Vietnam Journal of Biotechnology. **24**(1): 81-97, 2026.DOI: 10.15625/vjbt-23376

Research Article

INVESTIGATION OF SOYASAPOGENOLS FROM SOYBEAN SEED GERMS AS ANTI-APOPTOTIC AGENTS IN COLON CANCER USING AN *IN SILICO* APPROACH

Trong Luong Vu[1], Thi Ngoc Lan Nguyen[1], Quang Tan Tu[1], Duc Hung Nguyen[1] and Hoang Mau Chu[1,✉]

[1]*Thai Nguyen University of Education, 20 Luong Ngoc Quyen, Thai Nguyen, Vietnam*

✉To whom correspondence should be addressed. Email: chuhoangmau@tnu.edu.vn

Received: 30 August 2025; Accepted: 15 January 2026; Published online: 26 January 2026

ABSTRACT

Colorectal cancer is one of the most common malignancies worldwide, with 1.9 million new cases and 903,859 deaths recorded in 2022. Current treatments remain constrained by limited efficacy and safety concerns, emphasizing the need for new therapeutic options. Soybean [*Glycine max* (L.) Merr.] seed germs are rich in triterpenoids and represent a promising source of anticancer compounds. Preliminary phytochemical analysis confirmed the presence of alkaloids, flavonoids, terpenoids, cardiac glycosides, coumarins, saponins, and tannins, while steroids were absent, reflecting the metabolic diversity of this material. This study investigated the anticancer activity of soyasapogenol A and B, triterpenoids derived from soybean seed germs, by inhibiting the anti-apoptotic protein Bcl-2. *In silico* docking analysis revealed that soyasapogenol B exhibited a higher binding affinity and more stable interaction with the Bcl-2 protein (PDB: 6GL8) compared to soyasapogenol A and the reference drug paclitaxel. Molecular dynamics simulations over 100 nanoseconds supported the persistence of these interactions. ADMET profiling using pkCSM predicted that soyasapogenol B exhibits favorable pharmacokinetic properties, including high intestinal absorption, moderate distribution, and the absence of significant toxicities. Collectively, these findings identify soyasapogenol B as a potential Bcl-2 inhibitor with therapeutic relevance to colorectal cancer. The results provide a basis for further experimental validation and the development of soybean seed germs-derived compounds as novel anticancer agents.

Keywords: Anti-cancer, apoptosis, docking, *Glycine max*, *in silico*, soyasapogenol.

INTRODUCTION

Colorectal cancer is the third leading cause of cancer-related deaths worldwide, with an estimated 1.9 million new cases and 903,859 fatalities reported in 2022, accounting for nearly 10% of the global cancer burden. In Vietnam, it ranks fourth among the most common cancers in both sexes, with 16,835 new diagnoses and 8,454 deaths in the same year (Bray *et al.*, 2024). The continuing demand for safer and more effective treatments has driven efforts to develop targeted therapies (Zafar *et al.*, 2025). One

strategy with notable promise is the inhibition of angiogenesis, which restricts tumor growth by preventing the formation of new blood vessels. By blocking the vascular supply of oxygen and nutrients, angiogenesis inhibitors act indirectly on tumor progression, depriving malignant cells of resources essential for survival and dissemination (Lopes-Coelho *et al.*, 2021).

Therapeutic targeting of Bcl-2 family proteins represents a promising strategy in oncology, as these molecules serve as central regulators of the intrinsic apoptotic pathway. In many malignancies, elevated expression of Bcl-2 family members contributes to impaired apoptosis, thereby promoting survival of transformed cells and reducing sensitivity to anticancer interventions (Xu *et al*., 2023). Pharmacologic inhibition of anti-apoptotic proteins can restore apoptotic signaling in malignant cells and thereby promote programmed cell death. However, adaptive mechanisms may lead to acquired resistance during prolonged exposure, which can attenuate therapeutic efficacy. Consequently, rational combination regimens are frequently considered to mitigate resistance pathways and enhance overall clinical benefit (Cao *et al*., 2023). Recent evidence indicates that dysregulated Bcl-2 expression in colorectal cancer may constitute an early event, enabling the persistence of genetically damaged cells by suppressing apoptosis and thereby facilitating clonal expansion and tumor progression. Moreover, Bcl-2 overexpression has been associated with reduced responsiveness to chemotherapy, since many anticancer agents exert cytotoxic effects through induction of apoptotic cell death (Palabiyik Alperen, 2025; Ramesh and Medema, 2020).

Natural products derived from plants and microorganisms offer an abundant source of structural diversity and biological activity, serving as the basis for numerous clinically approved drugs, including antibiotics, anticancer agents, and antimalarials. While some natural compounds can be administered directly, many serve as scaffolds for semi-synthetic derivatives with improved pharmacological properties (Zeng *et al.*, 2024). Advances in computational and analytical platforms, including high-throughput screening, genomic tools, and artificial intelligence, have further accelerated the identification and optimization of bioactive molecules from these reservoirs (Bharate and Lindsley, 2024). Soyasapogenol A and B, the triterpenoid aglycones of soyasaponins in *Glycine max*, have been reported to exert cytotoxic activity against the colorectal cancer cell line HT-29 (Gurfinkel and Rao, 2003). However, the bioactivity of *G. max* sprouts remains uncharacterized, mainly presenting an opportunity for further investigation. Although studies on other natural compounds, such as magnolialide, have suggested pro-apoptotic activity, the underlying molecular mechanisms remain poorly understood. Computational approaches, including molecular docking, molecular dynamics simulations, and ADMET profiling, now provide powerful tools to explore compound–target interactions, binding stability, and pharmacokinetic behavior (Agu *et al.*, 2023; Arango *et al.*, 2026).

The present work applies these *in silico* methods to assess the anticancer potential of soyasapogenol A and B against colorectal cancer. The outcomes offer a foundation for subsequent experimental studies and the

possible development of soybean-derived compounds as therapeutic agents.

MATERIALS AND METHODS

Preparation of extract

Seeds from the *GmDREB* transgenic soybean line were germinated at room temperature for 72 h. The seedlings, reaching approximately 1.5 cm in length, were dehulled, dried at a temperature below 60°C, and subsequently ground into a fine powder. A 100 g portion of the dried material was extracted using ultrasound-assisted extraction with an ethanol–water solvent mixture (70:30, v/v) at 60°C for 45 min. The extraction was repeated three times, each with 250 mL of solvent. After removing the solvent, the obtained extract was subjected to analysis for its bioactive compounds (Hoang *et al*., 2023).

Qualitative phytochemical analysis

The identification of bioactive phytochemicals, encompassing alkaloids, flavonoids, steroids, terpenoids, glycosides, coumarins, saponins, and tannins, was performed employing established analytical techniques. A stock solution with a concentration of 2 mg/mL of the extract was prepared using the corresponding extraction solvent (Alqethami and Aldhebiani, 2021; Zumu *et al.*, 2024).

Collection of bioactive ligands

The molecular configurations of bioactive ligands derived from *G. max* were retrieved from the PubChem database (https://pubchem.ncbi.nlm.nih.gov). Specifically, these ligands, encompassing soyasapogenol A and soyasapogenol B, possess molecular formulas of $C_{30}H_{50}O_4$ and $C_{30}H_{50}O_3$, and molecular weights of 474.3709 and 458.3760 *m/z*, respectively. The two-dimensional (2D) structures of these ligands were generated in .sdf format utilizing ChemDraw Prime v23.1 (PerkinElmer, USA) and subsequently transformed into three-dimensional (3D) structures in .pdb format through Biovia Discovery Studio Visualizer v24.1 (Dassault Systèmes BIOVIA, USA). Energy minimization for the selected compounds was performed using Avogadro v1.2.0, employing the MMFF94 force field and the steepest descent algorithm (Hanwell *et al.*, 2012). Additionally, polar hydrogens and Gasteiger charges were incorporated, and all torsional angles were permitted to rotate using AutoDock Tools 1.5.7 (Center for Computational Structural Biology, USA) (Morris *et al.*, 2009). The 2D structure of the reference compound, paclitaxel, characterized by a molecular formula of $C_{47}H_{51}NO_{14}$ and a molecular weight of 853.3310 *m/z*, was acquired from the PubChem database in .sdf format and converted to .pdb format using Biovia Discovery Studio Visualizer (Figure 1).

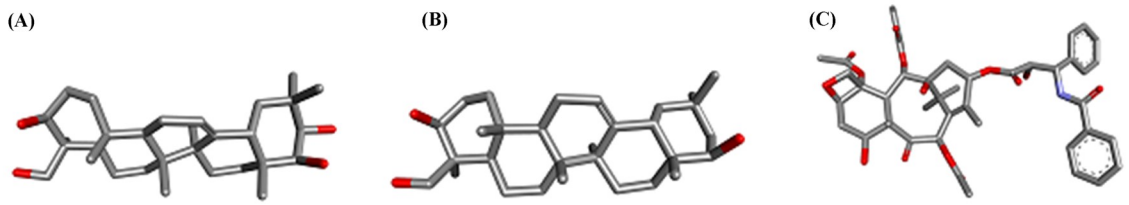

Figure 1. 3D Structures of selected ligands. (A) Soyasapogenol A, (B) Soyasapogenol B, and (C) Paclitaxel.

Molecular docking

The methodology for molecular docking is described in the previous study (Hoang *et al.*, 2025). The 3D configurations of all compounds were generated in .pdb format utilizing Biovia Discovery Studio Visualizer. Polar hydrogens were incorporated, Gasteiger charges were calculated, and all torsional angles were permitted to rotate. The 3D structure of the Bcl-2 target (protein ID: 6GL8) was obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) database (Casara *et al.*, 2018). Molecular docking of the protein and ligands was conducted using AutoDock Tools. The grid was configured with dimensions of x = 40, y = 40, and z = 40, employing a grid point spacing of 0.375 Å. The docking site was defined at coordinates x = 13.002 Å, y = 2.369 Å, and z = 12.299 Å within the binding pocket of the 6GL8 protein. The Lamarckian genetic algorithm was applied to identify the most energetically stable conformations for ligand-protein interactions.

Molecular dynamics simulation

Molecular dynamics simulations were conducted for the optimal docked conformation with the 6GL8 protein over a 100 nanosecond (ns) duration using GROMACS v2024.4 software (Van Der Spoel *et al.*, 2005). The protein structure was stabilized using Swiss-Pdb Viewer to ensure the integrity of atoms and residues (Guex and Peitsch, 1997). Ligand topologies were generated using SwissParam (Zoete *et al.*, 2011). The solvation environment was established using a triclinic simulation box tailored to the protein-ligand complex, employing the SPC water model. The system was neutralized with a 0.15 M sodium chloride solution. Energy minimization and system neutralization were performed over 50,000 steps. Molecular equilibration was achieved through a 200 ps NVT simulation (constant number of atoms, volume, and temperature), followed by a 200 ps NPT simulation (constant number of atoms, pressure, and temperature), maintaining the system at 300 K and 1.0 bar. Three independent 100-nanosecond simulations were executed with an integration time step of 2 femtoseconds (0.002 ps). Trajectory data were recorded at 10 ns intervals. Simulation results were evaluated using Grace software (Grace Development Team) to analyze parameters including root mean square deviation (RMSD), root mean square fluctuation (RMSF) of residues, radius of gyration (Rg), hydrogen bond counts (Hbonds), and solvent-accessible surface area (SASA). Structural alignment was conducted using UCSF Chimera 1.19 (Pettersen *et al.*, 2004) to assess the conformational stability of the docked compounds. The ligand conformations at zero ns and 100 ns were superimposed onto the protein's binding pocket to evaluate the consistency of interactions with critical amino acid residues, encompassing hydrogen bonds, van der Waals interactions, and hydrophobic contacts, throughout the simulation period. This analysis elucidated the stability and dynamic characteristics of the ligand-protein interactions over the 100 ns timeframe.

Drug likeness and ADMET prediction

The term ADMET, representing Absorption, Distribution, Metabolism, Excretion, and Toxicity, defines the pharmacokinetic characteristics of a pharmaceutical compound. These characteristics are essential for understanding the pharmacodynamic interactions, which

describe the mechanisms by which a compound elicits therapeutic effects within the body. Specifically, the ADMET profile governs the processes of absorption into the systemic circulation, distribution across tissues, metabolic transformation, excretion from the body, and potential toxicity (Hamidi *et al.*, 2013). The drug-likeness attributes of selected compounds were evaluated in this study utilizing the pkCSM database (Pires *et al.*, 2015).

RESULTS AND DISCUSSION

Qualitative phytochemical analysis

The extraction from soybean seed germs of the transgenic line TG1-10 yielded 3.9 g of crude extract. Phytochemical screening demonstrated the presence of several bioactive constituents, including tannins, terpenoids, flavonoids, alkaloids, phenolic compounds, and saponins (Table 1).

Table 1. Phytochemical test results of the extract of *G. max*.

Class of compound	Phytochemical test	Result
Alkaloid	Dragendorff's test	+
	Mayer's test	+
Flavonoid	Shinoda's test	+
Steroid	Liebermann-Burchard's test	-
Terpenoid	Liebermann-Burchard's test	+
Cardiac glycoside	Keller-Kilian's test	+
Coumarin	NaOH 10%	+
Saponin	Foam test	+
Tannin	$FeCl_3$ 5%	+

"+" Indicates the presence and "-" indicates the absence of phytochemicals.

As shown in Table 1, the extract of soybean seed germs of the transgenic line TG1-10 contained several classes of organic compounds, including alkaloids, flavonoids, terpenoids, cardiac glycosides, coumarins, saponins, and tannins, while steroids were absent. A previous study by Lisanti and Arwin (2019) similarly reported the presence of alkaloids, saponins, flavonoids, phenolics, triterpenoids, and glycosides but noted the absence of tannins and steroids. Based on the identification of terpenoids, particularly triterpenoids, in the extract, we selected soyasapogenol A and soyasapogenol B, two oleanane-type triterpenoid sapogenins derived from soyasaponins, previously isolated from soyasaponins and evaluated for their structure-activity relationship in colon cancer studies (Gurfinkel and Rao, 2003), as candidates for further investigation. Accordingly, molecular docking analyses were performed to assess their potential interactions and therapeutic mechanisms.

Molecular docking analysis

Molecular docking represents a pivotal methodology in structural molecular biology and computational drug design, employed to forecast the optimal binding conformation of a molecule, such as a pharmaceutical compound, to a protein target with a defined

3D structure. This simulation of molecular interactions facilitates the identification of prospective drug candidates, elucidates their binding mechanisms, and enhances their molecular properties (Meng *et al.*, 2011). Within molecular docking, the active site of a protein denotes the specific region where a ligand, such as a drug molecule, interacts with the protein to initiate a targeted interaction or chemical reaction. This site is critical for comprehending protein functionality and developing effective therapeutic agents (Agu *et al.*, 2023). Consequently, prior to executing molecular docking, the active sites of the 6GL8 protein were determined as an initial step. Structural visualization of the 6GL8 protein revealed active sites comprising Phe104, Tyr108, Asp111, Phe112, Gln118, Leu137, Gly145, Arg146, Ala149, Phe153, and Asp171 (Figure 2).

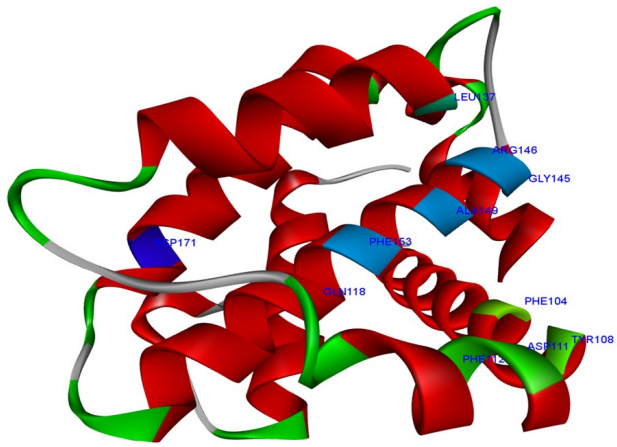

Figure 2. The active sites of the protein 6GL8.

Table 2. The interactions between the docked ligands and the protein 6GL8.

N°	Docked ligand	Binding energy (kcal/mol)	Hydrogen bond interaction	Van der Waals interaction	Hydrophobic interaction
1	Soyasapogenol A	-6.41	Arg146	Asp111, Met115, Leu137, Asn143, Gly145	Phe104, Tyr108, Phe112, Ala149
2	Soyasapogenol B	-7.14	Arg146	Val133, Leu137, Asn143, Glu152, Phe153	Phe104, Tyr108, Phe112, Met115, Arg146, Ala149
3	Paclitaxel	-6.54	Gln118, Glu136	Phe104, Tyr108, Asp111, Phe112, Thr132, Val133, Leu137, Arg146, Ala149, Glu152, Phe153	Leu119, Met115

The interactions between the selected ligands and the binding pockets of the 6GL8 protein are presented in Table 2. The amino acid residues participating in these interactions, along with their specific locations within the ligand-binding site,

were determined. Molecular docking revealed the presence of hydrogen bonds, van der Waals interactions, and hydrophobic interactions between the protein and the selected ligands. The detailed molecular interactions with the amino acid residues of the 6GL8 protein are illustrated in Figure 3.

Soyasapogenol A exhibited interactions with the 6GL8 protein through hydrogen bonds involving Arg146, van der Waals interactions with Asp111, Met115, Leu137, Asn143, and Gly145, and hydrophobic interactions with Phe104, Tyr108, Phe112, and Ala149. Among these, the amino acid residues Phe104, Tyr108, Asp111, Phe112, Leu137, Gly145, Arg146, and Ala149 are located within the active sites of the 6GL8 protein (Figure 3A). The binding energy for soyasapogenol A with the 6GL8 protein was calculated to be -6.41 kcal/mol.

Soyasapogenol B demonstrated interactions with the 6GL8 protein via hydrogen bonds with Arg146, van der Waals interactions with Val133, Leu137, Asn143, Glu152, and Phe153, and hydrophobic interactions with Phe104, Tyr108, Phe112, Met115, Arg146, and Ala149. The amino acid residues Phe104, Tyr108, Phe112, Leu137, Arg146, Ala149, and Phe153 are situated within the active sites of the 6GL8 protein (Figure 3B). The binding energy for soyasapogenol B with the 6GL8 protein was determined to be -7.14 kcal/mol.

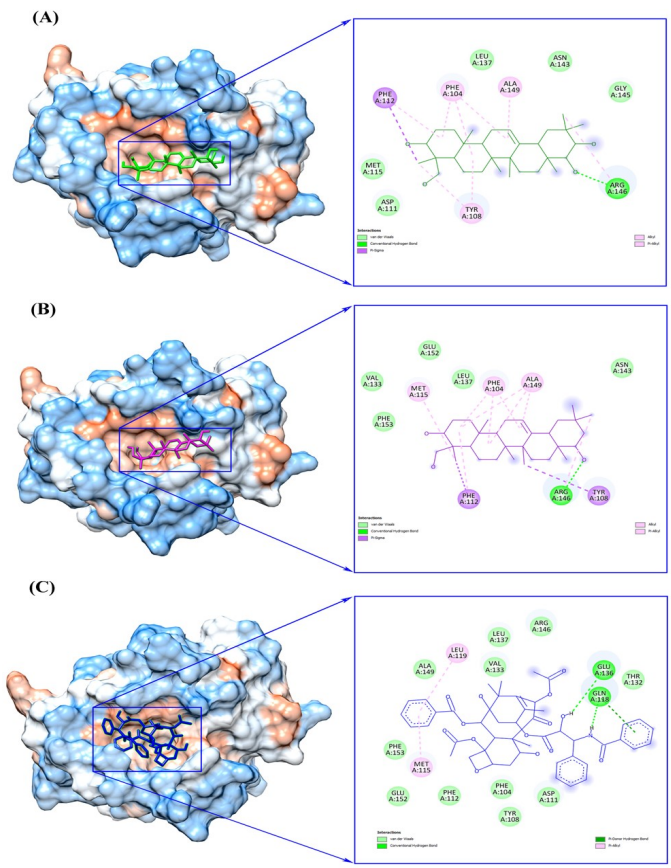

Figure 3. Molecular docking model and 2D interaction diagram of soyasapogenol A (A), soyasapogenol B (B), and paclitaxel (C) with the protein 6GL8.

The reference compound, paclitaxel, displayed interactions with the 6GL8 protein through hydrogen bonds with Gln118 and Glu136, van der Waals interactions with Phe104, Tyr108, Asp111, Phe112, Thr132, Val133, Leu137, Arg146, Ala149, Glu152, and Phe153, and hydrophobic interactions with Met115 and Leu119. The amino acid residues Phe104, Tyr108, Asp111, Phe112, Gln118, Leu137, Arg146, and Ala149 are located within the active sites of the 6GL8 protein (Figure 3C). The binding energy for paclitaxel with the 6GL8 protein was -6.54 kcal/mol.

All evaluated compounds exhibited hydrogen bonds, van der Waals interactions, and hydrophobic interactions with the 6GL8 protein. Hydrogen bonds and hydrophobic interactions are critical for stabilizing the ligand-protein complex, as they enhance the specificity and strength of binding (Patil *et al.*, 2010). Van der Waals interactions contribute to the stability of the protein-ligand complex by supporting the 3D conformation of both the protein and ligand, thereby augmenting overall binding affinity (Du *et al.*, 2016). Soyasapogenol B exhibited the most favorable binding energy (-7.14 kcal/mol) with 11 interactions, including 7 active site residues (Phe104, Tyr108, Phe112, Leu137, Arg146, Ala149, and Phe153). Soyasapogenol A showed a slightly less negative binding energy (-6.41 kcal/mol) with 10 interactions, including 8 active site residues (Phe104, Tyr108, Asp111, Phe112, Leu137, Gly145, Arg146, and Ala149). Paclitaxel, the reference compound, had the least negative binding energy (-6.54 kcal/mol) with 15 interactions, including 8 active site residues (Phe104, Tyr108, Asp111, Phe112, Gln118, Leu137, Arg146, and Ala149). A more negative binding energy indicates stronger binding affinity between the ligand and the receptor (Alsedfy *et al.*, 2024). Consequently, soyasapogenol B demonstrates the highest binding affinity and stability within the 6GL8 protein's binding pocket, followed by soyasapogenol A, with paclitaxel exhibiting the weakest affinity among the tested compounds.

Based on these findings, soyasapogenol A and soyasapogenol B were selected for further molecular dynamics studies to investigate their dynamic behavior within the 6GL8 protein's binding pocket, with paclitaxel serving as the reference compound. The ligand-protein complex structures obtained from the docking studies served as the starting point for these simulations.

Molecular dynamics simulation

Molecular dynamics simulations constitute essential computational methodologies employed post-docking to evaluate physical properties and elucidate reaction mechanisms. These simulations serve as a computational tool for observing and analyzing atomistic interactions between molecules and macromolecules at a detailed level. They enable refinement of structures derived from docking, assessment of binding conformations of docked compounds, and examination of conformational dynamics in proteins and other molecular entities (Mortier *et al.*, 2015). The analysis encompassed parameters such as RMSD, RMSF, Rg, Hbonds, and SASA. The total energy was determined to be -227,229 kJ/mol for soyasapogenol A and -227,503 kJ/mol for soyasapogenol B. The potential energy values were calculated as -282,292 kJ/mol for soyasapogenol A and -282,580 kJ/mol for soyasapogenol B. For the

reference compound paclitaxel, the total energy and potential energy were -225,831 kJ/mol and -280,908 kJ/mol, respectively. The system was maintained at an equilibrium temperature of 300 K.

RMSD measures the average distance between corresponding atoms in two superimposed molecular structures after optimal alignment, indicating their structural similarity or dissimilarity. A lower RMSD value signifies that the structures are more similar, while a higher value indicates greater differences, making it a crucial metric in fields like bioinformatics and computational chemistry for assessing structural stability and quality. RMSD also serves as a tool to assess system convergence and confirm attainment of equilibrium (Maiorov and Crippen, 1994; Schreiner *et al.*, 2012). The RMSD values for the complexes soyasapogenol A-6GL8, soyasapogenol B-6GL8, and paclitaxel-6GL8 demonstrated a consistent fluctuation pattern, averaging approximately 0.18 nm throughout the 100 ns simulation duration (Figure 4A). This restricted range of RMSD values indicates robust structural stability for the protein-ligand complexes involving soyasapogenol A, soyasapogenol B, and paclitaxel, with limited deviations from their initial configurations. Such stability reflects a well-equilibrated system, where molecular interactions and conformational dynamics achieved a steady state, suggesting preservation of functional integrity across the simulation period.

In molecular dynamics simulation, RMSF quantifies the average deviation of atomic positions from their mean positions over time, indicating the flexibility of different regions within a molecule. High RMSF values suggest high flexibility, while low values indicate rigidity or stability (Song *et al.*, 2024). In this study, RMSF analysis offers a residue-specific view of the flexibility and dynamic characteristics of the protein or protein-ligand complexes involving soyasapogenol A, soyasapogenol B, and paclitaxel throughout the simulation (Figure 4B). For all three complexes, RMSF values displayed a consistent pattern across the residue range from Gly7 to Gly203 within the 6GL8 protein. Throughout this segment, the RMSF remained within a range of approximately 0.05 nm, indicating minimal fluctuation. This low variability suggests a high degree of structural stability and restricted mobility for the residues in this region, likely due to strong stabilizing interactions or a well-defined secondary structure, contributing to the overall integrity of the protein-ligand complexes during the simulation.

The Rg is a valuable metric for assessing the compactness and overall dimensions of a molecular structure. It provides insights into how spread out the molecule's atoms are around its center of mass. A smaller Rg generally indicates a more compact structure, while a larger Rg suggests a more extended or loose conformation (Silverman *et al.*, 2024). The Rg analysis for the complexes is presented in Figure 4C. The Rg value reflects the global conformation of the protein, with lower values denoting greater compactness. The Rg values for the soyasapogenol A-6GL8, soyasapogenol B-6GL8, and paclitaxel-6GL8 complexes remained stable, fluctuating between 1.44 and 1.46 nm over the 100 ns simulation, suggesting that all complexes maintain a relatively compact protein structure.

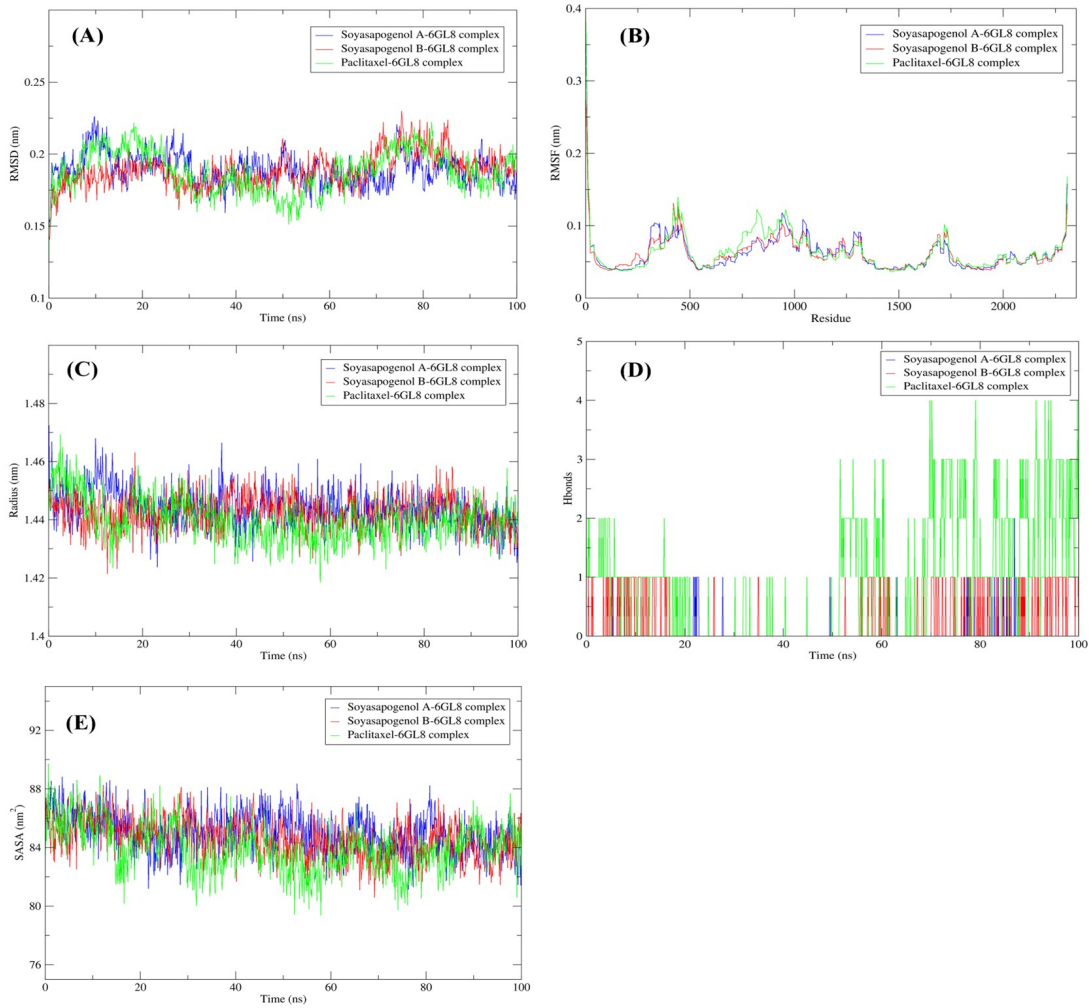

Figure 4. Results of molecular dynamics simulation for the bindings of soyasapogenol A (blue), soyasapogenol B (red) and paclitaxel (green) with the 6GL8 protein. (A) RMSD, (B) RMSF, (C) Rg, (D) Hbonds, and (E) SASA.

Hydrogen bond occupancy in molecular dynamics simulations refers to the fraction of time a specific hydrogen bond is formed and maintained during the simulation's trajectory. It essentially indicates how frequently a hydrogen bond persists between two atoms or groups of atoms over the simulated time period. A higher occupancy suggests a more stable and persistent hydrogen bond, while a lower occupancy indicates a more transient or less frequent interaction (Aulifa *et al.*, 2024). During the 100 ns simulation of the 6GL8 protein in complex with soyasapogenol A, soyasapogenol B, and paclitaxel, hydrogen bonds persisted throughout the period, signifying that all three ligands remained within the protein's binding pocket. Specifically, the soyasapogenol A-6GL8 complex displayed hydrogen bond numbers ranging from 1 to 2, indicating a stable yet limited hydrogen bonding profile. The soyasapogenol B-6GL8 complex maintained a single hydrogen bond throughout, reflecting a consistent but minimal interaction. Conversely, the paclitaxel-6GL8

complex exhibited hydrogen bond numbers fluctuating between 1 and 4, with the majority stabilizing at 1-4 bonds, suggesting a moderately variable hydrogen bonding capacity. These observations highlight the differing hydrogen bonding dynamics among the complexes, with soyasapogenol A and paclitaxel demonstrating greater variability compared to the uniform interaction observed with soyasapogenol B.

Drug likeness and *in silico* pharmacokinetics ADMET prediction

The ADMET predictions conducted using the pkCSM database evaluated the pharmacokinetic properties of soyasapogenol A, soyasapogenol B, and paclitaxel, focusing on their oral bioavailability. The results, detailed in Table 3, provide comprehensive insights into their absorption, distribution, metabolism, excretion, and toxicity profiles.

Table 3. ADMET properties of selected ligands using pkCSM database.

ADMET properties	Unit	Soyasapogenol A	Soyasapogenol B	Paclitaxel
Absorption				
Water solubility	(Log mol/L)	-5.057	-5.718	-3.158
Caco2 permeability	(Log Papp in 10-6 cm/s)	0.921	1.141	0.623
Intestinal absorption (human)	(% Absorbed)	93.01	92.764	100
Skin permeability	(Log Kp)	-3.406	-3.396	-2.735
P-glycoprotein substrate	Yes/No	No	No	Yes
P-glycoprotein I inhibitor	Yes/No	Yes	Yes	Yes
P-glycoprotein II inhibitor	Yes/No	Yes	Yes	Yes
Distribution				
Volume of distribution (VDss)	(Log L/kg)	-0.481	-0.042	1.458
Fraction unbound (human)	(Fu)	0.051	0.0	0.0
Blood-brain barrier (BBB) permeability	(Log BB)	-0.57	-0.052	-1.731
Central nervous system (CNS) permeability	(Log PS)	-1.896	-1.495	-3.95
Metabolism				
CYP2D6 substrate	Yes/No	No	No	No
CYP3A4 substrate	Yes/No	Yes	Yes	Yes
CYP1A2 inhibitor	Yes/No	No	No	No
CYP2C19 inhibitor	Yes/No	No	No	No
CYP2C9 inhibitor	Yes/No	No	No	No

CYP2D6 inhibitor	Yes/No	No	No	No
CYP3A4 inhibitor	Yes/No	No	No	Yes
Excretion				
Total clearance	(Log mL/min/kg)	0.174	0.104	-0.36
Renal organic cation transporter 2 (OCT2) substrate	Yes/No	Yes	No	No
Toxicity				
AMES toxicity	Yes/No	No	No	No
Max. tolerated dose (human)	(Log mg/kg/day)	-0.604	-0.769	0.199
hERG I inhibitor	Yes/No	No	No	No
hERG II inhibitor	Yes/No	No	Yes	Yes
Oral rat acute toxicity (LD50)	(mol/kg)	2.454	2.323	2.776
Oral rat chronic toxicity (LOAEL)	(Log mg/kg_Body Weight/day)	1.613	1.77	3.393
Hepatotoxicity	Yes/No	No	No	Yes
Skin sensation	Yes/No	No	No	No
Tetrahymena pyriformis toxicity	(Log μg/L)	0.319	0.402	0.285
Minnow toxicity	(Log mM)	0.2	-0.339	2.988

In the ADMET prediction, absorption is the process by which a chemical or drug enters the bloodstream from the site of administration, which influences its speed of action and concentration at its target. The absorption profiles of soyasapogenol A, soyasapogenol B, and paclitaxel offer insights into their pharmacokinetics. Intestinal absorption is 93.01% for soyasapogenol A, 92.764% for soyasapogenol B, and 100% for paclitaxel, indicating high gastrointestinal uptake, with paclitaxel showing complete absorption. The similar absorption rates of soyasapogenol A and B suggest strong membrane permeability, making them promising for drug development. Water solubility (log mol/L) is -5.057, -5.718, and -3.158, respectively, with paclitaxel's higher solubility aiding its absorption. Caco2 permeability (log Papp in 10^{-6} cm/s) is 0.921, 1.141, and 0.623, reflecting moderate to high permeability, with soyasapogenol B being highest. Skin permeability (log Kp) is -3.406, -3.396, and -2.735, indicating low transdermal potential. Soyasapogenol A and B are not P-glycoprotein substrates, unlike paclitaxel, but all inhibit P-glycoprotein I and II, potentially affecting drug efflux.

Accurate assessment of distribution is essential in drug development to predict target site accessibility and potential adverse

effects. The steady-state volume of distribution (VDss) quantifies the extent of drug dispersion from plasma into tissues, where a higher VDss indicates greater tissue penetration (Berezhkovskiy, 2004). The distribution characteristics of soyasapogenol A, soyasapogenol B, and paclitaxel reveal distinct profiles. The VDss (log L/kg) is -0.481 for soyasapogenol A, -0.042 for soyasapogenol B, and 1.458 for paclitaxel. Paclitaxel's high VDss suggests extensive tissue distribution, while soyasapogenol A and B predominantly remain in plasma. The fraction unbound in human plasma (Fu) is 0.051 for soyasapogenol A, 0 for soyasapogenol B, and 0 for paclitaxel, indicating high protein binding for soyasapogenol B and paclitaxel, potentially limiting their bioavailability. Blood-brain barrier permeability (log BB) is -0.57 for soyasapogenol A, -0.052 for soyasapogenol B, and -1.731 for paclitaxel, suggesting soyasapogenol B has the greatest potential to cross the blood-brain barrier (BBB). Similarly, central nervous system (CNS) permeability (log PS) is -1.896 for soyasapogenol A, -1.495 for soyasapogenol B, and -3.95 for paclitaxel, reinforcing soyasapogenol B's superior CNS penetration. These distribution properties highlight soyasapogenol B as a promising candidate for applications requiring CNS access.

Metabolism represents a fundamental aspect of ADMET profiling, essential for evaluating drug behavior during development. It encompasses enzymatic processes that transform compounds, facilitating their elimination from the body. These reactions predominantly occur in the liver, mediated by cytochrome P450 (CYP) enzymes, though other tissues such as the intestines and kidneys also contribute. CYP enzymes, notably CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4, are critical for metabolizing over 90% of drugs, influencing their efficacy, safety, and bioavailability (Abdelwahab *et al.*, 2025). The metabolism profiles of soyasapogenol A, soyasapogenol B, and paclitaxel indicate that none are substrates for CYP2D6, but all three are substrates for CYP3A4, suggesting significant hepatic metabolism via this enzyme. Regarding inhibitory effects, none of the compounds inhibit CYP1A2, CYP2C9, CYP2C19, or CYP2D6. However, paclitaxel uniquely inhibits CYP3A4, potentially affecting the metabolism of co-administered drugs processed by this enzyme. Soyasapogenol A and soyasapogenol B, lacking CYP3A4 inhibition, may reach therapeutic targets with reduced risk of drug-drug interactions, enhancing their potential as drug candidates.

Assessing excretion pathways is essential to identify compounds with efficient clearance, minimizing risks of accumulation and toxicity. Key parameters include total clearance and interactions with renal transporters, such as Organic Cation Transporter 2 (OCT2). The excretion profiles of soyasapogenol A, soyasapogenol B, and paclitaxel reveal total clearance values (log mL/min/kg) of 0.174, 0.104, and -0.36, respectively. Soyasapogenol A and B exhibit positive clearance rates, indicating moderate elimination, which may support sustained systemic exposure. Conversely, paclitaxel's negative clearance value suggests slower elimination, potentially prolonging its therapeutic effect but increasing the risk of accumulation. Regarding renal OCT2, soyasapogenol A is a substrate, suggesting potential renal transporter-mediated excretion, whereas soyasapogenol B and paclitaxel are not

substrates, indicating reliance on alternative clearance mechanisms.

Toxicity evaluation is an essential element of ADMET profiling, critical for identifying compounds with favorable safety profiles during drug development. Early assessment mitigates risks of adverse effects, enhancing the selection of viable candidates (Amorim *et al.*, 2024). The toxicity profiles of soyasapogenol A, soyasapogenol B, and paclitaxel were analyzed across multiple parameters. The AMES test, a standard assay for mutagenicity, indicates no toxicity for soyasapogenol A, soyasapogenol B, or paclitaxel. The maximum tolerated dose (human, log mg/kg/day) is -0.604 for soyasapogenol A, -0.769 for soyasapogenol B, and 0.199 for paclitaxel, suggesting paclitaxel has a higher tolerable dose. None of the compounds inhibit hERG I, but soyasapogenol B and paclitaxel inhibit hERG II, indicating potential cardiac risk. Oral rat acute toxicity (LD50, mol/kg) is 2.454, 2.323, and 2.776, respectively, reflecting comparable acute toxicity. Chronic toxicity (LOAEL, log mg/kg_bw/day) is 1.613 for soyasapogenol A, 1.77 for soyasapogenol B, and 3.393 for paclitaxel; the higher LOAEL of paclitaxel indicates a higher dose threshold for adverse effects, consistent with lower predicted chronic toxicity relative to the two soyasapogenols. Hepatotoxicity is absent for soyasapogenol A and B but present for paclitaxel. No compounds exhibit skin sensitization. Toxicity in *Tetrahymena pyriformis* (log µg/L) is 0.319, 0.402, and 0.285, and minnow toxicity (log mM) is 0.2, -0.339, and 2.988, respectively, suggesting low environmental toxicity for soyasapogenol B.

Based on the *in silico* ADMET evaluation results for soyasapogenol A, soyasapogenol B, and paclitaxel, it can be concluded that soyasapogenol B exhibits favorable pharmacokinetic properties, including non-toxicity, moderate distribution capacity, and high absorption. This study highlights soyasapogenol B as a potential candidate for further drug development, particularly due to its efficient pharmacokinetic profile.

CONCLUSION

Qualitative analysis of *G. max* extract revealed the presence of alkaloids, flavonoids, terpenoids, cardiac glycosides, coumarins, saponins, and tannins, with no steroids detected. This finding aligns with prior studies, except for the presence of tannins. Based on the identification of triterpenoids, soyasapogenol A and soyasapogenol B were selected for further investigation. Molecular docking and dynamics simulations were conducted on soyasapogenol A and soyasapogenol B, as well as paclitaxel, to assess their interactions with the 6GL8 protein. The docking results demonstrated that soyasapogenol B exhibits enhanced stability and favorable positioning within the 6GL8 protein binding pocket compared to soyasapogenol A and paclitaxel. Molecular dynamics simulations over 100 ns confirmed consistent binding interactions and a stable inhibition mode for all compounds. *In silico* ADMET predictions, utilizing the pkCSM database, further evaluated the oral bioavailability of these compounds. Soyasapogenol B fulfills key pharmacokinetic criteria, displaying high absorption, moderate distribution, and non-toxicity. These attributes position soyasapogenol B as a potent inhibitor of the anti-apoptotic protein Bcl-2, making it a promising candidate for the development of novel therapeutics targeting human colon cancer.

ACKNOWLEDGMENTS

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.02-2023.05.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Abdelwahab A. A., Elattar M. A., and Fawzi S. A. (2025). Advancing ADMET prediction for major CYP450 isoforms: graph-based models, limitations, and future directions. *BioMedical Engineering OnLine*, *24*(1), 93. https://doi.org/10.1186/s12938-025-01412-6

Agu P. C., Afiukwa C. A., Orji O. U., Ezeh E. M., Ofoke I. H., Ogbu C. O., *et al*. (2023). Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Scientific Reports*, *13*(1), 13398. https://doi.org/10.1038/s41598-023-40160-2

Alqethami A., and Aldhebiani A. Y. (2021). Medicinal plants used in Jeddah, Saudi Arabia: Phytochemical screening. *Saudi Journal of Biological Sciences*, *28*(1), 805–812. https://doi.org/10.1016/j.sjbs.2020.11.013

Alsedfy M. Y., Ebnalwaled A. A., Moustafa M., and Said A. H. (2024). Investigating the binding affinity, molecular dynamics, and ADMET properties of curcumin-IONPs as a mucoadhesive bioavailable oral treatment for iron deficiency anemia. *Scientific Reports*, *14*(1), 22027. https://doi.org/10.1038/s41598-024-72577-8

Amorim A. M. B., Piochi L. F., Gaspar A. T., Preto A. J., Rosário-Ferreira N., and Moreira I. S. (2024). Advancing drug safety in drug development: Bridging computational predictions for enhanced toxicity prediction. *Chemical Research in Toxicology*, *37*(6), 827–849. https://doi.org/10.1021/acs.chemrestox.3c00352

Arango J. P. B., Rodriguez D. Y. M., Cruz S. L., and Ocampo G. T. (2026). *In silico* evaluation of pharmacokinetic properties and molecular docking for the identification of potential anticancer compounds. *Computational Biology and Chemistry*, *120*, 108626. https://doi.org/10.1016/j.compbiolchem.2025.108626

Aulifa D. L., Al Shofwan A. A., Megantara S., Fakih T. M., and Budiman, A. (2024). Elucidation of molecular interactions between drug–polymer in amorphous solid dispersion by a computational approach using molecular dynamics simulations. *Advances and Applications in Bioinformatics and Chemistry*, 17, 1–19. https://doi.org/10.2147/AABC.S441628

Berezhkovskiy L. M. (2004). Determination of volume of distribution at steady state with complete consideration of the kinetics of protein and tissue binding in linear pharmacokinetics. *Journal of Pharmaceutical Sciences*, *93*(2), 364–374. https://doi.org/10.1002/jps.10539

Bharate S. B., and Lindsley C. W. (2024). Natural products driven medicinal chemistry. *Journal of Medicinal Chemistry*, *67*(23), 20723–20730. https://doi.org/10.1021/acs.jmedchem.4c02736

Bray F., Laversanne M., Sung H., Ferlay J., Siegel R. L., Soerjomataram I., *et al*. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229–263. https://doi.org/10.3322/caac.21834

Cao Q., Wu, X., Zhang Q., Gong J., Chen Y., You Y., *et al*. (2023). Mechanisms of action of the BCL-2 inhibitor venetoclax in multiple myeloma: a literature review. *Frontiers in Pharmacology*, *14*, 1291920. https://doi.org/10.3389/fphar.2023.1291920

Casara P., Davidson J., Claperon A., Toumelin-Braizat G. L., Vogler M., Bruno A., *et al*. (2018). S55746 is a novel orally active BCL-2 selective and potent inhibitor that impairs hematological tumor growth. *Oncotarget, 9*(28), 20075. https://doi.org/10.18632/oncotarget.24744

Du X., Li Y., Xia Y. L., Ai S. M., Liang J., Sang P., *et al*. (2016). Insights into protein–ligand interactions: Mechanisms, models, and methods. In *International Journal of Molecular Sciences*, *17*(2), 144. https://doi.org/10.3390/ijms17020144

Guex N., and Peitsch M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis*, *18*(15), 2714–2723. https://doi.org/10.1002/elps.1150181505

Gurfinkel D. M., and Rao A. V. (2003). Soyasaponins: The relationship between chemical structure and colon anticarcinogenic activity. *Nutrition and Cancer*, *47*(1), 24–33. https://doi.org/10.1207/s15327914nc4701_3

Hamidi M., Azadi A., Rafiei P., and Ashrafi H. (2013). A pharmacokinetic overview of nanotechnology-based drug delivery systems: an ADME-oriented approach. *Critical Reviews™ in Therapeutic Drug Carrier Systems*, *30*(5), 435-467. https://doi.org/10.1615/CritRevTherDrugCarrierSyst.2013007419

Hanwell M. D., Curtis D. E., Lonie D. C., Vandermeersch T., Zurek E., and Hutchison G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, *4*(1), 17. https://doi.org/10.1186/1758-2946-4-17

Hoang C. V., Tu T. Q., Nguyen H. D., and Chu M. H. (2025). *In silico* studies of saponins from *Hoya verticillata* var. verticillate with Important apoptosis potency. *Letters in Organic Chemistry*, *22*(11), 1–9. https://doi.org/10.2174/0115701786363779250528013022

Hoang C. V., Tu T. Q., Nguyen L. T. N., Nguyen H. D., Nguyen Q. H., and Chu M. H. (2023). Two New C21 steroidal glycosides from the leaves of *Hoya parasitica*. *Records of Natural Products*, *17*(6), 1046–1051. https://doi.org/10.25135/rnp.419.2307.2831

Lisanti E., and Arwin A. (2019). Phytochemical screening and proximate analysis of soybeans (*Glycine max*) variety Gamasugen 1 and Gamasugen 2 derived from gamma rays irradiation. *Journal of Physics: Conference Series*, *1402*(5), 55023. https://doi.org/10.1088/1742-6596/1402/5/055023

Lopes-Coelho F., Martins F., Pereira S. A., and Serpa J. (2021). Anti-angiogenic therapy: Current challenges and future perspectives. In *International Journal of Molecular Sciences*, *22*(7), 3765. https://doi.org/10.3390/ijms22073765

Maiorov V. N., and Crippen G. M. (1994). Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*, *235*(2), 625–634. https://doi.org/10.1006/jmbi.1994.1017

Meng X. Y., Zhang H. X., Mezei M., and Cui M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, *7*(2), 146-157. https://dx.doi.org/10.2174/157340911795677602

Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D. S., *et al*. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

Mortier J., Rakers C., Bermudez M., Murgueitio M. S., Riniker S., and Wolber G. (2015). The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug Discovery Today*, *20*(6), 686–702. https://doi.org/10.1016/j.drudis.2015.01.003

Palabiyik Alperen A. (2025). The role of Bcl-2 in controlling the transition between autophagy and apoptosis (Review). *Molecular Medicine Report, 32*(1), 172. https://doi.org/10.3892/mmr.2025.13537

Patil R., Das S., Stanley A., Yadav L., Sudhakar A., and Varma A. K. (2010). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *Plos One*, *5*(8), e12029. https://doi.org/10.1371/journal.pone.0012029

Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C., *et al*. (2004). UCSF Chimera—A visualization

system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pires D. E. V, Blundell T. L., and Ascher D. B. (2015). pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry*, *58*(9), 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104

Ramesh P., and Medema, J. P. (2020). BCL-2 family deregulation in colorectal cancer: potential for BH3 mimetics in therapy. *Apoptosis: An International Journal on programmed cell death*, *25*(5-6), 305–320. https://doi.org/10.1007/s10495-020-01601-9

Schreiner W., Karch R., Knapp B., and Ilieva N. (2012). Relaxation estimation of RMSD in molecular dynamics immunosimulations. *Computational and Mathematical Methods in Medicine*, *2012*(1), 173521. https://doi.org/10.1155/2012/173521

Silverman I., Gerber M., Shaykevich A., Stein Y., Siegman A., Goel S., *et al*. (2024). Structural modifications and kinetic effects of KRAS interactions with HRAS and NRAS: an *in silico* comparative analysis of KRAS mutants. *Frontiers in Molecular Biosciences*, *11*, 1436976. https://doi.org/10.3389/fmolb.2024.1436976

Song X., Bao L., Feng C., Huang Q., Zhang F., Gao X., *et al*. (2024). Accurate prediction of protein structural flexibility by deep learning integrating intricate atomic structures and cryo-em density information. *Nature Communications*, *15*(1), 5538. https://doi.org/10.1038/s41467-024-49858-x

Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., and Berendsen H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718. https://doi.org/10.1002/jcc.20291

Xu J., Dong X., Huang D. C., Xu P., Zhao Q., and Chen B. (2023). Current advances and future strategies for bcl-2 inhibitors: potent weapons against cancers. *Cancers*, *15*(20), 4957. https://doi.org/10.3390/cancers15204957

Zafar A., Khatoon S., Khan M. J., Abu J., and Naeem A. (2025). Advancements and limitations in traditional anti-cancer therapies: A comprehensive review of surgery, chemotherapy, radiation therapy, and hormonal therapy. *Discover Oncology*, *16*(1), 607. https://doi.org/10.1007/s12672-025-02198-8

Zeng T., Li J., and Wu R. (2024). Natural product databases for drug discovery: Features and applications. *Pharmaceutical Science Advances*, *2*, 100050. https://doi.org/10.1016/j.pscia.2024.100050

Zoete V., Cuendet M. A., Grosdidier A., and Michielin O. (2011). SwissParam: A fast force field generation tool for small organic molecules. *Journal of Computational Chemistry*, *32*(11), 2359–2368. https://doi.org/10.1002/jcc.21816

Zumu F. S., Akbor M. S., Amin A. Al Haque M. F., Sultana I., Faruq A. Al *et al*. (2024). Phytochemical screening and evaluation of antibacterial, antipyretic, hypoglycemic, and anxiolytic effects of *Adiantum philippense* leaf extracts. *Pharmacological Research - Natural Products*, *5*, 100108. https://doi.org/10.1016/j.prenap.2024.100108