


 Cite this: *RSC Adv.*, 2025, **15**, 36938

## Multi-omics pan-cancer profiling of CDK2 and *in silico* identification of plant-derived inhibitors using machine learning approaches

Md Ahad Ali, <sup>ab</sup> Hridhi Sarker, <sup>ac</sup> Tania Khan, <sup>d</sup> Humaira Sheikh, <sup>ae</sup> Ahmed Saif, <sup>f</sup> Farhad Bin Farid, <sup>d</sup> Sadia Afrin, <sup>a</sup> Most. Asha Khatun<sup>ag</sup> and Neeraj Kumar<sup>h</sup>

Cancer is a complex disease characterized by uncontrolled cell proliferation, often driven by dysregulated cyclin-dependent kinases (CDKs), particularly CDK2, which plays a crucial role in cell cycle progression. Aberrant CDK2 activity is associated with tumor growth and resistance to therapy, making CDK2 a promising therapeutic target. The main focus of this research is to integrate the multi-omics-based pan-cancer analysis of CDK2 to identify novel plant-derived inhibitors, bridging the prognostic and therapeutic relevance of CDK2 across various cancer types. In this study, to evaluate CDK2's expression, prognostic behavior, genetic alterations, and immune infiltrations, we performed pan-cancer analysis. The oncogenic analysis showed that CDK2 is significantly overexpressed in multiple tumor types and, in some cancers, which correlated with poor overall and disease-free survival, indicating its potential as a context-dependent prognostic biomarker. The involvement of CDK2 in key cell cycle and oncogenic pathways was investigated, highlighting its centrality in tumor proliferation networks. Additionally, cheminformatics and machine learning approaches were applied to screen phytocompounds from six medicinal plants, and the top phytocompounds ( $>\text{pIC}_{50} = 5.1$ ) were then subjected to molecular docking, pharmacodynamics, pharmacokinetics, and dynamics simulation studies. Docking results revealed that withanolide M, withanolide K, and ergosterol showed the highest binding affinities against CDK2, with scores of  $-10.2$ ,  $-10.1$ , and  $-9.9$  kcal mol $^{-1}$ , respectively. These lead phytocompounds exhibited high potency, excellent pharmacokinetic properties, and minimal predicted toxicity as compared with the control inhibitor of CDK2. The binding stability of the protein–ligand complexes was confirmed by dynamic simulations along with MM-GBSA calculations, with the results supporting our previously reported affinity score. Therefore, these phytocompounds could be potential CDK2 inhibitors, warranting exploration in future cancer research. Furthermore, additional experimental and clinical validations are required to confirm the efficacy and efficiency of these potential lead compounds.

Received 30th July 2025  
 Accepted 18th September 2025  
 DOI: 10.1039/d5ra05535k  
[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## 1. Introduction

Cancer is a prevalent, debilitating, and potentially fatal illness that impacts individuals and their families across diverse

sociodemographic categories, including age, sex, race, ethnicity, education, occupation, income, social class, spirituality, faith, and culture, in communities worldwide.<sup>1</sup> Despite significant advancements in its early detection, diagnosis, and treatment, cancer remains a leading cause of death globally. According to recent estimates, 19.3 million new cancer cases and nearly 10 million cancer-related deaths were reported in 2020, with projections indicating a 76.6% increase in incidence by 2050, reaching 35.3 million new cases. Likewise, global cancer-related mortality is expected to rise by 89.7%, reaching 18.5 million deaths by 2050. The burden of cancer is expected to be most severe in low- and middle-income countries (LMICs), where incidence and mortality rates will nearly triple compared to a moderate increase in high-income countries. This disproportionate burden is driven by factors such as aging populations, environmental and lifestyle changes, limited access to healthcare, and inadequate early detection programs. By 2030, nearly three-quarters of all cancer-related deaths will occur in

<sup>a</sup>Department of Computational Chemistry and Drug Design, Panacea Research Center, Bangladesh. E-mail: [ahad.chembd@gmail.com](mailto:ahad.chembd@gmail.com)

<sup>b</sup>Department of Chemistry, University of Rajshahi, Rajshahi, Bangladesh

<sup>c</sup>Department of Biochemistry and Molecular Biology, University of Rajshahi, Rajshahi, Bangladesh

<sup>d</sup>Department of Pharmacy, University of Development Alternative (UODA), Dhaka, Bangladesh

<sup>e</sup>Department of Chemistry, Gopalganj University of Science and Technology, Gopalganj, Bangladesh

<sup>f</sup>Department of Pharmacy, University of Rajshahi, Rajshahi, Bangladesh

<sup>g</sup>Department of Chemistry, Pabna University of Science and Technology, Pabna, Bangladesh

<sup>h</sup>Department of Pharmaceutical Chemistry, Bhupal Nobles' College of Pharmacy, Udaipur 313001, Rajasthan, India



LMICs with high mortality-to-incidence ratios, highlighting the urgent need for improved diagnostic and therapeutic strategies.<sup>2-4</sup>

The disease is driven by a complex interplay of molecular mechanisms, often described by the thirteen hallmarks of cancer: (I) sustaining proliferative signaling, (II) deregulating cellular energetics, (III) evading growth suppressors, (IV) resisting cell death, (V) enabling replicative immortality, (VI) inducing angiogenesis, (VII) activating invasion and metastasis, (VIII) avoiding immune destruction, (IX) tumor-promoting inflammation, (X) genome instability and mutation, (XI) unlocking phenotypic plasticity, (XII) nonmutational epigenetic reprogramming, and (XIII) senescent cells.<sup>5</sup> These hallmarks pose a formidable obstacle to effective treatment.<sup>6</sup> These hallmarks enable cancer cells to bypass normal regulatory mechanisms, allowing unchecked proliferation and tumorigenesis. Among these hallmarks, sustained proliferative signaling and evasion of growth suppressors directly contribute to uncontrolled cell cycle progression, a fundamental process in cancer development.<sup>7</sup> Cyclin-dependent kinases (CDKs) and other CDK-associated proteins regulate the cell cycle tightly, ensuring proper cell division. However, the dysregulation of these regulatory pathways leads to abnormal proliferation, genomic instability, and tumorigenesis.<sup>8,9</sup>

Cyclin-dependent kinase 2 (CDK2) is a key regulatory enzyme in the G1-to-S phase transition of the cell cycle. It forms complexes with cyclins E and A, driving DNA replication and cell cycle progression.<sup>10</sup> Under normal physiological conditions, CDK2 activity is precisely regulated by cyclin availability, cyclin-dependent kinase inhibitors (CKIs), and phosphorylation events.<sup>11</sup> However, in many cancers, CDK2 becomes hyperactivated due to genetic alterations, epigenetic modifications, or the dysregulation of its associated cyclins and inhibitors.<sup>9</sup> Several studies have reported CDK2 overexpression in several cancers, including breast, ovarian, prostate, leukemia, and lymphoma.<sup>12</sup> An elevated CDK2 expression has also been observed in oral cancers, where it serves as a negative prognostic indicator.<sup>13</sup> Additionally, an overexpression of cyclin E2, a regulatory partner of CDK2, has been linked to endocrine resistance in breast cancer cells.<sup>14</sup> To better understand the oncogenic relevance of CDK2 across different malignancies, pan-cancer analysis has emerged as a crucial tool.<sup>15</sup> This approach enables the comprehensive evaluation of CDK2 expression, mutation, and interaction patterns across diverse tumor types, helping identify universal and cancer-type-specific roles.<sup>16-19</sup> Such large-scale analysis provides valuable insights into the extent of CDK2 dysregulation across tumor types, reinforcing its critical role in promoting uncontrolled proliferation, resistance to apoptosis, and genomic instability. These insights aid in the prioritization of CDK2 as a therapeutic target and enhance the translational potential of CDK2 inhibitors in precision oncology. However, despite its potential, clinical success has been limited due to off-target effects, toxicity, and resistance mechanisms.<sup>20</sup> Recently, FDA-approved CDK4/6-targeted inhibitors have emerged, while several CDK2 inhibitors have been developed, none of which are yet FDA-approved. Some commonly used drugs, such as ribociclib, abemaciclib,

and palbociclib, which are used in hormone receptor-positive breast cancer, have shown efficacy but are associated with adverse effects, such as neutropenia, hepatotoxicity, and gastrointestinal toxicity.<sup>21</sup> Additionally, resistance to CDK4/6 inhibitors is increasingly observed, often due to the compensatory upregulation of CDK2 activity. This highlights the urgent need for selective CDK2 inhibitors to target tumors reliant on CDK2 hyperactivation. Recent studies have identified novel CDK2 inhibitors, such as INX-315, which demonstrated promising anti-cancer activity in preclinical models of CCNE1-amplified tumors.<sup>22</sup> The development of CDK2-selective inhibitors could provide a more targeted approach, overcome resistance mechanisms while minimizing off-target effects associated with pan-CDK inhibition.

The isolation and identification of chemical substances (phytocompounds) with their biological activities from natural sources have traditionally resulted in the discovery of new treatments, advancing the health and pharmaceutical sectors.<sup>23</sup> In the pharmaceutical industry, phytochemicals are the main resources for the development of any novel therapeutic compounds.<sup>24</sup> In the oncology sector, about 50% of contemporary therapeutic medications are derived from natural sources and possess the ability to combat cancer cells.<sup>25</sup> The IMPPAT is a database that contains information about the plant chemicals and therapeutics from traditionally used medicinal plants, particularly those from the Indian subcontinent, useful for facilitating *in silico* drug discovery. In contrast to other databases that concentrate on Chinese, Korean and other data that offer restricted phytochemical information, IMPPAT delivers extensive insights specifically pertinent to traditional Indian medicine.<sup>26</sup> Nowadays, the *in silico* approach is an effective technique for identifying potential CDK2 inhibitors before experimental validation through *in vitro* and *in vivo* studies, which will reduce our time and save the experimental cost as well.<sup>27</sup> Our integrative omics-ML framework aligns with the broader movement of AI-enabled biomedical discovery, where CNN-based approaches in clinical imaging, such as automated skin lesion and gastrointestinal abnormality classification,<sup>28,29</sup> exemplify the generalizability of machine learning across diverse data modalities—from imaging to genomics to drug discovery. As CDK2 plays an important role in cancer progression and considering the challenges associated with existing therapies, multi-omics guided ML-based computer-aided drug discovery (CADD) methodologies provide a rational framework for discovering selective CDK2 inhibitors with improved efficacy and reduced toxicity. To the best of our knowledge, no prior study has comprehensively combined pan-cancer validation of CDK2 expression with machine learning-based screening of natural compounds to identify potential lead inhibitors. This integrative approach remains largely unexplored in current literature.<sup>30-32</sup> Therefore, this study utilizes multi-omics-based Pan cancer analysis, machine learning-based bioactivity prediction of the collected compounds, and several CADD techniques, including molecular docking, molecular dynamics (MD) simulations, MM-GBSA, pharmacokinetics and SMILE-based quantitative structure-activity relationship (QSAR) modeling.<sup>33-36</sup> These computational techniques facilitate the



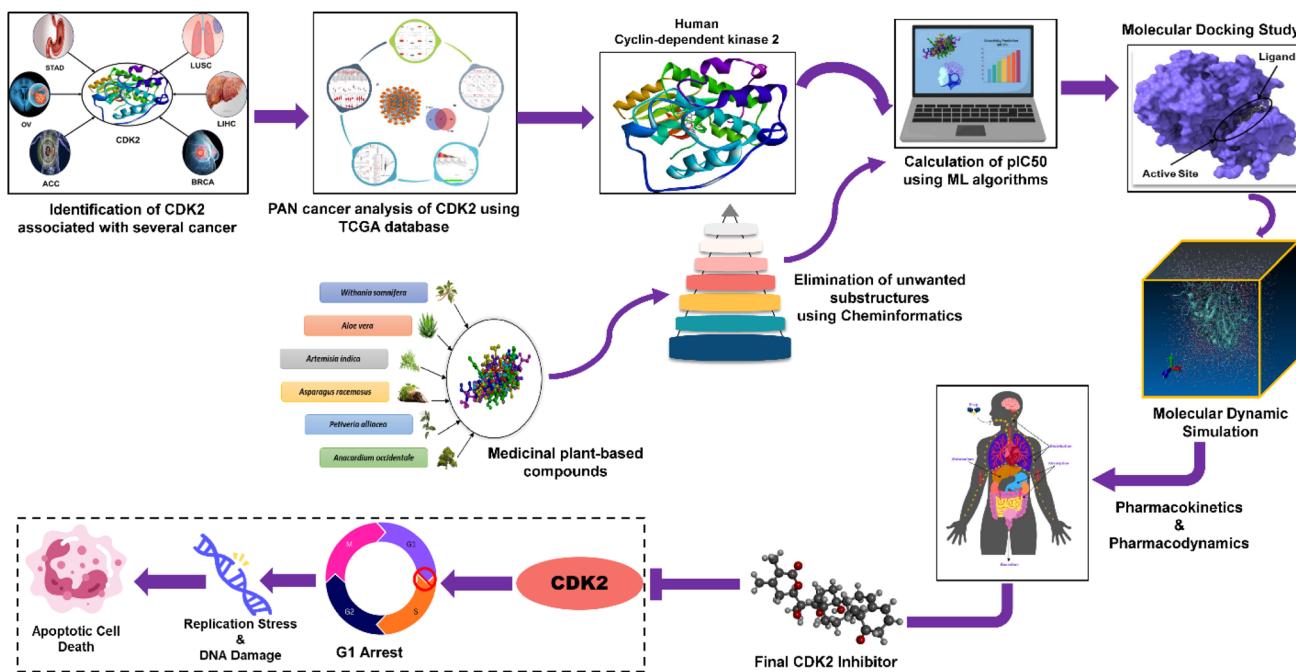


Fig. 1 Complete graphical representation/flowchart of this research.

identification of potential bioactive compounds that target the CDK2 protein, which will be good resources for the development of a novel and effective cancer therapy. The complete workflow of this study is shown in Fig. 1.

## 2. Materials and methods

To investigate the oncogenic relevance of CDK2 across a broad spectrum of human cancers, a comprehensive pan-cancer analysis was performed using integrative multi-omics data from publicly available datasets. This study included the expression analysis of transcriptomic & proteomic data, survival analysis, immune infiltration, and pathway enrichment (GO & KEGG) analyses.

### 2.1 Transcriptomic and proteomic expression profiling

The Tumor Immune Estimation Resource (TIMER 2.0) is an important database for analyzing the association between gene expression and tumor features using data from the Cancer Genome Atlas (TCGA).<sup>37</sup> In this study, TIMER2.0 (<https://timer.comp-genomics.org/>) was used to analyze the expression of CDK2 across different tumors and their corresponding normal tissues *via* the 'Gene\_DE' module of the TIMER 2.0 database.

We used the GEPIA2 (<https://gepia2.cancer-pku.cn>) database<sup>38</sup> to calculate the expression analysis box plot for certain tumors where normal controls were missing, which helps to compare the expression levels of tumors and their corresponding normal tissues using the data from TCGA and GTEx. For GEPIA2-based expression analysis, the parameters were set as follows: Log2FC cutoff = 1, *p*-value cutoff = 0.01, jitter size = 0.4. Gene expression values were log-transformed

using  $\log_2(\text{TPM} + 1)$  for visualization and statistical comparison. It should be noted that TIMER2.0 and GEPIA2 rely on different underlying datasets. TIMER2.0 is based on uniformly reprocessed TCGA RNA-seq data,<sup>39</sup> whereas GEPIA2 integrates RNA-seq data from both TCGA and GTEx processed through a unified pipeline.<sup>38</sup> Therefore, minor discrepancies in expression patterns across cancer types are expected between these two platforms. Furthermore, to generate violin plots depicting CDK2 expression patterns across pathological stages (I–IV) in various tumor types, the "Expression Analysis–Stage Plot" module was utilized, providing valuable insights into the role of CDK2 in cancer progression.

Again, the UALCAN (<https://ualcan.path.uab.edu/analysis-prot.html>) database,<sup>40</sup> based on the Clinical Proteomic Tumor Analysis Consortium (CPTAC), was used to assess the CDK2 protein expression across multiple cancers.

### 2.2 Survival and prognostic analysis

To conduct Kaplan–Meier survival analysis of the overall survival (OS) and disease-free survival (DFS) associated with CDK2 expression across all TCGA tumors,<sup>41</sup> the GEPIA2 database was used. High (red) and low (blue) expression cohorts were defined based on the 50% median cutoff value. To analyze the data statistically, we used the log-rank tests, and survival plots were generated using GEPIA2's 'Survival Analysis' module. The survival analysis is displayed with axis units in months in 'X' and the percentage in 'Y', with 95% confidence interval indicated by a dotted line.

### 2.3 Genetic alteration profiling

For Cancer Genomics, cBioPortal (<https://cbioportal.org>) is an interactive web platform designed for the exploration,



visualization, and analysis of multidimensional cancer genomics data.<sup>42</sup> In this study, the genetic alteration profiles of CDK2 were examined using the public database, cBioPortal. The TCGA Pan Cancer Atlas studies were conducted using the 'query by Gene' module in the quick search subsection, which includes 32 tumor types with a total of 10 957 patients' data comparison. The survival variant data, copy number alteration, and mutation types of CDK2 across all TCGA tumors were illustrated as 'Cancer Type Summary'. The "Mutations" module was utilized to generate a schematic diagram depicting the mutated site of CDK2.

#### 2.4 Immune cell infiltration analysis

The associations between CDK2 expression and tumor-infiltrating immune cells were analyzed *via* the 'Immune-Gene' module of TIMER 2.0. To calculate the immune cell infiltration specifically, the T follicular helper cells (TFH) and cancer-associated fibroblasts (CAFs) were calculated using six deconvolution algorithms, including MCP COUNTER, CIBERSORT, TIDE, XCELL, EPIC, and CIBERSORT-ABS. Spearman's rank correlation analysis, adjusted for tumor purity, was conducted to estimate both the *p*-values and partial correlation coefficients, and the results are displayed as scatter plots and heatmaps.

#### 2.5 Co-expression and functional enrichment analysis

We employed the STRING database<sup>43</sup> to systematically investigate the protein–protein interaction network of human CDK2. The analysis was performed by querying "CDK2" as the target protein with the organism specified as *Homo sapiens*. To ensure biologically relevant interactions, we configured the database parameters to display only experimentally validated interactions with an evidence-based network edge setting. The minimum required interaction score was maintained at 0.150 (low confidence threshold) to capture potential interactions while maintaining reliability. The analysis was limited to no more than 50 interactors in the first shell to focus on the most immediate interaction partners. The resulting interaction data were then imported into Cytoscape<sup>44</sup> for advanced network visualization and further topological analysis, enabling comprehensive examination of CDK2's protein interaction landscape. Venn diagram analysis was conducted using Jvnn to identify intersecting genes from STRING and GEPiA2. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed using the DAVID database.<sup>45</sup>

#### 2.6 Natural-plant-derived phytocompound library preparation

To identify potential CDK2 inhibitors, we sourced the data from the IMPPAT 3.0 database.<sup>26</sup> The traditionally used medicinal plants were selected as a source of phytocompounds. Data on 700 available phytocompounds were downloaded from the database for the different parts (roots, leaves and seeds) of these medicinal plants to create a compound library.

#### 2.7 Machine learning (ML) application to predict bioactivity ( $\text{pIC}_{50}$ )

The  $\text{IC}_{50}$  value is a critical measure of a drug's effectiveness, indicating the concentration required to inhibit a biological target by 50%.<sup>46</sup> Accurately predicting  $\text{IC}_{50}$  values allows researchers to assess a compound's potential effectiveness early in the drug-discovery process. Numerous studies conducted over the years have refined several methods, thus enabling the prioritization of the most promising drug candidates.<sup>47,48</sup> Thus, this study utilized machine learning methods to predict the bioactivities of filtered compounds against CDK2.

**2.7.1 Dataset preparation and curation.** To build a reliable dataset for analysis, we used bioactivity data and chemical structures from the ChEMBL database, a well-known resource for quantitative structure–activity relationship (QSAR) research. Specifically, we selected the CHEMBL301 dataset, which includes data on CDK2 inhibitors. Our objective was to create a regression model focused on CDK2 inhibitors; thus, we carefully curated the dataset by including only compounds with available  $\text{IC}_{50}$  values. The final dataset consisted of 9307 compounds that demonstrated activity against the target protein in laboratory settings. This meticulous filtering process resulted in a refined dataset of active inhibitors that laid the groundwork for further analyses.

**2.7.2 Generation of molecular descriptors.** Molecular descriptors are crucial in QSAR modeling as they capture the structural and physicochemical characteristics of compounds, enabling the identification of patterns. To gather this information, we employed PaDEL software to generate PubChem fingerprints for each compound, producing 881 molecular descriptors per compound. These descriptors provided an in-depth representation of the compounds' structural properties, which were vital for constructing predictive models aimed at estimating bioactivity.

**2.7.3 Data partitioning.** To enhance the robustness and generalizability of our model, we split the dataset into two subsets: a training set and a testing set. The training set, comprising 80% of the data, was used to train the model, while the remaining 20% was set aside for testing its performance. This partitioning approach minimizes overfitting and ensures that the model can make accurate predictions on new, unseen data.

**2.7.4 Model selection, feature optimization, and evaluation.** For predictive modeling, we chose LightGBM Regression, a highly efficient machine learning algorithm known for its accuracy, speed, and interpretability. To optimize the model's performance, we employed Recursive Feature Elimination (RFE), a technique that eliminates less important features, thereby reducing noise and improving interpretability. Additionally, we performed hyperparameter tuning using Grid Search in combination with 10-fold cross-validation. The key hyperparameters tuned included "num\_leaves," "learning\_rate," "n\_estimators," "max\_depth," and "feature\_fraction".<sup>49</sup>

We evaluated the model's predictive performance using several common metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and *R*-squared ( $R^2$ ). These metrics were computed using the following formulas:



Mean Absolute Error (MAE):

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Mean Squared Error (MSE):

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Root Mean Squared Error (RMSE):

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

R-squared ( $R^2$ ):

$$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2}$$

Here,  $y_i$  represents the observed  $IC_{50}$  values,  $\hat{y}_i$  denotes the predicted values, and  $\bar{y}$  is the mean of the observed values.

#### 2.7.5 Application of the model for bioactivity prediction.

After training and optimizing the LightGBM model, it was employed to predict the  $IC_{50}$  and  $pIC_{50}$  values of the filtered compounds targeting CDK2. The compounds with the highest predicted  $pIC_{50}$  values, indicating greater potency, were identified as the most promising candidates for further exploration. These compounds were then subjected to molecular docking simulations to assess their binding affinities and interactions with the CDK2 protein target.

#### 2.8 Target protein retrieval and preparation

The crystallographic structure of the target protein was retrieved from the RCSB protein database using the accession ID: 6GUE.<sup>50</sup> Then, the protein structure was prepared by removing existing heteroatoms, ligands, and water molecules from the crystallographic structure using BIOVIA Discovery Studio 2021.<sup>51</sup> Finally, the energy was minimized using Swiss-PDB Viewer (spdbv) version 4.1.0.<sup>52</sup> Fig. 2 depicts the 3D structure and active site of the target protein (6GUE).

#### 2.9 Molecular docking study

In CADD, molecular docking is a significant technique for lead optimization and identification, used to predict the potential interactions between two molecules, such as a protein and a ligand.<sup>53</sup> In this study, molecular docking was performed using the AutoDock Vina tools.<sup>54</sup> The binding site/pocket (Fig. 2) for the PDB structure was identified using the BIOVIA Discovery Studio 2021<sup>51</sup> to create a receptor grid box with dimensions of  $X = -7.1531$ ,  $Y = -23.0788$ , and  $Z = 22.6518$  to calculate a phytocompound's binding affinity. Finally, we observed the ligand interaction with the target receptor through PyMOL and BIOVIA Discovery Studio 2021.<sup>51</sup>

PDB ID: 6GUE

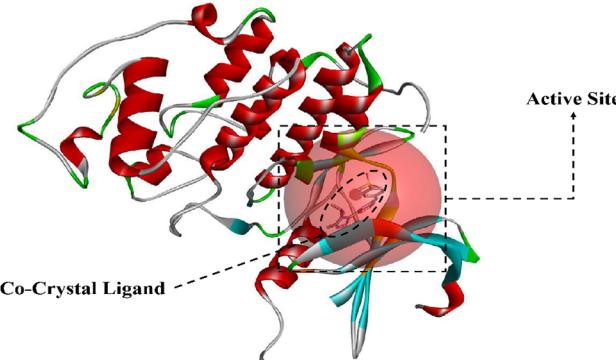


Fig. 2 The 3D-crystal structure of the 6GUE protein bound to the experimentally available ligand (FB8), with the active site highlighted (red spherical region).

#### 2.10 Molecular dynamics simulation

Molecular dynamics is the study of the dynamic behavior of protein-ligand complexes in a simulated physiological environment.<sup>55</sup> Molecular dynamics simulations of the protein-ligand complex were performed with Desmond v2020-4 (Schrödinger, Academic) on Linux using the OPLS3e force field and the SPC (simple point charge) water model in an orthorhombic periodic box, providing a solvent buffer of 10 Å minimum on all sides. The systems were first neutralized with counterions to remove any net charge, and then  $Na^+$  and  $Cl^-$  were added to achieve a 0.15 M ionic strength to emulate physiological conditions. Following system building, energy minimization was conducted to remove steric clashes, after which a restrained NVT heating phase at 300 K was applied, followed by NPT equilibration at 300 K and 1 atm to relax density and pressure. Long-range electrostatics were treated with PME, the bonds to hydrogens were constrained, and a 2-fs time step was used. Production simulations were run in the NPT ensemble at 300 K and 1 atm for 100 ns, saving 5000 frames per trajectory. Unless otherwise noted, three independent replicates with distinct initial velocity seeds were performed, and the results are reported as mean  $\pm$  SEM across replicates, with any single-trajectory case explicitly indicated. The choice of OPLS3e reflects its extended coverage and improved torsional and electrostatic parameters for drug-like ligands, alongside robust biomolecular parameters, making it well-suited for protein-ligand simulations in this workflow.

The stability and flexibility of the P-L system were then studied from the plots of the RMSD, RMSF, SASA, RoG, H-Bonds, DCCM, and PCA calculation from the trajectories analysis.

**2.10.1 Root mean square deviation (RMSD) analysis.** The PL complex stability can be predicted using RMSD calculations. The RMSD results were plotted using the following statistical equations:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^n (r_i(t_x) - r'_i(t_{ref}))^2}$$

where  $N$  = the selected atom number,  $t_{ref}$  = the reference time,  $t_x$  = the length of the recording intervals, and  $r$  = the location of



the particular atom in frame  $x$  after alignment with the reference frame.

#### 2.10.2 Root mean square fluctuation (RMSF) analysis.

RMSF measures the local conformational change within the residues of the protein structure. The RMSF value of a protein can be calculated using the given equation.

$$\text{RMSF} = \sqrt{\frac{1}{T} \sum_{t=1}^n (r_i(t) - r'_i(t_{\text{ref}}))^2}$$

where  $T$  = the calculated time from each trajectory.

#### 2.10.3 Free energy landscape (FEL) & principal component analysis (PCA).

Principal Component Analysis (PCA) was performed to reduce the dimensionality of the molecular dynamics (MD) data and identify dominant motions by extracting principal components (PCs), which represent the directions of maximum atomic displacement. The simulation trajectories, originally in the GROMACS format (.xtc), were analyzed using the R environment<sup>56</sup> and the Bio3D package (version 2.3.0)<sup>57</sup> to compute the principal components.

To further explore the conformational landscape, Free Energy Landscapes (FEL) were constructed using the structural parameters: Root Mean Square Deviation (RMSD) and Radius of Gyration ( $R_g$ ). The Geo-measure<sup>58</sup> plugin in PyMOL, operated within the Conda environment on a Linux OS, was utilized to generate 3D FEL plots. This approach facilitates the prediction of joint probability distributions in three-dimensional space, providing insights into the stability and structural transitions of the system.<sup>53</sup>

### 2.11 MM-GBSA calculation

Finally, the protein–ligand (PL) binding free energy for each snapshot was calculated using the gmx\_MMPBSA tool with the help of Gromacs trajectories, which were converted from the Desmond trajectories. To explore the biophysical underpinnings of PL interactions, the Molecular Mechanics–Generalized Born Surface Area (MM-GBSA) approach was employed for the top-ranked three ligands and a standard compound. MM-GBSA calculations were utilized to estimate the binding affinities of the PL complexes.

The following equation was utilized for the binding free energy ( $\Delta G_{\text{bind}}$ ) calculation:<sup>59</sup>

$$\Delta G_{\text{bind}} = \langle G_{\text{PL}} \rangle - \langle G_{\text{P}} \rangle - \langle G_{\text{L}} \rangle$$

where  $\Delta G_{\text{bind}}$  represents the binding free energy,  $\langle G_{\text{PL}} \rangle$  denotes the average free energy of the PL complex,  $\langle G_{\text{P}} \rangle$  denotes the free energy of the free protein, and  $\langle G_{\text{L}} \rangle$  denotes the free energy of the free ligand.

$\Delta G_{\text{bind}}$  can be expressed as follows:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{SOLV}} - T\Delta S$$

where  $\Delta E_{\text{MM}}$  represents the gas-phase molecular mechanics energy (including electrostatic interactions and van der Waals),  $\Delta G_{\text{SOLV}}$  is the solvation free energy change, and  $T\Delta S$  denotes the entropic contribution. The polar solvation energy was calculated using a linearized Poisson–Boltzmann model, while the

nonpolar contribution was estimated using a SASA-based approach.<sup>60</sup>

### 2.12 Pharmacokinetics and drug-likeness analysis

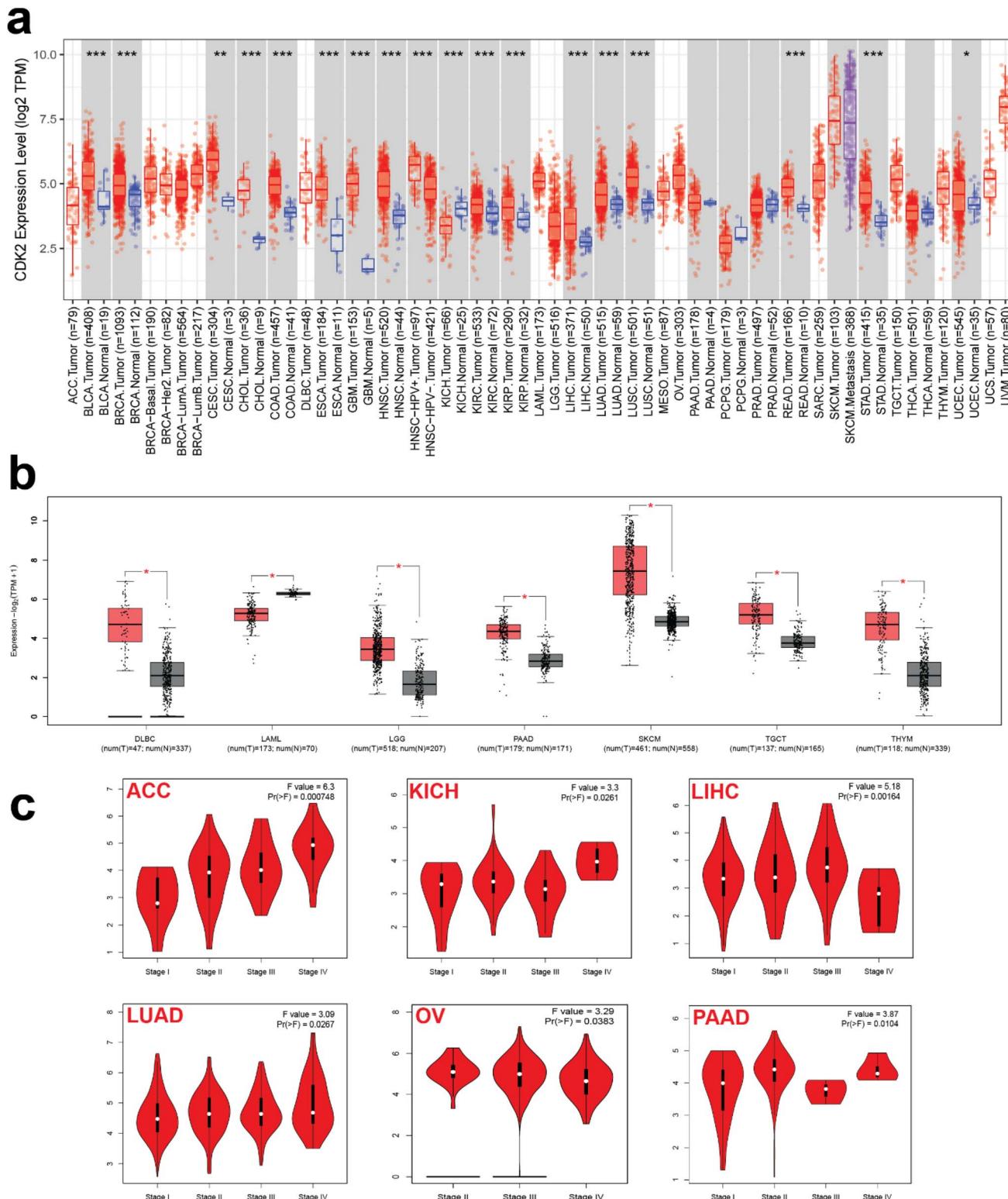
In drug discovery, pharmacokinetics (PK) plays a crucial role not only in filtering ligands from a large compound library but also in ensuring the safety of those compounds.<sup>61</sup> Primarily, Lipinski's Rule of Five (RO5)<sup>62</sup> is used to identify estimated lead compounds by excluding redundant phytochemicals. Lipinski's Rule of Five states that a compound is more likely to be orally active if it has no more than 5 HBD (hydrogen-bond donors), 10 HBA (hydrogen-bond acceptors), a molecular weight (MW) under 500 Da, and a  $\log P$  less than 5. Therefore, the phytochemicals that accept RO5 are considered for further testing. The drug-like characteristics of the small molecules were assessed using the SwissADME web tool (<https://www.swissadme.ch/>), which evaluates key drug-like features, such as adherence to Lipinski's rule of five, water solubility ( $\log S$ ), gastrointestinal (GI) absorption, and blood–brain barrier (BBB) permeability—critical factors in rational drug design.<sup>63,64</sup> In addition, pkCSM,<sup>65</sup> another online predictive tool, was used to evaluate important graph-based PK parameters, including human intestinal absorption, volume of distribution, and total clearance (excretion).

**2.12.1 Toxicity analysis.** Finally, the toxicity of small molecules is an important indicator in drug development, as they may have harmful effects on human organs, including hepatotoxicity, immunotoxicity, mutagenicity, carcinogenicity, and cytotoxicity. To predict these toxicological risks, the ProTox-III platform (<https://tox.charite.de/>) was employed to analyze the safety profiles of the selected compounds.<sup>66</sup> ProTox-III, pkCSM, and SwissADME utilize QSAR-based models developed from existing experimental datasets. Therefore, the accuracy and reliability of their predictions can be constrained by adjusting the quality and chemical diversity of the training data.

### 2.13 Pharmacodynamic property analysis

**2.13.1 SMILE-based 2D-QSAR prediction.** The ChemDes (<https://www.scbdd.com/chemdes/>) website provides the necessary datasets to calculate a variety of molecular descriptors by accessing the ChEMBL database, such as chiv5, bcutm1, MRVSA9, MRVSA6, PEOEVSAs, GATSv4, J, and diametert.<sup>67,68</sup> We created an Excel sheet with a standard equation known as the multiple linear regression (MLR) to calculate the  $\text{pIC}_{50}$  for the selected compounds. The following equation was taken from an earlier study<sup>2</sup> for the  $\text{pIC}_{50}$  calculation:

$$\begin{aligned} \text{pIC}_{50} (\text{activity}) = & -2.768483965 + 0.133928895 \times (\text{Chiv5}) \\ & + 1.59986423 \times (\text{bcutm1}) \\ & + (-0.02309681) \times (\text{MRVSA9}) \\ & + (-0.002946101) \times (\text{MRVSA6}) \\ & + (0.00671218) \times (\text{PEOEVSAs}) \\ & + (-0.15963415) \times (\text{GATSv4}) \\ & + (0.207949857) \times (J) \\ & + (0.082568569) \times (\text{diameter}) \end{aligned}$$



**Fig. 3** Comparative expression profiling of the CDK2 gene in various cancers and their pathological phases. (a) Boxplot showing the CDK2 mRNA expression (log<sub>2</sub> TPM) across various tumor types compared to their matched normal tissues using TIMER2.0. Red boxes indicate tumor tissues and blue boxes indicate normal tissues. Statistically significant differences are marked as \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$ . (b) GEPIA2-based validation highlights the significant CDK2 overexpression levels in DLBC, LAML, LGG, PAAD, SKCM, TGCT, and THYM, where TIMER lacks normal tissue data. For SKCM, the tumor tissue data were compared with the metastatic sample data, as normal tissue data were unavailable. (c) Pathological stage analysis using GEPIA2 shows a significant association between CDK2 expression and tumor stage in ACC, KICH, LIHC, LUAD, OV, and PAAD (\* $p < 0.05$ ).

### 3. Results

#### 3.1 Transcriptomics and proteomics expression analysis

The expression profiling of the cyclin-dependent kinase-2 or CDK2 was analyzed across different types of cancers using the TIMER2.0 database; the gene expression profiles are illustrated in Fig. 3a. CDK2 is significantly overexpressed in several tumor types compared to the expression levels in the respective normal tissues; the most significant overexpression in different tumor types is shown in Fig. 3a with three stars (\*\*\*)). In cervical squamous cell carcinoma (CESC), the CDK2 expression was significantly higher in tumors than in normal tissues ( $\log_{2}FC \approx 1.246$ ,  $adjP < 0.001$ ), indicating a  $\sim 2.4$ -fold increase, whereas UCEC exhibited a smaller but statistically significant elevation ( $P < 0.05$ ). The exact  $p$ -values for all cancer types are provided in Table S9.

Among the cancer types, we found seven that were not covered by TIMER2.0 for normal comparison; seven tumors showed statistically significant differences in CDK2 expression between tumor and normal tissues (Fig. 3b). To ensure transparency, the boxplots of the CDK2 expression across all tumor types generated by GEPIA2 are provided in SI file 2 [Fig. S4(A–

C)]. We further explored the relationship between the CDK2 expression and tumor progression using the module named “Pathological Stage Plot” within the GEPIA2 web tools. The analysis revealed a significant correlation between the CDK2 expression levels and the tumor stage ( $P < 0.05$ ) in several cancer types. These include ACC, KICH, LIHC, LUAD, OV, and PAAD (Fig. 3c). While TIMER2.0 and GEPIA2 generally showed consistent overexpression of CDK2 across multiple tumor types, minor differences were observed in certain cancers. This variation likely arises from the differences in the underlying data sets (TCGA-only in TIMER vs. TCGA + GTEx in GEPIA2).

Additionally, to explore the CDK2 expression at the protein level, we analyzed its proteomic profiles using the UALCAN database. The results revealed that the CDK2 protein expression was significantly higher in tumor tissues compared to their normal-tissue counterparts across eight cancer types. These included Clear cell renal cell carcinoma (ccRCC), Clear cell renal cell carcinoma (extended) (ccRCC-ext), Colon cancer (COAD), Glioblastoma multiforme (GBM), Head and neck squamous carcinoma (HNSC), Liver hepatocellular carcinoma (LIHC), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), and Pancreatic adenocarcinoma (PAAD) (Fig. 4).

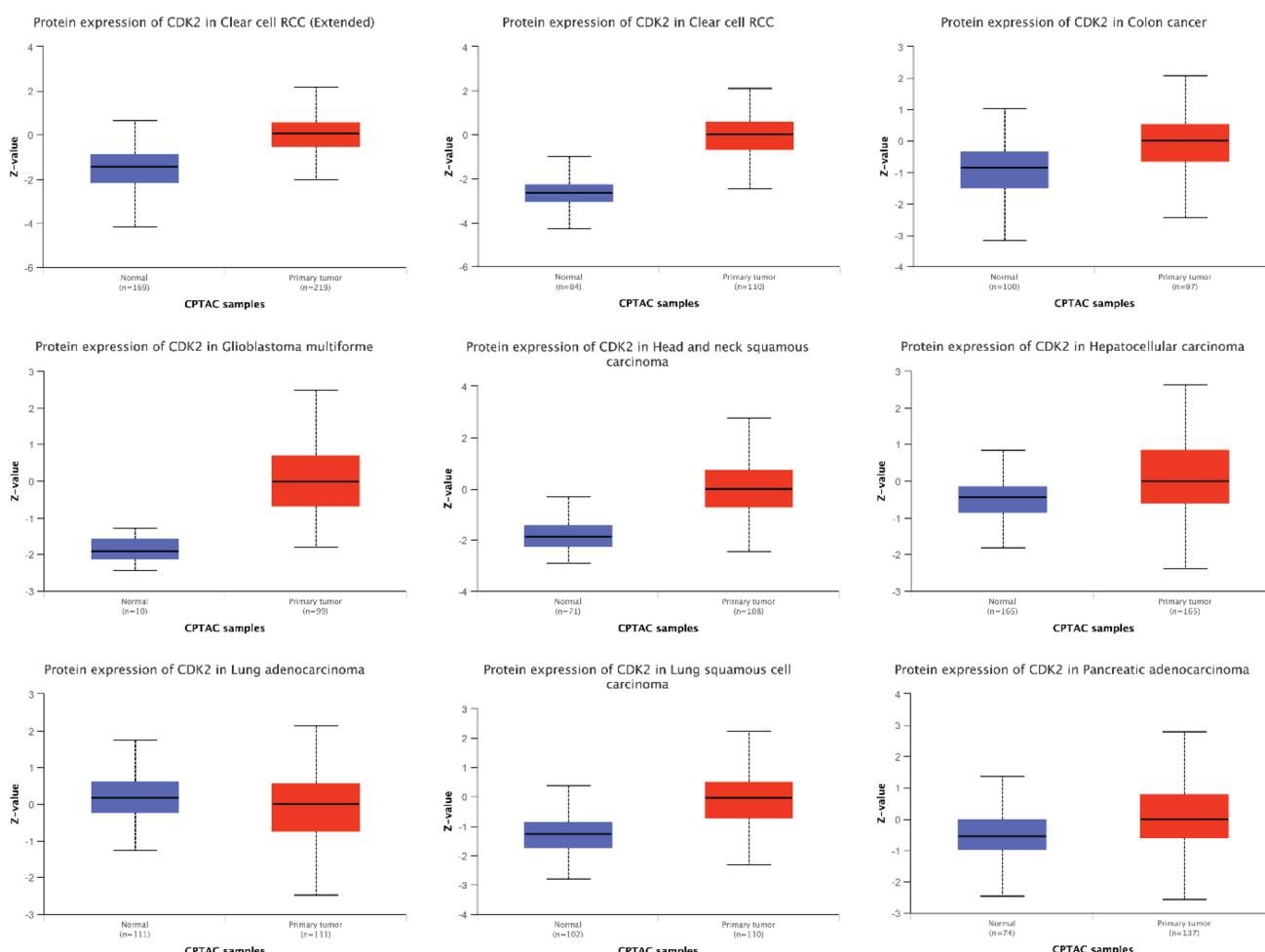


Fig. 4 CDK2 protein expression in normal (blue) and tumor (red) tissues across multiple cancer types. Using the UALCAN database, the CDK2 protein levels were found to be significantly higher in tumor tissues compared to their normal tissue counterparts in ccRCC, ccRCC-ext, COAD, GBM, HNSC, LIHC, LUAD, LUSC, and PAAD.



### 3.2 Prognostic and survival study

To explore the prognostic value of the CDK2 expression, we classified patients into high and low expression groups and assessed their disease-free survival (DFS) and overall survival (OS). As shown in Fig. 5, a high CDK2 expression correlated with poor survival in several cancers (ACC, KIRP, LGG, LIHC, MESO, PAAD, SKCM), while UVM and READ showed the opposite trend, suggesting tissue-specific prognostic roles. Interestingly, Li *et al.* (2001) reported that reduced CDK2 expression was correlated with poor prognosis in colorectal carcinoma, which stands in contrast to our general finding of high CDK2 = poor survival, further supporting tissue-specific prognostic roles.<sup>69</sup> For disease-free survival (DFS), a similar trend was observed in cancers, such as ACC, LGG, LIHC, and PAAD, where patients with a high CDK2 expression had shorter relapse-free periods. These findings suggest that elevated CDK2 levels may serve as a marker of poor prognosis in multiple tumor types.

### 3.3 Gene alteration analysis

The gene alteration summary (Fig. 6a) demonstrated that CDK2 alterations across TCGA cohorts were dominated by mutations and copy number amplifications, whereas deep deletions and multiple alterations were comparatively rare. In Endometrial

Cancer, CDK2 showed the highest alteration frequency (slightly above 4%), with both “mutations” and “amplifications” observed, reflecting its diverse genomic changes in this cancer type. In contrast, in CHOL, all detected alterations were “mutations,” with a total alteration frequency of almost 3%, suggesting a more mutation-specific alteration pattern for CDK2 in this cancer. Other cancer types, such as Ovarian Epithelial Tumor, Esophagogastric Cancer, and Sarcoma, also exhibited moderate alteration frequencies ranging between 2% and 3%. Additionally, cancers such as adrenocortical carcinoma, leukemia, mature B-cell neoplasm, hepatobiliary cancer, non-small cell lung cancer, and glioma displayed predominant amplification events. These findings imply that CDK2 copy number gains may play a key role in tumor progression within these cancer types (Fig. 6a).

The various types, positions, and frequencies of genetic alterations in CDK2 are illustrated in Fig. 6b. CDK2 mutations display clear tissue-specific patterns across different cancer types. Missense mutations, such as R265L, D68N, and R247H, are predominantly found in endometrial (UCEC) and ovarian cancers, suggesting the potential role of CDK2 in female reproductive tumors. In contrast, truncating mutations, like E51 and W291, are mainly observed in lung squamous cell carcinoma (LUSC) and esophageal adenocarcinoma, indicating

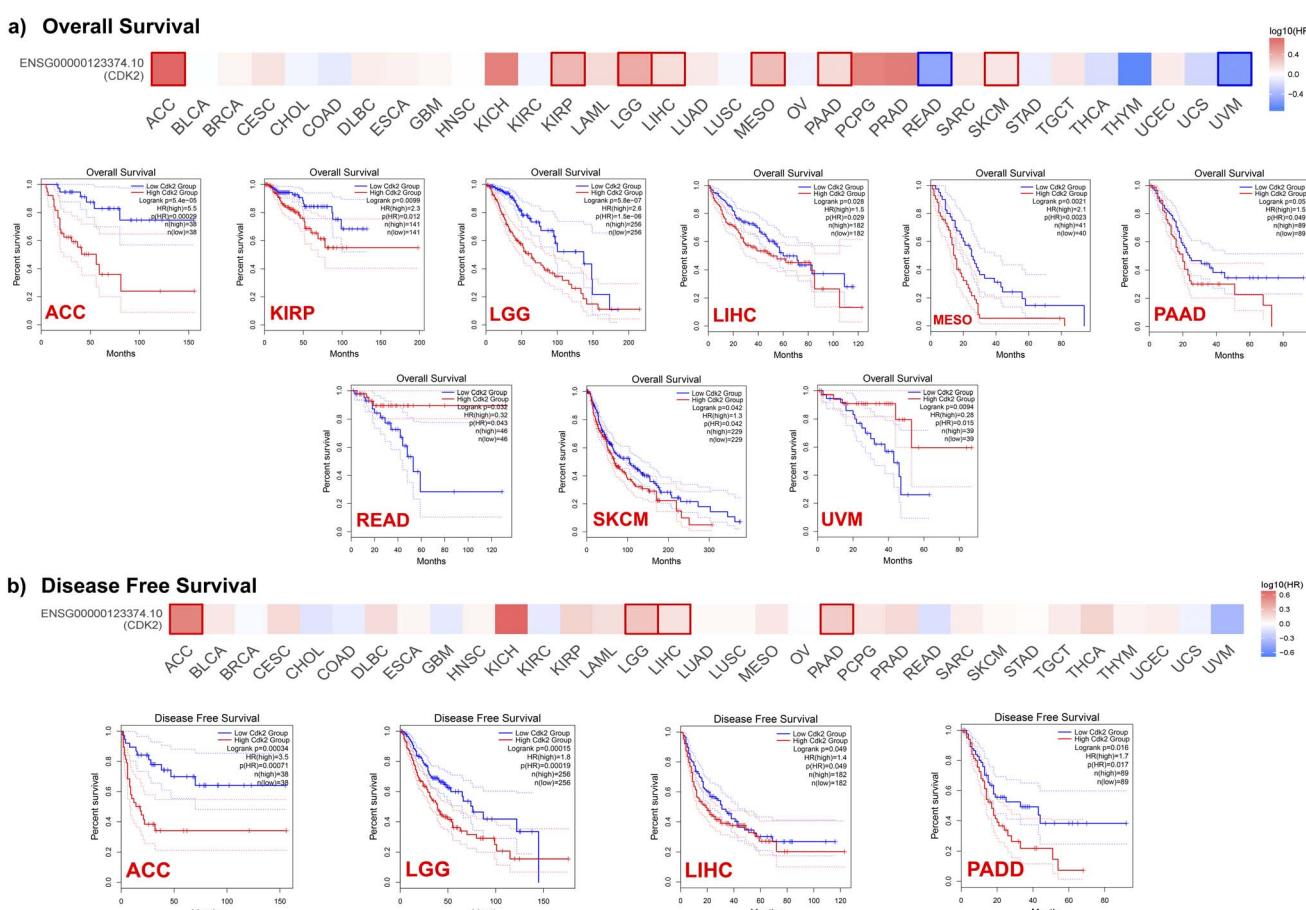


Fig. 5 CDK2 expression and patient survival across cancer types. (a) Overall survival (OS) and (b) disease-free survival (DFS) analyses across TCGA cohorts comparing high and low CDK2 expression groups.



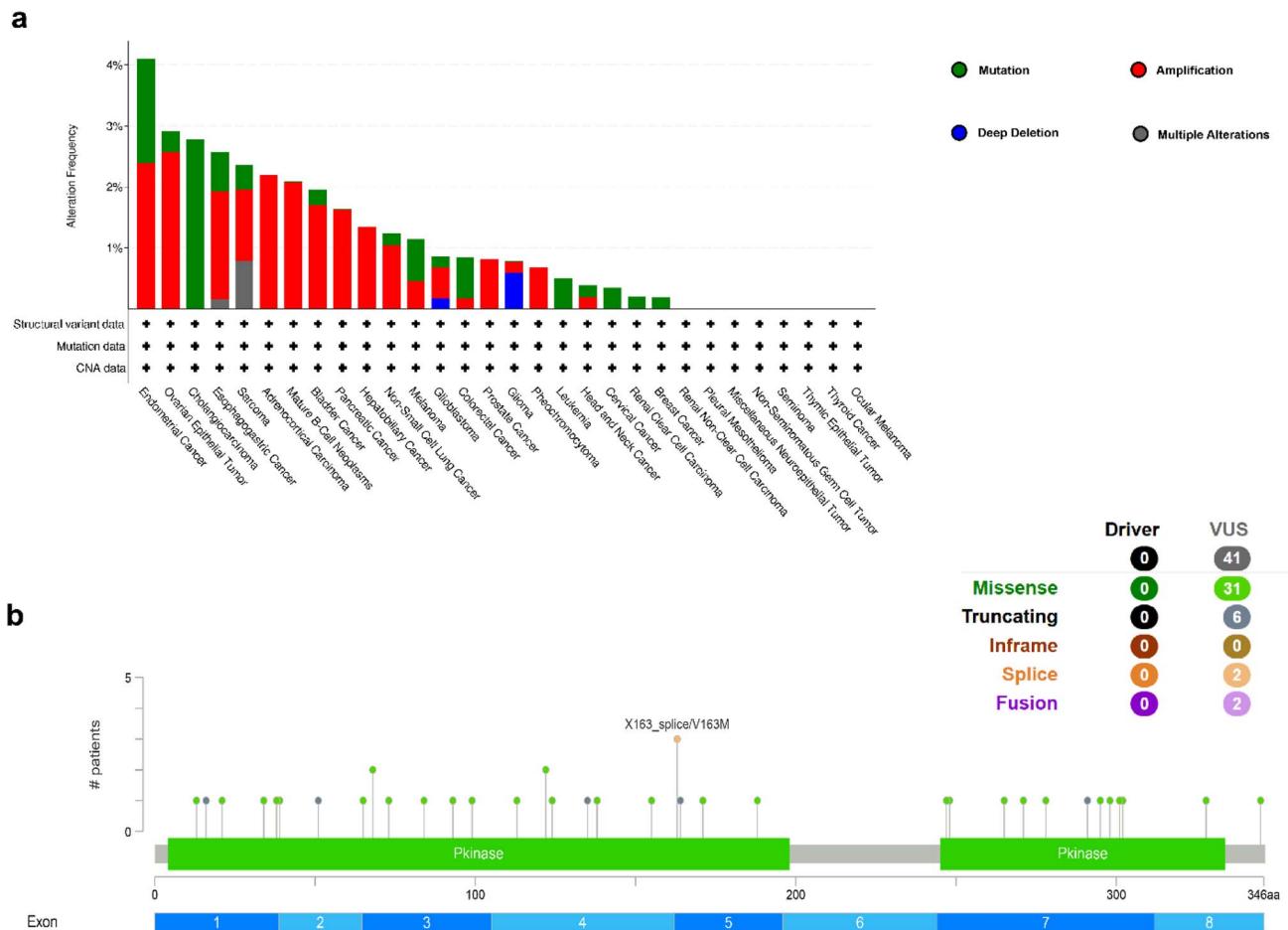


Fig. 6 Genetic alteration analysis of CDK2 using cBioPortal. (a) Frequency and types of CDK2 alterations across TCGA cancer types. (b) Genomic locations and classification of CDK2 mutations.

a different mode of gene disruption. Notably, rare splice-site alterations (X163\_splice) are detected in aggressive cancers, like glioblastoma (GBM), while CDK2 gene fusions (CDK2-ERBB3 and CDK2-PAN2) appear to be unique to sarcomas. Functional predictions highlight mutations like G13D in UCEC and P155H in serous endometrial cancer as having potentially high impact, possibly contributing to tumor development. Overall, this diverse mutation landscape underlines CDK2's dual role as both a broadly involved (pan-cancer) and tissue-specific cancer gene. Detailed mutation data are provided in Table S1.

#### 3.4 Immune infiltration data analysis

We investigated the association between CDK2 gene expression and immune cell infiltration across multiple cancer types using data from TCGA. As demonstrated in Fig. 7a, a positive correlation between CDK2 expression and T follicular helper cell (Tfh) infiltration was observed in LUSC, COAD, HNSC, LIHC, OV, STAD and THYM, whereas only one negative correlation was observed in UVM. In addition, we investigated the relationship between CDK2 expression and cancer-associated fibroblast (CAF) infiltration, as illustrated in Fig. 7b. Our

analysis revealed a positive correlation exclusively in LGG, with no significant negative correlations observed in any cancer type.

#### 3.5 Enrichment analysis of CDK2-related genes

To investigate the role of CDK2 in tumorigenesis, we first identified its protein–protein interaction network. Using STRING, we identified 50 proteins that interact with CDK2 (Fig. 8A, Cytoscape visualization). Next, we analyzed TCGA tumor expression data *via* GEPPIA2 to pinpoint the top 100 genes co-expressed with CDK2 (Table S2). Comparative analysis of these two datasets revealed no shared elements (Fig. 8B). We performed a combined GO and KEGG pathway enrichment analysis integrating both datasets (Table S3). The enrichment analysis of CDK2 and its associated genes demonstrated its significant involvement ( $P \leq 0.05$ ) in multiple cancer-related and regulatory pathways. KEGG pathway analysis revealed its enrichment not only in the canonical cell cycle pathway but also in cellular senescence, p53 signaling, and oncogenic signaling pathways: prostate cancer, small cell lung cancer, viral carcinogenesis, and infection-driven pathways (including HPV, HTLV-1, and Epstein–Barr virus). Biological process enrichment indicated CDK2's central role in G1/S transition, cell division,

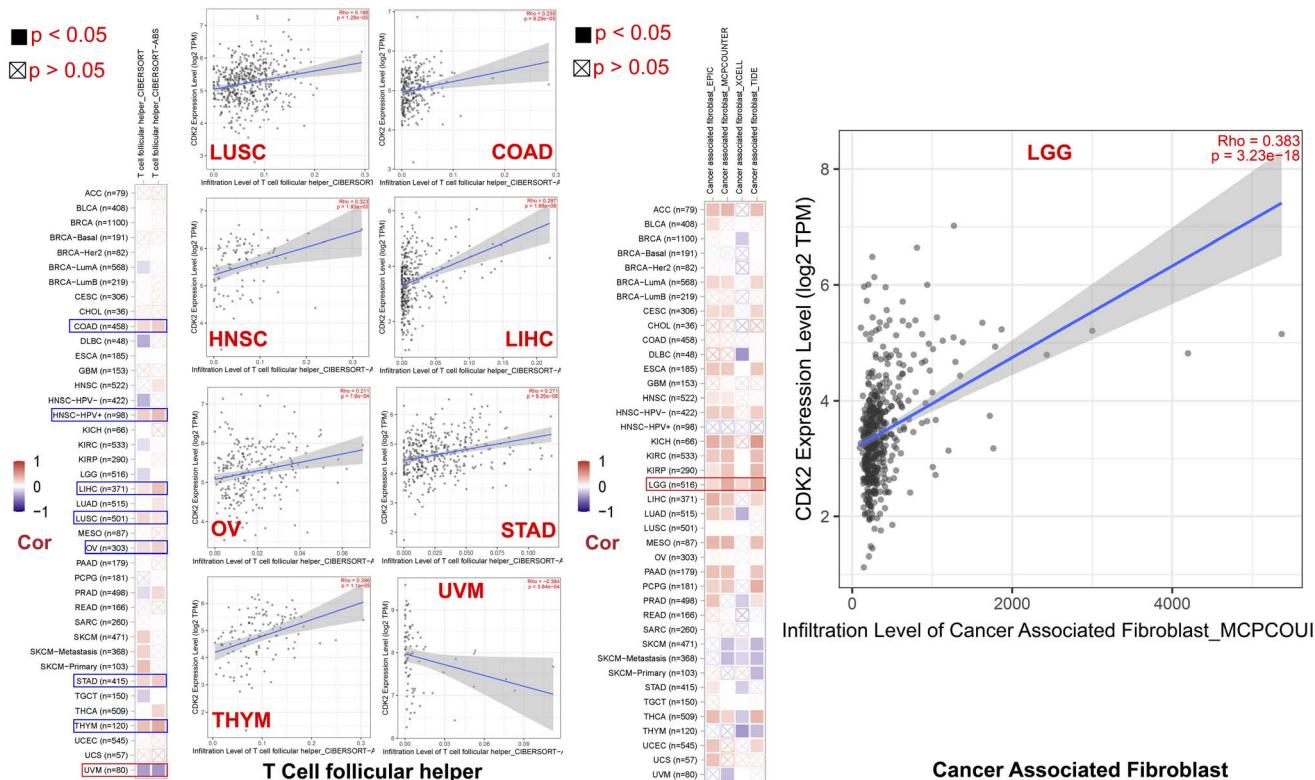


Fig. 7 Correlation of CDK2 expression with immune cell infiltration across cancers. (a) Positive correlations between CDK2 expression and T follicular helper cells in several cancers; only UVM showed a negative correlation. (b) Positive correlation between CDK2 and cancer-associated fibroblast infiltration observed only in LGG, with no negative correlations.

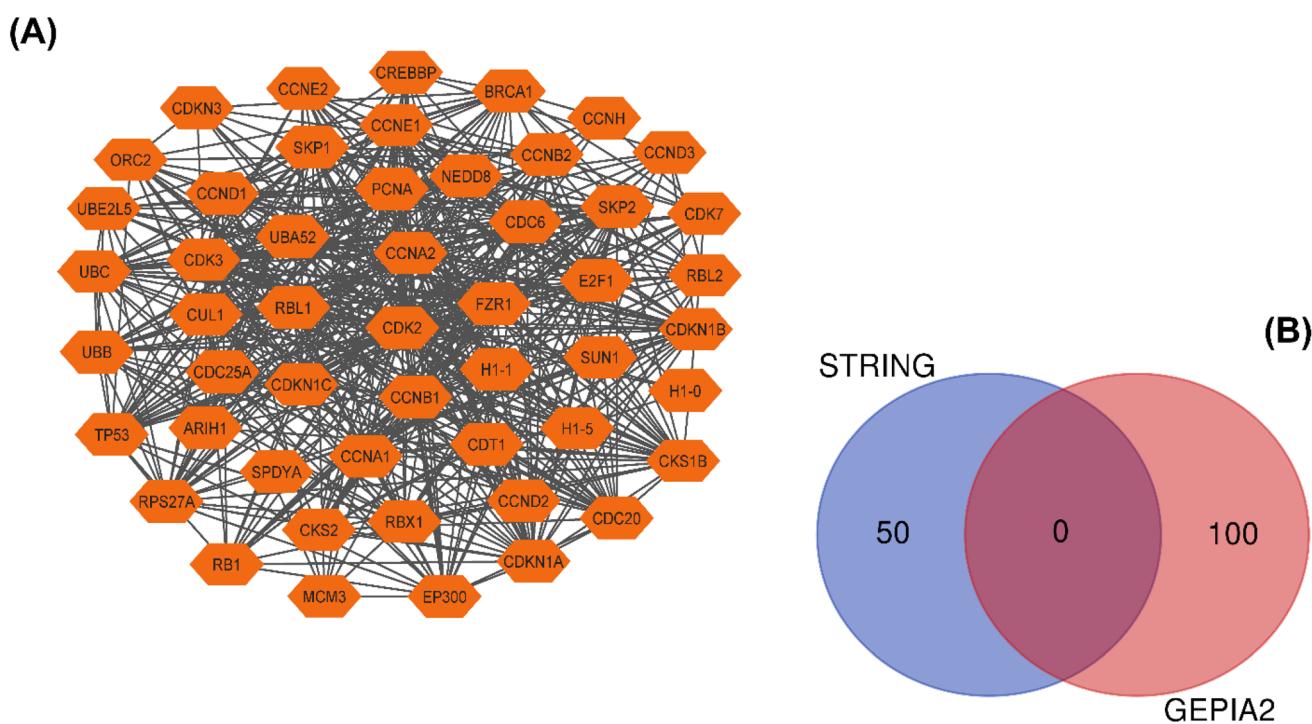


Fig. 8 (A) CDK2 protein–protein interaction network using the STRING database. (B) Venn diagram of the CDK2 interactors and co-expressed genes.



and regulation of cyclin-dependent kinases, underscoring its importance in cell cycle progression. Cellular component analysis localized CDK2 within the nucleus, nucleoplasm, cytosol, and as part of the cyclin-dependent protein kinase holoenzyme complex, while molecular function analysis highlighted its activity as a cyclin-dependent serine/threonine kinase with ubiquitin protein ligase binding capacity. Collectively, these findings reaffirm CDK2 as a critical regulator of cell cycle progression while simultaneously emphasizing its broader impact on oncogenesis and virus-associated signaling pathways.

### 3.6 Filtering of duplicate compounds and ligand-library preparation

First, we selected traditionally used natural medicinal plants, including *Aloe vera*, *Artemisia indica*, *Asparagus racemosus*, *Petiveria alliacea*, *Anacardium occidentale*, and *Withania somnifera*,<sup>70–75</sup> as the source of natural inhibitors. Then, we found and extracted approximately 764 phytocompounds from the IMPPAT database for the different parts of these plants. After that, we reduced the data volume to 467 (Table S4) on the basis of duplicate entries and structural unavailability in database records for further machine learning-based screening.

### 3.7 Screening through a machine learning (ML) approach

**3.7.1 Bioactivity prediction ( $\text{pIC}_{50}$ ) using ML.** Among the diverse machine learning techniques, LightGBM regression has emerged as a standout ensemble learning method widely recognized for its high accuracy, efficiency, and ability to handle large-scale datasets. Its popularity stems from its ability to construct robust predictive models by utilizing gradient-based decision trees, incorporating histogram-based optimization, and employing leaf-wise tree growth, which enhances both speed and model performance,<sup>76</sup> and it has previously been successfully applied in a study in the field of drug discovery.<sup>77</sup> In this study, we leveraged LightGBM regression to predict the bioactivity ( $\text{pIC}_{50}$ ) of compounds targeting a specific biological pathway.

**3.7.2 Model training and hyperparameter optimization.** The model training process involved the curated dataset, which included 9307 compounds with known  $\text{IC}_{50}$  values. During training, several hyperparameters were carefully fine-tuned to optimize the model's performance. These key hyperparameters included `n_estimators`, `max_depth`, `learning_rate`, `subsample`, `reg_alpha`, `reg_lambda`, `gamma`, and `colsample_bytree`. Each of these parameters was adjusted systematically to improve the model's predictive accuracy and generalization capability. This fine-tuning process aimed to strike a balance between the model's complexity and precision, ensuring reliable predictions while maintaining simplicity and preventing overfitting.

**3.7.3 Model-performance evaluation.** To evaluate the trained model's effectiveness, we used several well-established metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and  $R^2$ -squared ( $R^2$ ). These metrics offered a comprehensive assessment of how well the model's predictions aligned with the actual experimental

data. Specifically, the MAE measures the average absolute difference between predicted and observed values, while the MSE calculates the average squared differences, emphasizing larger deviations. The RMSE, derived from the MSE, presents prediction errors in the original units of the dependent variable, providing a more intuitive understanding of the model's accuracy. Lastly, the  $R^2$  metric evaluates the proportion of variance in the target variable explained by the model, with higher values indicating a better fit.

The model demonstrated impressive performance, with MAE, MSE, RMSE, and  $R^2$  values of 0.372, 0.385, 0.529, and 0.854, respectively, as shown in Fig. S1. These results suggest that the model is highly accurate, as reflected by its low error rates and substantial explanatory power.

**3.7.4 Application to compound screening.** After validating the model, it was applied to a dataset of 467 filtered compounds to predict their  $\text{pIC}_{50}$  values. Among these, 322 compounds had predicted  $\text{pIC}_{50}$  values exceeding 5.1, indicating good potency against the biological target, as presented in Table S5. These results underline the utility of the model in identifying promising candidates for further studies, such as molecular docking simulations or experimental validation, to confirm their potential as drug candidates.

### 3.8 Molecular docking and Re-docking analysis

A comprehensive molecular docking study was conducted to assess the binding affinity and interaction mechanisms between the target protein and selected phytocompounds obtained from the above-mentioned medicinal plants. In this study, 3 out of 322 compounds, *i.e.*, withanolide M, ergosterol, and withanolide K, demonstrated the highest binding affinities ( $\Delta G$ ) towards the receptor protein 6GUE, with  $-10.2$ ,  $-10.1$ , and  $-9.9$  kcal mol $^{-1}$  values, respectively (Table 1), while the experimental control ligand (FB8) showed a binding energy of  $-9.0$  kcal mol $^{-1}$ , indicating the better affinity of our lead compounds. Additionally, compounds with binding energies more negative than  $-7$  kcal mol $^{-1}$  were considered highly potent.<sup>66,78,79</sup> In addition, the binding energies ( $\Delta G$ ) of all the docked compounds are shown in Table S6. Furthermore, we validated the binding strength of our selected compounds and compared them with other CDK2's homology protein structures (such as CDK1, CDK4, CDK5 and CDK6) using the docking methodology. The results revealed that our selected compounds are also potent against these CDK proteins, with binding energies greater than  $-7.7$  kcal mol $^{-1}$  towards all target proteins (Table S8).

To validate the docking protocol, we performed re-docking analysis of the target protein (PDB: 6GUE) with the co-crystal ligand or native ligand of 6GUE, which is experimental inhibitor and reference compounds against this protein.<sup>80</sup> To compare the positions of the experimental ligand and the docked poses, we calculated the RMSD of the docking poses in comparison to that of the co-crystal pose of the ligand (FB8). In this study, the RMSD of the co-crystal ligand was calculated (Fig. S2) for the best binding poses with their corresponding protein, which afforded 1.1069 Å. In structure-based drug



Table 1 Binding affinity ( $\Delta G$ ) scores of the top-ranked phytocompounds against CDK2

Compound ID	Name of compounds	2D structure	Docking score (kcal mol <sup>-1</sup> )
IMPHY003143	Withanolide M		-10.2
IMPHY012566	Ergosterol		-10.1
IMPHY010667	Withanolide K		-9.9
FB8	Control_ligand		-9.0

design and docking validation, a re-docking RMSD  $< 2.0 \text{ \AA}$  between the re-docked ligand and its crystallographic (co-crystal) pose is generally considered a standard value, as it indicates that the docking protocol can reliably reproduce experimentally observed binding modes.

### 3.8.1 Interpretation of protein-ligand interaction analysis.

The interactions between the selected ligands and the receptor protein were analyzed using the BIOVIA Discovery Studio Visualizer tool. Fig. 9 and Table S7 represent the interactions of the receptor proteins and the phytochemicals, along with the types of interactions (such as hydrogen bond, hydrophobic bonds, and van der Waals bonds) and the bond distance between the amino acid residues of the receptor protein 6GUE and the selected compounds (withanolide M, ergosterol, and withanolide K) or the FB8/control. Fig. 9 illustrates that all the compounds, as well as the control ligands, have some common residues where they interact with the proteins through different bonds. In Fig. 9A, withanolide M exhibits multiple interactions with the target protein *via* conventional hydrogen bonds and

alkyl interactions (Table S7). Additionally, an unfavorable acceptor-acceptor interaction is observed at GLN A:131 (2.82  $\text{\AA}$ ). The presence of strong hydrogen bonds suggests potential stability in the ligand binding, while hydrophobic (alkyl and  $\pi$ -alkyl) interactions may contribute to enhanced affinity. However, the unfavorable acceptor-acceptor bond at GLN A:131 may slightly affect the ligand's binding efficiency, although a single such contact does not automatically invalidate a pose. The impact of such a contact depends on distance/geometry, surrounding dielectric effects, and water mediation, as well as whether other favorable interactions (H-bonds, hydrophobics) dominate the overall  $\Delta G$ .<sup>81</sup> Notably, withanolide M also shared common residues with the control ligand, including ALA A:144, PHE A:80, VAL A:18, ALA A:31, LEU A:134, VAL A:64, and ILE A:10, highlighting the conserved interaction hotspots. For the compound ergosterol, one pi-sigma bond was found to form only for the PHE A:80 (3.76  $\text{\AA}$ ) position, where pi-Alkyl and alkyl bonds have predominantly formed (Fig. 9B). These residues also overlap with the control ligand, again pointing to ALA



## Complex View 3D Interaction View 2D Interaction View

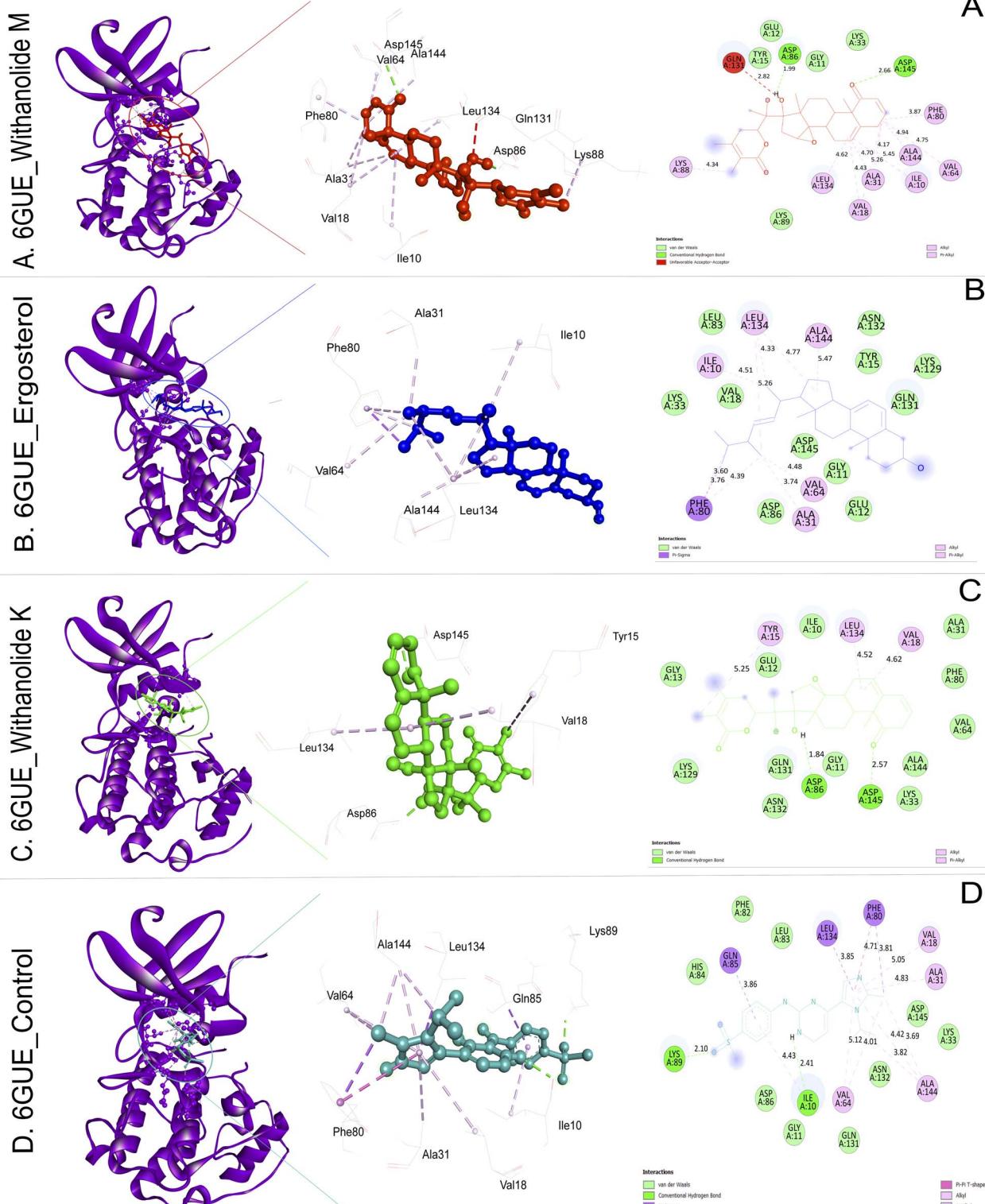


Fig. 9 3D (left) and 2D (right) diagrams of the interaction of receptor protein (A) withanolide M (red), (B) ergosterol (blue), (C) withanolide K (green), and (D) the control\_ligand (Cyan).

A:144, PHE A:80, ALA A:31, LEU A:134, VAL A:64, and ILE A:10 as crucial stabilizing contacts. In Fig. 9C, the interactions of the compound withanolide K (IMPHY010667) are described,

revealing conventional hydrogen bonds, pi-Alkyl and Alkyl bonds (Table S7). Similar to the other ligands and the control\_ligand, withanolide K exhibited common interactions at



LEU A:134 and VAL A:18 position. Finally, the control\_ligand (FB8) (Fig. 9D) displayed its own set of hydrogen and hydrophobic interactions with key residues of the receptor. Importantly, several of these residues overlapped with those identified for the phytochemicals, further validating the consistency of binding sites across both the natural ligands and the control.

### 3.9 MD simulation analysis

Molecular dynamics (MD) simulation can assess the stability of protein–ligand complexes within a controlled and artificial environment.<sup>82</sup> In this study, a 100 ns MD simulation was conducted to evaluate the conformational stability and steady-state nature of the protein–ligand complexes. The simulation was performed on the docked complex structure to examine its stability over time. The results of MD simulation are described based on the RMSF, RMSD, SASA, Rg, MM-GBSA, PCA, and FEL. Standard deviations (SD) represent the extent of fluctuation in each parameter, indicating the stability of the complexes during the simulations. Therefore, the SD values for RMSD, RMSF, SASA, Rg, and MM-GBSA are reported in Table 2 to illustrate the variation and stability of each protein–ligand complex.

**3.9.1 Root mean square deviation (RMSD).** The RMSD measures the structural stability and conformational change of the docked complex,<sup>83</sup> where high deviation indicates low stability and low RMSD indicates the stable performance of the system. The average RMSD values of withanolide M (black), withanolide K (green), and ergosterol (blue) with the receptor protein 6GUE were 2.23, 2.30, and 1.75 Å, respectively (Table 2), and the control (red) system showed a RMSD of 2.03 Å. Conversely, the free or unbound 6GUE (cyan) showed a higher RMSD (3.35 Å) than the bonded 6GUE complex (Fig. 10a), suggesting that our proposed compounds formed stable complexes with 6GUE.

**3.9.2 Root mean square fluctuation (RMSF).** The RMSF analysis was employed to assess the changes in the flexibility of particular amino acid residues in the receptor protein upon binding with the proposed phytocompounds (withanolide M, withanolide K, ergosterol, and control\_ligand). It helps to calculate the degree of fluctuation of a particular atom based on its mean position during the simulation period.<sup>47</sup> The average values of the RMSF of the free 6GUE and withanolide M complex showed similar trends, while withanolide K (green color) and control\_ligand (red color) showed similar fluctuations (Fig. 10b). Consequently, ergosterol (blue) showed an average

RMSF value of 1.06 Å, which is lower than that of the free protein, suggesting that the protein is more stable in its binding state with the ligands than in its free state.

**3.9.3 Solvent-accessible surface area (SASA).** The SASA describes the protein macromolecule surface that can be contacted by a solvent. It provides fundamental insights into how ligands interact with protein macromolecules and how their shapes change. In this study, all of our phytocompounds (withanolide M, withanolide K, ergosterol) and control\_ligand showed relatively consistent and low SASAs (Fig. 11b) of 227.54, 191.25, 192.64, and 201.46 Å<sup>2</sup>, respectively, which are comparatively better than the free protein's SASA of 506.08 Å<sup>2</sup>. These results provide a clear understanding of how the selected phytocompounds bind to the target protein and form a stable complex.

**3.9.4 The radius of gyration (RoG).** The radius of gyration (RoG) serves as a key indicator of molecular compactness, where a higher RoG corresponds to an extended or flexible conformation, while a lower RoG reflects a more tightly packed structure.<sup>84</sup> According to this study, ergosterol significantly showed the minimum fluctuation in the RoG plots (3.75 Å), indicating a high compactness and structural stability of its complex (Fig. 11a). Conversely, the RoG plots shows that the withanolide M (4.77 Å) and withanolide K (4.34 Å) follow the similar trend and remained same, and control\_ligand showed a fluctuation value (5.10 Å) similar to theirs, suggesting an extended, flexible conformation with greater RoG trajectory variability (Fig. 11a).

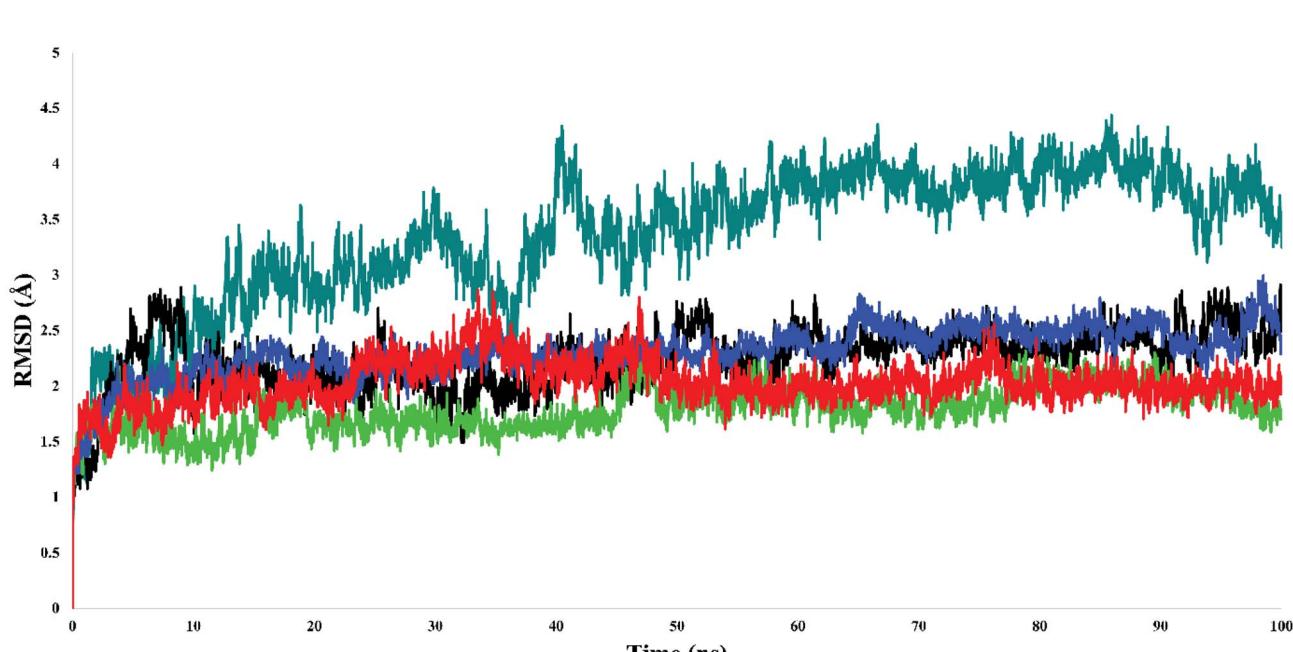
**3.9.5 Hydrogen bonding analysis.** Hydrogen bonds play a critical role in stabilizing ligand–protein interactions and significantly influence drug specificity, metabolism, and absorption.<sup>85,86</sup> They help to calculate the molecular interaction, strength, and stability between the ligand and protein. In Fig. 12, the 6GUE\_Withanolide K complex consistently exhibits the highest and most sustained number of hydrogen bonds between 2 and 1. Again, withanolide M and control\_ligand (black and red lines, respectively) showed irregular and few hydrogen bonds, rarely exceeding one bond over the simulation time, while ergosterol (blue line) did not form any H bonds throughout the simulation. These results suggest that withanolide K forms the most stable hydrogen bonding interactions with 6GUE, which may contribute to its enhanced binding affinity and complex stability. In contrast, ergosterol showed weaker hydrogen-bonding profiles.

**Table 2** The average values with standard deviations (SD) of RMSD, RMSF, RoG, SASA, H-bonds, and MM-GBSA-based binding free energy ( $\Delta G$ ) for all protein–ligand complexes

Compounds name	MD simulation analysis				
	RMSD (SD) (Å)	RMSF (SD) (Å)	SASA (SD) (Å <sup>2</sup> )	RoG (SD) (Å)	MM-GBSA (kcal mol <sup>-1</sup> )
Apo (CDK2)	3.35 (0.622)	1.13 (0.641)	506.08 (26.506)	7.43 (0.092)	—
Withanolide M	2.23 (0.278)	1.15 (1.121)	227.54 (40.951)	4.77 (0.044)	-34.12 (3.792)
Withanolide K	1.79 (0.199)	1.01 (0.630)	191.25 (26.635)	4.34 (0.044)	-32.80 (4.617)
Ergosterol	2.30 (0.231)	1.06 (0.701)	192.64 (33.223)	3.75 (0.034)	-33.06 (3.569)
Control_ligand	2.03 (0.203)	0.96 (0.581)	201.46 (29.868)	5.10 (0.167)	-30.33 (4.392)



(a)



(b)

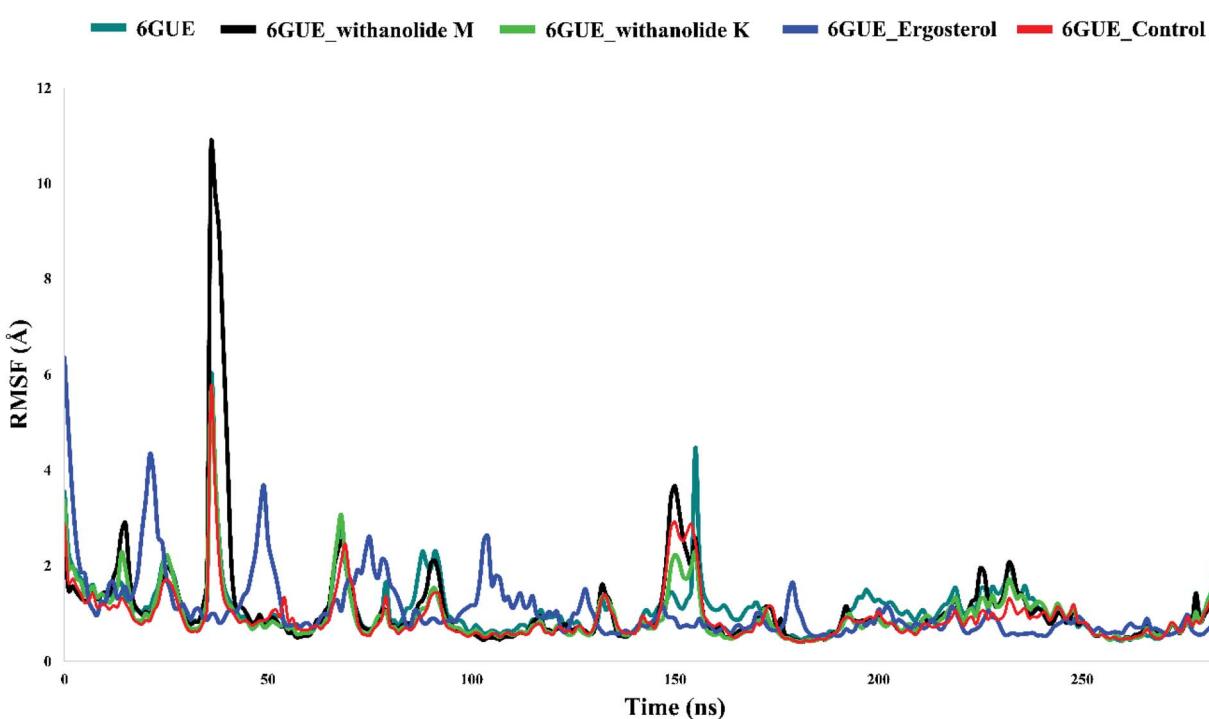


Fig. 10 Molecular dynamics simulation analysis of the (a) root mean square deviation (RMSD) and (b) root mean square fluctuation (RMSF), where the unbound 6GUE is denoted by cyan color, the complex of 6GUE with native control ligand in red, complex of withanolide M in black, withanolide K in green and ergosterol in blue.

### 3.10 MM-GBSA-based binding-energy analysis

The MM-GBSA binding free energies (BFEs) of withanolide M, withanolide K, ergosterol and the co-crystallized control ligand (FB8) against CDK2 (PDB ID: 6GUE) were analyzed over a 100 ns

MD simulation (Fig. 13). Our findings demonstrate the substantial binding affinities of the compounds, with  $\Delta G$  values reflecting the robust interactions between the potential ligands and the target protein. The calculated average BFEs of withanolide M, withanolide K, ergosterol and the co-crystallized



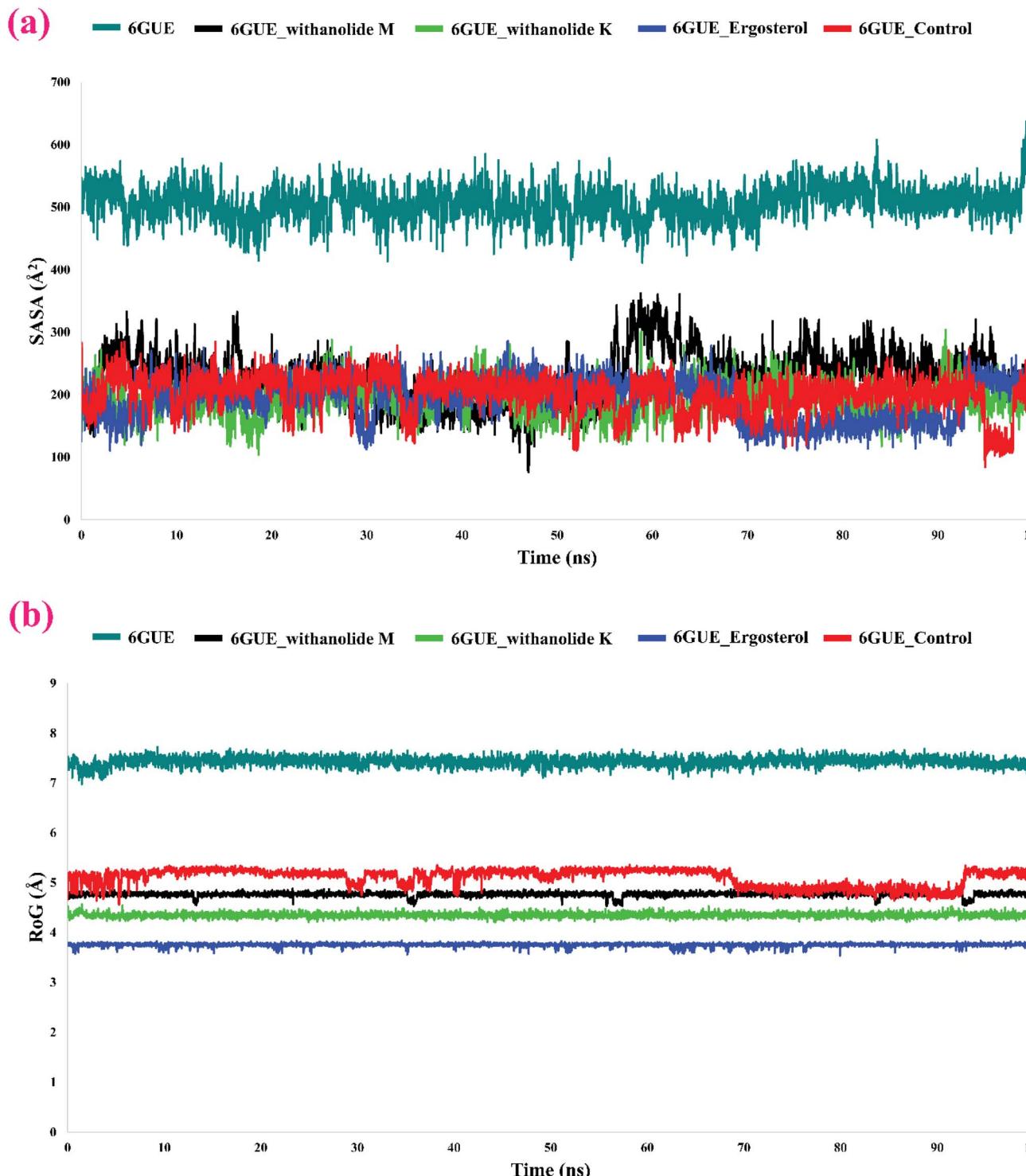


Fig. 11 Molecular dynamics simulation analysis of the (a) radius of gyration and (b) solvent accessible surface area over time, where the unbound 6GUE is denoted by cyan color, the complex of 6GUE with native control ligand in red, complex of withanolide M in black, withanolide K in green and ergosterol in blue.

control ligand (FB8) were  $-34.12$ ,  $-32.80$ ,  $-33.06$ , and  $-30.33$  kcal mol $^{-1}$ , respectively. Generally, a high negative binding energy represents a strong interaction at the protein-ligand interface.<sup>87</sup> Thus, the results signified that the phyto-compounds withanolide M, withanolide K, and ergosterol have higher negative energy compared to the control ligand.

### 3.11 Principal component analysis

Principal component analysis (PCA) of backbone atoms was used to compare the conformational sampling of the unbound 6GUE and its complexes with withanolide M (B), withanolide K (C), ergosterol (D) and the known control ligand FB8 (E). Rather



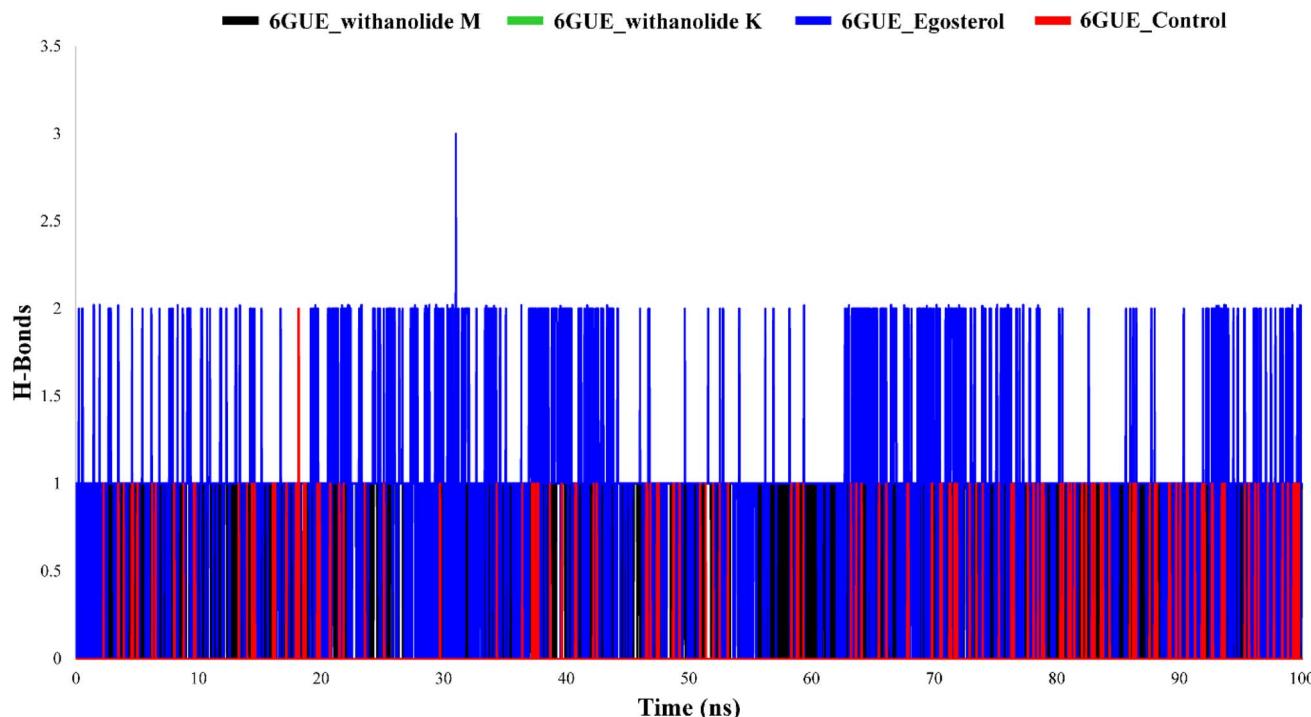


Fig. 12 Graphical representation of the protein–ligand contacts using H-bond calculations, showing withanolide M (orange), ergosterol (green), withanolide K (blue), and the control ligands (black).

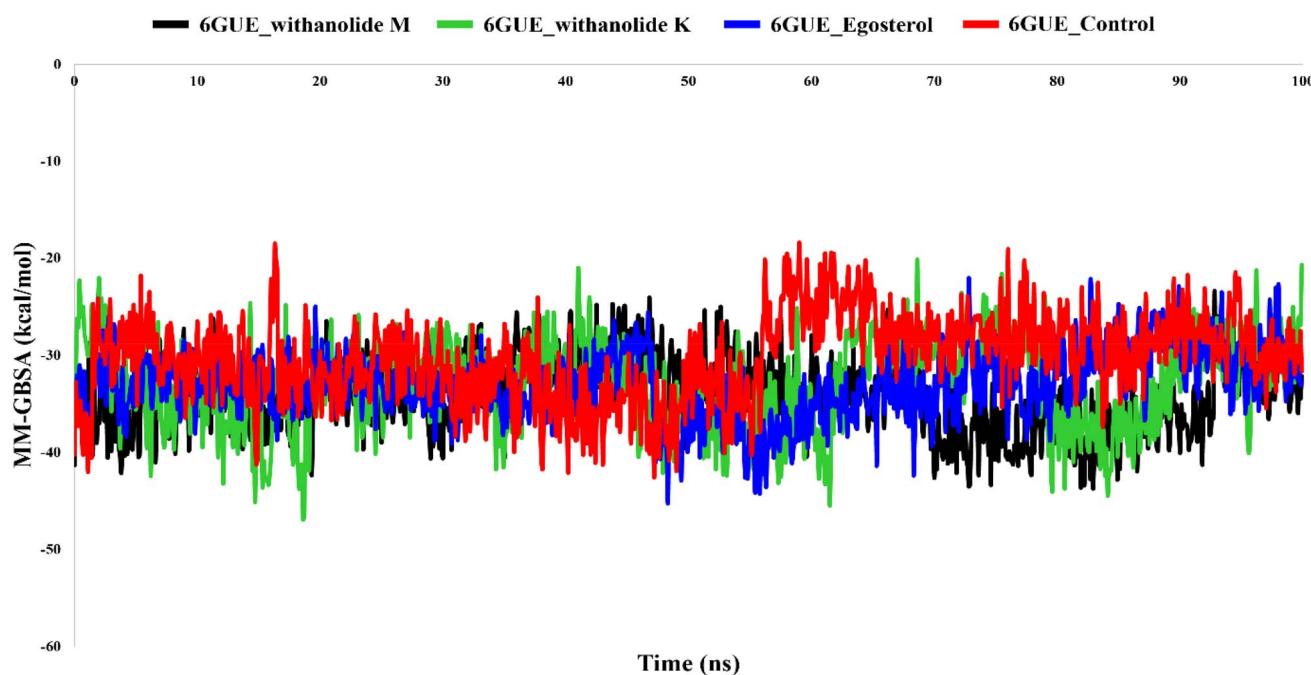
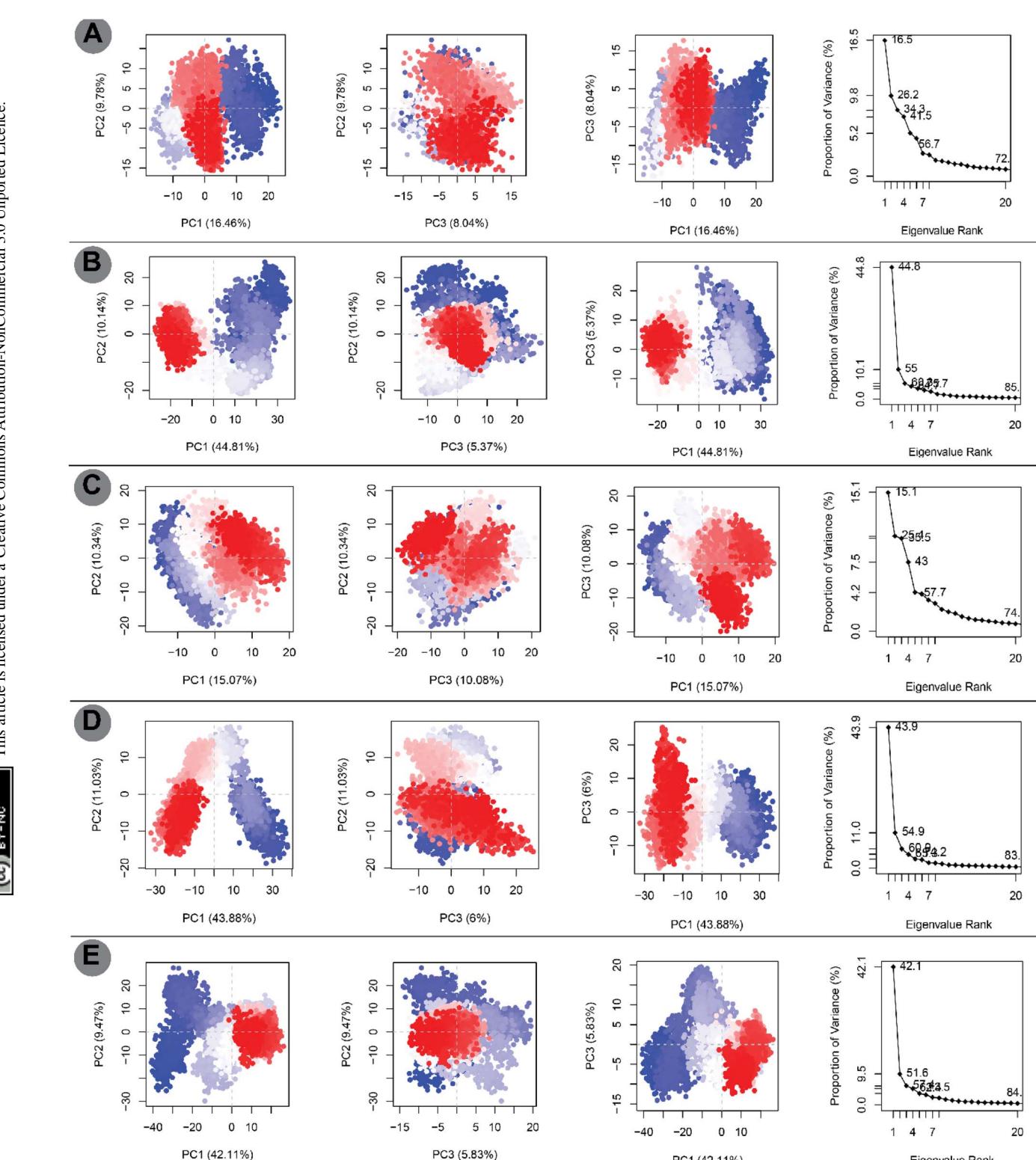


Fig. 13 MM-GBSA-based binding free energy calculation results of compounds withanolide M (orange), ergosterol (green), withanolide K (blue), and the control ligands (black) from the MD trajectories for 100 ns.

than relying on the percentage explained by PC1 alone, we evaluated the (i) variance explained by the top eigenvectors (scree plots), (ii) 2D projections (PC1–PC2 and PC1–PC3) to inspect cluster dispersion and overlap, and (iii) cluster

compactness (per-frame dispersion/convex hull area). The unbound protein (Fig. 14A) displays broad sampling across PCs, consistent with an increased intrinsic flexibility. Withanolide M (Fig. 14B) and ergosterol (Fig. 14D) both show relatively large





**Fig. 14** Principal Component Analysis (PCA) plots showing the dynamic behavior of the free unbound protein 6GUE (A) and three protein–ligand complexes: (B) 6GUE-withanolide M, (C) 6GUE-withanolide K, and (D) 6GUE-ergosterol, along with a known control ligand (E). Color mapping further clarified the fluctuation pattern across structures, with the red regions representing the least dynamic fluctuations (stable region), the white zones indicating moderate movements, and the blue areas highlighting the highest conformational change (dynamic region).

PC1 contributions, indicating dominant collective motions; however, their PC projections reveal broader sampling along PC1 (larger dispersion) rather than tight confinement. In

contrast, withanolide K (Fig. 14C) exhibits a compact and tightly clustered distribution in the PC projections with a smaller per-frame dispersion, indicating more restricted conformational

sampling while bound. The control ligand (Fig. 14E) shows intermediate behavior. Taken together with RMSD and RMSF analyses, the integrated picture is that withanolide K imposes the greatest conformational constraint on 6GUE (*i.e.*, most limited backbone sampling), whereas withanolide M and ergosterol induce large-scale motions (greater sampling along dominant modes). We, therefore, avoid equating high PC1% with increased stability or flexibility in isolation; instead, our conclusions are based on the combined evidence from PCA clustering and other parameters obtained from the MD data.

### 3.12 Dynamic cross-correlation matrix analysis

To further explore the dynamic behavior and correlated atomic motions, a Dynamic Cross-Correlation Matrix (DCCM) analysis was performed for the unbound protein and its ligand-bound complexes. The correlation coefficients range from  $-1$  (anti-correlated motion, magenta) to  $+1$  (correlated motion, cyan), with white regions representing uncorrelated motions.<sup>88</sup> The unbound 6GUE (Fig. 15A) showed moderately correlated motions predominantly along the diagonal, indicating internal residue-residue coherence in the localized regions. Upon ligand binding, substantial variations in dynamic correlations were observed. The withanolide M (Fig. 15B) and withanolide K (Fig. 15C) complexes showed enhanced correlated motions (cyan) among distant residue clusters. This suggests that these ligands promote a more organized, concerted movement across the protein, which can contribute to their overall stability and

potentially enhance their binding affinity. However, the ergosterol-bound system (Fig. 15D) displays pronounced anti-correlated motions (magenta) among different protein domains. This pattern suggests that ergosterol binding causes conformational switching, where one part of the protein moves in the opposite direction to another. This is not a simple stabilization but rather a ligand-triggered change in the protein's overall dynamics. The control ligand (Fig. 15E) shows scattered but weak cross-correlations, suggesting a less organized dynamic behavior compared to the observations for the other complexes. Overall, the DCCM analysis shows that withanolide M and withanolide K promote concerted, organized motions, while ergosterol induces a significant conformational switch. This detailed analysis reveals distinct mechanisms of action for each ligand, which were not captured in the initial interpretation.

### 3.13 Gibbs free energy landscape (FEL) analysis

To understand the conformational behavior and thermodynamic stability of a protein (6GUE) and protein–ligand complex (withanolide M, withanolide K, ergosterol, and the control ligand), the free energy landscape (FEL) analysis was conducted using the MD trajectories (Fig. 16). In this study, FEL plots were constructed using the RMSD (root mean square deviation) and RoG (radius of gyration) as reaction coordinates, capturing the conformational dynamics and thermodynamic stability of each system.

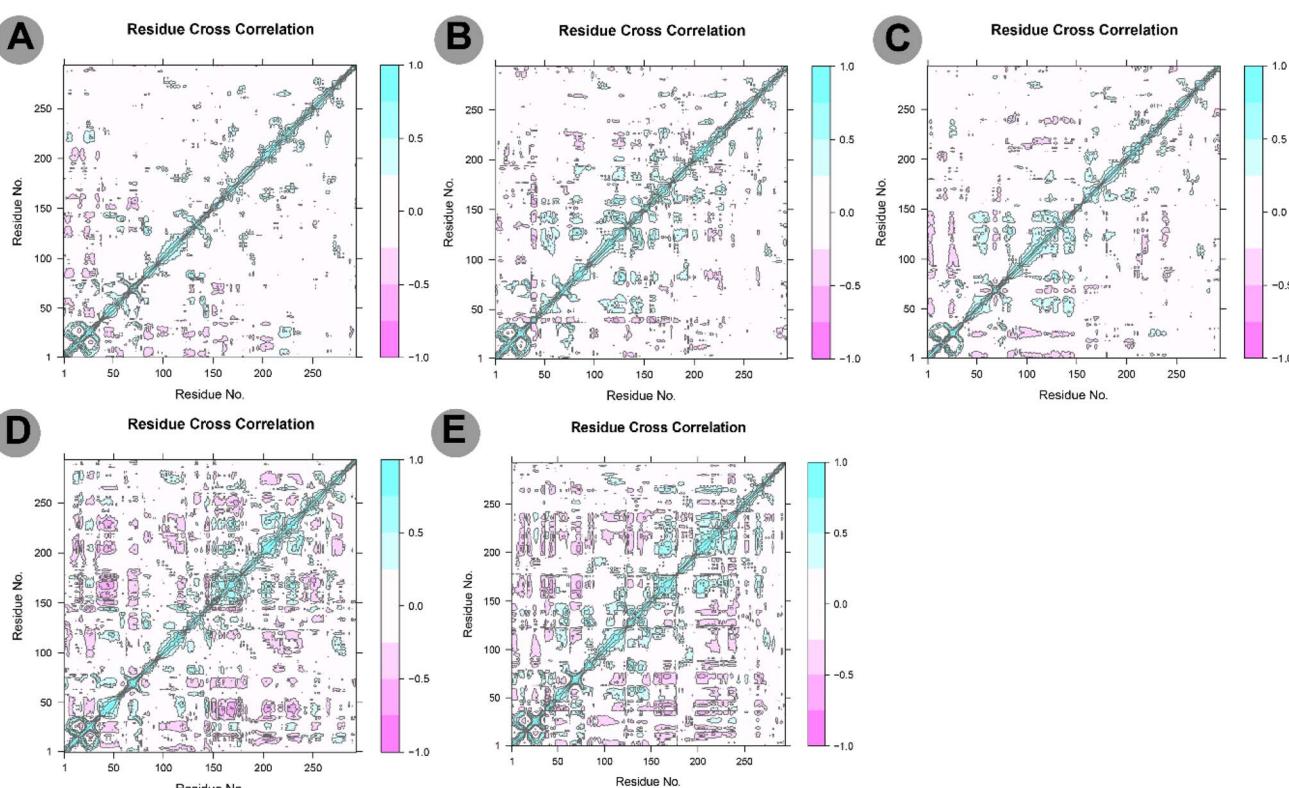


Fig. 15 DCCM analysis of the MD simulation results for the (A) free\_6GUE and the selected complexes, (B) 6GUE-withanolide M, (C) 6GUE-withanolide K, (D) 6GUE-ergosterol, and (E) 6GUE\_control.



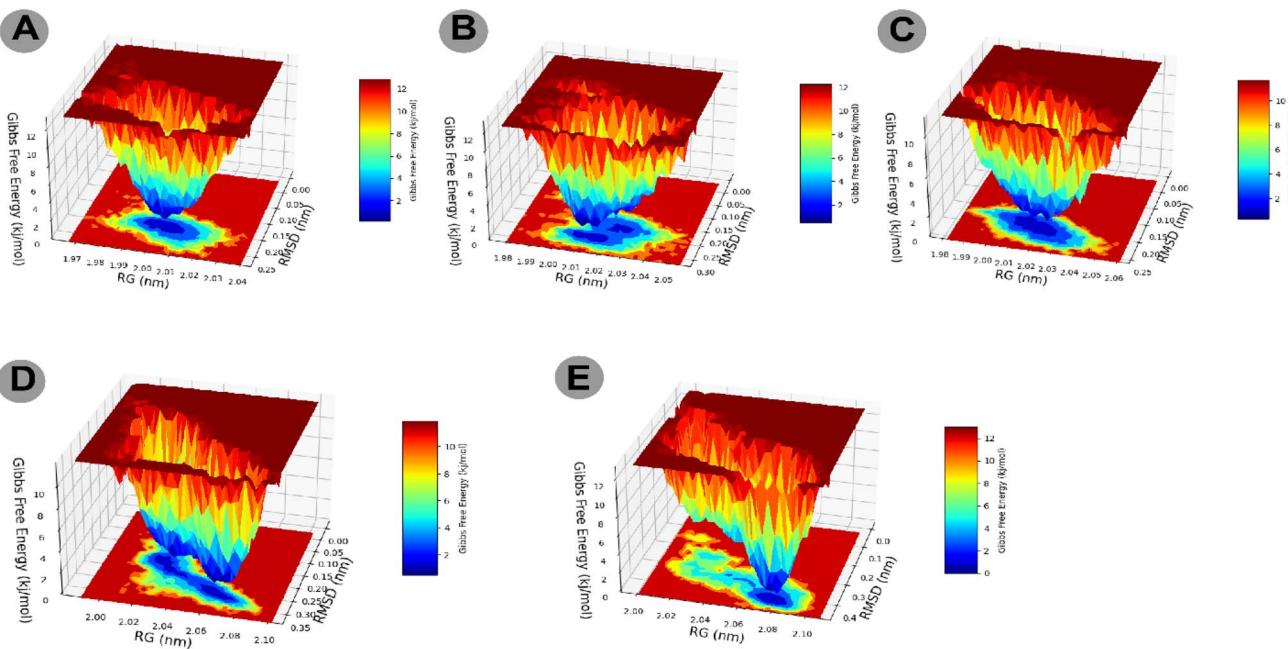


Fig. 16 Graphical representations of the Gibbs free energy landscape for (A) free\_6GUE and the selected complexes, (B) 6GUE-withanolide M, (C) 6GUE-withanolide K, (D) 6GUE-ergosterol, and (E) 6GUE\_control.

The unbound 6GUE system (Fig. 16A) showed a wide, shallow energy basin (deep blue), indicating a more flexible but less thermodynamically stable conformation compared to those of the ligand-bound forms. In comparison, the complexes of withanolide M (Fig. 16B) and the control ligand (Fig. 15E) both exhibit deep, well-defined energy basins. The deep blue region is concentrated and narrow, suggesting that these ligands constrain the protein to a few, thermodynamically favorable conformations. This indicates that these ligands induce a more rigid and stable structure. The withanolide K complex (Fig. 16C) shows a wider, slightly higher basin compared to the withanolide M and control systems. Conversely, the ergosterol complex (Fig. 16D) shows a broad, multi-basin landscape with several shallow minima. This indicates that ergosterol binding allows the protein to transition between multiple low-energy states, consistent with the conformational switching observed in the DCCM analysis.

### 3.14 Pharmacokinetics analysis

**3.14.1 ADME profiling of the selected phytocompounds.** The drug-like properties of the selected compounds (withanolide M & K, ergosterol) were evaluated based on Lipinski's rule of five, the number of rotatable bonds (nRotB), and the polar surface area or PSA (Table 3). Our study observed that none of our lead compounds violated the rule of 5, except withanolide M, for its high lipophilicity or LogP value (6.47). Also, according to Veber's rule, there was no violation in the physicochemical properties of the compounds, as their PSA values were less than  $140 \text{ \AA}^2$  and the nRotB was within the reference value.

The water solubility and GI absorption were found to be moderate and high for both withanolide M and withanolide K,

respectively; the control ligand absorption rate was high, and it was soluble in water. In contrast, ergosterol showed low absorption in gastrointestinal regions and poor water solubility. None of our lead compounds, as well as the control ligand, could penetrate the blood–brain barrier (BBB), which will help to reduce the side effects on the central nervous system (CNS). Their inhibition potential against major cytochrome P450 (CYP) isoenzymes was evaluated to assess their metabolic safety and potential for drug–drug interactions. Both withanolide M and K were predicted to be non-inhibitors of all major CYP isoenzymes assessed, suggesting no interference with hepatic metabolism and drug–drug interaction. In contrast, ergosterol and the control compound exhibited inhibitory activity against CYP2C9 and CYP2D6, respectively (Table 3), which are key isoenzymes involved in the metabolism of various clinically important drugs, indicating a possible risk of metabolic interactions and altered pharmacokinetics. The results suggest that withanolide M and K could be suitable candidate drugs due to their favorable metabolic profiles. The Pan-Assay Interference compounds (PAINS) help to identify the unwanted substructure of any compounds that could be problematic; none of our compounds showed any alerts that indicate frequent assay interference. This supports the reliability of their predicted bioactivity and reduces the likelihood of false-positive results in biological screening. All the compounds exhibited greater synthetic accessibility compared to the control compound, indicating their modest synthesis difficulties compared to the control ligand.

**3.14.2 *In silico* toxicity profiling of the selected compounds.** The *in silico* toxicological properties of the selected compounds, withanolide M, withanolide K, and ergosterol, were systematically evaluated across multiple toxicity endpoints



Table 3 Physicochemical, pharmacokinetics (ADME), and bioavailability of the selected lead compounds

Properties	Withanolide M	Withanolide K	Ergosterol	Control
Physico-chemical properties	MW (g mol <sup>-1</sup> )	468.58	470.6	396.65
	Consensus log P	3.31	3.11	6.47
	H-bond acceptors	6	6	1
	H-bond donors	2	3	1
Pharmacokinetics	Rotatable bonds	2	2	4
	Polar surface area	96.36	104.06	20.23
	Log S (ESOL)	-4.14	-4.1	-6.72
	GI absorption	High	High	Low
Drug likeness	BBB	No	No	No
	CYP1A2 inhibitor	No	No	No
	CYP2C19 inhibitor	No	No	No
	CYP2C9 inhibitor	No	No	Yes
	CYP2D6 inhibitor	No	No	No
	CYP3A4 inhibitor	No	No	Yes
	Lipinski's rule	Accepted	Accepted	Accepted
	Veber's rules	Accepted	Accepted	Accepted
Medi. Chemistry	Synth. accessibility	6.78	6.45	6.58
	Bioavailability	0.55	0.55	0.55
	PAINS	0	0	0

(including hepatotoxicity, neurotoxicity, cardiotoxicity, carcinogenicity, mutagenicity, and cytotoxicity), with a standard reference compound (FB8) comparative study (Table 4). According to the findings of this research, withanolide M was predicted to be a toxicological concern, as it exhibits cardiotoxicity, cytotoxicity, and carcinogenicity. In contrast, withanolide K demonstrated a more favorable safety profile, showing only cardiotoxicity, while it remained inactive across all other assessed endpoints. Similarly, ergosterol was found to be neurotoxic but did not show any other organ-specific or systemic toxic effects. Furthermore, the calculated Lethal Dose 50% (LD<sub>50</sub>) value supported the relative toxicity of the selected compounds (Table 4), where the lowest LD<sub>50</sub> value indicates the highest toxicity.<sup>89</sup> Then, the acute toxicity order according to this concept is ergosterol (10 mg kg<sup>-1</sup>) > withanolide M (34 mg kg<sup>-1</sup>) > withanolide K (400 mg kg<sup>-1</sup>) > control (500 mg kg<sup>-1</sup>). Thus, withanolide K had a substantially higher LD<sub>50</sub> of 400 mg kg<sup>-1</sup>, indicating comparatively low acute toxicity. The control ligand (FB8) showed the most favorable safety profile and was inactive across all endpoints, except for carcinogenicity.

### 3.15 Pharmacodynamics analysis

**3.15.1 Quantitative structure-activity relationship (QSAR) and pIC<sub>50</sub>.** The quantitative structure-activity relationship

(QSAR) is a well-developed computational methodology and equation routinely used to predict the biological profile of a compound. It has become an integral part of drug design for evaluating a predictive hypothesis based on a constructed model.<sup>90</sup> Generally, the standard range of pIC<sub>50</sub> for any compound to be considered a potential and bioactive compound should be greater than 4.0 but less than 10.<sup>68,91</sup> In Table 5, it has been reported that the pIC<sub>50</sub> values of our selected compounds, withanolide M, withanolide K, and ergosterol, and the control ligand FB8 are 5.786, 5.638, 6.264, and 5.640, respectively. Thus, the pIC<sub>50</sub> values of the selected compounds are within the acceptable range since the value is not more than 10.0 and not higher than 4.0.

## 4. Discussion

Cyclin-dependent kinase 2 (CDK2) is a pivotal enzyme that controls the G1/S phase transition in the cell cycle by activating key replication proteins.<sup>92</sup> In normal cells, the activity of CDK2 is tightly regulated to ensure proper cell cycle progression. However, in many cancers, this regulation is disrupted, leading to uncontrolled cell division, increased replication stress, and failure of critical cell cycle checkpoints.<sup>93,94</sup> In breast cancer, the overexpression of CDK2, often driven by cyclin E amplification,

Table 4 Physicochemical characteristics, water solubility, drug likeness, lipophilicity, GI absorption, and accessibility of specific synthesis of the three lead compounds, where red indicates toxicity and green indicates non-toxicity

Toxicity classification	Withanolide M	Withanolide K	Ergosterol	Control
Hepatotoxicity	Inactive	Inactive	Inactive	Inactive
Neurotoxicity	Inactive	Inactive	Active	Inactive
Cardiotoxicity	Active	Active	Inactive	Inactive
Carcinogenicity	Active	Inactive	Inactive	Active
Mutagenicity	Inactive	Inactive	Inactive	Inactive
Cytotoxicity	Active	Inactive	Inactive	Inactive
Predicted LD <sub>50</sub> (mg kg <sup>-1</sup> )	34	400	10	500



Table 5 Data table for the QSAR descriptors and  $\text{pIC}_{50}$  value calculation

ID	Chiv5	bctum1	MRVSA9	MRVSA6	PEOEVSA5	GATSV4	Diameter	J	$\text{pIC}_{50}$	ML-based $\text{pIC}_{50}$
Withanolide M	6.944	4.077	11.753	34.947	30.222	1.038	14	1.344	5.786	5.189
Withanolide K	6.365	4.016	11.753	34.947	30.725	1.038	14	1.484	5.638	5.153
Ergosterol	5.846	3.991	0	35.45	76.993	0.58	15	1.458	6.264	5.196
Control_ligand	3.269	4.623	21.473	48.55	0	1.068	13	1.501	5.640	

leads to the premature initiation of DNA replication and the bypass of regulatory checkpoints. This uncontrolled proliferation is strongly associated with tumor aggressiveness and poor patient outcomes.<sup>95</sup> In ovarian cancer, CDK2 helps sustain cell division even in the presence of DNA damage, which can make them more resistant to chemotherapy and lead to disease recurrence.<sup>96</sup> CDK2 plays a crucial role in lung and colorectal cancer by regulating cell cycle progression through the G1/S phase transition, and its overactivation supports the uncontrolled proliferation and survival of cancer cells.<sup>97</sup> High CDK2 levels are also associated with lymph node metastasis in early-stage colorectal carcinoma.<sup>69</sup> Therefore, targeting CDK2 represents a critical therapeutic strategy to suppress uncontrolled cell proliferation and potentially improve patient outcomes. Although various CDK4/6 inhibitors have been approved for clinical use, selective CDK2 inhibitors, such as BLU-222, remain investigational, with their clinical translation still in progress.<sup>98</sup> In this study, we explored a comprehensive *in silico* approach to identify potent CDK2 inhibitors derived from traditional-medicinal-plant-derived phytochemicals. Our work combines machine learning-based bioactivity prediction with structure-based and dynamics-based drug discovery tools to address this gap and propose novel, safe, and effective CDK2-targeting candidates.

In this study, we investigated the oncogenic relevance of CDK2 by integrating transcriptomic, proteomic, genomic, and immunological analyses across various cancer types. Our findings reinforce the pivotal role of CDK2 in malignancies, aligning with its known function in G1/S transition and replication control. Gene expression analyses from TIMER2.0 and GEPIA2 (ref. 38 and 39) revealed that CDK2 is significantly overexpressed in multiple tumor types, including breast, lung, liver, and colon cancers. Protein-level data from UALCAN further confirmed the elevated CDK2 levels in these cancers, supporting its active role in tumor biology. Notably, a high CDK2 expression was correlated with poor overall and disease-free survival results in cancers, such as ACC, KIRP, LGG, LIHC, MESO, PAAD, and SKCM, suggesting its potential as a prognostic biomarker.<sup>99</sup> Interestingly, our general observation that a high CDK2 expression indicates poor survivability is contradicted in the case of READ tumors, where Li *et al.* (2001)<sup>69</sup> found that a decreased CDK2 expression level was associated with poor prognosis. This suggests that in READ tumors, a low CDK2 expression may actually signal more aggressive disease progression. Combined with the opposite trend in UVM, the results underscore the context-dependent prognostic role of CDK2, possibly mediated by tumor-specific biology, microenvironmental factors, or mutation landscapes. Genomic

alteration analysis revealed frequent mutations and amplifications, particularly in endometrial and ovarian cancers, with missense mutations like R265L and D68N potentially contributing to oncogenic activation.<sup>100</sup> Interestingly, amplification-driven expression appeared prominently in aggressive cancers, such as glioma and non-small cell lung cancer. CDK2 expression also showed positive associations with tumor-promoting immune infiltrates, particularly T follicular helper cells and cancer-associated fibroblasts,<sup>101,102</sup> hinting at its involvement in modulating the tumor microenvironment. Furthermore, interaction and co-expression analyses suggested that CDK2 functions as more than just a canonical cell cycle regulator. Its connectivity with multiple oncogenic signaling pathways indicates that CDK2 serves as a central hub integrating proliferative signals, checkpoint regulation, and tumor-promoting networks. This centrality highlights its potential as a key driver of uncontrolled proliferation in cancer, consistent with its established role in G1/S transition, while also pointing to its broader implications in oncogenesis and therapeutic targeting. Collectively, these findings not only highlight the pan-cancer overexpression and prognostic significance of CDK2 but also underscore its potential as a therapeutic target. Given the current lack of selective CDK2 inhibitors, our comprehensive *in silico* screening of phytochemical libraries aims to bridge this gap by identifying novel, plant-derived candidates with promising bioactivity and safety profiles for future development.

Additionally, we first retrieved 764 phytocompounds from the IMPPAT database and refined the dataset by removing duplicates and entries lacking structural information, resulting in 467 compounds. A subsequent screening using a LightGBM-based machine learning model identified 322 compounds with predicted  $\text{pIC}_{50}$  values greater than 5.1, highlighting their potential as promising therapeutic leads. To assess the binding affinities of the selected compounds with CDK2, molecular docking was performed following the virtual screening of these top-ranked compounds. In such studies, it is generally understood that lower (more negative) binding energies indicate stronger and more stable interactions between a ligand and its target protein.<sup>103</sup> Among the tested compounds, withanolide M, ergosterol, and withanolide K demonstrated the most favorable binding scores ( $-10.2$ ,  $-10.1$ , and  $-9.9$  kcal mol<sup>-1</sup>, respectively), outperforming the control compound FB8 ( $-9.0$  kcal mol<sup>-1</sup>). Additionally, other CDK proteins bind to the selected compounds with good binding affinity scores, not less than  $-7.7$  kcal mol<sup>-1</sup>. Also, the control (FB8)'s binding energy towards the homologue's protein of CDK2 was lower than those of the top-ranked three compounds. These results suggest that the top-ranked candidates may have



a stronger affinity for the active site of CDK2 compared to the reference compound. Furthermore, the detailed protein–ligand interaction analysis provided critical insights into how the top compounds engage with the CDK2 binding site. The ligand–protein interaction analysis revealed that withanolide M, ergosterol, and withanolide K share several common residues with the control FB8, including key sites, such as ALA A:144, PHE A:80, VAL A:18, ALA A:31, LEU A:134, VAL A:64, and ILE A:10. These conserved contacts, involving hydrogen bonding and hydrophobic interactions, highlight critical hotspots for stabilizing ligand binding. Although an unfavorable acceptor–acceptor interaction was observed with withanolide M, the overall interaction profiles suggest strong binding affinities supported by favorable hydrogen and hydrophobic contacts. Collectively, the overlap between phytochemicals and the control underscores the reliability of these residues as potential anchoring points for inhibitor design.

Molecular dynamics (MD) simulation is a structure-based computational technique used to analyze the dynamic behavior and interaction of protein–ligand complexes in an environment that mimics physiological conditions. In MD simulations, a virtual biological system is created to continuously monitor atomic-level interactions over time. This approach provides insights into the conformational stability, rigidity, and adaptability of the complexes under cellular-like stress conditions.<sup>104</sup> To explore these properties in our study, we performed 100 ns MD simulations of the protein–ligand complexes. Dynamic parameters, such as RMSD, RMSF,  $R_g$ , SASA, Hydrogen bonding, MM/GBSA binding energy, PCA, DCCM, and FEL, were analyzed to understand the structural behavior and dynamic stability of the complexes. RMSD measures the average structural deviation of a protein–ligand complex over time, where lower values indicate greater structural stability.<sup>105</sup> Among the tested compounds, withanolide K showed the lowest RMSD (1.79 Å), indicating that it had the most stable interaction, with slightly better stability than the standard inhibitor FB8 (2.03 Å). Withanolide M (2.23 Å) and ergosterol (2.30 Å) exhibited RMSD values close to that of the control, reflecting comparable stability. These trends are represented in Fig. 10. Moreover, the RMSD of the bound ligand during the simulation was also considered in this study. The findings revealed that the selected compounds show comparatively lower RMSD than the control ligand, but the stability of the control ligand and proposed phytocompounds is almost the same (Fig. S3). The RMSD values of the control ligands vary from ~2.0 Å to ~2.5 Å, approximately, whereas the selected compound shows a similar trend, with values ranging from 0.2 Å to 1.0 Å. This value reflects how much the ligands deviate from their initial docked position, rather than the stability of the protein backbone. Root mean square fluctuation analysis provides insights into molecular flexibility, where reduced fluctuations correlate with structural rigidity<sup>106</sup> and elevated values indicate dynamic regions. In this study, all three compounds exhibited RMSF profiles that were reasonably comparable to that of the control (FB8, 0.96 Å), suggesting that they maintained a degree of structural stability throughout the simulation. In particular, ergosterol (1.06 Å) showed the closest

profile to the control. The SASA metric evaluates molecular surface exposure, where increased values correlate with structural instability, while decreased values demonstrate enhanced packing efficiency between solvent molecules and protein components.<sup>78</sup> In this case, the complexes formed by ergosterol and withanolide K with the receptor exhibited slightly lower SASA values than that formed by the control ligand FB8, indicating more compact and stable interactions. On the other hand, the complex formed by withanolide M showed a higher SASA value, suggesting that it is less tightly packed and possibly less stable compared to the control. RoG serves as an essential biophysical metric for evaluating the structural density and conformational stability of protein–ligand complexes.<sup>107</sup> The complexes formed by ergosterol, withanolide K, and withanolide M with the receptor exhibited lower RoG values compared to the complex formed by the control, indicating improved compactness and structural stability, as represented in Fig. 11a. Hydrogen bonds play a crucial role in stabilizing ligand–protein interactions.<sup>108</sup> Compared to the complex formed by the control ligand FB8, the complex formed by withanolide K exhibited a higher and more consistent number of hydrogen bonds throughout the simulation, indicating stronger and more stable binding. Withanolide M showed a similar or slightly weaker profile than the control, while ergosterol did not form any hydrogen bonds, suggesting its significantly weaker interaction stability, as illustrated in Fig. 11. Here, the hydrogen bonds act like strict rules in binding; they require specific distances and angles, which limit how freely the ligand and protein side chains can move. When these restrictions are relaxed, the ligand has more ways to fit into the pocket, and the protein side chains also gain more room to adjust, making binding easier in flexible or expanding sites.<sup>109,110</sup> In systems where both the pocket and the ligand are adaptable, this flexibility works together: the ligand can explore alternate binding poses that match the pocket as it reshapes, allowing good accommodation even without relying heavily on precise polar interactions.<sup>111</sup> MM-GBSA binding free energy (BFE) analysis provides insight into the strength and stability of protein–ligand interactions, where more negative  $\Delta G$  values indicate a stronger binding affinity.<sup>112</sup> In this study, withanolide M, ergosterol, and withanolide K all demonstrated stronger binding energies compared to the control. These findings suggest that all three phytocompounds exhibit more favorable binding interactions with CDK2 than the control ligand, as illustrated in Fig. 12.

PCA was performed to assess the large-scale conformational dynamics of 6GUE in its free form and when bound to different ligands.<sup>113</sup> The PCA results revealed that the unbound form of 6GUE (apo) showed a moderate level of flexibility, with PC1 accounting for 16.46% of the total variance, consistent with a dynamic structure in the absence of a ligand. Upon ligand binding, the conformational sampling patterns changed, and the trajectories formed more compact clusters in the PCA plots, suggesting restricted motion and enhanced stabilization of the complexes. Notably, the withanolide K complex exhibited a more balanced variance across PC1 (15.07%), PC2 (10.34%), and PC3 (10.08%), accompanied by a relatively dense clustering,



indicating constrained but flexible motions within a narrower conformational space. In contrast, withanolide M, ergosterol, and the control ligand FB8 displayed higher variance in PC1 ( $>40\%$ ), which is indicative of dominant large-scale motions, although they still form single dominant clusters that suggest stabilization in alternative conformational states. Overall, these patterns suggest that while withanolide K restricts conformational sampling in a more compact manner, the other ligands may drive structural rearrangements toward new stable conformations rather than simple destabilization. DCCM analysis revealed distinct differences in correlated and anti-correlated movements across the studied complexes.<sup>114</sup> The withanolide M and withanolide K complexes enhanced correlated motions across distant residues, suggesting a more organized and concerted protein movement. In contrast, the ergosterol complex induced strong anti-correlated motions, indicative of a ligand-triggered conformational switch. The control complex showed weaker, scattered correlations, suggesting less coordinated dynamics. FEL analysis provided further insights into the thermodynamic stability of the compounds. The control complex and the withanolide M complex exhibited the deepest and most well-defined energy minima, indicating that they have the highest thermodynamic stability. The withanolide K complex displayed a slightly broader energy basin, reflecting its greater flexibility compared to the most stable complexes. In contrast, the ergosterol-bound complex showed a broader distribution with multiple shallow energy minima, signifying higher conformational variability and lower stability. Based on these findings, the overall stability of the complexes can be ranked as withanolide M > control > withanolide K > unbound protein > ergosterol.

The drug-likeness and ADMET evaluation suggest that both withanolide M and K possess favorable pharmacokinetic and physicochemical profiles, making them promising candidates for further development. Despite withanolide M's elevated  $\log P$ , its compliance with Lipinski's and Veber's rules indicates an overall acceptable balance between permeability and solubility. Notably, the high gastrointestinal absorption of withanolide K and withanolide M indicates their favorable oral bioavailability.<sup>115</sup> Moreover, their inability to penetrate the blood-brain barrier minimizes the risk of unwanted CNS-related side effects, supporting their safety.<sup>116</sup> In contrast, ergosterol's poor water solubility and low GI absorption raise concerns about its oral bioavailability and overall drug-likeness. Furthermore, the metabolic safety profiles of withanolide M and withanolide K are strengthened by their predicted non-inhibition of key cytochrome P450 isoenzymes, reducing the likelihood of hepatic interference or drug-drug interactions. This is particularly important as drug candidates that inhibit enzymes like CYP2C9 or CYP2D6 often face clinical complications.<sup>117</sup> The inhibitory behavior of ergosterol and the control compound toward these enzymes highlights a potential limitation in their metabolic compatibility. The absence of PAINS alerts and good synthetic accessibility further support withanolide M and K as promising and reliable drug candidates with relatively safe pharmacokinetic profiles. The *in silico* toxicity assessment further reinforced the safety profiles of the

lead compounds. Withanolide K demonstrated the most favorable safety index, being inactive for all evaluated endpoints except mild cardiotoxicity. Withanolide M showed a comparatively higher toxicological risk, being active for cytotoxicity, carcinogenicity, and cardiotoxicity. Ergosterol exhibited neurotoxicity and had the lowest  $LD_{50}$  value ( $10 \text{ mg kg}^{-1}$ ), indicating that it has the highest level of acute toxicity among the tested compounds. In contrast, the control ligand (FB8) presented the overall best safety profile, showing no major organ-specific toxicities except for carcinogenicity, and it had the highest  $LD_{50}$  value ( $500 \text{ mg kg}^{-1}$ ), supporting its low acute toxicity. In this study,  $pIC_{50}$  values were calculated using a multiple linear regression (MLR) model with selected descriptors (Chiv5, bcutm1, MRVSA9, MRVSA6, PEOFVSA5, GATSv4, J, Diameter). These descriptors capture key molecular features, such as topology, polarity, size, and electronic distribution, which are critical for ligand-protein interactions.<sup>118</sup> Similar sets of descriptors have been successfully applied in previous QSAR-based studies, supporting the reliability of our predictions.<sup>119-122</sup> QSAR analysis revealed that all selected compounds, including the control ligand, exhibited  $pIC_{50}$  values within the optimal bioactivity range (4–10), indicating their potential as biologically active and promising drug candidates.

In summary, our integrated computational pipeline successfully identified three promising CDK2 inhibitors from phytochemical sources, with withanolide K emerging as the lead candidate due to its strong binding affinity, dynamic stability, favorable pharmacokinetic properties, and low toxicity. These findings offer valuable direction for future experimental validation, with the goal of advancing phytochemical-based therapeutic development against CDK2-driven cancers. In future investigations, the integration of deep learning-based generative models and multi-target screening could further enhance compound discovery and prioritize selectivity, thereby complementing our current computational framework.

## 5. Conclusion

This study commenced with a comprehensive pan-cancer analysis to establish the oncogenic significance of CDK2 across multiple tumor types, revealing its consistent over-expression, context-dependent association with prognosis, and potential involvement in immune modulation. Building on this foundation, an integrative computational approach was applied to identify potential CDK2 inhibitors from traditionally used medicinal plants. Using machine learning-based  $pIC_{50}$  prediction, molecular docking, molecular dynamics simulations, and ADMET/toxicity assessments, withanolide K emerged as the most promising candidate due to its predicted high binding affinity, structural stability, and favorable pharmacokinetic properties. These findings highlight the potential of natural compounds, particularly withanolide K, as candidate CDK2 inhibitors and underscore the relevance of CDK2 as a viable anticancer target. However, these results are based solely on *in silico* analyses; experimental validation through *in vitro* and *in*



*vivo* studies is essential to confirm the efficacy, specificity, and safety of these compounds.

## 6. Limitations of this study

This study employed a range of computational approaches, including pan-cancer analysis, machine learning-based screening, molecular docking, molecular dynamics (MD) simulations, MM-GBSA free energy calculations, and pharmacokinetic/pharmacodynamic predictions. While these methods provide valuable preliminary insights, the binding affinity estimations from docking and MM-GBSA cannot fully capture the complexity of biological interactions occurring *in vivo*. In particular, the tumor microenvironment (TME) may significantly influence the compound activity and therapeutic response, which is not accounted for in our computational framework. Moreover, pharmacokinetic predictions are based on *in silico* models and therefore require experimental confirmation. To establish the clinical relevance of the identified compounds (withanolide K, withanolide M, and ergosterol) as potential CDK2 inhibitors, further validation using TME-relevant *in vitro* models (e.g., 3D spheroid or co-culture systems) and *in vivo* studies will be essential to evaluate their efficacy, bioavailability, and possible off-target effects.

## Author contributions

Conceptualization – Md. Ahad Ali methodology – Md. Ahad Ali, Hridhhi Sarker, Ahmed Saif, and Humaira Sheikh data curation – Tania Khan, Humaira Sheikh, Sadia Afrin, and Most. Asha Khatun formal analysis – Md. Ahad Ali, Sadia Afrin, and Humaira Sheikh visualization – Md. Ahad Ali, Hridhhi Sarker, Sadia Afrin, and Most. Asha Khatun validation – Md. Ahad Ali, Hridhhi Sarker, and Ahmed Saif project administration – Md. Ahad Ali software & resources – Ahmed Saif, Neeraj Kumar and Farhad Bin Farid supervision – Md. Ahad Ali writing – original draft, Md. Ahad Ali and Hridhhi Sarker writing – review & editing, Tania Khan and Md. Ahad Ali.

## Conflicts of interest

All authors declare no conflict of interests.

## Data availability

The data supporting this article have been included as part of the Supplementary information (SI). Supplementary information is available at DOI: <https://doi.org/10.1039/d5ra05535k>.

## References

- 1 P. Jean-Pierre and B. C. McDonald, in *Neuroepidemiology of Cancer and Treatment-Related Neurocognitive Dysfunction in Adult-Onset Cancer Patients and Survivors*, 2016, pp. 297–309.
- 2 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA. Cancer J. Clin.*, 2021, **71**, 209–249, DOI: [10.3322/caac.21660](https://doi.org/10.3322/caac.21660).
- 3 H. M. Bizuayehu, K. Y. Ahmed, G. D. Kibret, A. F. Dadi, S. A. Belachew, T. Bagade, T. K. Tegegne, R. L. Venchiariutti, K. T. Kibret, A. H. Hailegebireal, *et al.*, *Global Disparities of Cancer and Its Projected Burden in 2050, JAMA Netw. Open*, 2024, **7**, e2443198, DOI: [10.1001/jamanetworkopen.2024.43198](https://doi.org/10.1001/jamanetworkopen.2024.43198).
- 4 C. S. Pramesh, R. A. Badwe, N. Bhoo-Pathy, C. M. Booth, G. Chinnaswamy, A. J. Dare, V. P. de Andrade, D. J. Hunter, S. Gopal, M. Gospodarowicz, *et al.*, *Priorities for Cancer Research in Low- and Middle-Income Countries: A Global Perspective, Nat. Med.*, 2022, **28**, 649–657, DOI: [10.1038/s41591-022-01738-x](https://doi.org/10.1038/s41591-022-01738-x).
- 5 M. Sibai, S. Cervilla, D. Grases, E. Musulen, R. Lazcano, C.-K. Mo, V. Davalos, A. Fortian, A. Bernat, M. Romeo, *et al.*, *The Spatial Landscape of Cancer Hallmarks Reveals Patterns of Tumor Ecological Dynamics and Drug Sensitivity, Cell Rep.*, 2025, **44**, 115229, DOI: [10.1016/j.celrep.2024.115229](https://doi.org/10.1016/j.celrep.2024.115229).
- 6 D. Hanahan, *Hallmarks of Cancer: New Dimensions, Cancer Discov.*, 2022, **12**, 31–46, DOI: [10.1158/2159-8290.CD-21-1059](https://doi.org/10.1158/2159-8290.CD-21-1059).
- 7 D. Hanahan and R. A. Weinberg, *The Hallmarks of Cancer, Cell*, 2000, **100**, 57–70, DOI: [10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- 8 D. Hanahan and R. A. Weinberg, *Hallmarks of Cancer: The Next Generation, Cell*, 2011, **144**, 646–674, DOI: [10.1016/j.cell.2011.02.013](https://doi.org/10.1016/j.cell.2011.02.013).
- 9 S. Ghafouri-Fard, T. Khoshbakht, B. M. Hussen, P. Dong, N. Gassler, M. Taheri, A. Baniahamad and N. A. Dilmaghani, *A Review on the Role of Cyclin Dependent Kinases in Cancers, Cancer Cell Int.*, 2022, **22**(1), 325, DOI: [10.1186/s12935-022-02747-z](https://doi.org/10.1186/s12935-022-02747-z).
- 10 S. Tadesse, E. C. Caldon, W. Tilley and S. Wang, *Cyclin-Dependent Kinase 2 Inhibitors in Cancer Therapy: An Update, J. Med. Chem.*, 2019, **62**, 4233–4251, DOI: [10.1021/acs.jmedchem.8b01469](https://doi.org/10.1021/acs.jmedchem.8b01469).
- 11 C. J. Sherr and J. M. Roberts, *CDK Inhibitors: Positive and Negative Regulators of G1-Phase Progression, Genes Dev.*, 1999, **13**, 1501–1512, DOI: [10.1101/gad.13.12.1501](https://doi.org/10.1101/gad.13.12.1501).
- 12 I. House, M. Valore-Caplan, E. Maris and G. S. Falchook, *Cyclin Dependent Kinase 2 (CDK2) Inhibitors in Oncology Clinical Trials: A Review, J. Immunother. Precis. Oncol.*, 2025, **8**, 47–54, DOI: [10.36401/JIPO-24-22](https://doi.org/10.36401/JIPO-24-22).
- 13 M. Mihara, S. Shintani, Y. Nakahara, A. Kiyota, Y. Ueyama, T. Matsumura and D. T. W. Wong, *Overexpression of CDK2 Is a Prognostic Indicator of Oral Cancer Progression, Japanese J. Cancer Res.*, 2001, **92**, 352–360, DOI: [10.1111/j.1349-7006.2001.tb01102.x](https://doi.org/10.1111/j.1349-7006.2001.tb01102.x).
- 14 C. E. Caldon, C. M. Sergio, J. Kang, A. Muthukaruppan, M. N. Boersma, A. Stone, J. Barraclough, C. S. Lee, M. A. Black, L. D. Miller, *et al.*, *Cyclin E2 Overexpression Is Associated with Endocrine Resistance but Not Insensitivity to CDK2 Inhibition in Human Breast Cancer Cells, Mol. Cancer Ther.*, 2012, **11**, 1488–1499, DOI: [10.1158/1535-7163.MCT-11-0963](https://doi.org/10.1158/1535-7163.MCT-11-0963).



15 L. A. Aaltonen, F. Abascal, A. Abeshouse, H. Aburatani, D. J. Adams, N. Agrawal, K. S. Ahn, S.-M. Ahn, H. Aikata, R. Akbani, *et al.*, Pan-Cancer Analysis of Whole Genomes, *Nature*, 2020, **578**, 82–93, DOI: [10.1038/s41586-020-1969-6](https://doi.org/10.1038/s41586-020-1969-6).

16 L. K. Gopi and B. L. Kidder, Integrative Pan Cancer Analysis Reveals Epigenomic Variation in Cancer Type and Cell Specific Chromatin Domains, *Nat. Commun.*, 2021, **12**, 1419, DOI: [10.1038/s41467-021-21707-1](https://doi.org/10.1038/s41467-021-21707-1).

17 Y. Fu, A. W. Jung, R. V. Torne, S. Gonzalez, H. Vöhringer, A. Shmatko, L. R. Yates, M. Jimenez-Linan, L. Moore and M. Gerstung, Pan-Cancer Computational Histopathology Reveals Mutations, Tumor Composition and Prognosis, *Nat. Cancer*, 2020, **1**, 800–810, DOI: [10.1038/s43018-020-0085-8](https://doi.org/10.1038/s43018-020-0085-8).

18 L. Han, Y. Yuan, S. Zheng, Y. Yang, J. Li, M. E. Edgerton, L. Diao, Y. Xu, R. G. W. Verhaak and H. Liang, The Pan-Cancer Analysis of Pseudogene Expression Reveals Biologically and Clinically Relevant Tumour Subtypes, *Nat. Commun.*, 2014, **5**, 3963, DOI: [10.1038/ncomms4963](https://doi.org/10.1038/ncomms4963).

19 M. D. M. Leiserson, F. Vandin, H.-T. Wu, J. R. Dobson, J. V. Eldridge, J. L. Thomas, A. Papoutsaki, Y. Kim, B. Niu, M. McLellan, *et al.*, Pan-Cancer Network Analysis Identifies Combinations of Rare Somatic Mutations across Pathways and Protein Complexes, *Nat. Genet.*, 2015, **47**, 106–114, DOI: [10.1038/ng.3168](https://doi.org/10.1038/ng.3168).

20 Y. Zeng, X. Ren, P. Jin, Z. Fan, M. Liu, Y. Zhang, L. Li, M. Zhuo, J. Wang, Z. Li, *et al.*, Inhibitors and PROTACs of CDK2: Challenges and Opportunities, *Expert Opin. Drug Discov.*, 2024, **19**, 1125–1148, DOI: [10.1080/17460441.2024.2376655](https://doi.org/10.1080/17460441.2024.2376655).

21 J. Cicenas and J. Simkus, CDK Inhibitors and FDA: Approved and Orphan, *Cancers*, 2024, **16**, 1555, DOI: [10.3390/cancers16081555](https://doi.org/10.3390/cancers16081555).

22 C. Dietrich, A. Trub, A. Ahn, M. Taylor, K. Ambani, K. T. Chan, K.-H. Lu, C. A. Mahendra, C. Blyth, R. Coulson, *et al.*, INX-315, a Selective CDK2 Inhibitor, Induces Cell Cycle Arrest and Senescence in Solid Tumors, *Cancer Discov.*, 2024, **14**, 446–467, DOI: [10.1158/2159-8290.CD-23-0954](https://doi.org/10.1158/2159-8290.CD-23-0954).

23 C. R. Pye, M. J. Bertin, R. S. Lokey, W. H. Gerwick and R. G. Linington, Retrospective Analysis of Natural Products Provides Insights for Future Discovery Trends, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 5601–5606, DOI: [10.1073/pnas.1614680114](https://doi.org/10.1073/pnas.1614680114).

24 U. Anand, N. Jacobo-Herrera, A. Altemimi and N. Lakhssassi, A Comprehensive Review on Medicinal Plants as Antimicrobial Therapeutics: Potential Avenues of Biocompatible Drug Discovery, *Metabolites*, 2019, **9**, 258, DOI: [10.3390/metabo9110258](https://doi.org/10.3390/metabo9110258).

25 S. T. Asma, U. Acaroz, K. Imre, A. Morar, S. R. A. Shah, S. Z. Hussain, D. Arslan-Acaroz, H. Demirbas, Z. Hajrulai-Musliu, F. R. Istanbullugil, *et al.*, Natural Products/Bioactive Compounds as a Source of Anticancer Drugs, *Cancers*, 2022, **14**, 6203, DOI: [10.3390/cancers14246203](https://doi.org/10.3390/cancers14246203).

26 K. Mohanraj, B. S. Karthikeyan, R. P. Vivek-Ananth, R. P. B. Chand, S. R. Aparna, P. Mangalapandi and A. Samal, IMPPAT: A Curated Database of Indian Medicinal Plants, Phytochemistry and Therapeutics, *Sci. Rep.*, 2018, **8**, 4329, DOI: [10.1038/s41598-018-22631-z](https://doi.org/10.1038/s41598-018-22631-z).

27 S. Alsenan, I. Al-Turaiki and A. Hafez, A Recurrent Neural Network Model to Predict Blood–Brain Barrier Permeability, *Comput. Biol. Chem.*, 2020, **89**, 107377, DOI: [10.1016/j.compbiochem.2020.107377](https://doi.org/10.1016/j.compbiochem.2020.107377).

28 I. Iqbal, M. Younus, K. Walayat, M. U. Kakar and J. Ma, Automated Multi-Class Classification of Skin Lesions through Deep Convolutional Neural Network with Dermoscopic Images, *Comput. Med. Imaging Graph.*, 2021, **88**, 101843, DOI: [10.1016/j.compmedimag.2020.101843](https://doi.org/10.1016/j.compmedimag.2020.101843).

29 I. Iqbal, K. Walayat, M. U. Kakar and J. Ma, Automated Identification of Human Gastrointestinal Tract Abnormalities Based on Deep Convolutional Neural Network with Endoscopic Images, *Intell. Syst. with Appl.*, 2022, **16**, 200149, DOI: [10.1016/j.iswa.2022.200149](https://doi.org/10.1016/j.iswa.2022.200149).

30 P. Solanki, S. Abdul Amin and A. Manhas, Integrating Machine Learning with *In Silico* Studies and Quantum Chemistry: Exploring Novel Compounds through Multiscale Screening Targeting the CDK2 Enzyme, *Comput. Biol. Med.*, 2025, **196**, 110712, DOI: [10.1016/j.compbomed.2025.110712](https://doi.org/10.1016/j.compbomed.2025.110712).

31 T. T. Liu, R. Li, C. Huo, J. P. Li, J. Yao, X. L. Ji and Y. Q. Qu, Identification of CDK2-Related Immune Forecast Model and CeRNA in Lung Adenocarcinoma, a Pan-Cancer Analysis, *Front. Cell Dev. Biol.*, 2021, **9**, 682002, DOI: [10.3389/fcell.2021.682002](https://doi.org/10.3389/fcell.2021.682002).

32 W. Filgueira de Azevedo, Predicting Inhibition of CDK2 with SANDRoS: The Application of Machine Learning to Navigate the Scoring Function Space, *Curr. Med. Chem.*, 2024, **31**, DOI: [10.2174/0109298673313727240819070317](https://doi.org/10.2174/0109298673313727240819070317).

33 T. M. Okyay, I. Yilmaz and M. Koldas, Machine Learning-Based Bioactivity Prediction of Porphyrin Derivatives: Molecular Descriptors, Clustering, and Model Evaluation, *Photochem. Photobiol. Sci.*, 2025, **24**, 923–937, DOI: [10.1007/s43630-025-00733-8](https://doi.org/10.1007/s43630-025-00733-8).

34 A. Rácz, L. M. Mihalovits, D. Bajusz, K. Héberger and R. A. Miranda-Quintana, Molecular Dynamics Simulations and Diversity Selection by Extended Continuous Similarity Indices, *J. Chem. Inf. Model.*, 2022, **62**, 3415–3425, DOI: [10.1021/acs.jcim.2c00433](https://doi.org/10.1021/acs.jcim.2c00433).

35 F. En-nahli, H. Hajji, M. Ouabane, M. Aziz Ajana, C. Sekatte, T. Lakhli and M. Bouachrine, ADMET Profiling and Molecular Docking of Pyrazole and Pyrazolines Derivatives as Antimicrobial Agents, *Arab. J. Chem.*, 2023, **16**, 105262, DOI: [10.1016/j.arabjc.2023.105262](https://doi.org/10.1016/j.arabjc.2023.105262).

36 M. Abdullahe, S. E. Adeniji, D. E. Arthur and S. Musa, Quantitative Structure-Activity Relationship (QSAR) Modelling Study of Some Novel Carboxamide Series as New Anti-Tubercular Agents, *Bull. Natl. Res. Cent.*, 2020, **44**, 136, DOI: [10.1186/s42269-020-00389-7](https://doi.org/10.1186/s42269-020-00389-7).

37 K. Chang, C. J. Creighton, C. Davis, L. Donehower, J. Drummond, D. Wheeler, A. Ally, M. Balasundaram, I. Birol, Y. S. N. Butterfield, *et al.*, The Cancer Genome Atlas Pan-Cancer Analysis Project, *Nat. Genet.*, 2013, **45**, 1113–1120, DOI: [10.1038/ng.2764](https://doi.org/10.1038/ng.2764).



38 Z. Tang, B. Kang, C. Li, T. Chen and Z. Zhang, GEPIA2: An Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis, *Nucleic Acids Res.*, 2019, **47**, W556–W560, DOI: [10.1093/nar/gkz430](https://doi.org/10.1093/nar/gkz430).

39 T. Li, J. Fu, Z. Zeng, D. Cohen, J. Li, Q. Chen, B. Li and X. S. Liu, TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells, *Nucleic Acids Res.*, 2020, **48**, W509–W514, DOI: [10.1093/NAR/GKAA407](https://doi.org/10.1093/NAR/GKAA407).

40 D. S. Chandrashekhar, S. K. Karthikeyan, P. K. Korla, H. Patel, A. R. Shovon, M. Athar, G. J. Netto, Z. S. Qin, S. Kumar, U. Manne, *et al.*, UALCAN: An Update to the Integrated Cancer Data Analysis Platform, *Neoplasia*, 2022, **25**, 18–27, DOI: [10.1016/j.neo.2022.01.001](https://doi.org/10.1016/j.neo.2022.01.001).

41 M. Goel, P. Khanna and J. Kishore, Understanding Survival Analysis: Kaplan-Meier Estimate, *Int. J. Ayurveda Res.*, 2010, **1**, 274, DOI: [10.4103/0974-7788.76794](https://doi.org/10.4103/0974-7788.76794).

42 J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, *et al.*, Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal, *Sci. Signal.*, 2013, **6**, pl1, DOI: [10.1126/scisignal.2004088](https://doi.org/10.1126/scisignal.2004088).

43 D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, *et al.*, The STRING Database in 2023: Protein–Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest, *Nucleic Acids Res.*, 2023, **51**, D638–D646, DOI: [10.1093/nar/gkac1000](https://doi.org/10.1093/nar/gkac1000).

44 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. C. Ideker, A Software Environment for Integrated Models of Biomolecular Interaction Networks, *Genome Res.*, 2003, **13**, 2498–2504, DOI: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303).

45 G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki, DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome Biol.*, 2003, **4**, R60, DOI: [10.1186/gb-2003-4-9-r60](https://doi.org/10.1186/gb-2003-4-9-r60).

46 W. Caldwell, Z. Yan, W. Lang and A. Masucci, The IC<sub>50</sub> Concept Revisited, *Curr. Top. Med. Chem.*, 2012, **12**, 1282–1290, DOI: [10.2174/156802612800672844](https://doi.org/10.2174/156802612800672844).

47 A. Saif, M. T. Islam, M. O. Raihan, N. Yousefi, M. A. Rahman, H. Faridi, A. R. Hasan, M. M. Hossain, R. M. Saleem, G. M. Albadrani, *et al.*, Pan-Cancer Analysis of CDC7 in Human Tumors: Integrative Multi-Omics Insights and Discovery of Novel Marine-Based Inhibitors through Machine Learning and Computational Approaches, *Comput. Biol. Med.*, 2025, **190**, 110044, DOI: [10.1016/j.combiomed.2025.110044](https://doi.org/10.1016/j.combiomed.2025.110044).

48 H. Gubler, U. Schopfer and E. Jacoby, Theoretical and Experimental Relationships between Percent Inhibition and IC<sub>50</sub> Data Observed in High-Throughput Screening, *J. Biomol. Screen.*, 2013, **18**, 1–13, DOI: [10.1177/1087057112455219](https://doi.org/10.1177/1087057112455219).

49 L. Yang and A. Shami, On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice, *Neurocomputing*, 2020, **415**, 295–316, DOI: [10.1016/j.neucom.2020.07.061](https://doi.org/10.1016/j.neucom.2020.07.061).

50 H. M. Berman, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**, 235–242, DOI: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235).

51 BIOVIA, D.S., *Discovery Studio Visualizer V21.1.0.20298*. BIOVIA, Dassault Systèmes, 2005.

52 N. Guex and M. C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: An Environment for Comparative Protein Modeling, *Electrophoresis*, 1997, **18**, 2714–2723, DOI: [10.1002/clps.1150181505](https://doi.org/10.1002/clps.1150181505).

53 M. A. Ali, H. Sheikh, M. Yaseen, M. O. Faruqe, I. Ullah, N. Kumar, M. A. Bhat and M. N. H. Mollah, Exploring the Therapeutic Potential of Petiveria Alliacea L. Phytochemicals: A Computational Study on Inhibiting SARS-CoV-2's Main Protease (Mpro), *Molecules*, 2024, **29**(11), 2524, DOI: [10.3390/molecules29112524](https://doi.org/10.3390/molecules29112524).

54 O. Trott and A. J. Olson, AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading, *J. Comput. Chem.*, 2010, **31**, 455–461, DOI: [10.1002/jcc.21334](https://doi.org/10.1002/jcc.21334).

55 R. Yousif, H. M. Mohamed, M. A. Almogaddam, K. M. Elamin, S. R. M. Ibrahim, B. E. Ainousah, A. M. Alraddadi, E. A. Awad and A. A. Alzain, Computational Screening Campaign Reveal Natural Candidates as Potential ASK1 Inhibitors: Pharmacophore Modeling, Molecular Docking, MMGBSA Calculations, ADMET Prediction, and Molecular Dynamics Simulation Studies, *Sci. African*, 2025, **28**, e02634, DOI: [10.1016/j.sciaf.2025.e02634](https://doi.org/10.1016/j.sciaf.2025.e02634).

56 R Core Team Award Number, *W81XWH-12-2-0022 TITLE: Prehospital Use of Plasma for Traumatic Hemorrhage* PRINCIPAL INVESTIGATOR: Bruce D. Spiess CONTRACTING ORGANIZATION: Virginia Commonwealth University, Richmond, VA 23298 REPORT DATE: June 2014 TYPE OF REPORT, 2014, ISBN, 3900051070.

57 B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. D. Caves, Bio3d: An R Package for the Comparative Analysis of Protein Structures, *Bioinformatics*, 2006, **22**, 2695–2696, DOI: [10.1093/bioinformatics/btl461](https://doi.org/10.1093/bioinformatics/btl461).

58 L. P. Kagami, G. M. das Neves, L. F. S. M. Timmers, R. A. Caceres and V. L. Eifler-Lima, Geo-Measures: A PyMOL Plugin for Protein Structure Ensembles Analysis, *Comput. Biol. Chem.*, 2020, **87**, 107322, DOI: [10.1016/j.combiolchem.2020.107322](https://doi.org/10.1016/j.combiolchem.2020.107322).

59 D. A. Pearlman, Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to P38 MAP Kinase, *J. Med. Chem.*, 2005, **48**, 7796–7807, DOI: [10.1021/jm050306m](https://doi.org/10.1021/jm050306m).

60 N. Kumar, R. Srivastava, A. Prakash and A. M. Lynn, Structure-Based Virtual Screening, Molecular Dynamics Simulation and MM-PBSA toward Identifying the Inhibitors for Two-Component Regulatory System Protein NarL of *Mycobacterium Tuberculosis*, *J. Biomol. Struct. Dyn.*, 2020, **38**, 3396–3410, DOI: [10.1080/07391102.2019.1657499](https://doi.org/10.1080/07391102.2019.1657499).

61 E. Lionta, G. Spyrou, D. Vassilatis and Z. Cournia, Structure-Based Virtual Screening for Drug Discovery: Principles,



Applications and Recent Advances, *Curr. Top. Med. Chem.*, 2014, **14**, 1923–1938, DOI: [10.2174/1568026614666140929124445](https://doi.org/10.2174/1568026614666140929124445).

62 C. A. Lipinski, Lead- and Drug-like Compounds: The Rule-of-Five Revolution, *Drug Discov. Today Technol.*, 2004, **1**, 337–341, DOI: [10.1016/j.ddtec.2004.11.007](https://doi.org/10.1016/j.ddtec.2004.11.007).

63 A. Daina, O. Michielin and V. Zoete, SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules, *Sci. Rep.*, 2017, **7**, 42717, DOI: [10.1038/srep42717](https://doi.org/10.1038/srep42717).

64 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings 1PII of Original Article: S0169-409X(96)00423-1, *Adv. Drug Deliv. Rev.*, 2001, **46**, 3–26, DOI: [10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0).

65 D. E. V. Pires, T. L. Blundell and D. B. Ascher, PkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures, *J. Med. Chem.*, 2015, **58**, 4066–4072, DOI: [10.1021/acs.jmedchem.5b00104](https://doi.org/10.1021/acs.jmedchem.5b00104).

66 M. E. K. Talukder, M. Aktaruzzaman, N. H. Siddiquee, S. Islam, T. A. Wani, H. M. Alkahtani, S. Zargar, M. O. Raihan, M. M. Rahman, S. Pokhrel, *et al.*, Cheminformatics-Based Identification of Phosphorylated RET Tyrosine Kinase Inhibitors for Human Cancer, *Front. Chem.*, 2024, **12**, 1407331, DOI: [10.3389/fchem.2024.1407331](https://doi.org/10.3389/fchem.2024.1407331).

67 H. Nour, O. Abchir, S. Belaïdi, F. A. Qais, S. Chtita and S. Belaaouad, 2D-QSAR and Molecular Docking Studies of Carbamate Derivatives to Discover Novel Potent Anti-Butyrylcholinesterase Agents for Alzheimer's Disease Treatment, *Bull. Korean Chem. Soc.*, 2022, **43**, 277–292, DOI: [10.1002/bkcs.12449](https://doi.org/10.1002/bkcs.12449).

68 S. Akash, M. R. Islam, M. M. Rahman, M. S. Hossain, M. A. K. Azad and R. Sharma, Investigation of the New Inhibitors by Modified Derivatives of Pinocembrin for the Treatment of Monkeypox and Marburg Virus with Different Computational Approaches, *Biointerface Res. Appl. Chem.*, 2023, **13**(6), 534, DOI: [10.33263/BRIAC136.534](https://doi.org/10.33263/BRIAC136.534).

69 J.-Q. Li, H. Miki, M. Ohmori, F. Wu and Y. Funamoto, Expression of Cyclin E and Cyclin-Dependent Kinase 2 Correlates with Metastasis and Prognosis in Colorectal Carcinoma, *Hum. Pathol.*, 2001, **32**, 945–953, DOI: [10.1053/hupa.2001.27116](https://doi.org/10.1053/hupa.2001.27116).

70 B. Thomas, M. Soladoye, T. Adegboyega, G. Agu and O. Popoola, Antibacterial and Anti-Inflammatory Activities of Anacardium Occidentale Leaves and Bark Extracts, *Niger. J. Basic Appl. Sci.*, 2015, **23**, 1, DOI: [10.4314/njabas.v23i1.1](https://doi.org/10.4314/njabas.v23i1.1).

71 R. Kumar, A. K. Singh, A. Gupta, A. Bishayee and A. K. Pandey, Therapeutic Potential of Aloe Vera—A Miracle Gift of Nature, *Phytomedicine*, 2019, **60**, 152996, DOI: [10.1016/j.phymed.2019.152996](https://doi.org/10.1016/j.phymed.2019.152996).

72 M. Nigam, M. Atanassova, A. P. Mishra, R. Pezzani, H. P. Devkota, S. Plygun, B. Salehi, W. N. Setzer and J. Sharifi-Rad, Bioactive Compounds and Health Benefits of Artemisia Species, *Nat. Prod. Commun.*, 2019, **14**(7), DOI: [10.1177/1934578X19850354](https://doi.org/10.1177/1934578X19850354).

73 M. K. Alam, M. O. Hoq and M. S. Uddin, Therapeutic Use of Withania Somnifera, *Asian J. Med. Biol. Res.*, 2016, **2**, 148–155, DOI: [10.3329/ajmbr.v2i2.29004](https://doi.org/10.3329/ajmbr.v2i2.29004).

74 R. Singh, Geetanjali Asparagus Racemosus : A Review on Its Phytochemical and Therapeutic Potential, *Nat. Prod. Res.*, 2016, **30**, 1896–1908, DOI: [10.1080/14786419.2015.1092148](https://doi.org/10.1080/14786419.2015.1092148).

75 A. Lateef, B. I. Folarin, S. M. Oladejo, P. O. Akinola, L. S. Beukes and E. B. Gueguim-Kana, Characterization, Antimicrobial, Antioxidant, and Anticoagulant Activities of Silver Nanoparticles Synthesized from Petiveria Alliacea L. Leaf Extract, *Prep. Biochem. Biotechnol.*, 2018, **48**, 646–652, DOI: [10.1080/10826068.2018.1479864](https://doi.org/10.1080/10826068.2018.1479864).

76 M. Li, H. Chen, H. Zhang, M. Zeng, B. Chen and L. Guan, Prediction of the Aqueous Solubility of Compounds Based on Light Gradient Boosting Machines with Molecular Fingerprints and the Cuckoo Search Algorithm, *ACS Omega*, 2022, **7**, 42027–42035, DOI: [10.1021/acsomega.2c03885](https://doi.org/10.1021/acsomega.2c03885).

77 M. T. Islam, M. Aktaruzzaman, A. Saif, A. R. Hasan, M. M. H. Sourov, B. Sikdar, S. Rehman, A. Tabassum, S. Abeed-Ul-Haque, M. H. Sakib, *et al.*, Identification of Acetylcholinesterase Inhibitors from Traditional Medicinal Plants for Alzheimer's Disease Using *In Silico* and Machine Learning Approaches, *RSC Adv.*, 2024, **14**, 34620–34636, DOI: [10.1039/d4ra05073h](https://doi.org/10.1039/d4ra05073h).

78 N. H. Siddiquee, M. I. Hossain, F. M. Priya, S. B. Azam, M. E. K. Talukder, D. Barua, S. Malek, N. Saha, S. Muntaha, R. Paul, *et al.*, Nature's Defense against Emerging Neurodegenerative Threats: Dynamic Simulation, PCA, DCCM Identified Potential Plant-Based Antiviral Lead Targeting Borna Disease Virus Nucleoprotein, *PLoS One*, 2024, **19**, e0310802, DOI: [10.1371/journal.pone.0310802](https://doi.org/10.1371/journal.pone.0310802).

79 F. Ahammad, R. Alam, R. Mahmud, S. Akhter, E. K. Talukder, A. M. Tonmoy, S. Fahim, K. Al-Ghamdi, A. Samad and I. Qadri, Pharmacoinformatics and Molecular Dynamics Simulation-Based Phytochemical Screening of Neem Plant (*Azadirachta Indica*) against Human Cancer by Targeting MCM7 Protein, *Brief. Biotinform.*, 2021, **22**(5), bbab098, DOI: [10.1093/bib/bbab098](https://doi.org/10.1093/bib/bbab098).

80 D. J. Wood, S. Korolchuk, N. J. Tatum, L.-Z. Wang, J. A. Endicott, M. E. M. Noble and M. P. Martin, Differences in the Conformational Energy Landscape of CDK1 and CDK2 Suggest a Mechanism for Achieving Selective CDK Inhibition, *Cell Chem. Biol.*, 2019, **26**, 121–130e5, DOI: [10.1016/j.chembiol.2018.10.015](https://doi.org/10.1016/j.chembiol.2018.10.015).

81 X. Du, Y. Li, Y. L. Xia, S. M. Ai, J. Liang, P. Sang, X. L. Ji and S. Q. Liu, Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods, *Int. J. Mol. Sci.*, 2016, **17**(2), 144, DOI: [10.3390/ijms17020144](https://doi.org/10.3390/ijms17020144).

82 R. S. Katiyar and P. K. Jha, Molecular Simulations in Drug Delivery: Opportunities and Challenges, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2018, **8**(4), e1358, DOI: [10.1002/wcms.1358](https://doi.org/10.1002/wcms.1358).



83 N. H. Siddiquee, M. I. Hossain, M. E. K. Talukder, S. A. A. Nirob, M. Shourav, I. Jahan, U. H. A. Tamanna, P. Das, R. Akter, M. Hasan, *et al.*, In-Silico Identification of Novel Natural Drug Leads against the Ebola Virus VP40 Protein: A Promising Approach for Developing New Antiviral Therapeutics, *Informatics Med. Unlocked*, 2024, **45**, 101458, DOI: [10.1016/j.imu.2024.101458](https://doi.org/10.1016/j.imu.2024.101458).

84 A. S. Abouzied, S. Alqarni, K. M. Younes, S. M. Alanazi, D. M. Alrsheed, R. K. Alhathal, B. Huwaimel and A. M. Elkashlan, Structural and Free Energy Landscape Analysis for the Discovery of Antiviral Compounds Targeting the Cap-Binding Domain of Influenza Polymerase PB2, *Sci. Rep.*, 2024, **14**, 25441, DOI: [10.1038/s41598-024-69816-3](https://doi.org/10.1038/s41598-024-69816-3).

85 S. Bharadwaj, A. Dubey, U. Yadava, S. K. Mishra, S. G. Kang and V. D. Dwivedi, Exploration of Natural Compounds with Anti-SARS-CoV-2 Activity *via* Inhibition of SARS-CoV-2 Mpro, *Brief. Bioinform.*, 2021, **22**, 1361–1377, DOI: [10.1093/bib/bbaa382](https://doi.org/10.1093/bib/bbaa382).

86 F. A. D. M. Opo, M. M. Rahman, F. Ahammad, I. Ahmed, M. A. Bhuiyan and A. M. Asiri, Structure Based Pharmacophore Modeling, Virtual Screening, Molecular Docking and ADMET Approaches for Identification of Natural Anti-Cancer Agents Targeting XIAP Protein, *Sci. Rep.*, 2021, **11**, 4049, DOI: [10.1038/s41598-021-83626-x](https://doi.org/10.1038/s41598-021-83626-x).

87 A. Kumar, M. Dutt, B. Dehury, G. Sganzerla Martinez, C. L. Swan, A. A. Kelvin, C. D. Richardson and D. J. Kelvin, Inhibition Potential of Natural Flavonoids against Selected Omicron (B.1.19) Mutations in the Spike Receptor Binding Domain of SARS-CoV-2: A Molecular Modeling Approach, *J. Biomol. Struct. Dyn.*, 2025, **43**, 1068–1082, DOI: [10.1080/07391102.2023.2291165](https://doi.org/10.1080/07391102.2023.2291165).

88 N. T. P. Ngidi, K. E. Machaba and N. N. Mhlongo, In Silico Drug Repurposing Approach: Investigation of Mycobacterium Tuberculosis FadD32 Targeted by FDA-Approved Drugs, *Molecules*, 2022, **27**(3), 668, DOI: [10.3390/molecules27030668](https://doi.org/10.3390/molecules27030668).

89 E. O. Erhiehie, C. P. Ihekwereme and E. E. Ilodigwe, Advances in Acute Toxicity Testing: Strengths, Weaknesses and Regulatory Acceptance, *Interdiscip. Toxicol.*, 2018, **11**, 5–12, DOI: [10.2478/intox-2018-0001](https://doi.org/10.2478/intox-2018-0001).

90 R. A. Lewis and D. Wood, Modern <scp>2D QSAR</Spc> for Drug Discovery, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 505–522, DOI: [10.1002/wcms.1187](https://doi.org/10.1002/wcms.1187).

91 F. M. M. Ahamed, S. Chinnam, M. Challa, G. Kariyanna, A. Kumer, S. Jadoun, A. Salawi, G. Al-Sehemi, U. Chakma, M. A. Al Mashud, *et al.*, Molecular Dynamics Simulation, QSAR, DFT, Molecular Docking, ADMET, and Synthesis of Ethyl 3-((5-Bromopyridin-2-Yl)Imino)Butanoate Analogues as Potential Inhibitors of SARS-CoV-2, *Polycycl. Aromat. Compd.*, 2024, **44**, 294–312, DOI: [10.1080/10406638.2023.2173618](https://doi.org/10.1080/10406638.2023.2173618).

92 J.-I. Suzuki, M. Isobe, R. Morishita, M. Aoki, S. Horie, Y. Okubo, Y. Kaneda, Y. Sawa, H. Matsuda, T. Ogihara, *et al.*, Prevention of Graft Coronary Arteriosclerosis by Antisense Cdk2 Kinase Oligonucleotide, *Nat. Med.*, 1997, **3**, 900–903, DOI: [10.1038/nm0897-900](https://doi.org/10.1038/nm0897-900).

93 R. Fagundes and L. K. Teixeira, Cyclin E/CDK2: DNA Replication, Replication Stress and Genomic Instability, *Front. Cell Dev. Biol.*, 2021, **9**, 774845, DOI: [10.3389/fcell.2021.774845](https://doi.org/10.3389/fcell.2021.774845).

94 S. Tadesse, A. T. Anshabo, N. Portman, E. Lim, W. Tilley, C. E. Caldon and S. Wang, Targeting CDK2 in Cancer: Challenges and Opportunities for Therapy, *Drug Discov. Today*, 2020, **25**, 406–413, DOI: [10.1016/j.drudis.2019.12.001](https://doi.org/10.1016/j.drudis.2019.12.001).

95 S. Akli, C. S. Van Pelt, T. Bui, L. Meijer and K. Keyomarsi, Cdk2 Is Required for Breast Cancer Mediated by the Low-Molecular-Weight Isoform of Cyclin E, *Cancer Res.*, 2011, **71**, 3377–3386, DOI: [10.1158/0008-5472.CAN-10-4086](https://doi.org/10.1158/0008-5472.CAN-10-4086).

96 V. E. Brown, S. L. Moore, M. Chen, N. House, P. Ramsden, H.-J. Wu, S. Ribich, A. R. Grassian and Y. J. Choi, CDK2 Regulates Collapsed Replication Fork Repair in CCNE1-Amplified Ovarian Cancer Cells *via* Homologous Recombination, *NAR Cancer*, 2023, **5**(3), zcad039, DOI: [10.1093/narcan/zcad039](https://doi.org/10.1093/narcan/zcad039).

97 Y. Dobashi, M. Shoji, S.-X. Jiang, M. Kobayashi, Y. Kawakubo and T. Kameya, Active Cyclin A-CDK2 Complex, a Possible Critical Factor for Cell Proliferation in Human Primary Lung Carcinomas, *Am. J. Pathol.*, 1998, **153**, 963–972, DOI: [10.1016/S0002-9440\(10\)65638-6](https://doi.org/10.1016/S0002-9440(10)65638-6).

98 A. P. Dommer, V. Kumarasamy, J. Wang, T. N. O'Connor, M. Roti, S. Mahan, K. McLean, E. S. Knudsen and A. K. Witkiewicz, Tumor Suppressors Condition Differential Responses to the Selective CDK2 Inhibitor BLU-222, *Cancer Res.*, 2025, **85**, 1310–1326, DOI: [10.1158/0008-5472.CAN-24-2244](https://doi.org/10.1158/0008-5472.CAN-24-2244).

99 V. Calleja, P. Leboucher and B. Larijani, Protein Activation Dynamics in Cells and Tumor Micro Arrays Assessed by Time Resolved Förster, in *Resonance Energy Transfer*, 2012, pp. 225–246.

100 P. Paul, A. K. Malakar and S. Chakraborty, The Significance of Gene Mutations across Eight Major Cancer Types, *Mutat. Res. Mutat. Res.*, 2019, **781**, 88–99, DOI: [10.1016/j.mrrev.2019.04.004](https://doi.org/10.1016/j.mrrev.2019.04.004).

101 N. Gutiérrez-Melo and D. Baumjohann, T Follicular Helper Cells in Cancer, *Trends Cancer*, 2023, **9**, 309–325, DOI: [10.1016/j.trecan.2022.12.007](https://doi.org/10.1016/j.trecan.2022.12.007).

102 O. E. Franco, A. K. Shaw, D. W. Strand and S. W. Hayward, Cancer Associated Fibroblasts in Cancer Pathogenesis, *Semin. Cell Dev. Biol.*, 2010, **21**, 33–39, DOI: [10.1016/j.semcd.2009.10.010](https://doi.org/10.1016/j.semcd.2009.10.010).

103 M. S. Hossen, A. Akter, M. Azmal, M. Rayhan, K. S. Islam, M. M. Islam, S. Ahmed and M. Abdullah-Al-Shoeb, Unveiling the Molecular Basis of Paracetamol-Induced Hepatotoxicity: Interaction of N-Acetyl-p-Benzoquinone Imine with Mitochondrial Succinate Dehydrogenase, *Biochem Biophys Rep.*, 2024, **38**, 101727, DOI: [10.1016/j.bbrep.2024.101727](https://doi.org/10.1016/j.bbrep.2024.101727).

104 S. Das, A. D. Talukdar, D. Nath and M. D. Choudhury, Discovery of Anticancer Therapeutics: Computational Chemistry and Artificial Intelligence-Assisted Approach, in *Computational Methods in Drug Discovery and Repurposing for Cancer Therapy*, Elsevier, 2023, pp. 19–41.



105 I. Aier, P. K. Varadwaj and U. Raj, Structural Insights into Conformational Stability of Both Wild-Type and Mutant EZH2 Receptor, *Sci. Rep.*, 2016, **6**, 34984, DOI: [10.1038/srep34984](https://doi.org/10.1038/srep34984).

106 Z. K. Bagewadi, T. M. Yunus Khan, B. Gangadharappa, A. Kamalapurkar, S. Mohamed Shamsudeen and D. A. Yaraguppi, Molecular Dynamics and Simulation Analysis against Superoxide Dismutase (SOD) Target of *Micrococcus luteus* with Secondary Metabolites from *Bacillus licheniformis* Recognized by Genome Mining Approach, *Saudi J. Biol. Sci.*, 2023, **30**(9), 103753, DOI: [10.1016/j.sjbs.2023.103753](https://doi.org/10.1016/j.sjbs.2023.103753).

107 M. B. Hasan, M. J. Rahman, R. Das, T. Akter, M. S. Hosen, U. M. P. Kona and M. Uzzaman, Physicochemical, Biological, and Toxicological Studies of Pyridine and Its Derivatives: An in-Silico Approach, *Discov. Chem.*, 2025, **2**, 70, DOI: [10.1007/s44371-025-00147-6](https://doi.org/10.1007/s44371-025-00147-6).

108 T. Lippert and M. Rarey, Fast Automated Placement of Polar Hydrogen Atoms in Protein–Ligand Complexes, *J. Cheminform.*, 2009, **1**, 13, DOI: [10.1186/1758-2946-1-13](https://doi.org/10.1186/1758-2946-1-13).

109 C. Gao, M. S. Park and H. A. Stern, Accounting for Ligand Conformational Restriction in Calculations of Protein–Ligand Binding Affinities, *Biophys. J.*, 2010, **98**, 901–910, DOI: [10.1016/j.bpj.2009.11.018](https://doi.org/10.1016/j.bpj.2009.11.018).

110 A. Stank, D. B. Kokh, J. C. Fuller and R. C. Wade, Protein Binding Pocket Dynamics, *Acc. Chem. Res.*, 2016, **49**, 809–815, DOI: [10.1021/acs.accounts.5b00516](https://doi.org/10.1021/acs.accounts.5b00516).

111 Y. Fu, J. Zhao and Z. Chen, Insights into the Molecular Mechanisms of Protein–Ligand Interactions by Molecular Docking and Molecular Dynamics Simulation: A Case of Oligopeptide Binding Protein, *Comput. Math. Methods Med.*, 2018, **2018**, 3502514, DOI: [10.1155/2018/3502514](https://doi.org/10.1155/2018/3502514).

112 X. Zhang, H. Perez-Sanchez and F. C. Lightstone, A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin, *Curr. Top. Med. Chem.*, 2017, **17**, 1631–1639, DOI: [10.2174/1568026616666161117112604](https://doi.org/10.2174/1568026616666161117112604).

113 C. C. David and D. J. Jacobs, in *Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins*, 2014, pp. 193–226.

114 S. Ahamad, D. Gupta and V. Kumar, Targeting SARS-CoV-2 Nucleocapsid Oligomerization: Insights from Molecular Docking and Molecular Dynamics Simulations, *J. Biomol. Struct. Dyn.*, 2022, **40**, 2430–2443, DOI: [10.1080/07391102.2020.1839563](https://doi.org/10.1080/07391102.2020.1839563).

115 G. D. Anderson and R. P. Saneto, Current Oral and Non-Oral Routes of Antiepileptic Drug Delivery, *Adv. Drug Deliv. Rev.*, 2012, **64**, 911–918, DOI: [10.1016/j.addr.2012.01.017](https://doi.org/10.1016/j.addr.2012.01.017).

116 X. Zhang, T. Liu, X. Fan and N. Ai, *In Silico* Modeling on ADME Properties of Natural Products: Classification Models for Blood–Brain Barrier Permeability, Its Application to Traditional Chinese Medicine and *In Vitro* Experimental Validation, *J. Mol. Graph. Model.*, 2017, **75**, 347–354, DOI: [10.1016/j.jmgm.2017.05.021](https://doi.org/10.1016/j.jmgm.2017.05.021).

117 J. Vrbanac and R. Sauter, ADME in Drug Discovery, in *A Comprehensive Guide to Toxicology in Nonclinical Drug Development*, Elsevier, 2017, pp. 39–67.

118 K. Roy, On Some Aspects of Validation of Predictive Quantitative Structure–Activity Relationship Models, *Expert Opin. Drug Discov.*, 2007, **2**, 1567–1577, DOI: [10.1517/17460441.2.12.1567](https://doi.org/10.1517/17460441.2.12.1567).

119 T. Balasubramaniyam, A. G. V. Sreekala, V. K. Nathan, K.-I. Oh, J. Z. Kubiak and S. Rampogu, *In Silico* Investigation on the Discovery of Synthesized Nucleoside-Based Antivirals against Monkeypox and SARS-CoV-2 Virus, *Silico Res. Biomed.*, 2025, **1**, 100051, DOI: [10.1016/j.insi.2025.100051](https://doi.org/10.1016/j.insi.2025.100051).

120 S. S. Swain, S. R. Singh, A. Sahoo, P. K. Panda, T. Hussain and S. Pati, Integrated Bioinformatics–Cheminformatics Approach toward Locating Pseudo-potential Antiviral Marine Alkaloids against <scp>SARS-CoV-2-Mpro</Scp>, *Proteins: Struct., Funct., Bioinf.*, 2022, **90**, 1617–1633, DOI: [10.1002/prot.26341](https://doi.org/10.1002/prot.26341).

121 V. P. Waman, N. Sen, M. Varadi, A. Daina, S. J. Wodak, V. Zoete, S. Velankar and C. Orengo, The Impact of Structural Bioinformatics Tools and Resources on SARS-CoV-2 Research and Therapeutic Strategies, *Brief. Bioinform.*, 2021, **22**, 742–768, DOI: [10.1093/bib/bbaa362](https://doi.org/10.1093/bib/bbaa362).

122 A. Ullah, F. A. Shahid, M. U. Haq, M. Tahir ul Qamar, M. Irfan, B. Shaker, S. Ahmad, F. Alrumaihi, K. S. Allemailem and A. Almatroudi, An Integrative Reverse Vaccinology, Immunoinformatic, Docking and Simulation Approaches towards Designing of Multi-Epitopes Based Vaccine against Monkeypox Virus, *J. Biomol. Struct. Dyn.*, 2023, **41**, 7821–7834, DOI: [10.1080/07391102.2022.2125441](https://doi.org/10.1080/07391102.2022.2125441).

