens and Kollman charges added to the proteins. Specific docking pocket (binding sites of interest) was interested in the BH3 motifs of *BCL2*, *BCL2L11 (BIM)*, *BAK1 (BAK)*, and *BAX*, corresponding to amino acid sequences 93–107, 46–76, 74–88, and 59–73, respectively. The binding affinities, expressed in kcal/mol, were used to identify the most favourable protein-ligand interactions, with lower (more negative) energy scores indicating stronger binding. Additionally, an RMSD value of 0.0 was required to confirm optimal docking positions. Visualization of the docking results was conducted using BIOVIA Discovery Studio Visualizer (Discovery Studio 4.5).

2.8. Molecular dynamics (MD) simulations

Molecular dynamics (MD) simulations were conducted to evaluate the behavior of docked protein-ligand complexes in a physiological environment. The MD simulation methodology follows protocols detailed in our previous work [24]. In summary, protein-ligand complexes were pre-processed and solvated in a physiological salt condition of 0.15 M [25]. Energy minimization was performed to convergence using the steepest descent algorithm. Subsequently, the systems underwent equilibration under NVT and NPT ensembles, followed by 5-ns production MD simulations. Energy and structural parameters were extracted using GROMACS built-in functions [26], with visualizations generated in Python. For clarity, a moving average with a window of 10 data points was applied to line plots.

3. Results

3.1. Prediction of pathogenic nsSNPs and their impact on protein stability in BCL-2 family proteins

A total of 94 nsSNPs were identified as pathogenic. Variants in *BID, PMAIP1 (NOXA)*, and *BBC3 (PUMA)* were excluded from further analysis, as they did not meet the threshold of being classified as deleterious by at least five computational tools (Supplementary Table S1). Stability assessments revealed that 31 nsSNPs were pathogenic, deleterious, and destabilizing. These variants were subsequently prioritized for detailed structural analysis (Supplementary Table S2).

3.2. Evolutionary conservation analysis

Fifteen highly conserved nsSNPs were identified with conservation scored of 8 or 9: rs376149674, rs960653284, rs758817904, rs1466732626, rs569276903, rs746711568, rs764437421, rs779690846, rs1375767408, rs1582066443, rs367558446, rs768643044, rs1319541919, rs1370070128, rs2038330314. Among these, eight nsSNPs (rs960653284, rs758817904, rs1466732626, rs569276903, rs746711568, rs764437421, rs779690846, rs2038330314) were predicted to be functional and exposed, whereas the remaining seven (rs376149674, rs1375767408, rs1582066443, rs367558446, rs367558446, rs1319541919, rs1370070128) were classified as structural and buried within the protein core (Supplementary Table S3). These results underscore the detrimental effects of these 15 pathogenic nsSNPs on the structure and/or function of *BCL-2* family proteins. The amino acid conservation profiles of the *BCL-2* family proteins are presented in Supplementary Fig. S1.

3.3. HOPE analysis

The Project HOPE server revealed the structural variations caused by mutations in *BCL-2* family proteins, highlighting changes in size, charge, hydrophobicity, and spatial structure compared to their wild-type counterparts (Table 1). Of the 17 analyzed mutations, eight (G233D, W58R, G54S, G54C, G54D, V46A, V46G and G107D) were located within domain regions, while five (R12G, R127C, R88C, R127P, and G175D) were situated in motif regions. Among the mutations, residues R12G, R102C, R102P, R127C, R88C, R127P, F111L, V34G, W58R, V46A, and V46G were smaller than their wild-type counterparts, resulted in empty spaces in the protein core, disrupting hydrogen bonding, and/or impairing proper folding. Conversely, mutations G233D, G175D, G54S, G54C, G54D, and G107D involved residues larger than their wild-type counterparts, leading to their localization on the protein surface and the loss of hydrophobic interactions, with G54C notably gaining hydrophobic interactions. Additionally, charge alterations were observed: G233D, G175D, G54D, and G107D introduced negative charges, potentially repelling ligands or other residues, while W58R introduced a positive charge in a buried residue, destabilizing the protein. Mutations such as R12G, R102C, R102P, R127C, R127P, and R88C caused the loss of charges, potentially diminishing the protein's capacity for intermolecular or intramolecular interactions. These findings demonstrate that mutations in key residues can significantly disrupt the structural integrity and inter- or intramolecular interactions of *BCL-2* family proteins, underscoring their potential to impair protein function.

3.4. Visualization of harmful substitutions using Mutation3D server

The Mutation3D server identified harmful substitutions in the *BCL2*, *BCL2L11 (BIM)*, *BAK1 (BAK)*, *BAX* proteins, encoded by the *BCL2*,

Table 1
HOPE analysis of amino acid changes and their impact on the structure of *BCL2, BCL2L11 (BIM)*, *BAK1 (BAK)*, and *BAX* protein.

Amino acid change	Size of residue		Charge of residue		Change of polarity	Change of hydrophobicity	Protein folding	Loss of interactions	Location
	Wild type	Mutant	Wild type	Mutant					
BCL2									
G233D	Small	Large	Neutral	Negative	Nonpolar to Polar	Decreased Hydrophobicity	Affected	Yes	Transmembrane domain
R12G	Large	Small	Positive	Neutral	Polar to Nonpolar	Increased Hydrophobicity	Affected	Yes	BH4 motif
BCL2L11 (BIM)									
R102C	Large	Small	Positive	Neutral	Polar to Nonpolar	Increased Hydrophobicity	Affected	Yes	Not stated
R102P	Large	Small	Positive	Neutral	Polar to Nonpolar	Increased Hydrophobicity	Affected	Yes	Not stated
BAK1 (BAK)									
R127C	Large	Small	Positive	Neutral	Polar to Nonpolar	Increased Hydrophobicity	Affected	Yes	BH1 motif
R88C	Large	Small	Positive	Neutral	Polar to Nonpolar	Increased Hydrophobicity	Affected	Yes	BH3 motif
R127P	Large	Small	Positive	Neutral	Polar to Nonpolar	Increased Hydrophobicity	Affected	Yes	BH1 motif
G175D	Small	Large	Neutral	Negative	Nonpolar to Polar	Decreased Hydrophobicity	Affected	Yes	BH2 motif
F111L	Large	Small	Neutral	Neutral	Nonpolar to Nonpolar	No changes	Affected	Yes	Not stated
V34G	Large	Small	Neutral	Neutral	Nonpolar to Nonpolar	Decreased Hydrophobicity	Affected	Yes	Not stated
BAX									
W58R	Large	Small	Neutral	Positive	Nonpolar to Polar	Decreased Hydrophobicity	Affected	Yes	BCL domain
G54S	Small	Large	Neutral	Neutral	Nonpolar to polar	No changes	Affected	Yes	BCL domain
G54C	Small	Large	Neutral	Neutral	Nonpolar to Nonpolar	Increased Hydrophobicity	Affected	Yes	BCL domain
G54D	Small	Large	Neutral	Negative	Nonpolar to Polar	Decreased Hydrophobicity	Affected	Yes	BCL domain
V46A	Large	Small	Neutral	Neutral	Nonpolar to Nonpolar	No changes	Affected	No	BCL domain
V46G	Large	Small	Neutral	Neutral	Nonpolar to Nonpolar	Decreased Hydrophobicity	Affected	Yes	BCL domain
G107D	Small	Large	Neutral	Negative	Nonpolar to Polar	Decreased Hydrophobicity	Affected	Yes	BCL domain

BCL2L11, *BAK1*, *BAX* genes, respectively (Table 2). The R12G mutation was localized to the BH4 domain (PF02180) of the *BCL2* protein, while R127C, R88C, R127P, G175D, and F111L were mapped to the BCL-2 domain (PF00452) of the *BAK1 (BAK)* protein. Additionally, G107D was identified in the BCL-2 domain (PF00452) of the *BAX* protein. Of the 17 analyzed mutations, seven (R12G, R127C, R88C, R127P, G175D, F111L, and G107D) were located within domain regions, classifying them as high-risk mutations for the *BCL2*, *BAK1 (BAK)*, and *BAX* proteins. Furthermore, two mutation clusters were identified: one comprising four substitutions (R88C, F111L, R127C, and R127P) in the *BAK1 (BAK)* protein and another comprising four substitutions (W58R, G54S, G54C, and G54D) in the *BAX* protein (Supplementary Fig. S2).

Table 2
Mutation cluster prediction by Mutation3D.

SNP ID	Amino acid change	Domain	Mutation cluster
BCL2			
rs376149674	G233D	Not stated	–
rs960653284	R12G	BH4	–
BCL2L11 (BIM)			
rs758817904	R102C	Not stated	–
rs1466732626	R102P	Not stated	–
BAK1 (BAK)			
rs569276903	R127C	*BCL-2*	Cluster
rs746711568	R88C	*BCL-2*	Cluster
rs764437421	R127P	*BCL-2*	Cluster
rs779690846	G175D	*BCL-2*	–
rs1375767408	F111L	*BCL-2*	Cluster
rs1582066443	V34G	Not stated	–
BAX			
rs367558446	W58R	Not stated	Cluster
rs768643044	G54S	Not stated	Cluster
rs768643044	G54C	Not stated	Cluster
rs1319541919	G54D	Not stated	Cluster
rs1370070128	V46A	Not stated	–
rs1370070128	V46G	Not stated	–
rs2038330314	G107D	*BCL-2*	–

"-" means no cluster.

3.5. Prediction of 3D structural alteration in BCL-2 family proteins using I-TASSER

Protein models with the highest C-score were selected, and their TM-scores, RMSD values, ERRAT values, and Ramachandran plot analysis were documented (Supplementary Table S4). The Ramachandran plot analysis revealed that over 80 % of residues in the native and variant mutant proteins were located within the favoured and additionally allowed regions, with the exception of the rs758817904 (R102C) variant. Structural comparisons between native and mutant proteins were conducted via superimposition using Chimera, highlighting key structural deviations (Supplementary Figs. S3–S4).

3.6. Protein-ligand docking

The binding interactions between ligands and target proteins are summarized in Supplementary Tables S5–S7, with visual representations provided in Supplementary Figs. S5–S7. Among the studied ligands, Tocotrienol exhibited the most pronounced changes in docking scores when interacting with mutated *BCL2*, *BCL2L11 (BIM)*, *BAK1*

(BAK), *BAX* proteins, followed by D-α-Tocopherol, while Fluorouracil demonstrated the smallest differences. Notably, the R12G-Tocopherol complex showed the most significant change, achieving a docking score of −7.8 kcal/mol compared to −3.8 kcal/mol for the wild-type complex, indicating a stronger binding affinity for the mutant protein at the BH3 motif docking site. This enhanced interaction was attributed to altered binding modes: the wild-type *BCL2*-Tocopherol complex formed hydrogen bonds (Arg-107), Pi-Cation interactions (Arg-106), and Alkyl interactions (Leu-217), while the R12G-Tocopherol complex engaged in Alkyl and Pi-Pi stacked interactions involving residues Pro-204 and Trp-214. A similar trend was observed for Tocotrienol, with G233D and R12G mutations showing the largest differences in docking scores. The G233D-Tocotrienol complex achieved a docking score of −5.6 kcal/mol (wild-type: −3.5 kcal/mol), while the R12G-Tocotrienol complex achieved −7.9 kcal/mol (wild-type: −3.5 kcal/mol) at the BH3 motif docking site. For G233D, the altered binding involved the formation of a hydrogen bond with His-94 and Alkyl and Pi-Alkyl interactions with Lys-17 and Pro-91. Similarly, the R12G-Tocotrienol complex displayed a hydrogen bond with Pro-204 and Alkyl and Pi-Alkyl interactions with Trp-214. Both mutations disrupted the original hydrogen bonding (Arg-98), Pi-Cation interactions (Arg-106), and Alkyl interactions (Met-206) observed in the wild-type *BCL2*-Tocotrienol complex at the BH3 motif docking site.

3.7. MD simulations

MD simulations were performed to investigate the structural behavior and intermolecular interactions of the docked protein-ligand complexes over 5 ns in equilibrated systems. Across the simulations, the macromolecule structures remained largely consistent with their original configurations (Fig. 3A–C), exhibiting RMSD below 1 nm, and a radius of gyration of approximately 2 nm.

Analysis of average protein-ligand distances (Fig. 3D) revealed that D-α-Tocopherol and Tocotrienol maintained strong binding affinities with all *BCL-2* family proteins and most of their variants, with no significant changes in binding stability between the wild-type and variants. However, the G233D mutation in the *BCL2* protein displayed an exception, as this variant weakened intermolecular interactions with D-α-Tocopherol (Fig. 3B). This disruption, primarily affecting the Lennard-Jones potential, led to the ligand escaping the binding pocket. In contrast, Fluorouracil exhibited weak binding across all wild-type *BCL-2* family proteins, with the exception of *BIM*. Notably, no significant changes in binding stability were observed in the R127P mutant of *BAK1*, the V46G and G107D mutants of *BAX*, or both *BCL2* mutants when compared to their respective wild-types.

Interestingly, several variants restored protein-ligand interactions for Fluorouracil, which exhibited weak binding to wild-type proteins. For example, the F111L, G175D, R127C, R88C, and V34G mutations in *BAK1*, as well as the G54C, G54D, G54S, V46A, and W58R mutations in *BAX*, stabilized Fluorouracil binding (Fig. 3F). Conversely, both R102 mutations in *BIM* weakened Fluorouracil binding, even though wild-type *BIM* did not form hydrogen bonds with Fluorouracil at residue R102.

Correlation analysis of intermolecular interaction energies and average protein-ligand distances (Fig. 3E) demonstrated that the Lennard-Jones potential predominantly influenced ligand binding across all three compounds and their corresponding protein variants. In contrast, the Coulomb potential played a supplemental role, particularly contributing to the binding strength of Fluorouracil. A comprehensive hydrogen bonding map of the protein-ligand complexes throughout the MD simulations is provided in Supplementary Fig. S8.

4. Discussion

This study identified pathogenic nsSNPs in *BCL-2* family proteins and evaluated their potential implications for CRC treatment. Using a comprehensive *in-silico* approach, a suite of bioinformatic tools were employed to predict pathogenic nsSNPs associated with CRC (Fig. 2). A total of 94 nsSNPs were classified as highly deleterious based on pathogenicity analyses performed using at least five algorithms. These predictions were based on the scores generated by the selected *in-silico* tools. Protein stability analyses, conducted using I-Mutant2.0, MutPred2, and the PredictSNP suite, revealed that 31 nsSNPs were associated with reduced stability relative to the wild-type structure, indicating potential disruption of protein function. Protein stability is predominantly influenced by the physicochemical properties of amino acid residues, including polarity, charge, and hydrophobicity [27]. Hydrophobic interactions are particularly critical, contributing approximately 60 % of a protein's stability, as reported by Pace et al. (2011) [28]. Amino acid substitutions, such as R12G (rs960653284), R102C (rs758817904), R102P (rs1466732626), R127C (rs569276903), R88C (rs746711568), and R127P (rs764437421), increased hydrophobicity, which may disrupt hydration and electrostatic interactions [29], potentially leading to protein misfolding and degradation.

To assess evolutionary conservation among the identified nsSNPs, ConSurf analysis were performed. Comparative analyses of conservation scores obtained from ConSurf and I-mutant revealed that 15 nsSNPs — rs376149674 (G233D), rs960653284 (R12G), rs758817904 (R102C), rs1466732626 (R102P), rs569276903 (R127C), rs746711568 (R88C), rs764437421 (R127P), rs779690846 (G175D), rs1375767408 (F111L), rs1582066443 (V34G), rs367558446 (W58R), rs768643044 (G54S, G54C), rs1319541919 (G54D), rs1370070128, (V46A, V46G) and rs2038330314 (G107D)— were both highly conserved (scores of 8 or 9) and associated with decreased protein stability (Fig. 4). These findings suggest that these variants may have profound impacts on the structural integrity and functional activity of *BCL-2* family proteins.

The HOPE server analysis revealed that the nsSNPs were located in conserved domains of *BCL-2* family proteins. The G233D mutation (rs376149674) in the *BCL2* protein was mapped to the transmembrane domain (TMD) (Fig. 5), while mutations R12G (rs960653284), R88C (rs746711568), R127C (rs569276903), R127P (rs764437421), and G175D (rs779690846) were identified in BH1, BH2, BH3, or BH4 motifs. These motif annotations were confirmed using the UniProt database. Additionally, mutations W58R (rs367558446), G54S (rs768643044), G54C (rs768643044), G54D (rs1319541919), V46A (rs1370070128), V46G (rs1370070128), and G107D (rs2038330314) were associated with the BCL domain.

The TMB plays a critical role in the anti-apoptotic activity and homooligomerization of *BCL2* within the outer mitochondrial membrane [30,31]. The G233D mutation (rs376149674) could disrupt the domain's confirmation, potentially compromising protein function. This observation is consistent with the findings of Anand & Vardhanan (2021), who reported that the G233D mutation (rs376149674) may cause repulsion between residues or ligands, leading to incorrect protein conformation and perturbation of the local structure [32]. The native glycine residue provides flexibility to the protein, and its substitution can have significant structural and functional consequences.

A domain typically refers to a larger, independently folding structural unit within a protein, often composed of multiple motifs or functional regions, while a motif generally describes a smaller, recurring sequence or structural pattern that contributes to specific biological functions [33]. Both motifs and domains are integral to defining protein function and three-dimensional structure. The BH4 motif in *BCL2* is pivotal for its anti-apoptotic activity [34], whereas the intact BH3 motif in *BAK* and *BAX* is essential for their pro-apoptotic activity and interactions with anti-apoptotic members of the *BCL-2* family [35]. In this study, the BH3 motif was selected as the docking site to investigate whether mutations in this critical domain affect the binding interactions of selected ligands, providing insight into their potential impact on apoptosis regulation. The BH3 domain, recognized as a potent death domain, plays a central role in protein-protein interactions and the regulation of cell death [36]. Disruption of BH3 binding is sufficient to

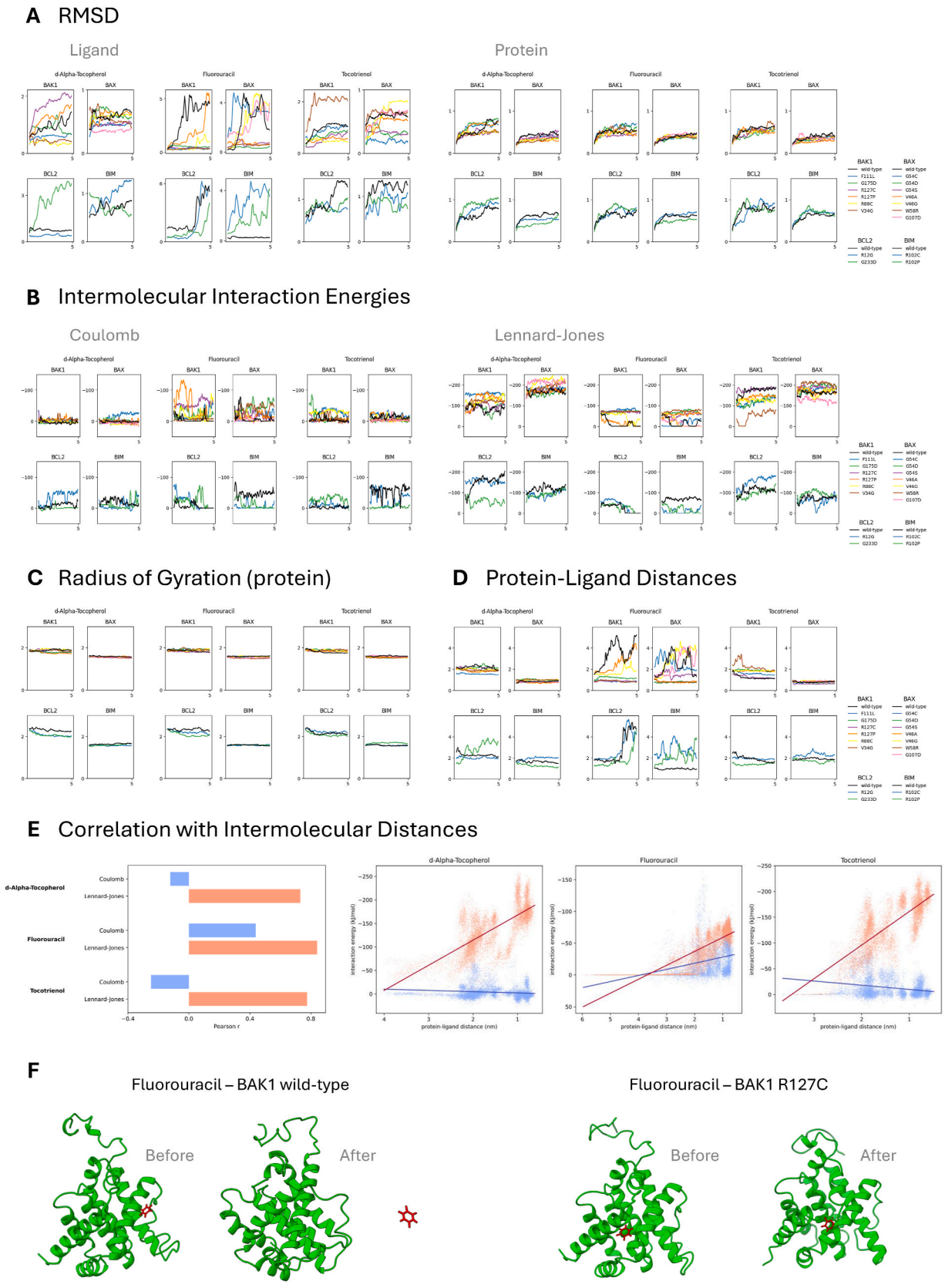

Fig. 3. Molecular dynamics (MD) simulation of d-Alpha Tocopherol, Fluorouracil, and Tocotrienol against BCL-2 family proteins and their variants. (A) RMSD of simulated ligands and proteins, (B) Coulomb and Lennard-Jones interaction energies between proteins and ligands, (C) radius of gyration of proteins, and (D) average protein-ligand distances, throughout the simulation. (E) Correlation of intermolecular interaction energies and protein-ligand distances to inspect the predominant energy models. (F) Structural visualization example of variants that gained binding strength compared to wild type.

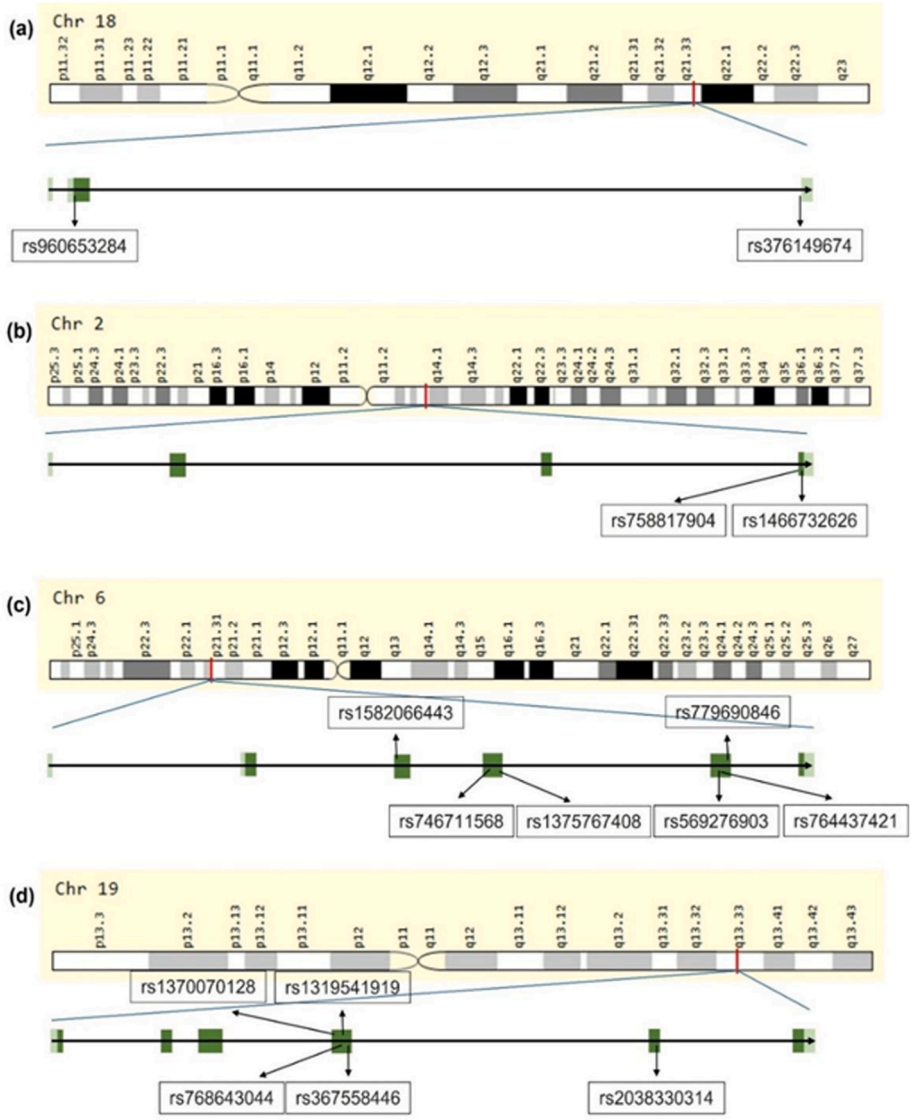

Fig. 4. Diagrammatic position of the nsSNPs on (a) *BCL2* gene located on the long arm of chromosome 18 (q21.33), (b) *BCL2L11 (BIM)* gene located on the long arm of chromosome 2 (q13), (c) *BAK1 (BAK)* gene located on the short arm of chromosome 6 (p21.31), and (d) *BAX* gene located on the long arm of chromosome 19 (q13.33). The untranslated regions are represented by light green boxes, exons are depicted by dark green boxes, and introns are shown by black horizontal lines.

impair the formation of protein-protein heterodimers [37]. BH3 mimetics, which selectively target anti-apoptotic *BCL-2* family proteins, have demonstrated significant potential in inducing apoptosis in hematologic malignancies and solid tumors [38]. These findings underscore the indispensable role of the BH3 domain in apoptosis and maintaining cellular homeostasis.

The selected ligands – Fluorouracil, D-α-Tocopherol, and Tocotrienol – are well-recognized for their therapeutic efficacy in CRC treatment through mechanisms such as apoptosis induction, DNA replication inhibition, antioxidation, anti-inflammation, anti-proliferation, and anti-angiogenesis [13,39,40]. These ligands were chosen for molecular docking studies based on their established relevance to CRC therapy and the promising therapeutic potential of Vitamin E derivatives. Fluorouracil remains the cornerstone of CRC chemotherapy, widely employed as a first-line treatment due to its ability to inhibit tumor growth through thymidylate synthase inhibitor, which disrupts DNA synthesis and induces apoptosis in proliferating cancer cells [13]. Recent research highlights the anticancer potential of TRFs, which include both Tocopherols and Tocotrienols, with Tocotrienols demonstrating superior antiproliferative, pro-apoptotic, and anti-metastatic effects compared to other Vitamin E forms [14,39]. Tocotrienols have been shown to modulate key oncogenic pathways, such as NF-kB and STAT3, which play critical roles in CRC progression [39]. Furthermore, they have been linked to the inhibition of tumor cell survival and metastasis and have demonstrated the ability to enhance the efficacy of chemotherapeutic agents when used in combination therapies [14]. While Vitamin E derivatives are traditionally recognized for their antioxidant properties, emerging evidence underscores their broader role in regulating apoptosis and inhibiting cancer cell survival [40]. Given their higher bioactivity, tocotrienols are gaining attention as potential adjuncts to conventional cancer therapies, particularly in CRC.

Molecular docking analyses revealed that these ligands interact with both wild-type and mutated *BCL-2* family proteins, with varying binding affinities. Among the tested ligands, Tocotrienol exhibited the most pronounced variations in docking scores when interacting with mutated *BCL2*, *BCL2L11 (BIM)*, *BAK1 (BAK)*, and *BAX* proteins, suggesting alterations in binding dynamics and a potential capacity to inhibit protein function. Key interactions at the active site, including hydrogen bonds, van der Waals forces, and Pi interactions, were crucial in stabilizing the ligand-protein complexes [41]. Additionally, the docking results revealed diverse interaction types, such as Pi-sigma and Pi-alkyl interactions, which may disrupt the binding of proteins critical to the

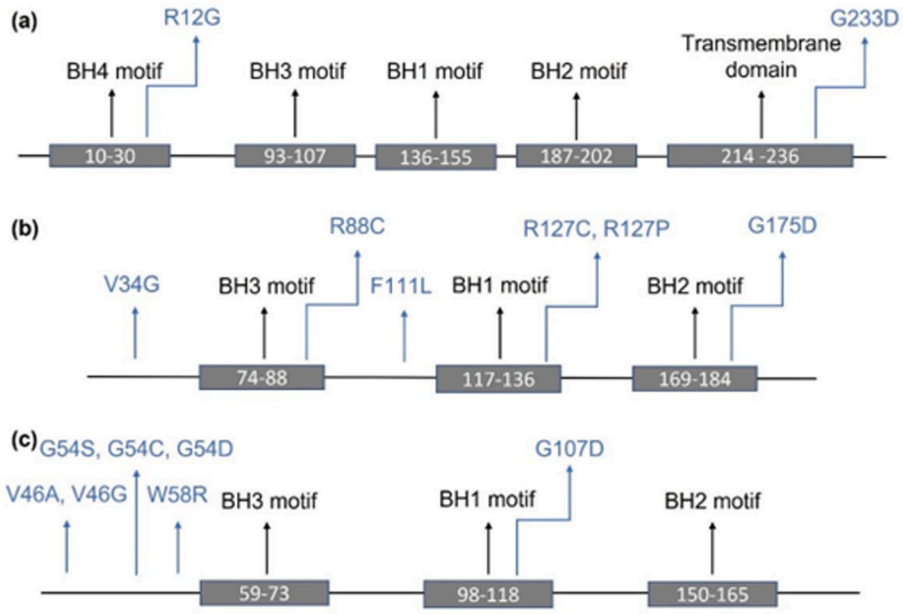

Fig. 5. Domain and motifs within the (a) *BCL2* protein, (b) *BAK1 (BAK)* protein, and (c) *BAX* protein. The positions of mutations are shown with respect to the TMD, BH1 motif, BH2 motif, BH3 motif, and BH4 motif. The domain and motifs are shown in grey horizontal bars and the mutations are depicted by blue arrows. The scaling of the mutations' positions, the domain, and the motifs are approximate.

apoptosis pathway, thereby reducing interactions among *BCL-2* family proteins. These findings provide insights into the potential mechanisms by which these ligands influence apoptosis regulation in CRC at the BH3 motif binding site.

To evaluate the impact of the three drugs on *BCL-2* family proteins, MD simulations were conducted to assess the stability of wild-type and mutant proteins bound to Fluorouracil, D-α-Tocopherol, and Tocotrienol at the BH3 motif binding site. RMSD analysis from the MD simulations provided insights into structural dynamics, with higher RMSD values indicating greater structural fluctuations. While minor fluctuations were primarily observed in the ligands, the proteins displayed stable behavior, with RMSD values consistently below 1 nm, suggesting minimal deviation from their initial structures. The simulations demonstrated that D-α-Tocopherol and Tocotrienol, both identified as strong binders, enhanced the conformational stability of *BCL-2* family proteins by guiding ligand conformations towards local energy minima. In contrast, Fluorouracil, a weak binder, exhibited less pronounced effects, on protein stability but still showed binding stability in certain conformations of the *BCL-2* family proteins. Future studies could benefit from longer MD simulations to allow ligands to explore a broader conformational space, though this approach is computationally expensive and was beyond the scope of this study due to resource limitations.

This study, primarily utilizing bioinformatics tools and web servers based on advanced mathematical and statistical algorithms, emphasizes the necessity for subsequent *in-vitro* and *in-vivo* validation, alongside genome-wide association studies (GWAS), to confirm the identified findings. The pathogenic nsSNPs identified herein may impact the efficacy of conventional CRC therapies, underscoring the potential for incorporating nsSNP screening into clinical practice. Such integrative approaches could facilitate the development of personalized treatment strategies, tailored to the genetic profiles of CRC patients, thereby optimizing therapeutic outcomes.

5. Conclusion

This study investigated the genetic variations in *BCL-2* family proteins by identifying pathogenic nsSNPs associated with CRC, with a focus on their role in apoptosis regulation. The use of diverse bioinformatics tools to predict the impact of these pathogenic nsSNPs strengthens the robustness of the findings. A total of fifteen potentially pathogenic nsSNPs (rs376149674, rs960653284, rs758817904, rs1466732626, rs569276903, rs746711568, rs764437421, rs779690846, rs1375767408, rs1582066443, rs367558446, rs768643044, rs1319541919, rs1370070128, rs2038330314) were identified, with their effects on protein stability and localization within functionally important regions characterized in detail. These mutations in *BCL2*, *BCL2L11 (BIM)*, *BAK1 (BAK)*, *BAX* proteins are shown to disrupt homo- and heterodimerization, potentially contributing to tumorigenesis and CRC pathogenesis. Molecular docking and MD simulations further revealed that D-α-Tocopherol and Tocotrienol stabilizing effects on *BCL-2* family proteins, particularly in their mutated forms, while Fluorouracil showed weaker, selective binding stability. These findings provide a foundation for development of targeted therapeutic strategies in CRC. By refining the list of missense SNPs through stringent filters, this study offers a framework for future investigations into their functional roles in disease progression. Nonetheless, experimental validation through *in-vitro*, *in-vivo* studies or GWAS is crucial, paving the way for future research, including clinical trials.

CRediT authorship contribution statement

Amanda Shen-Yee Kong: Writing – original draft, Methodology, Investigation, Formal analysis. **Yong Chiang Tan:** Methodology, Investigation, Formal analysis. **Hin-Yee Thew:** Methodology, Investigation, Formal analysis. **Kok-Song Lai:** Writing – review & editing. **Swee-Hua Erin Lim:** Writing – review & editing. **Sathiya Maran:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Hwei-San Loh:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to acknowledge the Malaysian Ministry of Higher Education (MoHE) for supporting this research work under the Fundamental Research Grant Scheme (FRGS) [FRGS/1/2020/SKK0/UNIM/01/1] and the University of Nottingham Malaysia for providing AKSY a scholarship to study her MRes program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2025.101957.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

[1] P. Karousi, et al., Identification of two novel circular RNAs deriving from BCL2L12 and investigation of their potential value as a molecular signature in colorectal cancer, Int. J. Mol. Sci. 21 (22) (2020) 8867.
[2] Cancer, I.A.f.R.o. GLOBOCAN 2020: estimated cancer incidence, mortality, and prevalence worldwide in 2020. [cited 2022 December 23]; Available from: https://gco.iarc.fr/.
[3] M.R. Abu Hassan, et al., Incidence and mortality rates of colorectal cancer in Malaysia, Epidemiol Health 38 (2016) e2016007.
[4] O.O. Ogunwobi, F. Mahmood, A. Akingboye, Biomarkers in colorectal cancer: current research and future prospects, Int. J. Mol. Sci. 21 (15) (2020).
[5] Q. Huang, et al., High expression of anti-apoptotic protein Bcl-2 is a good prognostic factor in colorectal cancer: result of a meta-analysis, World J. Gastroenterol. 23 (27) (2017) 5018–5033.
[6] C.F.A. Warren, M.W. Wong-Brown, N.A. Bowden, BCL-2 family isoforms in apoptosis and cancer, Cell Death Dis. 10 (3) (2019) 177.
[7] A.S.-Y. Kong, S. Maran, H.-S. Loh, Navigating the interplay between BCL-2 family proteins, apoptosis, and autophagy in colorectal cancer, Adv. Cancer Biol.-Metastasis 11 (2024) 100126.
[8] L.T. Rozario, T. Sharker, T.A. Nila, In silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function, PLoS One 16 (6) (2021) e0252932.
[9] S.W. Lim, et al., Functional and structural analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the MYB oncoproteins associated with human cancer, Sci. Rep. 11 (1) (2021) 24206.
[10] L. Alessandro, et al., Identification of NRAS diagnostic biomarkers and drug targets for endometrial cancer—an integrated in silico approach, Int. J. Mol. Sci. 23 (22) (2022) 14285.
[11] A. Sharma, et al., Identification of high-risk single nucleotide polymorphisms (SNPs) of epidermal growth factor receptor (EGFR) and their interaction with various TKI drugs, Eurasian J. Med. Oncol. 7 (2023) 334–344.
[12] H. Venkata Subbiah, P. Ramesh Babu, U. Subbiah, Determination of deleterious single-nucleotide polymorphisms of human LYZ C gene: an in silico study, J. Genet. Eng. Biotechnol. 20 (1) (2022) 92.
[13] S. Blondy, et al., 5-Fluorouracil resistance mechanisms in colorectal cancer: from classical pathways to promising processes, Cancer Sci. 111 (9) (2020) 3142–3154.
[14] S.-Y. Tham, et al., Tocotrienols modulate a life or death decision in cancers, Int. J. Mol. Sci. 20 (2) (2019) 372.
[15] N.A. Khoruddin, et al., Pathogenic nsSNPs that increase the risks of cancers among the Orang Asli and Malays, Sci. Rep. 11 (1) (2021) 16158.
[16] J. Remali, et al., In silico analysis on the functional and structural impact of Rad50 mutations involved in DNA strand break repair, PeerJ 8 (2020) e9197.
[17] J. Bendl, et al., PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations, PLoS Comput. Biol. 10 (1) (2014) e1003440.
[18] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (Web Server issue) (2005) W306–W310.
[19] M.J. Meyer, et al., mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome, Hum. Mutat. 37 (5) (2016) 447–456.
[20] H. Venselaar, et al., Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinf. 11 (1) (2010) 548.
[21] J. Yang, et al., The I-TASSER Suite: protein structure and function prediction, Nat. Methods 12 (1) (2015) 7–8.
[22] M.S. Hossain, A.S. Roy, M.S. Islam, In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions, Sci. Rep. 10 (1) (2020) 14542.
[23] N. Alizadehmohajer, et al., Using in silico bioinformatics algorithms for the accurate prediction of the impact of spike protein mutations on the pathogenicity, stability, and functionality of the SARS-CoV-2 virus and analysis of potential therapeutic targets, Biochem. Genet. 61 (2) (2023) 778–808.
[24] H.Y. Thew, et al., Probing the anti-Aβ42 aggregation and protective effects of prenylated xanthone against Aβ42-induced toxicity in transgenic Caenorhabditis elegans model, Chem. Biol. Interact. 394 (2024) 110978.
[25] D. Gao, et al., The destructive mechanism of Aβ(1-42) protofibrils by norepinephrine revealed via molecular dynamics simulations, Phys. Chem. Chem. Phys. 24 (33) (2022) 19827–19836.
[26] M.J. Abraham, et al., GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25.
[27] A.S. Panja, S. Maiti, B. Bandyopadhyay, Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges, Sci. Rep. 10 (1) (2020) 1822.
[28] C.N. Pace, et al., Contribution of hydrophobic interactions to protein stability, J. Mol. Biol. 408 (3) (2011) 514–528.
[29] A. Zahra, et al., In silico study to predict the structural and functional consequences of SNPs on biomarkers of ovarian cancer (OC) and BPA exposure-associated OC, Int. J. Mol. Sci. 23 (3) (2022).
[30] V. Andreu-Fernández, et al., Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes, Proc. Natl. Acad. Sci. U. S. A. 114 (2) (2017) 310–315.
[31] L.A. Caro-Gómez, et al., Exploring the conformational space of bcl-2 protein variants: dynamic contributions of the flexible loop domain and transmembrane region, Molecules 24 (21) (2019) 3896.
[32] P. Anand, Y.S. Vardhanan, Deleterious missense single nucleotide polymorphism alters the structural conformation of human anti-apoptotic Bcl-2 protein, Curr. Sci. 120 (4) (2021) 666.
[33] D.A. Korasick, J.M. Jez, Protein domains: structure, function, and methods, in: R. A. Bradshaw, G.W. Hart, P.D. Stahl (Eds.), Encyclopedia of Cell Biology, second ed., Academic Press, Oxford, 2023, pp. 106–114.
[34] A. Aouacheria, et al., Evolution of Bcl-2 homology motifs: homology versus homoplasy, Trends Cell Biol. 23 (3) (2013) 103–111.
[35] F. Aguilar, et al., Discovery of diverse human BH3-only and non-native peptide binders of pro-apoptotic BAK indicate that activators and inhibitors use a similar binding mode and are not distinguished by binding affinity or kinetics, bioRxiv (2022) 2022, 05.07.491048.
[36] A. Kelekar, C.B. Thompson, Bcl-2-family proteins: the role of the BH3 domain in apoptosis, Trends Cell Biol. 8 (8) (1998) 324–330.
[37] J.-L. Diaz, et al., A common binding site mediates heterodimerization and homodimerization of bcl-2 family members, J. Biol. Chem. 272 (17) (1997) 11350–11355.
[38] L. Zhang, et al., BH3 mimetic sensitivity of colorectal cancer cell lines in correlation with molecular features identifies predictors of response, Int. J. Mol. Sci. 22 (8) (2021) 3811.
[39] A. Abraham, et al., Vitamin E and its anticancer effects, Crit. Rev. Food Sci. Nutr. 59 (17) (2019) 2831–2838.
[40] S.K. Park, B.G. Sanders, K. Kline, Tocotrienols induce apoptosis in breast cancer cell lines via an endoplasmic reticulum stress-dependent increase in extrinsic death receptor signaling, Breast Cancer Res. Treat. 124 (2) (2010) 361–375.
[41] D. Chen, et al., Regulation of protein-ligand binding affinity by hydrogen bond pairing, Sci. Adv. 2 (3) (2016) e1501240.