TRENDS IN SCIENCES 2026; 23(11): 13841
https://doi.org/10.48048/tis.2026.13841 RESEARCH ARTICLE

System-Level Computational Investigation of Indonesian Multi-Target Phytochemicals Against Oncogenic Drivers in Lung Cancer Therapy

Alfa Marzelino[1,4,†], Dona Suzana[2,3,†,*], Dina Melia Oktavilantika[3], Jonathan Timothy Oei[4], Patricia Evelyna Bitin[4], Giovanni Effendy[4], Ionna Wijaya[4], Hanna Haviva Picaulima[4], Elizabeth Aurelia[4], Laurin Alrysia[4], Militia Difa Praycilia Tubagus[4] and Renata Kharis Enggrasari[4]

[1]Bioinformatics Research Center, Indonesian Institute of Bioinformatics (INBIO Indonesia), East Java 65145, Indonesia
[2]Faculty of Medicine, Muhammadiyah University Prof. Dr. HAMKA, Jakarta 13830, Indonesia
[3]Faculty of Health Science and Pharmacy, Gunadarma University, Depok 16242, Indonesia
[4]Department of Biomedical Science, Calvin Institute of Technology, Jakarta 10610, Indonesia

([*]Corresponding author's e-mail: dona.suzana@uhamka.ac.id)

Received: 16 March 2026, Revised: 7 May 2026, Accepted: 14 May 2026, Published: 15 June 2026

Abstract

Lung cancer has become a major health problem worldwide. Existing treatments, such as surgery, chemotherapy, and radiation, have limitations and significant side effects for humans. Therefore, new therapeutic agents that are more specific and safer are needed. The exploration of new drugs derived from Indonesia's plant biodiversity has great potential for treating lung cancer. Various technical and computational approaches, such as network pharmacology, virtual screening, molecular docking, and molecular dynamics, are used to identify potential drug compounds. Further validation was conducted using drug-likeness and pharmacokinetic profile assessments to assess the suitability of the compounds as drug candidates. AKT1, EGFR, KRAS, and SRC were selected as therapeutic targets, with the aim of inhibiting oncogenic proteins that regulate signaling pathways. The target proteins were selected based on network pharmacology and enrichment analysis, with high centrality values. Cross-validation was also performed using the TCGA database to compare with wet-lab data. Through a series of virtual screenings, the best targets were obtained, namely CID: 44425461 (AKT1), 76336754 (EGFR), 11081540 (KRAS), and 16757188 (SRC). This research is an early stage in drug discovery and also supports the development of Indonesian herbal-based drugs. Through various computational techniques, it can be concluded that Indonesian herbal compounds have the potential to be used as lung cancer drugs. Although computational techniques in this study suggested promising results, further *in vitro*/*in vivo* validation is required to confirm the biological evidence.

Keywords: Lung cancer, Network pharmacology, Molecular docking, Molecular dynamics, Natural products, Indonesian herbal medicine

Introduction

Cancer is a major healthcare issue worldwide, accounting for a total of 9.7 million deaths and 20 million new cases in 2022. Among all types of cancer, lung cancer remains the leading cause of cancer-related mortality, standing at approximately 18.7%, followed by colorectal, liver, female breast, and stomach cancers [1]. In 2022, nearly 2.5 million new cases and 1.8 million deaths of lung cancer were reported, highlighting the substantial financial burden imposed on

[†]These authors have equal contributions to this work.

welfare systems across regions and genders [2,3]. Beyond the rising prevalence of cancer, treatment costs have also increased. Li *et al.* observed that the average cost of lung cancer treatment per patient in China rose from USD 7,600 during 2005 - 2011 to USD 10,400 during 2012 - 2014 [4]. Driven by rising medication costs and an increasing number of patients, total health expenditure on lung cancer treatment in the Netherlands surged by 67.5% between 2016 and 2021, reaching €1.22 billion in 2021 [2]. Furthermore, the OECD has estimated increases in depression cases and declines in workforce productivity due to cancer, demonstrating its significant socio-economic impact [3].

To inhibit cancer progression, conventional treatments, namely surgery, chemotherapy, and radiation, have been extensively studied and implemented. However, these treatments exhibit significant limitations: no targeting ability, low specificity, cytotoxicity, poor bioavailability, rapid drug clearance, and drug resistance [5,6]. Chemotherapy, in particular, is often ineffective over the long term due to drug resistance mechanisms, including enhanced drug efflux, genetic modifications, increased DNA repair capacity, and elevated xenobiotic metabolism. Notably, drug resistance contributes to over 90% cancer-related mortality [7]. Chemotherapy is also frequently associated with adverse effects such as diarrhoea, vomiting, constipation, mucositis, and anaemia [8]. As a result, its 5-year survival rate is only 19% for all stages combined [9]. These highlight the urgency to explore novel therapeutic agents with improved specificity, efficacy, and safety.

Many cancer medications target oncogene pathways in lung cancer, including v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS), epidermal growth factor receptor (EGFR), AKT serine/threonine kinase 1 (AKT1), and proto-oncogene tyrosine kinase c-SRC (SRC). As one of the most mutated proteins in cancer, KRAS overactivation triggers downstream pathways such as RAS/MEK/ERK and PI3K/AKT/mTOR pathways, thereby leading to increased cellular proliferation, differentiation, anti-apoptosis, and invasion [10]. For non-small cell lung cancer (NSCLC), EGFR mutation is the key mediator of lung cancer initiation and progression through autocrine and/or paracrine signalling [11]. Both KRAS and EGFR activate PI3K-AKT signalling, promoting cell survival, proliferation, and metabolism. This hyperactivation provides greater resistance to chemotherapy and radiotherapy [10-15]. EGFR and AKT also interact with SRC, a protein tyrosine kinase that regulates cell migration, adhesion, and invasion, ultimately leading to metastasis [16].

As an alternative approach, plant-derived bioactive compounds have attracted increasing attention due to their high molecular diversity, novel biofunctionalities, and tumor-suppressive effects [17,18]. This study aims to identify novel bioactive compounds derived from Jamu, a traditional Indonesian herbal medicine, that may serve as safer, more effective alternatives to conventional cancer therapies. Jamu, widely used as a household remedy, predominantly uses medicinal plants from the Zingiberaceae family and contains active metabolites [19]. For example, curcumin has been shown to inhibit cancer-related signalling pathways, including the PI3K/Akt/mTOR, Ras/Raf/ERK, and JAK1,2/STAT3 pathways, thus reducing cell proliferation and inducing apoptosis [20]. Although clinical trials investigating Jamu as a cancer treatment remain limited, one study has reported a reduction in tumor size in patients with fibroadenoma mammae, a benign breast tumor [21].

To evaluate their therapeutic potential, *in silico* methods offer a cost-effective, time-saving approach. These include virtual screening of phytochemicals, network pharmacology, enrichment analysis, molecular docking, druglikeness analysis, and molecular dynamics simulations. Therefore, it allows identification of multiple targets and pathways, the interactions between targets and phytochemicals, and the overall potential of these compounds as novel antitumor drugs.

Materials and methods

Network pharmacology

Data mining

Target proteins associated with lung cancer were predicted using six databases, namely PharmGKB (https://www.pharmgkb.org/) [22], Comparative Toxicogenomics Database (https://ctdbase.org/) [23], Online Mendelian Inheritance in Man (https://www.omim.org/) [24], GeneCards (https://www.genecards.org/) [25], The Human Protein Atlas (https://www.proteinatlas.org/) [26], and STRING (https://string-db.org/) [27]. Data mining is based on

specific lung cancer in the target "homo sapiens". The data is processed and grouped into a single dataset of potential lung cancer gene pools for further analysis.

Protein-Protein Interaction (PPI) construction

Based on the protein pool obtained from data mining of six databases, protein-protein interactions were constructed. The PPI network pharmacology was constructed using a combination of the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) web server (https://string-db.org/) and Cytoscape [27-29]. The STRING database provided the initial protein-protein interaction relationships. Meanwhile, Cytoscape was used for clustering, filtering, and visualization of protein centrality [30]. The protein set was clustered using ClusterOne to obtain protein groups with the lowest *p*-value. Proteins with the best connections were filtered based on the highest degree centrality (DC), betweenness centrality (BC), and closeness centrality (CC) values.

Enrichment analysis

Functional enrichment analysis was performed to identify the relevance of biological pathways of potential target proteins. Gene function annotation was followed by mapping specific biological pathways to predict their potential in lung cancer. Enrichment analysis was divided into two main parts, namely gene ontology (GO) and KEGG pathway analysis. GO annotation was performed using Metascape (https://metascape.org) with the MCODE clustering model [31]. Disease enrichment analysis was performed using DisGeNET annotations and data from several other databases, including PaGenBase and the Transcription Factor Targets database. Cross-validation was performed using SRplot (https://www.bioinformatics.com.cn/en) to support the biological pathway analysis of target proteins [32,33]. SRplot provides GO analysis based on biological process (BP), cellular component (CC), and molecular function (MF). In addition, the analysis also includes biological pathway prediction using the KEGG pathway database.

The Cancer Genome Atlas Analysis (TCGA)

The potential of therapeutic proteins in lung cancer was evaluated using the Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/) [34]. The data used were biological experimental data taken from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx). This analysis examined gene expression in normal and cancer patients. Target proteins were used as input with the keyword "Lung Adenocarcinoma" (LUAD). Several proteins, including AKT1, EGFR, KRAS, and SRC, were evaluated based on RNA-sequence data.

Molecular docking

Protein preparation

Molecular docking simulations were performed in three main stages: protein and ligand preparation, simulation, and visualization. Molecular docking preparation was performed using MGLTools v1.5.7, including water removal, pure protein extraction, gridbox generation, hydrogen addition, and charging. At this stage, re-docking was also performed to evaluate the best protein model resulting from network pharmacology [35]. The analysis was performed on the AKT1 (PDB code: 3QKL) [36], EGFR (PDB code: 3W33) [37], KRAS (PDB code: 6OIM) [38] and SRC (PDB code: 2BDF) [39] proteins. The protein models were taken from the RCSB Protein Data Bank using the x-ray diffraction method and a resolution of < 3 Å.

Ligand data mining and preparation

Meanwhile, candidate compounds were taken from the KNApSAcK database using an entry list of all traditional Indonesian plants called Jamu [40]. Data mining was performed using an API to obtain all herbal plant name data. A total of 700 herb species were identified from the KNApSAcK database. From the plant list, compound information was mined from PubChem, yielding 37,000 chemical compounds [41]. As a preliminary step, compounds with a molecular weight < 500 Da were selected to meet the small-molecule requirement of Lipinski's rules. Through this stage, a total of 21,000 compounds were obtained in sdf file format and converted and optimized into PDBQT file format.

Molecular docking simulation and visualization

The simulation process was performed on four target proteins; each protein was docked against 21,000

compounds using the Autodock Vina algorithm [42,43]. Molecular docking is an initial screening process to test the binding affinity of drug candidates with target proteins. The simulation was performed using a Perl 5 v32 script to run it automatically. Drug candidates were filtered based on the highest-ranked docking compounds and compared with the control. Then, the best docking results were visualized in PyMOL and Biovia Discovery Studio to identify 2D and 3D structural interactions.

Pharmacokinetics and toxicology prediction

The best docking compounds are revalidated using pharmacokinetic and toxicological information. Drug candidates are evaluated using the pKCSM webserver (https://biosig.lab.uq.edu.au/pkcsm/) and following Lipinski's rules [44]. This analysis is performed by examining drug properties to assess bioavailability and human safety. Predictions are made to assess the absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of candidate compounds using SMILES input. Bioavailability predictions follow Lipinski's rules: molecular weight (MW) ≤ 500 Da, lipophilicity (Log P) ≤ 5, hydrogen bond donors (HBD) ≤ 5, and hydrogen bond acceptors (HBA) ≤ 10.

Permeability prediction

The permeability of potential candidate compounds across membranes was analyzed using the Permeability of Molecules across Membranes (PerMM) web server (https://permm.phar.umich.edu/) [45]. The best compounds from the previous screening were used as input in PDB format. Experimental conditions included 310 K and pH 7.4. This analysis included the calculation of membrane binding affinities using the Black Lipid Membrane (BLM) model.

Molecular dynamics simulation

Molecular dynamics simulations were performed using a combination of CHARMM-GUI (https://www.charmm-gui.org/) and GROMACS [46,47]. CHARMM-GUI was used to prepare the protein-ligand complex using the solution builder feature. Compounds not listed in the RCSB Protein Data Bank and that could not be parameterized using CGenFF were constructed using the Ligand Reader and Modeler modules [48]. The parameters used in the preparation stage were 310K, 0.15% NaCl, the AMBER force field, and a simulation time of 200 nanoseconds. The output from CHARMM-GUI was an initial topology file that was used as input for a molecular dynamics simulation in GROMACS. The molecular dynamics simulation in GROMACS included minimization, equilibration, and final simulation for predicting the stability of the protein-ligand system. The results were visualized using XMGrace tools to view the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and hydrogen bond (HB) values.

MM/GBSA scoring

Free energy binding analysis was performed using the MM/GBSA method [46,49]. Free energy calculations were analyzed based on molecular dynamics simulation trajectories generated by GROMACS. Free energy calculations were performed using the external tool gmx_MMPBSA to generate van der Waals, electrostatic, solvation (GB), and non-polar energy values. MM/GBSA graph visualization was performed using Matplotlib in the Python programming language.

Free Energy Landscape (FEL)

Free energy landscape (FEL) calculations were performed using the output trajectory (xtc) from molecular dynamics simulations of each protein-ligand complex [50]. The components required for FEL calculations include the trajectory (xtc) file, topology (tpr) file, and index (ndx) file. The simulation was performed for 100 ns using “gmx covar” to generate eigenvalues (eigenval.xvg) and eigen vectors (eigen.trr). These results were used to identify the atomic motion covariance values in the complex system. Next, the main components (PC1 and PC2) were projected using the “gmx anaeig” tool to analyze the conformation distribution of the system in PCA space. The projections were then reanalyzed using “gmx sham” to create Gibbs free energy maps (kJ/mol) and visualized in eps and pdf files using “gmx xpm2p” and “imageMagick”. Gibbs free energy calculations were performed using the following formula:

$$\Delta G_i = -kT\left(\frac{P_i(r)}{P_{max}(r)}\right) \quad (1)$$

In the following equation, k and T represent the Boltzmann constant and simulation temperature, respectively. The probability of each state is calculated based on the reaction coordinate distribution of the MD trajectory used. The maximum probability state is used as a reference in determining ΔG.

Principal Component Analysis (PCA)

Principal Component Analysis is a method for calculating essential dynamics to evaluate the stability of protein-ligand complexes. This analysis is based on a covariance matrix constructed from α-carbon (Cα) fluctuations. PCA is performed using the Bio3D package in RStudio software to analyze trajectory files (xtc) resulting from GROMACS molecular dynamics simulations [51,52]. The trajectory file (xtc) contains pure protein, cleaned of water molecules. PCA is computed by eigenvalue decomposition of the covariance matrix to identify the main components with the greatest variance. The PCA model used follows the following mathematical formula:

$$M = T_K P_K^T + R \quad (2)$$

M is the matrix product, T_K represents sample correlations, P_K is the variable correlations, K is the number of parameters, and R is the residual matrix. Matrix product information is used in evaluating the stability of interactions between proteins and ligands during simulation. Eigenvector calculations are performed using the best PC1, PC2, and PC3 values for vector values in the matrix. Covariance matrix calculations follow the following formula.

$$C = VAV^T \quad (3)$$

C is the covariance matrix, V represents the matrix of eigenvectors, and A corresponds to the diagonal matrix of eigenvalues. All of these parameters are used in the analysis of the dispersion of each PC to identify the stability of proteins and ligands in a complex.

Dynamic Cross Correlation Matrix (DCCM)

The dynamics cross-correlation matrix (DCCM) was computed using the Bio3D (dccm) package in RStudio [51,52]. This analysis was conducted to identify antiparallel interactions between Cα and other atoms. DCCM values range from –1 to +1, indicating parallel interactions for positive values and antiparallel interactions for negative values. The following parameters were used to analyze internal dynamic communication within proteins and ligand binding to protein dynamics. DCCM calculations were performed using the following mathematical formula:

$$DCCM_{ij} = \frac{\langle \vec{d}i \,.\, \vec{d}j \rangle}{\sqrt{\langle di^2 \rangle \langle dj^2 \rangle}} \quad (4)$$

The vector $\vec{d}i$ indicates the shift of the residual to i from the average position, while $\vec{d}j$ represents the shift of the residual to j.

Results

Identification of potential targets through network pharmacology analysis

The first step in the network pharmacology analysis for determining the target protein was to conduct a protein-protein interaction (PPI) analysis. **Figure 1** shows the best candidates based on clustering and filter results. The STRING PPI generated 1,658 nodes, 57,876 edges, and provided a visualization of the network before filtering. Then, ClusterOne was used to identify protein complexes in the PPI networks that exhibit strong biological relationships. These clusters are formed based on strong biological linkages among network nodes. **Figure 1(A)** shows the results of the top cluster with the lowest p-value of 9.67×10^{-6} and a total of 223 proteins/nodes. The clustering results were screened using three main parameters: degree > 150, betweenness centrality > 0.005, and closeness centrality > 0.5. After applying these parameters, the results identified 25 nodes with 296 degrees, indicating strong connectivity across centrality and biological relationships. Further validation using GOlorize (**Figure 1(C)**) showed that the target genes AKT1, EGFR, KRAS, and SRC are significantly correlated with general cancer mechanisms. These 4 genes regulate protein kinase activity, signaling processes, and cell migration regulation associated with metastasis.

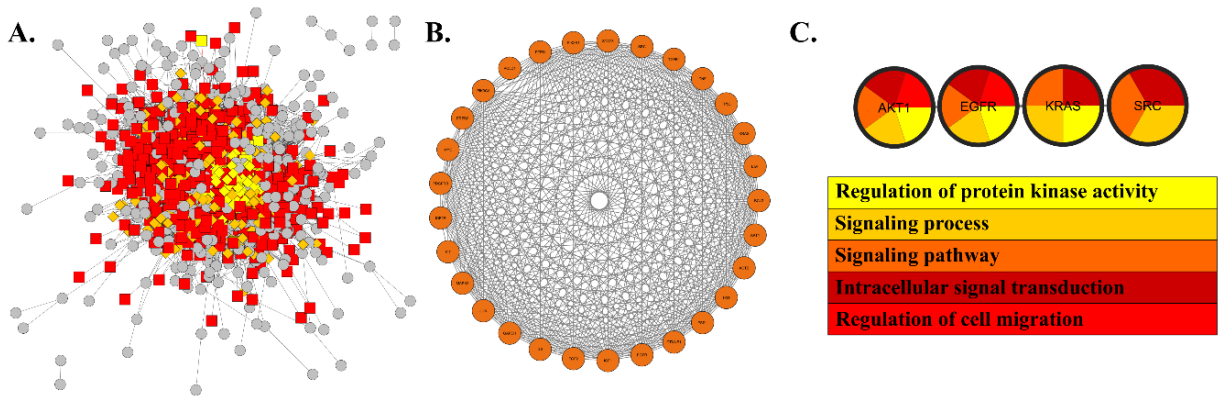

Figure 1 Protein-protein interaction (PPI) network using Cytoscape. (A) ClusterOne clustering calculation; (B) Protein-protein interaction (PPI) network; (C) Gene annotation by GOlorize.

Gene Ontology (GO) and KEGG pathway enrichment analysis

Next, an enrichment analysis was conducted using two independent tools: Metascape and SRplot. The Metascape analysis (**Figure 2(A)**) revealed that the target genes are significantly associated with cancer pathways and the role of microRNAs in cancer. In addition, the target genes are involved in several other biological regulatory processes, such as proliferation, locomotion, and signaling. GO enrichment was also conducted by using an SRplot for cross-validation (**Figure 2(B)**). The SRplot gene function annotations were divided into 3 categories: biological process (BP), cellular component (CC), and molecular function (MF). In the BP category, most genes were associated with positive regulation pathways for kinase activity, peptidyl-tyrosine phosphorylation, and the MAPK cascade. For the CC category, the highest scores were obtained for the phosphatidylinositol 3-kinase complex, membrane rafts, and focal adhesions. The MF category showed the highest scores for receptor-ligand activity, signaling receptor activator activity, and cytokine receptor binding. The analysis showed the significance of the candidate genes in cancer and their signaling pathways.

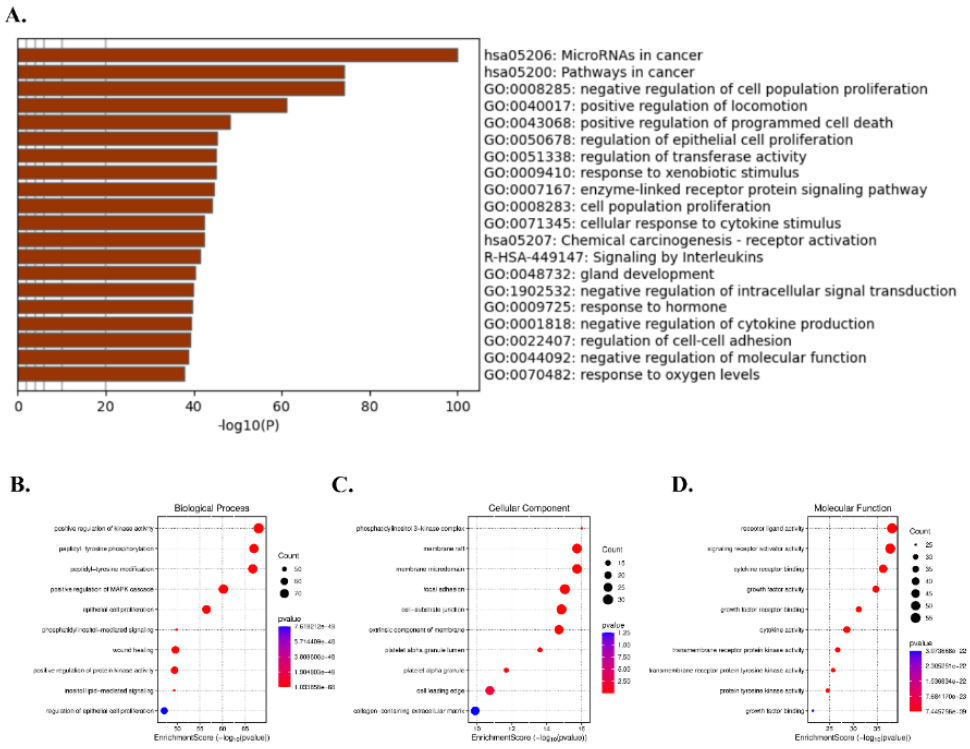

Figure 2 Functional enrichment analysis of the identified target genes. (A) Gene Ontology (GO) enrichment analysis of the target genes using Metascape; (B) GO Biological Process (BP), (C) GO cellular Component, (D) Molecular Function (MF) enrichment analysis visualized by SRplot.

In addition, KEGG enrichment analysis was performed to map the gene pool's involvement in specific biological pathways and complement the GO and pathway analyses. The results of the analysis are mapped using an SRplot algorithm (**Table 1**). The results revealed that the candidate genes are linked to proteoglycan effectors and EGFR tyrosine kinase, both of which modulate the PI3K-Akt signaling pathway and are potentially involved in cancer pathways. Therefore, the candidate genes are closely linked to cancer signaling pathways and are potential oncogenes. The most significant biological pathways were selected based on p-values and q-values. The *p*-value represents the probability of linking a gene pool to a pathway as a whole. Meanwhile, the q-value is the *p*-value corrected for multiple comparisons using the false discovery rate (FDR) method. This adjustment was used to increase the confidence level of the best biological pathways.

Table 1 KEGG pathway enrichment analysis results of the identified target genes.

ID	Description	Gene Ratio	*p-value*	*q-value*
hsa04151	PI3K-Akt signaling pathway	93/216	2.14614E-76	9.71409E-75
hsa05205	Proteoglycans in cancer	69/216	1.88015E-64	4.25507E-63
hsa01521	EGFR tyrosine kinase inhibitor resistance	50/216	1.49564E-63	2.25657E-62
hsa04933	AGE-RAGE signaling pathway in diabetic complications	51/216	4.22544E-58	4.78142E-57
hsa04010	MAPK signaling pathway	73/216	1.03774E-56	9.39428E-56

Clinical relevance validation using the TCGA database

Gene expression analysis using the TCGA database (**Figure 3**) showed that AKT1, EGFR, KRAS, and SRC were expressed at higher levels in tumor tissue than in normal tissue. The data were based on comparisons of gene expression levels between tumor tissue (T) and normal tissue (N), with sample sizes of 483 and 347, respectively. **Figure 3** shows the difference in the expression of the target genes between tumor and normal tissue (**Figures 3(A) - 3(D)**). The boxplot shows that the 4 target genes are overexpressed in patients with lung adenocarcinoma (LUAD). Therefore, as this study's findings validate the TCGA, they help healthcare workers identify the most promising therapeutic targets during the virtual screening of drug compound candidates.

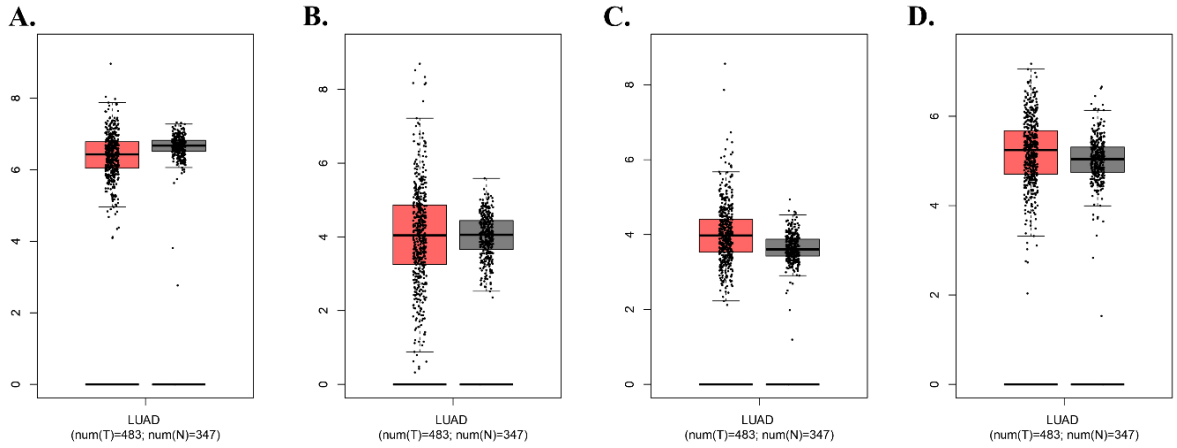

Figure 3 mRNA expression analysis of core target genes based on The Cancer Genome Atlas (TCGA) database. mRNA expression level of (A) AKT1, (B)EGFR, (C) KRAS, and (D) SRC.

Molecular docking simulation analysis

The first stage of the molecular docking screening process compared the control co-crystallized ligand with the drug candidate compounds for parameters such as binding affinity, RMSD, and interaction. **Figures 4(A) - 4(D))** shows that the control co-crystallized ligand binds stably to the binding site of each target protein. The interaction between the amino acid residue and the control compound acts as the coordinate template for screening attachment. There are several observable interactions in this protein-ligand complex, including hydrogen bonds, hydrophobic interactions, and the α interaction, indicating that multiple bonds contribute to the stability of the complex. This information was used to create a standard re-docking process template to validate and test the feasibility of drug candidate compounds.

Meanwhile, the drug candidate compounds (**Figures 4(E) - 4(H)**) bind at the same coordinates as the control co-crystallized ligand. The interaction strength was reevaluated based on binding affinity and RMSD values for each protein complex (**Table 2**). The second of 10 modes for ligand pose and binding orientation was selected using the AutoDock Vina algorithm. The second mode was chosen as the binding pose because the first mode represents a self-docking conformation similar to the control co-crystallized ligand. This self-docking conformation impacts the RMSD value in the first mode, making it always 0. Thus, the 2nd mode was selected to indicate an alternative pose orientation at the specified binding site. The results of all molecular docking simulations for all compounds can be seen in Supplementary Material 2, which includes the docking results (.log) and ligand pose visualizations (_out.pdbqt) for each protein.

Next, the visualization of the drug candidate compounds indicates the bonds that formed with the residues. **Figures 4(E) - 4(H))** shows these bond combinations, including hydrogen bonds and hydrophobic interactions that stabilize protein-ligand complexes. Overall, the molecular docking results indicate that the AKT1, EGFR, KRAS, and SRC protein complexes can form stable protein-ligand bonds with the candidate compounds.

Table 2 Docking scores of candidate compounds against selected protein targets.

Protein	Ligands	IUPAC Name	Binding affinity (kcal/mol)	RMSD	Hydrogen bond
AKT1 PDB ID: 3QKL	Control (SMR)	N-(2-ethoxyethyl)-N-[(2S)-2-hydroxy-3-[(7S)-spiro[8,9-dihydro-3H-pyrano[3,2-e]indazole-7,3'-piperidine]-1'-yl]propyl]-2,6-dimethylbenzenesulfonamide	–9.5	1.64	Ala230, Glu228
	CID: 44425461	(9R,17R)-17-hydroxy-2,16-dimethyl-15-[(1R,4R,5R)-1-methyl-8-methylidene-7-oxo-2,6-dioxabicyclo[3.3.1]nonan-4-yl]-8-oxatetracyclo[9.7.0.03,9.012,16]octadeca-2,4-dien-7-one	–12.3	2.19	Phe161
EGFR PDB ID: 3W33	Control (W19)	4-[4-(1-benzothiophen-4-yloxy)-3-chloroanilino]-N-(2-hydroxyethyl)-8,9-dihydro-7H-pyrimido[4,5-b]azepine-6-carboxamide	–11.2	1.37	Met793
	CID: 76336754	(1S,5S,7S,9R,11S,12S)-5-hydroxy-2,16-dimethyl-15-[(1R,4R,5R)-1-methyl-8-methylidene-7-oxo-2,6-dioxabicyclo[3.3.1]nonan-4-yl]-8-oxapentacyclo[9.7.0.02,7.07,9.012,16]octadecan-3-one	–11.1	1.80	Asp855, Thr790
KRAS PDB ID: 6OIM	Control (MOV)	6-fluoro-7-(2-fluoro-6-hydroxyphenyl)-4-[(2S)-2-methyl-4-propanoylpiperazin-1-yl]-1-(4-methyl-2-propan-2-yl-3-pyridinyl)pyrido[2,3-d]pyrimidin-2-one	–7.8	1.25	Glu63
	CID: 11081540	(3aS,5aR,9S,9aS,9bS)-9-[[(3R,3aR,6aR,9aR,9bR)-3-hydroxy-6,9-dimethylidene-2-oxo-3a,4,5,6a,7,8,9a,9b-octahydroazuleno[4,5-b]furan-3-yl]methyl]-5a,9-dimethyl-3-methylidene-4,5,7,8,9a,9b-hexahydro-3aH-benzo[g][1]benzofuran-2,6-dione	–9.9	2.084	Gln61
SRC PDB ID: 2BDF	Control (24A)	[[4-[[2-(4-aminocyclohexyl)-9-ethylpurin-6-yl]amino]phenyl]-hydroxyphosphoryl]methylphosphonic acid	–8.7	1.87	Thr338
	CID: 16757188	(1S,9R,11S,12S,15R,16S)-2,16-dimethyl-15-[(1R,4R,5R)-1-methyl-8-methylidene-7-oxo-2,6-dioxabicyclo[3.3.1]nonan-4-yl]-8-oxatetracyclo[9.7.0.03,9.012,16]octadeca-2,4-dien-7-one	–10.7	2.25	Cys277

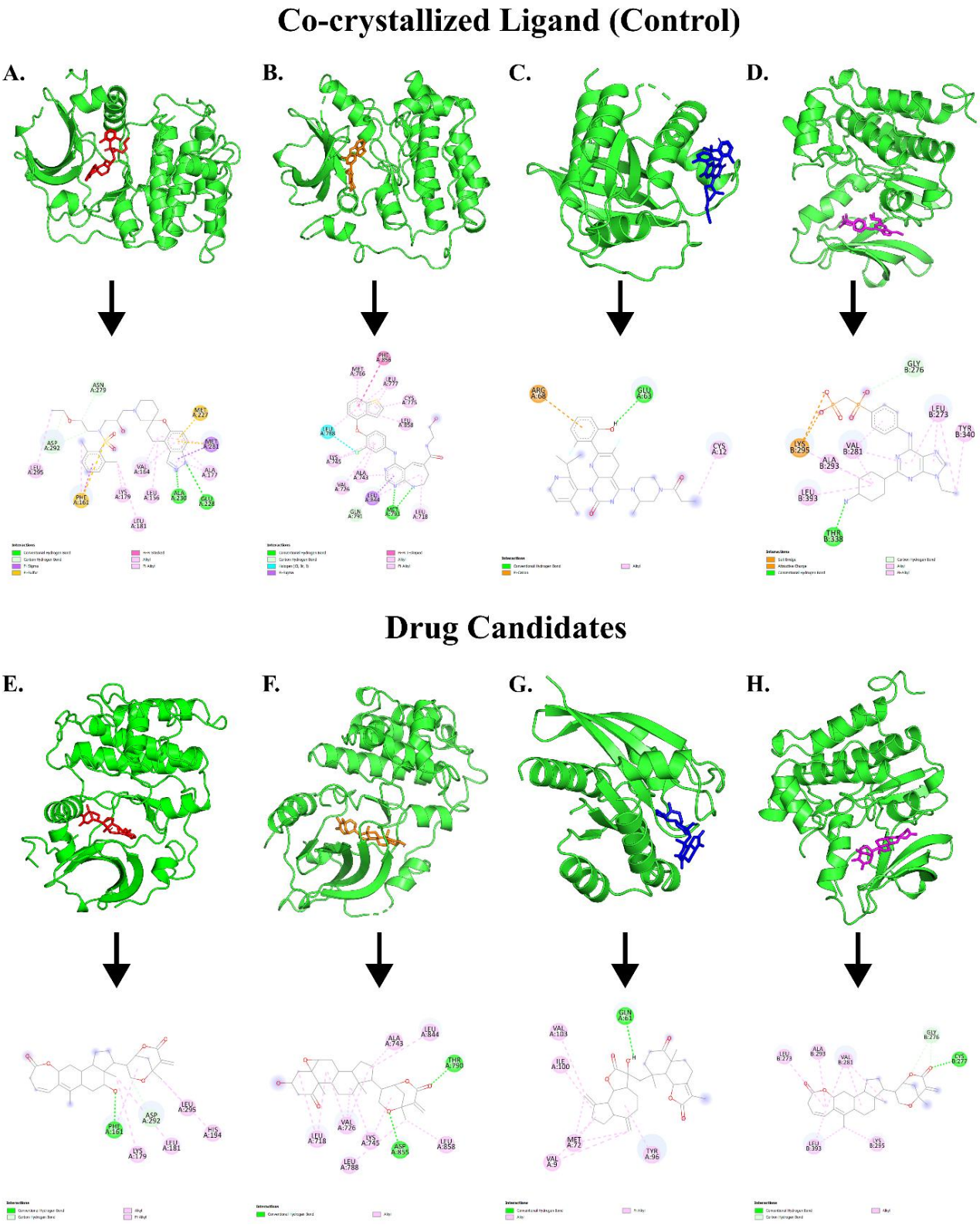

Figure 4 Molecular docking visualization of control inhibitors and candidate compounds bound to target proteins. Control inhibitors: (A) SMR-AKT1 complex, (B) W19-EGFR complex, (C) MOV-KRAS complex, and (D) 24A-SRC complex. Candidate compounds: (E) CID: 44425461-AKT1 complex, (F) CID: 76336754-EGFR complex, (G) CID: 11081540-KRAS complex, and (H) CID: 16757188-SRC complex.

Drug-likeness and pharmacokinetic profile evaluation

Further validation of drug suitability was performed using pharmacokinetic analysis to assess the bioavailability of candidate drug compounds. Pharmacokinetic predictions from pKCSM are shown in **Table 3**, which includes four compounds from the entire AKT1, EGFR, KRAS, and SRC complexes. All candidates met the Lipinski rules without any violations. This indicates that the candidate compounds have good potential for absorption by the gastrointestinal tract. The Lipinski rules are used as an initial screening step to select drug candidates based on physicochemical properties. Still, they are not the basis for the final determination of drug compounds as patent criteria.

Table 3 Drug-likeness evaluation of selected compounds based on Lipinski's rule of five.

CID	Molecular weight (g/mol)	AlogP	H-Bond aceptor	H-Bond donor	Rotatable bonds
44425461	468.59	3.884	6	1	1
76336754	470.60	3.593	6	1	1
16757188	452.59	4.913	5	0	1
11081540	494.62	4.464	6	1	2

ADME and toxicity prediction analysis

Pharmacokinetic and toxicological predictions for drug candidate compounds are shown in **Table 4**. The results include values for absorption, distribution, metabolism, excretion, and toxicity parameters. All compounds (CID: 44425461, 76336754, 16757188, and 11081540) showed high Caco-2 permeability and intestinal absorption above 97%. This indicates good oral absorption potential of the drug candidate compounds. Distribution parameters indicate low to moderate BBB and CNS penetration, suggesting reduced potential for CNS side effects. Meanwhile, the CYP3A4 and CYP1A2 substrate metabolism pathways indicate control of the risk of drug interactions in the metabolic pathway. The total clearance value is also within the moderate range, and the majority of the compound is not a renal OCT2 substrate. This analysis also emphasizes four toxicity parameters: AMES, hERG I, hERG II, and hepatotoxicity, which reinforce the compound's suitability as a candidate for further drug development.

Table 4 Pharmacokinetic and toxicological prediction of selected ligands.

Category	Parameter	Unit	Pubchem Compound ID (CID)			
			44425461	**76336754**	**16757188**	**11081540**
Absorption	Caco2 permeability	Log Papp in 10^{-6} cm/s	1.159	1.237	1.310	0.969
	Intestinal absorption	%	97.381	99.525	98.722	99.225
	Water solubility	Log mol/L	–5.388	–4.891	–5.545	–5.486
Distribution	BBB permeability	Log BB	–0.337	–0.342	0.249	–0.309
	CNS permeability	Log PS	–2.486	–2.672	–1.511	–2.564
Metabolism	CYP3A4 substrate	Yes/No	Yes	Yes	Yes	Yes
	CYP1A2 inhibitor	Yes/No	No	No	No	No
Excretion	Total clearance	Log ml/min/kg	0.38	0.251	0.339	0.223
	Renal OCT2 substrate	Yes/No	No	No	No	Yes
Toxicity	AMES toxicity	Yes/No	No	No	No	No
	hERG I inhibitor	Yes/No	No	No	No	No
	hERG II inhibitor	Yes/No	No	No	No	No
	Hepatotoxicity	Yes/No	No	No	No	No

Molecular dynamics simulation analysis

The molecular dynamics simulation results are shown in **Figure 5** for the protein-ligand complexes AKT1 (A), EGFR (B), KRAS (C), and SRC (D). The graph shows a molecular dynamics simulation for 200 nanoseconds and compares the control co-crystallized ligand with the best candidate compound. Stability was evaluated based on the following parameters: RMSD, RMSF, Rg, and the number of hydrogen bonds during the simulation. The RMSD backbone profiles across all systems showed relatively stable fluctuations at the beginning of the simulation, consistent with the control co-crystallized ligand. However, the EGFR-CID 76336754 complex showed increased fluctuations after 100 nanoseconds. The other candidate compounds (CID 44425461, 11081540, and 16757188) showed stable attachment, as their RMSD graphs resembled those of the control co-crystallized ligand, with no significant fluctuations. Meanwhile, RMSF results show that residual fluctuations are localized in flexible regions. The EGFR-CID 76336754 complex also showed high graphic spikes on certain residues, but most of its other residues remained stable. The radius of gyration (Rg) results for all compounds also show a pattern similar to the control. All drug candidate compounds show good stability. The KRAS-CID 11081540 complex showed a brief spike at a specific time, but the system returned to a stable pattern after a while. Meanwhile, from the hydrogen bond parameter, it can be seen that all complex systems maintained stable hydrogen bonds throughout the simulation. The results of the molecular dynamics trajectories are shown in Supplementary File 3.

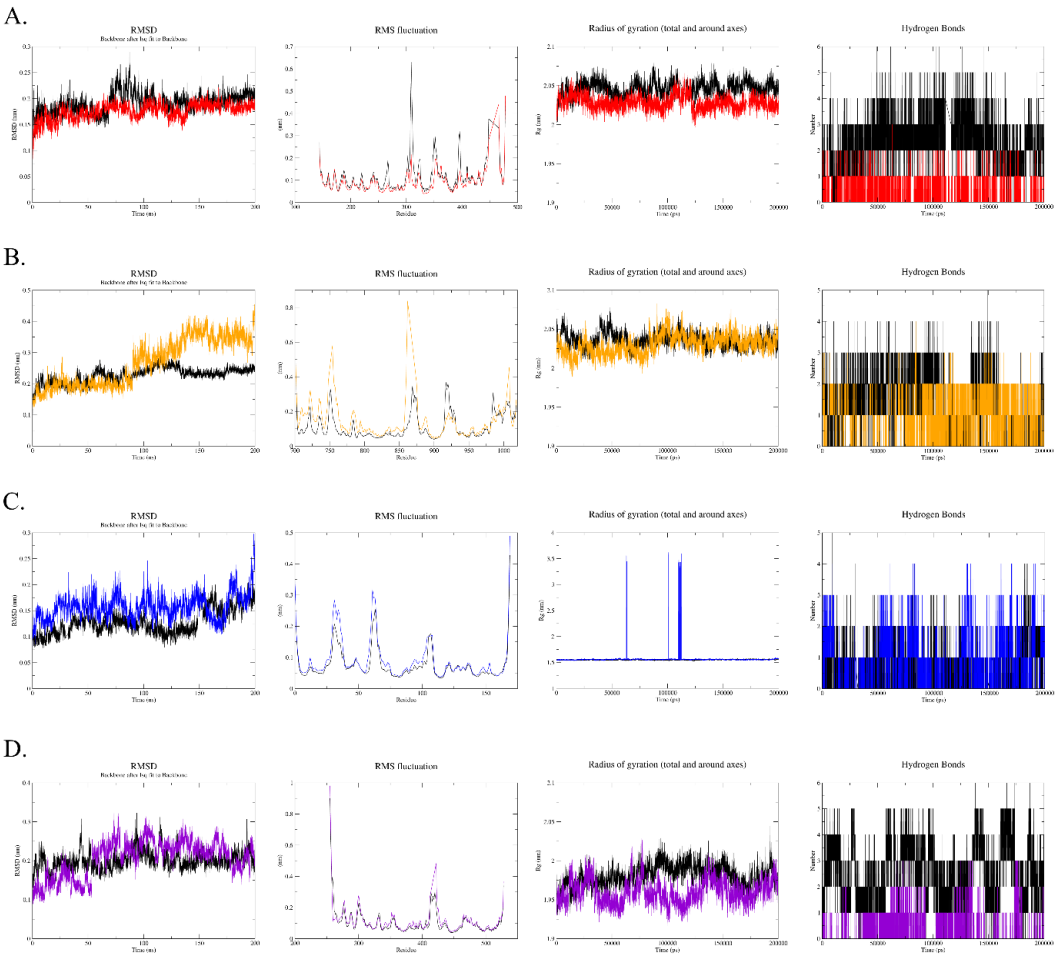

Figure 5 Molecular dynamics simulation analysis of protein-ligand complexes, including root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and hydrogen bond profiles. (A) AKT1 complexes with the control inhibitor SMR (black) and the candidate compound CID: 44425461 (red). (B) EGFR complexes with the control inhibitor W19 (black) and the candidate compound CID: 76336754 (orange). (C) KRAS complexes with the control inhibitor MOV (black) and the candidate compound CID: 11081540 (blue). And (D) SRC complexes with the control inhibitor 24A (black) and the candidate compound CID: 16757188 (purple).

MM/GBSA binding free energy analysis

The post-molecular dynamics analysis includes an MM/GBSA calculation. **Figure 6** shows the ratio of binding free energy (ΔG Total) between the control (black line) and the drug candidate compound. The MM/ computational simulation was performed for 200 nanoseconds and divided into 8,000 frames. **Figure 6(A)** shows the ΔG total value for the AKT1 complex. The figure shows that the control has a more negative total binding free energy value than the CID 44425461 compound. This finding indicates that the control has a stronger affinity than the test compound. However, the graph shows that the CID 44425461 compound exhibits a stable, consistent pattern throughout the simulation. **Figures 6(B)** shows the total binding free energy of the EGFR complex. The CID 76336754 compound showed a lower total ΔG than the control. Thus, in some simulations, the test compound showed stronger binding, characterized by a more negative ΔG. Next, the incline at the 75 - 100 nanosecond range indicates a phase of conformational adjustment in the system, after which it reaches a stable state.

Furthermore, given the important roles of KRAS and SRC in the signaling pathway, the MM/GBSA evaluation was also performed on the protein complex, and the results are shown in **Figures 6(C)** and **6(D)**. The graph shows that the CID compounds 11081540 and 16757188 have higher ΔG than the control. The ΔG gap is 5 -10 kcal/mol, indicating a difference in bond stability between the control system and the test compound complex. This difference is influenced by several factors, such as van der Waals forces and electrostatic interactions, which help determine the stability of protein-ligand complexes. Compared directly with the test compound, the control remains more stable. However, the entire system of test compounds could reach stability at a specific time without any significant graphic spikes and maintain it until the end of the simulation.

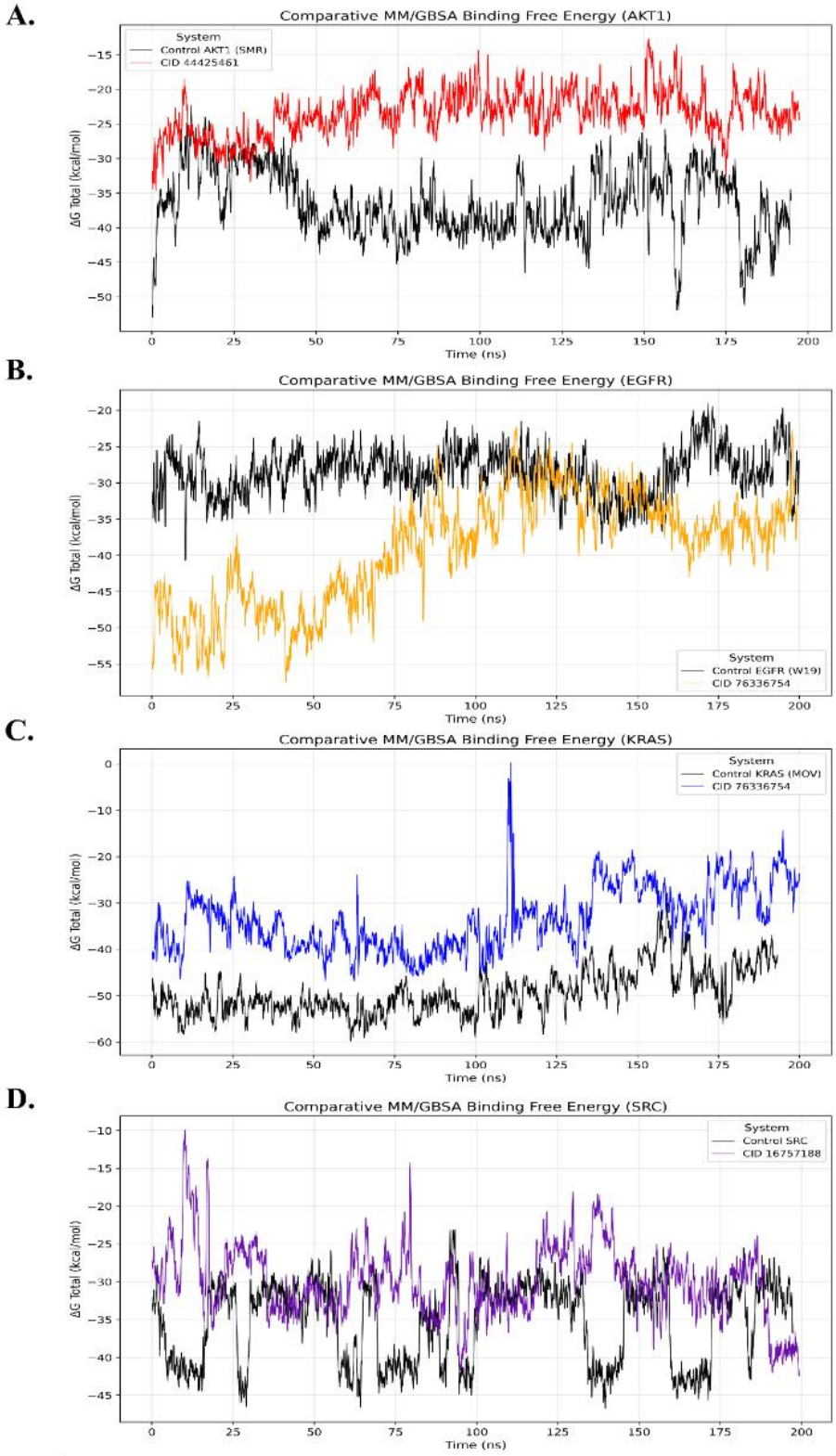

Figure 6 Molecular mechanics generalized Born surface area binding free energy analysis of protein-ligand complexes. (A) AKT1 complex, (B) EGFR complex, (C) KRAS complex, (D) SRC complex.

Membrane permeability assessment analysis (PerMM)

The membrane permeability prediction showed the differences in interaction profiles between the control and test compounds (**Figure 7**). The control co-crystallized ligand (**Figures 7(A) - 7(C)**) maintained consistent orientation and dynamics when passing through the lipid membrane. The control and test

compounds crossed both leaflet membranes with a distance between leaflets of 30 Å. **Figure 7(H)** shows the binding free energy profile of the control ligand. It shows energy valleys along the sides and a peak in the middle of the membrane. The control compound w19 showed the highest value at 6 kcal/mol. This finding indicates that it is more difficult for the w19 compound to penetrate the hydrophobic core than smr (1 kcal/mol), which has the lowest profile.

Meanwhile, the results for the test compounds were significantly different from those of the control. Compounds CID 16757188, 11081540, 76336754, and 44425461 show values below 0 kcal/mol in the middle of the graph. The test compounds CID 44425461 and 11081540 have the best balance and value, with valleys that are not too negative and resistance in the middle being close to Z = 0. Meanwhile, the CID 16757188 compound showed the deepest valley, indicating the ligand could be easily trapped in the membrane. Compound CID 76336754 also shows an increase in its Z values (0 - 12 Å), indicating that it was difficult for this compound to penetrate the membrane due to a high energy barrier. Overall, the moderate binding free energy value shows good ligand dynamics within the lipid bilayer environment. Therefore, the compound could interact well with the membrane without any indications of excessive energy entrapment.

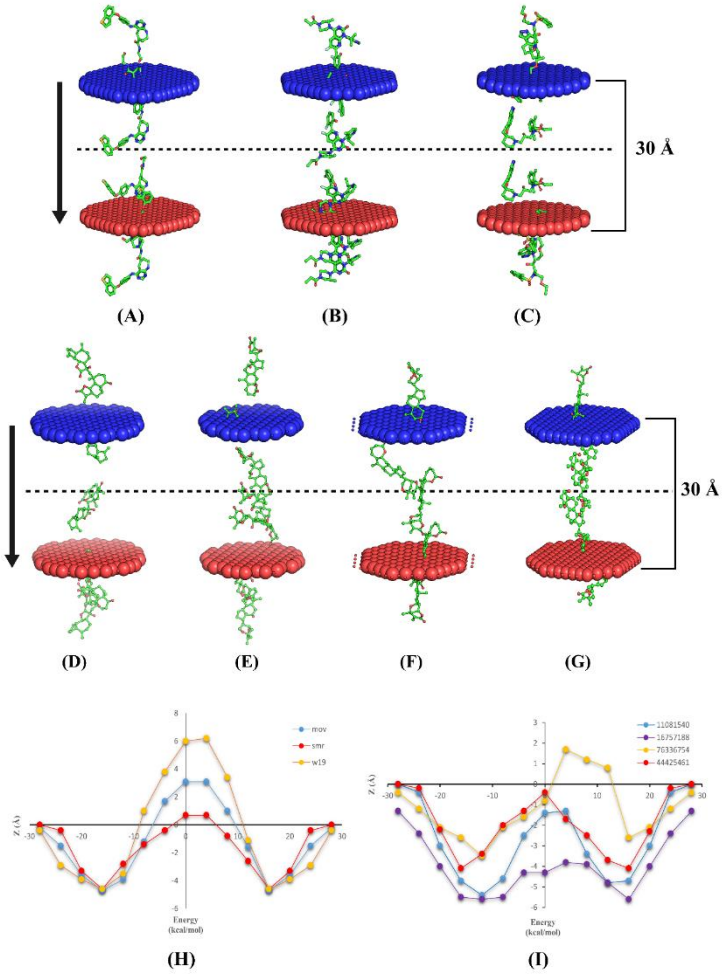

Figure 7 Membrane permeability analysis of control inhibitors and candidate compounds using the perMM approach. Control inhibitors: (A) EGFR, (B) KRAS, and (C) AKT1 complexes. Candidate compounds: (D) CID: 11081540, (E) CID: 44425461, (F) CID: 16757188, (G) CID: 76336754. Energy profiles: Free energy profile of membrane permeation for (H) control inhibitors and (I) candidate compounds.

Free Energy Landscape (FEL) analysis

The free energy landscape (FEL) analysis, based on PC1 and PC2 projections, shows the distribution of energy states between the control system and the test compounds. The results are presented as a heatmap, with color gradients indicating the distribution of Gibbs

energy across the system's conformation. The energy distribution is shown in **Figure 8** for the FEL of the control system with the test compounds in each complex system. **Figures 8(A)** and **8(B)** show that the energy basin for the AKT1 protein is relatively spread across the control system. Meanwhile, the test compound CID 44425461 exhibited a mostly centralized system, accompanied by a reduced conformational distribution pattern. Next, the EGFR complex (**Figures 8(C)** and **8(D)**) shows a fragmented control system with multiple main basins. Similarly, the test compound CID 76336754 had an energy distribution that was more concentrated in a single dominant basin. The test compounds also showed several transition areas around the main basin.

Furthermore, the KRAS system complex (**Figures 8(E)** and **8(F)**) shows a clear energy difference between the control and the test compound, CID 11081540. In the control, the distribution is broader, whereas the test compound shows a narrowing of the low-energy region with a single main basin. The FEL of the SRC system (**Figures 8(G)** and **8(H)**) shows that the control exhibits two distinct low-energy regions. The test compound CID 16757188 has a more focused distribution in a single dominant basin, with a limited distribution pattern. Overall, the FEL results show distinct conformational distributions between the control and test compounds. The test compounds showed a potentially stable dynamic potential, as their energy distributions are centered on the dominant basin.

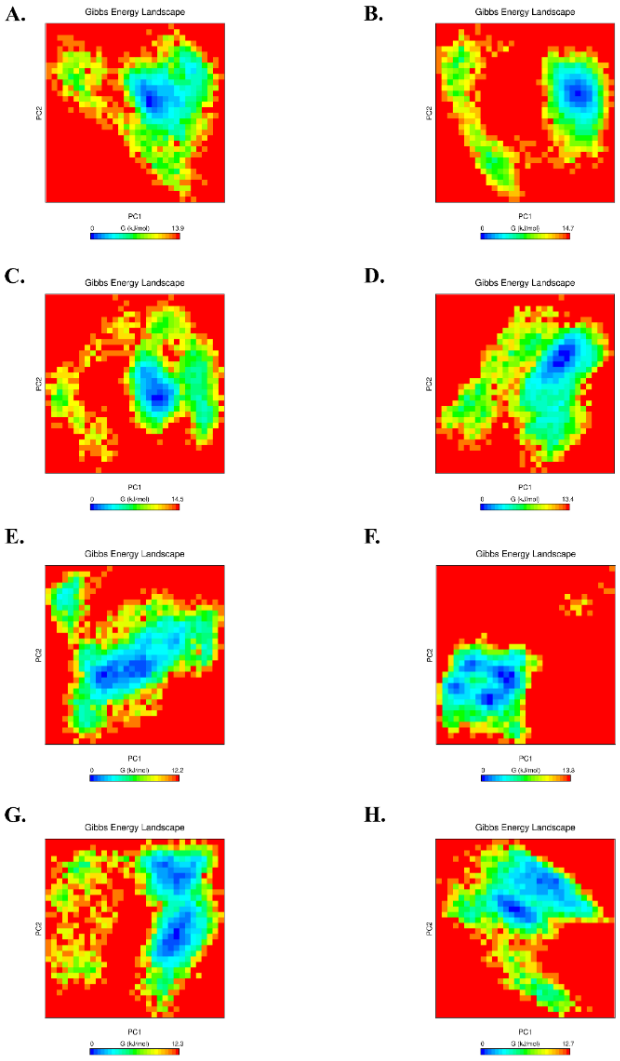

Figure 8 Free energy landscape (FEL) plots of protein-ligand complexes. (A - B) AKT1 control and AKT1- CID: 44425461 complexes, (C - D) EGFR control (W19) and CID: 76336754 complexes, (E-F) KRAS control (MOV) and CID: 11081540 complexes, and (G-H) SRC control (24A) and CID: 16757188 complexes.

Principal component analysis and dynamic cross-correlation matrix analysis

The stability of the protein-ligand complex was further analyzed using principal component analysis (PCA) and the dynamic cross-correlation matrix (DCCM). This test was performed to identify the difference in dynamics between the control system and the test compound complexes. In the AKT1 system complex, comparing the control (**Figure 9(A)**) and the CID 44425461 (**Figure 9(B)**) complexes showed that the PCA distribution was more dispersed for the test compound. The PCA analysis showed that the first three main components (PC1-PC3) accounted for 33.9% of the total variance in the control conformation. Meanwhile, the CID 44425461 compound showed a value of 24.26%. This decrease in cumulative variance percentage in the test compound indicates a higher index of conformational flexibility. The DCCM results for CID 44425461 indicate a pattern of distribution to various degrees of structural freedom due to the interaction between the AKT1 complex and CID 44425461.

Meanwhile, for the other three complexes, PCA analysis showed that ligand binding increased the variance explained by the three main components (PC1-PC3) relative to the control. In contrast, the test compound CID 44425461 showed a lower covariance than the control. In the EGFR complex, the test compound CID 76336754 showed a cumulative variance of 55.65%, compared with 35.09% for the control (**Figures 9(C)** and **9(D)**). Then, in the KRAS complex, the cumulative variance for the control was 35.77%, and for the test compound, CID 11081540, it was 42.04% (**Figures 9(E)** and **9(F)**). Meanwhile, in the SRC complex, the test compound CID 16757188 showed a stable, high-variance profile of 50.19%, compared with the control at 48.79% (**Figures 9(G)** and **9(H)**). This finding shows that CID 76336754 and 16757188 exhibit a 50% variance profile, indicating stability and regularity in their dynamics.

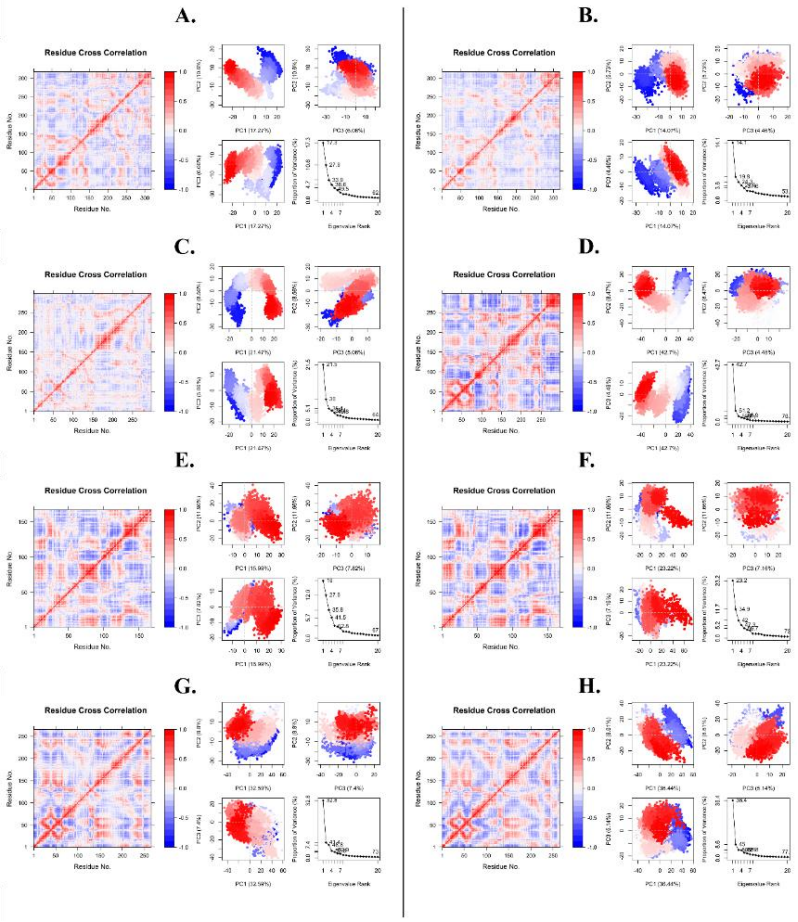

Figure 9 Principal component analysis (PCA) and dynamic cross-correlation matrix (DCCM) of protein-ligand complexes. (A - B) AKT1 control and AKT1- CID: 44425461 complexes, (C - D) EGFR control (W19) and CID: 76336754 complexes, (E - F) KRAS control (MOV) and CID: 11081540 complexes, and (G - H) SRC control (24A) and CID: 16757188 complexes.

Discussions

Lung cancer is among the most prevalent types of malignant cancers and remains the leading cause of cancer-related death globally [53]. This disease is clinically classified into two major subtypes: Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) [54,55]. Despite the advancement of treatment strategies, current treatments, such as chemotherapy, radiation, and immunotherapy, remain frequently associated with significant adverse side effects [56,57]. These side effects compromise the patients' quality of life during treatments. Therefore, there is a need for more effective therapeutic agents with reduced toxicity that can mitigate the adverse effects of current treatments. In this context, traditional Eastern medicine has been an interesting alternative, gaining increasing potential as a complementary approach through the holistic use of plant-derived compounds [19,58]. Accumulating evidence suggests that such interventions confer both preventive and therapeutic benefits. However, the efficacy of traditional Eastern medicine still varies.

Indonesia is one of the world's most biodiverse countries. The country has diverse plant biodiversity, making it an abundant source of plant-derived natural products [59,60]. These herbal plants contain a wide range of bioactive compounds [61]. They are typically processed and regulated in accordance with national regulations to produce the widely known traditional medicine referred to as *jamu* [19,40,62]. Therefore, there is a high potential for discovering novel therapeutic agents for lung cancer derived from Indonesian herbal plants. However, a large proportion of these plants remains insufficiently explored. Many studies have shown the health benefits of these plants, but they lack comprehensive scientific validation. Thus, optimization strategies to identify and refine bioactive compounds from Indonesian medicinal plants could provide a promising approach to developing novel lung cancer therapeutics.

Previous studies have explored the functional annotation of the herbal plants using *in silico* approaches [63,64]. Advances in computational technology have enabled high-throughput drug screening by integrating diverse bioinformatics techniques [65,66]. These computational methodologies opened new avenues in drug discovery by combining multiple analytical methods and algorithms [51,67,68]. In this study, the researchers applied these integrative bioinformatics approaches to address the research objective of evaluating the potential of Indonesian herbal plants (*jamu*) in lung cancer treatment.

The two main therapeutic strategies commonly employed in lung cancer treatment aim to activate tumor suppressor genes or inhibit oncogenes [69-71]. Oncogene inhibition aims to modulate dysregulated gene expression involved in cell cycle progression and uncontrolled cellular proliferation [72]. By suppressing proliferative signaling and oncogene overexpression, cancer cells may be indirectly driven toward apoptosis [73]. Conversely, the activation of tumor suppressor genes can also directly enhance pro-apoptotic pathways and contribute to the elimination of malignant cells [70,74].

In this study, the therapeutic strategy was selected by identifying optimal protein targets using an integrated approach combining network pharmacology, enrichment analysis, and TCGA. Based on this multi-layered analysis, four key target proteins (AKT1, EGFR, KRAS, and SRC) were selected. These protein targets were selected based on their high centrality metrics, including degree, betweenness, and closeness [68,75]. The network pharmacology analysis identified several additional potential targets, such as STAT3. However, the TCGA expression data showed lower expression of STAT3 protein levels in lung cancer samples compared to normal tissues. Therefore, this protein was excluded. Other potential target proteins, such as the tumor suppressor genes TP53 and PTEN, were considered. Although tumor suppressor genes are an effective therapeutic strategy, this study primarily focuses on oncogenic proteins as potential therapeutic targets. Thus, TP53 and PTEN were also excluded.

Next, the functional gene annotation revealed that the potential gene pool is strongly associated with cancer-related signaling pathways. The calculation was performed using both SRplot and Metascape to enhance the reliability of the enrichment analysis [31-33]. The consistency of the enrichment outcomes across these two platforms indicates that the identified biological pathways are robust and not dependent on a single analytical method. The use of two independent

platforms also served as a cross-validation strategy to strengthen the pathway analysis and identification of specific genes associated with lung cancer. The overall data mining result revealed that the identified genes play significant roles in cancer-associated pathways, including cell proliferation, apoptosis, locomotion, and metastasis. These findings are further supported by the SRplot's cross-validation results, which highlighted several key signaling pathways, including kinase activation and the MAPK cascade [76]. The cellular component analysis also identified significant enrichment in components such as the phosphatidylinositol 3-kinase complex, membrane rafts, and focal adhesions. These cellular components function as signaling microdomains that activate AKT-MAPK-SRC, thereby regulating critical cellular signaling events [77]. Additionally, adhesion and membrane rafts also represent key cellular contact sites involved in the regulation of cancer cell migration and invasion, thereby facilitating metastatic progression [78].

The candidate compounds were initially screened against the selected target proteins, AKT1, EGFR, KRAS, and SRC using molecular docking. This method aims to evaluate the binding affinity of candidate compounds to the protein models' active binding sites [79,80]. Protein structures were retrieved from the RCSB Protein Data Bank, with selection criteria focused on the highest structural quality, prioritizing low-resolution values and X-ray diffraction methods. Visualization of the X-ray crystallographic structures confirmed the high quality of the three-dimensional protein models [81,82]. Model validation was performed by re-docking the respective co-crystallized ligands into each protein binding site. **Table 1** presents the co-crystallized ligand re-docking results for each protein model, serving as controls to evaluate the reliability and suitability of the protein models for docking-based screening. The results demonstrated strong binding affinities for the control complexes: AKT1 (–9.5 kcal/mol), EGFR (–11.2 kcal/mol), KRAS (–7.8 kcal/mol), and SRC
(–8.7 kcal/mol). These negative values indicate strong ligand-protein interactions [43,83]. Based on these results, the binding sites of the co-crystallized ligands were used as reference coordinates for molecular docking-based compound screening.

Molecular docking analysis was conducted to identify the most promising candidate compounds. A total of 24,000 compounds were screened based on binding affinity, RMSD, and interaction profiles, with the most stable binding to the target proteins. The screening strategy was based on oncogene protein inhibition, using the co-crystallized ligand as a reference template to define binding-site coordinates. The top-ranked candidate compounds that exhibit the highest binding affinities relative to other screened molecules are summarized in **Table 1**. Meanwhile, the results of the 100 best compounds in each complex are shown in Supplementary Material 1. The best compounds were selected based on their pharmacological properties to yield candidate compounds that are readily absorbed by the body and safe for consumption.

Next, the top-ranked docking hits were further selected based on drug-likeness and safety criteria through pharmacokinetic and ADMET evaluations. These analyses aim to assess oral bioavailability and to predict the potential of candidate compounds to be effectively absorbed *in vivo* [65,84]. Lipinski's Rule of Five was applied as an initial screening criterion: molecular weight < 500 Da, hydrogen bond donors < 5, hydrogen bond acceptors < 10, and logP < 5. Lipinski's rule is widely used for early assessment of oral drug candidacy, particularly for predicting membrane permeability [85]. Compounds that satisfy Lipinski's rule are generally more likely to exhibit favorable gastrointestinal absorption [85]. These screened compounds were further evaluated using ADMET parameters to cross-validate their pharmacokinetic profiles and safety characteristics [44]. Toxicity parameters are crucial considerations in selecting candidate drug compounds. This analysis comprised Ames toxicity, hepatotoxicity, hERG I, and hERG II to evaluate mutagenicity, liver dysfunction, and potential cardiotoxicity in the candidate compounds [44]. The selected candidate compounds must be free from any indication of these four toxicities. The results of this evaluation serve as an initial screening to fully support the selection of drugs with better pharmacokinetic and safety profiles.

In addition to Lipinski's rule and ADMET profiling, the membrane permeability of the selected compounds was further validated using perMM [45,86]. This analysis was conducted to predict molecular

permeability across lipid membranes as an indicator of oral absorption potential. **Figure 7** exhibits the passive diffusion profiles of the top-ranked candidate compounds across the cellular membrane. The results in **Figures 7(H)** and **7(I)**) show the permeability values for the control and candidate compounds in the preliminary pharmacokinetic evaluation.

However, the perMM analysis for the EGFR control compound could not be analyzed due to a lack of computational and training datasets for molecules containing phosphorus (P) atoms [87]. This lack of data is due to insufficient experimental data needed for accurate phosphorus parameterization in the perMM membrane permeability model [87]. In this note, several clinically used anticancer agents, such as cyclophosphamide, ifosfamide, and melflufen, contain phosphorus atoms [88]. Therefore, future development of the perMM model, incorporating more comprehensive datasets, such as phosphorus-containing compounds, would enhance predictive accuracy. However, it should be emphasized that the perMM analysis in this study was applied as a complementary evaluation rather than a sole determinant of drug candidacy.

As a subsequent step to address the inherent limitations of molecular docking, molecular dynamics (MD) simulations were performed. In this study, molecular docking was employed as an efficient high-throughput screening approach [43,89]. However, this technique is limited by its static nature and its exclusion of explicit water-molecule effects, which are critical in intracellular environments [67,90]. Molecular docking yields a single-frame conformation, whereas molecular dynamics simulations provide a time-resolved representation of the system's dynamic behavior [51,91]. Thus, molecular dynamics simulations can be used to further assess the conformational dynamics and stability of protein-ligand complexes [92]. The compounds selected for molecular dynamics simulation analysis were the top candidates identified from the molecular docking, pharmacokinetic, and ADMET screenings (CID: 11081540, 16757188, 76336754, and 44425461).

Molecular dynamics simulations were conducted to assess complex stability using four primary parameters: root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and hydrogen bond [52]. The stability assessment across these parameters indicated that all systems on each protein reached equilibrium during the early stages of simulation. The AKT1 - CID 4425461 and SRC - CID 16757188 complexes exhibited RMSD, RMSF, and Rg values relatively comparable to or lower than those of the control complexes, indicating that ligand binding did not disrupt global protein stability. In contrast, the EGFR-CID 76336754 complex displayed significant RMSD fluctuations after 100 ns. This fluctuation suggests conformational adjustment of the protein upon ligand binding. The KRAS - CID 11081540 complex maintained a stable RMSD range with moderate fluctuations. Although transient spikes in Rg were observed, the overall Rg profile remained stable. These results indicate temporary structural adjustments rather than significant destabilization. Collectively, these observations suggest that the EGFR - CID 76336754 complex warrants further assessment, as it indicates conformational instability during molecular dynamics simulations. Nevertheless, additional post-molecular dynamics analysis is required to validate the results further.

Thus, the MM/GBSA analysis was conducted to estimate binding free energy and confirm interaction stability during molecular dynamics simulations [49]. To assess the difference in stability levels between complexes, particularly EGFR - CID 76336754, a complementary quantitative assessment was needed. The MM/GBSA calculation results showed that all compounds in each complex maintained a relatively stable profile throughout the simulation. The pattern shown on the graph does not exhibit a substantial gap relative to the provided control reference. In the MM/GBSA calculations, a more negative ΔG_binding value reflects stronger ligand-protein binding affinity [93,94]. Therefore, smaller ΔG fluctuations also reflect better consistency [95]. The candidate compound assessed on the KRAS protein exhibited a slight increase in ΔG toward zero, but this effect was temporary, and the system rapidly stabilized. The system indicates an adjustment in ligand orientation within the binding pocket, resulting in a restored pattern showing a continuing stable energy.

Further computational optimization using MM/GBSA calculations was performed on the co-crystallized EGFR ligand through targeted atom-

specific modifications. However, the presence of pseudo atoms (LPH) prevented the MM/GBSA calculations from running properly. Therefore, these pseudo atoms were removed from the protein-ligand complex [49,96]. This removal was performed on the index (ndx) and topology (top) files to generate a system compatible with MM/GBSA calculations without altering the core geometry of the complex. Several atom modifications were required to ensure the computational system runs properly. However, the atom-specific modifications were kept to a minimum to ensure the computational system ran normally without causing significant differences in the results.

The post-molecular dynamics analyses also included free energy landscape (FEL) calculations to identify the most stable conformational states during the molecular dynamics simulations [50]. The FEL analysis complements the MMGBSA results by providing a specific analysis of the dynamic stability of the protein-ligand complex. The MMGBSA was used to assess ligand-protein binding, focusing on binding free energy (ΔG) and domain enthalpy [46]. Meanwhile, the FEL focuses on the free-energy distribution during molecular dynamics simulations [51]. **Figure 8** shows the comparisons between control ligands and the best candidate compounds on each protein target. Compound CID: 44425461, in complex with AKT1, exhibited conformational switching behavior. The compound CID: 11081540, in complex with KRAS, and the compound CID: 76336754, in complex with EGFR, demonstrated good conformational stability. Meanwhile, compound CID: 16757188 in complex with SRC showed moderate stability, characterized by a broader and elongated basin. This pattern suggests conformational rearrangement within a relatively stable structure and ongoing internal adjustments during the simulations.

The final stability validation was performed using PCA and DCCM analyses to strengthen the analysis of the best candidate compounds. The PCA and DCCM analyses provide additional insights into protein dynamics when binding to test compounds. The results demonstrate the potential significance of candidate compounds as drugs, based on more restricted protein conformations, including a more organized residue correlation pattern that indicates well-coordinated protein fluctuations in a specific conformation [97]. These results were then combined with other stability analyses, including RMSD, RMSF, Rg, MM/GBSA, and FEL, to demonstrate strong binding stability for the four test compounds.

Overall, the results indicate that compounds CID 44425461, 76336754, 11081540, and 16757188 are potential drug candidates targeting the AKT1, EGFR, KRAS, and SRC proteins, respectively. The various computational and bioinformatics techniques used have identified that these candidate compounds in each complex have the potential to be drug candidates for lung cancer. Each compound was directly compared with a control co-crystallized ligand, serving as the comparison and primary template for each parameter. Based on the molecular docking, molecular dynamics, MM/GBSA, FEL, PCA, and DCCM tests, the test compounds demonstrated good stability in the complex system.

Although this study comprehensively incorporated various computational approaches, several limitations should be noted. All research results are predictions based on *in silico* approaches, such as molecular docking and molecular dynamics. While these approaches are effective for the early stages of drug discovery, they do not yet fully represent the actual biological complexity under physiological conditions. In addition, some prediction methods, such as ADMET and perMM, still rely on specific datasets and algorithm parameters, limiting the accuracy of predictions for complex natural compounds. This study relies heavily on existing databases to develop algorithms and make predictions that closely resemble real-world conditions.

Beyond the computational limitations, additional biological and translational considerations are also important in evaluating the therapeutic potential of candidate compounds for lung cancer treatment. Although the integrated analysis conducted in this study provides an optimal strategy for the early stages of drug discovery, further *in vitro* and *in vivo* studies are still needed to validate the biological activity, safety, and therapeutic relevance of the identified compounds [98,99]. Experimental studies can also be used to strengthen the evidence and confidence level of the research conducted. Several experimental validations can be performed, including enzyme inhibition assays, cell-based assays, ADME experiments, and *in vivo* validation [100,101]. This research provides the

foundation and initial steps for conducting early-stage screening based on biomolecular analysis using bioinformatics techniques.

Nevertheless, this research can be enhanced by improving the therapeutic targets. Future studies can also explore other therapeutic targets and other strategies, such as activating tumour suppressor genes and blocking the locomotion pathway, which could be used in lung cancer treatment [102]. Potential protein targets that have been explored, such as STAT3, p53, PTEN, MMP9, and ERBB2/HER2 [103], can still be further investigated to enrich and expand new opportunities for drug discovery research in lung cancer.

Conclusions

This concludes the entire series of studies that have been conducted and answers the research objective of examining the stability of protein-ligand complexes in the AKT1, EGFR, KRAS, and SRC targets. The best candidate compounds are CID 44425461 (AKT1), 76336754 (EGFR), 11081540 (KRAS), and SRC (16757188). The candidate compounds have the potential to be oral drugs, having passed pharmacokinetics and ADMET tests. Therefore, it can be concluded that these four compounds are the best compounds from the screening and are predicted to be safe for oral consumption. A series of tests has also been used to evaluate stability, and it has been concluded that the four compounds can bind to target proteins, with the aim of inhibition based on the controls used.

Declaration of Generative AI in Scientific Writing

During the preparation of this manuscript, the authors used DeepL and Grammarly to improve the readability and language of the manuscript. These tools were used only for language refinement and grammar checking. All content, interpretation, and conclusions were developed and verified by the authors, who take full responsibility for the final version of the manuscript.

CRediT Author Statement

Alfa Marzelino: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing-Original Draft, Writing- Review and Editing, Visualization, Supervision, Project Administration, Funding Acquisition. **Dona Suzana:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing-Original Draft, Writing- Review and Editing, Visualization, Supervision, Project Administration, Funding Acquisition. **Dina Melia Oktavilantika:** Conceptualization, Methodology, Investigation, Writing-Original Draft, Writing-Review and Editing, Funding Acquisition. **Jonathan Timothy Oei:** Methodology, Data Curation, Writing-Original Draft, Writing-Review and Editing, Funding Acquisition. **Patricia Evelyna Bitin:** Software, Writing-Original Draft, Writing Review and Editing. **Giovanni Effendy:** Software, Writing-Original Draft, Writing Review and Editing. **Ionna Wijaya:** Software, Writing-Original Draft, Writing Review and Editing. **Hanna Haviva Picaulima:** Software, Writing-Original Draft, Writing Review and Editing. **Elizabeth Aurelia:** Software, Writing-Original Draft, Writing Review and Editing. **Laurin Alrysia:** Software, Writing-Original Draft, Writing Review and Editing. **Militia Difa Praycilia Tubagus:** Software, Writing-Original Draft, Writing Review and Editing. **Renata Kharis Enggrasari:** Software, Writing-Original Draft, Writing Review and Editing.

References

[1] F Bray, M Laversanne, H Sung, J Ferlay, RL Siegel, I Soerjomataram and AJ DVM. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 2024; **74(3)**, 229-263.

[2] KD Nijs, HJD Koning, CVD Aalst and KT Haaf. Medical costs of lung cancer by stage, histology and first-line treatment modality in the Netherlands (2012 - 2021). *European Journal of Cancer* 2024; **208**, 114231.

[3] OECD. *Tackling the Impact of Cancer on Health, the Economy and Society*. OECD Publishing, Paris, 2024.

[4] J Li, X Li, H Zhang and M Yang. POSC95 the economic burden of lung cancer in China: A systematic literature review. *Value in Health* 2022; **25(1)**, 105.

[5] A Zafar, S Khatoon, MJ Khan, J Abu and A Naeem. Advancements and limitations in traditional anti-cancer therapies: A comprehensive

review of surgery, chemotherapy, radiation therapy, and hormonal therapy. *Discover Oncology* 2025; **16**, 607.

[6] C Sulaiman, BP George, I Balachandran and H Abrahamse. Cancer and traditional medicine: An integrative approach. *Pharmaceuticals* 2025; **18(5)**, 644.

[7] PK Dhanyamraju. Drug resistance mechanisms in cancers: Execution of pro-survival strategies. *Journal of Biomedical Research* 2024; **38(2)**, 95-121

[8] MS Shrmeka, MF Semman, BT Moges, FN Dereja and AW Garedo. Chemotherapy-related adverse drug reaction and associated factors among adult cancer patient attending Jimma medical center oncology unit, Southwest Ethiopia. *PLoS One* 2025; **20(5)**, 0321785.

[9] National Cancer Institute, Available at: https://seer.cancer.gov/statfacts/html/lungb.html, accessed April 2026.

[10] K Choucair, H Imtiaz, MH Uddin, M Nagasaka, MN Al-Hallak, PA Philip, B El-Rayes, BC Pasche and AS Azmi. Targeting KRAS mutations: Orchestrating cancer evolution and therapeutic challenges. *Signal Transduction and Targeted Therapy* 2025; **10**, 385.

[11] Y Dong, L Khan and Y Yao. Immunological features of EGFR-mutant non-small cell lung cancer and clinical practice: A narrative review. *Journal of the National Cancer Center* 2024; **4(4)**, 289-298.

[12] MS Ayoup, I Shawki, H Abdel-Hamid, DA Ghareeb, A Masoud, MF Harras, M El-Atawy, NS Alharbi and MMF Ismail. Targeting EGFR/PI3K/AKT/mTOR signaling in lung and colon cancers: Synthesis, antitumor evaluation of new 1,2,4-oxidiazoles tethered 1,2,3-triazoles. *RSD Advances* 2024; **14(24)**, 16713-16726

[13] J Zhan and M Jücker. The role of PI3K/AKT/mTOR signaling in tumor radioresistance and advances in inhibitor research. *International Journal of Molecular Sciences* 2025; **26(14)**, 6853.

[14] RS Iver, SR Needham, I Gadadas, BM Davis, SK Roberts, RCH Man, LC Zanetti-Dominiques, DT Clarke, GO Fruhwirth, PJ Parker, DJ Rolfe, FL Gervasio and ML Martin-Fernandez. Drug-resistant EGFR mutations promote lung cancer by stabilizing interfaces in ligand-free kinase-active EGFR oligomers. *Nature Communications* 2024; **15**, 2130.

[15] J Cabon, D Lerouge, S Thureau, J Balosso, M Guigo, R Gervais, C Dubos, P Dô, P Demontrond, H Curcio, S Deshaves, J Madelaine, DL Ferreira, L Jacson, K Lecrosnier, F Chevalier, F Guisier, E Dantoing, A Leconte, ..., M Césaire. Impact of KRAS G12C mutation on the efficacy of chemoradiotherapy in patients with unresectable stage II or III non-small cell lung cancer. *Frontiers in Oncology* 2025; **15**, 1675825.

[16] R Ramos, C Sousa and N Vale. Src and Abl as therapeutic targets in lung cancer: Opportunities for drug repurposing. *Pharmaceuticals* 2025; **18(10)**, 1426.

[17] PC Situmorang, S Ilyas, SE Nugraha, RA Syahputra and NMANA Rahman. Prospects of compounds of herbal plants as anticancer agents: A comprehensive review from molecular pathways. *Frontiers in Pharmacology* 2024; **15**, 1387866.

[18] K Kangra, S Kakkar, V Mittal, V Kumar, N Aggarwal, H Chopra, T Malik and V Garg. Incredible use of plant-derived bioactives as anticancer agents. *RSC Advances* 2025; **15(3)**, 1721-1746.

[19] ME Gondokesumo and AS Muthmainnah. Benefits of Indonesian saintific jamu in the treatment of various disease: A review. *Indonesian Journal of Pharmaceutical Science and Technology* 2024; **6(1)**, 71-88.

[20] SF Ameer, MY Mohamed, QA Elzubair, EAM Sharif and WN Ibrahim. Curcumin as a novel therapeutic candidate for cancer: Can this natural compound revolutionize cancer treatment? *Frontiers in Oncology* 2024; **14**, 1438040.

[21] Z Zulkarnain, PRW Astana, A Triyono, D Ardiyanto, F Novianto, U Fitriani, U Nisa and Saryanto. The potential of jamu to relieve clinical symptoms and reduce the tumor size of patients with Fibro Adenoma Mammae (FAM) at Rumah Riset Jamu (RRJ) Hortus Medicus Tawangmangu: A pilot study. *IOP Conference Series Earth and Environmental Science* 2021; **913**, 012086.

[22] CF Thorn, TE Klein and RB Altman. *PharmGKB: The pharmacogenomics knowledge base*. Humana Press, New Jersey, 2013, p. 311-320.

[23] AP Davis, TC Wiegers, RJ Johnson, D Sciaky, J Wiegers and CJ Mattingly. Comparative Toxicogenomics Database (CTD): Update 2023. *Nucleic Acids Research* 2023; **51(1)**, 1257-1262.

[24] A Hamosh, AF Scott, JS Amberger, CA Bocchini and VA McKusick. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* 2005; **30(1)**, 514-517.

[25] G Stelzer, N Rosen, I Plaschkes, S Zimmerman, M Twik, S Fishilevich, TI Stein, R Nudel, I Lieder, Y Mazor, S Kaplan, D Dahary, D Warshawsky, Y Guan-Golan, A Kohn, N Rappaport, M Safran and D Lancet. The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics* 2016; **54(1)**, 1.30.1-1.30.33.

[26] PJ Thul and C Lindskog. The human protein atlas: A spatial map of the human proteome. *Protein Science* 2018; **27(1)**, 233-244.

[27] D Szklarczyk, R Kirsch, M Koutrouli, K Nastou, F Mehryary, R Hachilif, AL Gable, T Fang, NT Doncheva, S Pyysalo, P Bork, LJ Jensen and CV Mering. The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* 2023; **51(1)**, 638-646.

[28] P Shannon, A Markiel, O Ozier, NS Baliga, JT Wang, D Ramage, N Amin, B Schwikowski and T Ideker. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003; **13(11)**, 2498-2504.

[29] L Li, L Yang, L Yang, C He, Y He, L Chen, Q Dong, H Zhang, S Chen and P Li. Network pharmacology: A bright guiding light on the way to explore the personalized precise medication of traditional Chinese medicine. *Chinese Medicine* 2023; **18**, 146.

[30] M Li, D Li, Y Tang, F Wu and J Wang. Cytocluster: A cytoscape plugin for cluster analysis and visualization of biological networks. *International Journal of Molecular Sciences* 2017; **18(9)**, 1880.

[31] Y Zhou, B Zhou, L Pache, M Chang, AH Khodabakhshi, O Tanaseichuk, C Benner and SK Chanda. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* 2019; **10**, 1523.

[32] S Xu, E Hu, Y Cai, Z Xie, X Luo, L Zhan, W tang, Q Wang, B Liu, R Wang, W Xie, T Wu, L Xie, and G Yu. Using cluster profiler to characterize multiomics data. *Nature Protocols* 2024; **19**, 3292-3320.

[33] SP Nambiar and SC Banuru. A comprehensive review of bioinformatics tools and applications revolutionising aquatic animal health management. *Journal of Fish Health* 2025; **5(2)**, 86-110.

[34] Z Tang, C Li, B Kang, G Gao, C Li and Z Zhang. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research* 2017; **45(1)**, 98-102.

[35] S Forli, R Huey, ME Pique, MF Sanner, DS Goodsell and AJ Olson. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* 2016; **11**, 905-919.

[36] NC Kallan, KL Spencer, JF Blake, R Xu, J Heizer, JR Bencsik, IS Micthell, SL Gloor, M Martinson, T Risom, SD Gross, TH Morales, WI Wu, GPA Vigers, BJ Brandhuber and NJ Skelton. Discovery and SAR of spirochromane Akt inhibitors. *Bioorganic & Medicinal Chemistry Letters* 2011; **21(8)**, 2410-2414.

[37] Y Kawakita, M Seto, T Ohashi, T Tamura, T Yusa, H Miki, H Iwata, H Kamiguchi, T Tanaka, S Sogabe, Y Ohta, and T Ishikawa. Design and synthesis of novel pyrimido [4,5-*b*] azepine derivatives as HER2/EGFR dual inhibitors. *Bioorganic & Medicinal Chemistry* 2013; **21(8)**, 2250-2261.

[38] J Canon, K Rex, AY Saiki, C Mohr, K Cooke, D Bagal, K Gaida, T Holt, CG Knutson, N Koppada, BA Lanman, J Werner, AS Rapaport, TS Miguel, R Ortiz, T Osgood, JR Sun, X Zhu, JD McCarter, LP Volak, ..., JR Lipford. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* 2019; **575**, 7781.

[39] D Dalgarno, T Stehle, S Narula, P Schelling, MRV Schravendijk, S Adams, L Andrade, J Keats, M Ram, L Jin, T Grossman, I MacNeil, C Metcalf, W Shakespeare, Y Wang, T Keenan, R Sundaramoorthi, R Bohacek, M Weigele and T Sawyer. Structural basis of Src tyrosine kinase inhibition with a new class of potent and selective trisubstituted purine-based compounds. C*hemical Biology & Drug Design* 2005; **67(1)**, 46-57.

[40] FM Afendi, T Okada, M Yamazaki, A Hirai-Morita, Y Nakamura, K Nakamura, S Ikeda, H Takahashi, M Altaf-Ul-Amin, LK Darusman, K Saito and S Kanaya. KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology* 2012; **53(2)**, 1.

[41] S Kim, J Chen, T Cheng, A Gindulyte, J He, S He, Q Li, BA Shoemaker, PA Thiessen, B Yu, L Zaslavsky, J Zhang and EE Bolton. PubChem 2025 update. *Nucleic Acids Research* 2025; **53(1)**, 1516-1525.

[42] O Trott and AJ Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 2009; **31(2)**, 455-461.

[43] J Eberhardt, D Santos-Martins, AF Tillack and S Forli. AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* 2021; **61(8)**, 3891-3898.

[44] DEV Pires, TL Blundell and DB Ascher. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry* 2015; **58(9)**, 4066-4072.

[45] AL Lomize, JM Hage, K Schnitzer, K Golobokov, MB LaFaive, AC Forsyth and ID Pogozheva. PerMM: A web tool and database for analysis of passive membrane permeability and translocation pathways of bioactive molecules. *Journal of Chemical Information and Modeling* 2019; **59(7)**, 3094-3099.

[46] I Henrique, EB Botelho, TJDS Gomes, R Kist, RA Caceres and FB Zanchi. Visual dynamics: A WEB application for molecular dynamics simulation using GROMACS. *BMC Bioinformatics* 2023; **24**, 107.

[47] S Kim, J Lee, S Jo, CL Brooks, HS Lee and W Im. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *Journal of Computational Chemistry* 2017; **38(21)**, 1879-1886.

[48] S Jo, T Kim, VG Iyer and W Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* 2008; **29(11)**, 1859-1865.

[49] MS Valdés-Tresanco, ME Valdés-Tresanco, PA Valiente and E Moreno. gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. *Journal of Chemical Theory and Computation* 2021; **17(10)**, 6281-6291.

[50] AS Abouzied, S Alqarni, KM Younes, SM Alanazi, DM Alrsheed, RK Alhathal, B Huwaimel and AM Elkashlan. Structural and free energy landscape analysis for the discovery of antiviral compounds targeting the cap-binding domain of influenza polymerase PB2. *Scientific Reports* 2024; **14**, 25441.

[51] MJ Abraham, T Murtola, R Schulz, S Páll, JC Smith, B Hess and E Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015; **1-2**, 19-25.

[52] FR Mahendra, I Prakoso, A Marzelino, MR Fadhillah, Syukriyansyah, MMA Hasibuan, JF Suparningtyas, M Kristiadi, AG Setiawan, FI Rahman, AN Sari, NR Raysendria, I Kurniawan, W Arozal and K Kusmardi. Optimisation of peptides targeting reverse transcriptase HIV-1 using QSAR, machine learning, and computational approaches. *Frontiers in Pharmacology* 2025; **16**, 1707377.

[53] B Smolarz, H Łukasiewicz, D Samulak, E Piekarska, R Kołaciński and H Romanowicz. Lung cancer—epidemiology, pathogenesis, treatment and molecular aspect (Review of Literature). *International Journal of Molecular Sciences* 2025; **26(5)**, 2049.

[54] K Tahayneh, M Idkedek and FA Akar. NSCLC: Current evidence on its pathogenesis, integrated treatment, and future perspectives. *Journal of Clinical Medicine* 2025; **14(3)**, 1025.

[55] KB Kim. The new way to define key oncogenic drivers of small cell lung cancer. *Development & Reproduction* 2023; **27(1)**, 1-7.

[56] U Anand, A Dey, AKS Chandel, R Sanyal, A Mishra, DK Pandey, VD Falco, A Upadhyay, R Kandimalla, A Chaudhary, JK Dhanjal, S Dewanjee, J Vallamkondu and JMPDL Lastra. Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes & Diseases* 2023; **10(4)**, 1367-1401.

[57] B Katta, C Vijayakumar, S Dutta, B Dubashi and VPN Ramakrishnaiah. The incidence and severity of patient-reported side effects of chemotherapy in routine clinical care: A prospective observational study. *Cureus* 2023; **15(4)**, 38301.

[58] M Loy and JK Chen. Integrating traditional East Asian medicine dietary and herbal therapies for supporting cancer treatment and survivorship in the West. *Perspectives on Integrative Medicine* 2025; **4(3)**, 171-180.

[59] A Austin, JH Colyson, FG Rohmatulloh, W Destriani and MM Rosyati. Potential of Indonesian medicinal plant biodiversity as CHK1 inhibitor agent for cancer treatment by bioinformatics and computational chemistry. *Indonesian Journal of Computational Biology* 2023; **2(1)**, 1-10.

[60] J Sun, B Liu, H Rustiami, H Xiao, X Shen and K Ma. Mapping Asia plants: Plant diversity and a checklist of vascular plants in Indonesia to Southeast. *Plants* 2024; **13(16)**, 2281.

[61] A Permatasanti and W Hidayat. Potential of Indonesian herbal as an anti-cancer therapy: A systemic review of *in vitro* studies. *Cancer Management and Research* 2023; **15**, 837-850.

[62] R Sagita, S Widhiawati and A Setyawati. Herbal therapy for cancer patients. *Jurnal Pustaka Keperawatan* 2025; **4(1)**, 62-77.

[63] VS Kulkarni, V Alagarsamy, VR Solomon, PA Jose and S Murugesan. Drug repurposing: An effective tool in modern drug discovery. *Russian Journal of Bioorganic Chemistry* 2023; **49(2)**, 157-166.

[64] NT Telang. Natural products as drug candidates for breast cancer (Review). *Oncology Letters* 2023; **26(2)**, 349.

[65] CA Lipinski, F Lombardo, BW Dominy and PJ Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* 2001; **23(1-3)**, 3-25.

[66] S Zhang, K Liu, Y Liu, X Hu and X Gu. The role and application of bioinformatics techniques and tools in drug discovery. *Frontiers in Pharmacology* 2025; **16**, 1547131.

[67] C Aguiar and I Camps. Molecular docking in drug discovery: Techniques, applications, and advancements. *Current Medicinal Chemistry* 2024; **32(28)**, 5924-5938.

[68] X Chen, M Xu and Y An. Identifying the essential nodes in network pharmacology based on multilayer network combined with random walk algorithm. *Journal of Biomedical Informatics* 2021; **114**, 103666.

[69] KC DeGeorge, BRD Jr, JS Testa and JL Rothstein. Inhibition of oncogene-induced inflammatory chemokines using a farnesyltransferase inhibitor. *Journal of Inflammation* 2008; **8**, 3.

[70] AN Sari, A Elwakeel, JK Dhanjal, V Kumar, D Sundar, SC Kaul and R Wadhwa. Identification and characterization of Mortaparib ^{Plus} — A Novel triazole derivative that targets Mortalin-p53 interaction and inhibits cancer-cell proliferation by Wild-Type p53-Dependent and independent mechanisms. *Cancers* 2021; **13(4)**, 835.

[71] R Stojchevski, EA Sutanto, R Sutanto, N Hadzi-Petrushev, M Mladenov, SR Singh, JK Sinha, S Ghosh, B Yarlagadda, KK Singh, P Verma, S Sengupta, R Bhaskar and D Avtanski. Translational advances in oncogene and tumor-suppressor gene research. *Cancers* 2025; **17(6)**, 1008.

[72] UN Wasko, J Jiang, TC Dalton, A Curiel-garcia, AC Edwards, Y Wang, B Lee, M Orlen, S Tian, CA Stalnecker, K Drizyte-Miller, M Menard, J Dilly, SA Sastra, CF Palermo, MC Hasselluhn, AR Decker-Farrel, S Chang, L Jiang, X Wei, …, KP Olive. Tumour-selective activity of RAS-GTP inhibition in pancreatic cancer. *Nature* 2024; **629**, 927-936.

[73] X Tian, PR Srinivasan, V Tajiknia, AFSS Uruchurtu, AA Seyhan, BA Carneiro, ADL Cruz,

M Pinho-Schwermann, A George, S Zhao, J Strandberg, FD Cristofano, S Zhang, L Zhou, AG Raufi, A Navaraj, Y Zhang, N Verovkina, M Ghandali, D Ryspayeva and WS El-Deiry. Targeting apoptotic pathways for cancer therapy. *Journal of Clinical Investigation* 2024; **134(14)**, 179570.

[74] A Sgro, J Cursons, C Waryah, EA Woodward, M Foroutan, R Lyu, GCT Yeoh, PJ Leedman and P Blancafort. Epigenetic reactivation of tumor suppressor genes with CRISPRa technologies as precision therapy for hepatocellular carcinoma. *Clinical Epigenetics* 2023; **15**, 73.

[75] AL Hopkins. Network pharmacology: The next paradigm in drug discovery. *Nature Chemical Biology* 2008; **4**, 682-690.

[76] C Braicu, M Buse, C Busuioc, R Drula, D Gulei, L Raduly, A Rusu, A Irimie, AG Atanasov, O Slaby, C Ionescu, and I Berindan-Neagoe. A comprehensive review on MAPK: A promising therapeutic target in cancer. *Cancers* 2019; **11**, 1618.

[77] B Li, K Tan, AR Lao, H Wang, H Zheng and L Zhang. A comprehensive review of artificial intelligence for pharmacology research. *Frontiers in Genetics* 2024; **15**, 1450529.

[78] Z Cheng, H Cui, Y Wang, J Yang, C Lin, X Shi, Y Zou, J Chen, X Jia and L Su. The advance of the third-generation EGFR-TKI in the treatment of non-small cell lung cancer (Review). *Oncology Reports* 2023; **51(1)**, 16.

[79] PC Agu, CA Afiukwa, OU Orji, EM Ezeh, IH Ofoke, CO Ogbu, EI Ugwuja and PM Aja. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Scientific Reports* 2023; **13**, 13398.

[80] I Fatima, A Rehman, Y Ding, P wang, Y Meng, HU Rehman, DA Warraich, Z Wang, L Feng and M Liao. Breakthroughs in AI and multi-omics for cancer drug discovery: A review. *European Journal of Medicinal Chemistry* 2024; **280**, 116925.

[81] L Wu, D Wei, W Chen, C Wu, Z Lu, S Li and W Liu. Comprehensive potential of artificial intelligence for predicting PD-L1 expression and EGFR mutations in lung cancer: A systematic review and meta-analysis. *Journal of Computer Assisted Tomography* 2025; **49(1)**, 101-112.

[82] T Yin, Y Chen, Y Wang, H Su, C Duan and J Liu. Protein structure prediction powered by artificial intelligence: From biochemical foundations to practical applications. *Frontiers in Molecular Biosciences* 2026; **13**, 1767821.

[83] SM Zaidh, HT Vengateswaran, M Habeeb, KB Aher, GB Bhavar, N Irfan and KNVC Lakshmi. Network pharmacology and AI in cancer research uncovering biomarkers and therapeutic targets for RALGDS mutations. *Scientific Reports* 2025; **15**, 10938.

[84] X Li, Y Lin, S Lin, J Huang and Z Ruan. Advancements in understanding cardiotoxicity of EGFR- TKIs in non-small cell lung cancer treatment and beyond. *Frontiers in Pharmacology* 2024; **15**, 1404692.

[85] M Makarem and PA Jänne. Top advances of the year: Targeted therapy for lung cancer. *Cancer* 2024; **130(19)**, 3239-3250.

[86] C Gao, W Wang, T Liu, X Li, Y Yu and J Wu. Annual review of EGFR inhibitors in 2024. *European Journal of Medicinal Chemistry* 2025; **292**, 117677.

[87] Y Wang, Y Zhang, X Zhan, Y He, Y Yang, L Cheng and D Alghazzawi. Machine learning for predicting protein properties: A comprehensive review. *Neurocomputing* 2024; **597**, 128103.

[88] B Ye and Q Xiao. Natural compounds for non-small cell lung cancer treatment: Focus on the EGFR signaling pathway. *Frontiers in Pharmacology* 2026; **17**, 1758414.

[89] S Boutry, R Helaers, T Lenaerts and M Vikkula. Rare variant association on unrelated individuals in case–control studies using aggregation tests: Existing methods and current limitations. *Briefings in Bioinformatics* 2023; **24(6)**, 412.

[90] B Polini, C Ricardi, A Bertolini, V Carnicelli, G Rutigliano, F Saponaro, R Zucchi and G Chiellini. T1AM/TAAR1 system reduces inflammatory response and β-Amyloid toxicity in human microglial HMC3 cell line. *International Journal of Molecular Sciences* 2023; **24(14)**, 11569.

[91] C Stavrogiannis, F Sofos and TE Karakasidis. Machine learning-enhanced molecular dynamics:

Current state, challenges and perspectives. *Archives of Computational Methods in Engineering* 2026; **33**, 6125-6147.

[92] MF Malekzada, SH Mosawi, N Fani and S Nazir. Integrating molecular docking and molecular dynamics simulation approaches for investigation of the affinity and interactions of the Curcumin with phosphatase and tensin homolog (PTEN) and mutated PTEN. *Journal of Molecular Structure* 2024; **1318**, 139306.

[93] K Fujita and Y Urano. Activity-based fluorescence diagnostics for cancer. *Chemical Reviews* 2024; **124(7)**, 4021-4078.

[94] R Kumar and A Parashar. Atomistic simulations of pristine and nanoparticle reinforced hydrogels: A review. *WIREs Computational Molecular Science* 2023; **13(4)**, 1655.

[95] M Taylor and J Ho. MM/GBSA prediction of relative binding affinities of carbonic anhydrase inhibitors: Effect of atomic charges and comparison with Autodock4$_{Zn}$. *Journal of Computer-Aided Molecular Design* 2023; **37**, 167-182.

[96] C Zhao, T Wang and R Luo. Extending MMPBSA for membrane proteins: Addressing P2Y$_{12}$R conformational changes upon ligand binding. *Biophysical Journal* 2025. https://doi.org/10.1016/j.bpj.2025.05.023

[97] FR Mahendra, GM Gholam, M Muhammad, MMA Hasibuan, A Rafiqi, Z Hafizh, R Pratama, A Laksmi, IM Artika and N Waras. Unveiling potential therapeutic targets for colon cancer: A comprehensive bioinformatics analysis of miRNA-Mediated regulation. *Pharmaceutical Sciences Asia* 2025; **52(1)**, 11-29.

[98] J Banerjee, AK Tiwari and S Banerjee. Drug repurposing for cancer. *Progress in Molecular Biology and Translational Science* 2024; **207**, 123-150.

[99] K Singh, JK Gupta, DK Chanchal, MG Shinde, S Kumar, D Jain, ZM Almarhoon, AM Alshahrani, D Calina, J Sharifi-Rad and A Tripathi. Natural products as drug leads: Exploring their potential in drug discovery and development. *Naunyn-Schmiedeberg's Archive of Pharmacology* 2024; **398**, 4673-4687.

[100] ANM Ansori, MH Widyananda, Y Antonius, AAA Murtadlo, VD Kharisma, PA Wiradana, S Sahadewa, FD Durry, N Maksimiuk, M Rebezov and R Zainul. A review of cancer-related hypercalcemia: Pathophysiology, current treatments, and future directions. *Journal of Medicinal and Pharmaceutical Chemistry Research* 2023; **6(7)**, 944-952.

[101] MRT Alifiansyah, MA Herdiansyah, RC Pratiwi, RP Pramesti, NW Hafsyah, AP Rania, JERP Putra, PA Cahyono, Litazkiyyah, SK Muhammad, AAA Murtadlo, VD Kharisma, ANM Ansori, V Jakhmola, PK Ashok, JM Kalra, H Purnobasuki and IA Pratiwi. QSAR of acyl alizarin red biocompound derivatives of rubia tinctorum roots and its ADMET properties as anti-breast cancer candidates against MMP-9 protein receptor: *In silico* study. *Food Systems* 2024; **7(2)**, 312-320.

[102] E Martindah, AR Khairullah, R Widiastuti, A Andriani, R Damayanti, I Mustofa, S Mulyati, M Sukmanadi, BWK Wardhani, RZ Ahmad, AO Akintunde, BP Pratama, I Melati, DAA Kurniasih, ANM Ansori, W Yulianti, M Nirwan and S Wibowo. Oxidative stress: Molecular mechanisms, role in pathology, and therapeutic implications. *Trends in Sciences* 2025; **22(12)**, 11462.

[103] MH Widyananda, SK Pratama, ANM Ansori, Y Antonius and VD Kharisma. Quercetin as an anticancer candidate for glioblastoma multiforme by targeting AKT1, MMP9, ABCB1, and VEGFA: An *in silico* study. *Karbala Internasional Journal of Modern Science* 2023; **9(3)**, 10.

Supplementary Materials

Supplementary Material 1 Top 50 KEGG Pathway Enrichment Analysis Results.

ID	Description	Gene Ratio	p-value	q-value
hsa04151	PI3K-Akt signaling pathway	93/216	*2.15e-76*	*9.71e-75*
hsa05205	Proteoglycans in cancer	69/216	*1.88e-64*	*4.26e-63*
hsa01521	EGFR tyrosine kinase inhibitor resistance	50/216	*1.50e-63*	*2.26e-62*
hsa04933	*AGE-RAGE signaling pathway in diabetic complications*	51/216	*4.23e-58*	*4.78e-57*
hsa04010	MAPK signaling pathway	73/216	*1.04e-56*	*9.39e-56*
hsa04015	Rap1 signaling pathway	64/216	*6.69e-56*	*5.05e-55*
hsa04014	Ras signaling pathway	63/216	*5.60e-51*	*3.62e-50*
hsa05163	Human cytomegalovirus infection	61/216	*7.59e-50*	*4.29e-49*
hsa05215	Prostate cancer	46/216	*2.30e-48*	*1.16e-47*
hsa05210	Colorectal cancer	43/216	*2.56e-48*	*1.16e-47*
hsa04012	ErbB signaling pathway	40/216	*1.53e-43*	*5.81e-43*
hsa04510	Focal adhesion	54/216	*1.54e-43*	*5.81e-43*
hsa05223	Non-small cell lung cancer	37/216	*4.28e-42*	*1.49e-41*
hsa01522	Endocrine resistance	41/216	*5.08e-42*	*1.64e-41*
hsa05212	Pancreatic cancer	37/216	*5.88e-41*	*1.77e-40*
hsa05161	Hepatitis B	48/216	*6.30e-41*	*1.78e-40*
hsa05417	Lipid and atherosclerosis	53/216	*1.08e-40*	*2.87e-40*
hsa05218	Melanoma	36/216	*2.18e-40*	*5.48e-40*
hsa05226	Gastric cancer	46/216	*4.18e-40*	*9.96e-40*
hsa05224	Breast cancer	45/216	*4.93e-39*	*1.12e-38*
hsa04630	JAK-STAT signaling pathway	47/216	*6.76e-39*	*1.46e-38*
hsa05225	Hepatocellular carcinoma	47/216	*1.25e-38*	*2.57e-38*
hsa05142	Chagas disease	39/216	*4.07e-38*	*8.01e-38*
hsa05167	Kaposi sarcoma-associated herpesvirus infection	49/216	*5.66e-38*	*1.07e-37*
hsa05214	Glioma	35/216	*6.50e-38*	*1.18e-37*
hsa04926	Relaxin signaling pathway	42/216	*1.15e-37*	*2.00e-37*
hsa04066	HIF-1 signaling pathway	39/216	*8.55e-37*	*1.43e-36*
hsa04935	Growth hormone synthesis, secretion and action	40/216	*3.69e-36*	*5.96e-36*

ID	Description	Gene Ratio	p-value	q-value
hsa04668	TNF signaling pathway	39/216	*3.01e-35*	*4.70e-35*
hsa05220	Chronic myeloid leukemia	33/216	*1.85e-34*	*2.79e-34*
hsa05230	Central carbon metabolism in cancer	32/216	*2.60e-34*	*3.80e-34*
hsa05213	Endometrial cancer	30/216	*2.99e-34*	*4.22e-34*
hsa04062	Chemokine signaling pathway	45/216	*2.25e-33*	*3.09e-33*
hsa05235	PD-L1 expression and PD-1 checkpoint pathway in cancer	34/216	*3.24e-33*	*4.32e-33*
hsa04917	Prolactin signaling pathway	31/216	*1.03e-32*	*1.33e-32*
hsa05135	Yersinia infection	39/216	*2.01e-32*	*2.52e-32*
hsa05219	Bladder cancer	25/216	*1.99e-31*	*2.44e-31*
hsa04722	Neurotrophin signaling pathway	36/216	*5.55e-31*	*6.61e-31*
hsa04919	Thyroid hormone signaling pathway	36/216	*1.08e-30*	*1.25e-30*
hsa05418	Fluid shear stress and atherosclerosis	38/216	*1.31e-30*	*1.48e-30*
hsa04068	FoxO signaling pathway	37/216	*1.72e-30*	*1.90e-30*
hsa05211	Renal cell carcinoma	29/216	*8.09e-30*	*8.71e-30*
hsa04370	VEGF signaling pathway	27/216	*5.05e-29*	*5.31e-29*
hsa05166	Human T-cell leukemia virus 1 infection	43/216	*3.97e-28*	*4.08e-28*
hsa04660	T cell receptor signaling pathway	34/216	*4.47e-28*	*4.50e-28*
hsa05206	MicroRNAs in cancer	49/216	*1.79e-27*	*1.76e-27*
hsa05231	Choline metabolism in cancer	31/216	*2.16e-27*	*2.08e-27*
hsa04072	Phospholipase D signaling pathway	36/216	*2.70e-27*	*2.55e-27*
hsa05221	Acute myeloid leukemia	27/216	*3.34e-27*	*3.08e-27*
hsa05131	Shigellosis	44/216	*6.74e-27*	*6.11e-27*

Supplementary Material 2 and 3

Raw data from docking (Supplementary Material 2) and molecular dynamics trajectory files (Supplementary Material 3) can be viewed via the following link https://zenodo.org/records/18626691?token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImJkYTIwOTlhLWNhMTMtNDAyMS04YmY5LTdmZmRlNGU1MjAzZSIsImRhdGEiOnt9LCJyYW5kb20iOiJiZTlmMDc0NmM0NzgzMDZmODkzZWIyYThlYWVlNTkzMSJ9.TgVrmT8a4QYNqqZVvPd2_crx_Fje4kA2Jb39mx3LUk4hksP-sEwq_keg9ma9f27FcFlEitXC-lH0Fn-cRlzr6w