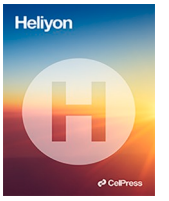
Heliyon 10 (2024) e40345

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Computational design of CDK1 inhibitors with enhanced target affinity and drug-likeness using deep-learning framework

Zuokun Lu [a,b,*], Jiayuan Han [a], Yibo Ji [a], Bingrui Li [a], Aili Zhang [a,**]

[a] Food and Pharmacy College, Xuchang University, Xuchang, 461000, Henan, China
[b] Key Laboratory of Biomarker-Based Rapid Detection Technology for Food Safety of Henan Province, Xuchang University, Xuchang, 461000, Henan, China

ARTICLE INFO

Keywords:
Cyclin dependent kinase 1 (CDK1)
Recurrent neural network
CDK1 inhibitors
Molecular docking
MD simulations

ABSTRACT

Cyclin Dependent Kinase 1 (CDK1) plays a crucial role in cell cycle regulation, and dysregulation of its activity has been implicated in various cancers. Although several CDK1 inhibitors are currently in clinical trials, none have yet been approved for therapeutic use. This research utilized deep learning techniques, specifically Recurrent Neural Networks with Long Short-Term Memory (LSTM), to generate potential CDK1 inhibitors. Molecular docking, evaluation of molecular properties, and molecular dynamics simulations were conducted to identify the most promising candidates. The results showed that the generated ligands exhibited substantial improvements in target affinity and drug-likeness. Molecular docking results showed that the generated ligands had an average binding affinity of -10.65 ± 0.877 kcal/mol towards CDK1. The Quantitative Estimate of Drug-likeness (QED) values for the generated ligands averaged 0.733 ± 0.10, significantly higher than the 0.547 ± 0.15 observed for known CDK1 inhibitors ($p < 0.001$). Molecular dynamics simulations further confirmed the stability and favorable interactions of the selected ligands with the CDK1 complex. The identification of novel CDK1 inhibitors with improved binding affinities and drug-likeness properties could potentially fill the gap in the ongoing development of CDK inhibitors. However, it is imperative to note that extensive experimental validation is required prior to advancing these generated ligands to subsequent stages of drug development.

1. Introduction

Cyclin Dependent Kinase 1 (CDK1) is a crucial protein involved in cell cycle regulation. Dysregulation of CDK1 activity has been linked to various types of cancer, including breast, lung, colon, and prostate cancer [1]. CDK1 belongs to the cyclin-dependent kinase (CDK) family, which plays a key role in the progression of the cell cycle. It controls the transition from G2 phase to mitosis by phosphorylating downstream target proteins involved in mitotic events such as spindle formation, chromosome condensation, and nuclear envelope breakdown [2]. Dysregulation of CDK1 activity can lead to abnormal cell cycle progression and genomic instability, which are common features in cancer cells. Overexpression or hyperactivation of CDK1 has been observed in various types of cancer,

* Corresponding author. Food and Pharmacy College, Xuchang University, Xuchang, 461000, Henan, China.
** Corresponding author. Key Laboratory of Biomarker-Based Rapid Detection Technology for Food Safety of Henan Province, Xuchang University, Xuchang, 461000, Henan, China.
E-mail addresses: Luzuokun@xcu.edu.cn (Z. Lu), Zhangaili@xcu.edu.cn (A. Zhang).

https://doi.org/10.1016/j.heliyon.2024.e40345
Received 17 February 2024; Received in revised form 20 September 2024; Accepted 11 November 2024
Available online 14 November 2024

promoting uncontrolled cell division and tumor growth [3].

Currently, several CDK inhibitors have been approved by regulatory authorities and are available on the market. These include palbociclib, ribociclib, and abemaciclib, which are used for the treatment of specific types of advanced or metastatic breast cancer [4]. In addition to their established use in breast cancer, ongoing clinical research is exploring the potential of CDK inhibitors in other cancer types, including lung, ovarian, and hematological malignancies [5–8]. Furthermore, research into specific CDK inhibitors targeting CDK7, CDK9, and CDK12 is advancing, offering new possibilities for treating a diverse range of cancers [9]. These inhibitors exhibit unique mechanisms and therapeutic targets, thereby expanding the therapeutic arsenal available for cancer treatment. However, it is important to note that targeting CDK alone may not be sufficient for effective cancer treatment. Combination therapies, such as CDK inhibitors with other targeted agents or standard chemotherapy, may be necessary to overcome resistance mechanisms and enhance therapeutic efficacy [10]. Targeting CDK1 with selective inhibitors offers a potential therapeutic approach for the treatment of these malignancies [11]. Selective inhibitors of CDK1 have shown promising anticancer activity in preclinical models and clinical trials. By inhibiting CDK1 activity, these inhibitors can induce cell cycle arrest and prevent cancer cells from proliferating [12].

Deep-learning frameworks have gained significant attention in the field of drug discovery due to their ability to generate novel drug-like molecules [13]. These frameworks utilize various generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs), to learn patterns and generate new molecules with desired properties [14–16]. He et al. (2021) demonstrated the application of machine learning in predicting active molecules against breast cancer cells [17], while Pang et al. (2023) conducted a SAR and QSAR study on CDK4 inhibitors using machine learning methods [18]. Zhang et al. (2022) identified novel CDK9 inhibitors using a combination of ligand- and structure-based approaches, demonstrating the potential of these inhibitors for the treatment of leukemia [19]. Similarly, Kaveh et al. (2024) derived general structure-activity/selectivity relationship patterns for different subfamilies of CDK inhibitors using machine learning methods [20]. These studies highlight the potential of computational approaches in accelerating the development of novel therapeutic agents for cancer treatment, which can be further validated through experimental approaches [13,21–25].

In this research, an in-depth study employed a deep learning approach using Recurrent Neural Networks (RNNs) for generating new potential ligands capable of acting as inhibitors of CDK1. To overcome the issue of the vanishing gradient problem encountered in vanilla RNNs, a Long Short-Term Memory (LSTM) network was specifically chosen [26]. By implementing LSTM, the model can capture long-range dependencies and mitigate the vanishing gradient problem, thus enhancing the accuracy and effectiveness of the generated ligands [23,27,28].

Molecular docking was performed through virtual screening to identify the top candidates with high affinity against CDK1. Subsequently, an evaluation of the ligands' molecular properties was conducted to study the biological activity and pharmacokinetic profile of the generated ligands. To assess the stability and binding interactions of the ligand-protease complex, molecular dynamics (MD) simulations were carried out. Furthermore, a comparative analysis was conducted between the properties of the newly generated ligands and drugs currently undergoing clinical trials. By leveraging AI and computational methods, this research seeks to contribute to the advancement of CDK inhibitor development, providing novel compounds with enhanced properties that could potentially improve therapeutic outcomes in cancer treatment. However, it is important to note that the generated molecules should undergo rigorous experimental validation to assess their biological activity, toxicity, and pharmacokinetic properties before progressing into drug development stages.

2. Materials and methods

2.1. Data preparation and de novo ligands generation

The transfer learning approach applied in this case involved utilizing a pre-trained model called LSTM_Chem [29]. This innovative approach involves the use of generative recurrent neural networks (RNNs) containing long short-term memory (LSTM) cells for de novo molecular design. By accurately capturing the syntax of molecular representations in SMILEs strings, this computational model has the capability to generate molecules that resemble the training data.

In the first stage, the model was trained on a dataset to learn the latent features of chemical molecules. This dataset contained a diverse range of ligands obtained from ChEMBL [30] (https://www.ebi.ac.uk/chembl). We removed duplicates, salts, and stereochemical information, resulting in a final dataset of 459323 unique ligands. The SMILES strings were constrained to have lengths between 34 and 74 tokens to avoid the vanishing gradient problem. In the second stage, a separate dataset was used to fine-tune the model specifically for generating ligands with chemical features of inhibitors for CDK1. The dataset for fine-tuning was obtained from Chembl33, with cyclin-dependent kinase 1 (CDK1) as the target (Supplementary materials: Table S1). Compounds with an IC50 value of less than 10 μM were selected. Unlike the training dataset, there was no restriction on the string length of the SMILES in the fine-tuning dataset.

The dataset of CDK1-targeting small molecules was used to fine-tuning the LSTM-based generative deep learning model called "lstm_chem". The model was trained for 22 epochs on the training dataset, followed by 12 epochs of fine-tuning on the fine-tuning dataset. Using the trained lstm_chem model, 5000 new small molecule candidates were generated. The generated SMILES representations were then subjected to several checks, including uniqueness, validity (ensuring compliance with chemical rules), and originality (determining if they differed significantly from the training dataset).

2.2. Technical implementation

In order to evaluate the properties of the generated molecules, such as physicochemical descriptors, pharmacokinetic properties, drug-like nature, and medicinal chemistry friendliness, we employed the Quantitative Estimate of Drug-Likeness (QED) [31] in python module and the free online website SwissADME (http://www.swissadme.ch/) [32].

The molecular properties such as molecular weight, logP, number of hydrogen bond donors, and number of rotatable bonds, along with the uniqueness, validity, and originality of SMILES strings, were computed using RDKit (https://www.rdkit.org/). The statistical data gathering was performed with the assistance of scikit-learn libraries (www.scikit-learn.org). Additionally, the graphs were generated using matplotlib (www.matplotlib.org) and seaborn (www.seaborn.pydata.org).

2.3. Molecular docking

The crystal structure of the human CDK1/CyclinB1/CKS2 with inhibitor was retrieved from the RCSB Protein Data Bank (PDB ID: 4Y72). AutoDock Vina v1.2.3 was used for virtual screening [33,34]. The ligands generated were converted from SMILES to pdbqt format using OpenBabel version 3.1.0 command line [35]. The native ligand (LZ9 inhibitor) and water molecules were removed to prepare the target using pymol. A grid box was defined with its center points and dimensions set to target the active site of the main protease protein at x = 28, y = −72, z = 186 (Å), and X = 30, Y = 30, Z = 30 (Å) for its dimensions. The binding affinities of the compounds were calculated and ranked based on their lowest values, indicating the best binding affinities.

2.4. Refinement and iteration

Molecules with low affinity values (lower than the average) in the docking results and high QED values (higher than the average) were chosen. These selected molecules were used to refine the lstm_chem model, generating a set of 5000 new molecules. The newly generated molecules underwent molecular docking and property calculations once again. This entire fine-tuning process was repeated for two cycles.

2.5. Molecular dynamics simulations and analysis

Four compounds with binding affinity lower than −12.0 kcal/mol and QED value larger than 0.65 were selected for MD simulation. A known CDK1 inhibitor, LZ9, was used as a control. The MD simulation was conducted using the GROMACS (version 2023.3) on a Linux host (Ubuntu 20.04) with GPU (GEFORCE RTX 3060) support. The Amber99sb force field was employed to generate force field parameters and define atom types. The ACPYPE [36] command line was utilized to create topology and coordinate files for the selected ligands. The protein complexes were solvated in a triclinic box with TIP3P water molecules, and the system was neutralized by adding 0.15 M NaCl.

The system underwent minimization using the steepest descent algorithm, followed by equilibration using NVT and NPT ensembles with a duration of 100 ps each. A v-rescale Berendsen thermostat and a C-rescale barostat were applied for temperature and pressure control, respectively. The systems were gradually heated from 0 to 300 K during equilibration, and the pressure was set to 1 atm for the NVT and NPT ensembles.

After the equilibration, the molecular dynamics simulation was conducted for a duration of 100 ns, with each time step set to 2 fs in the periodic boundary conditions. During the simulation, trajectories were saved for further analysis. The root mean square deviations

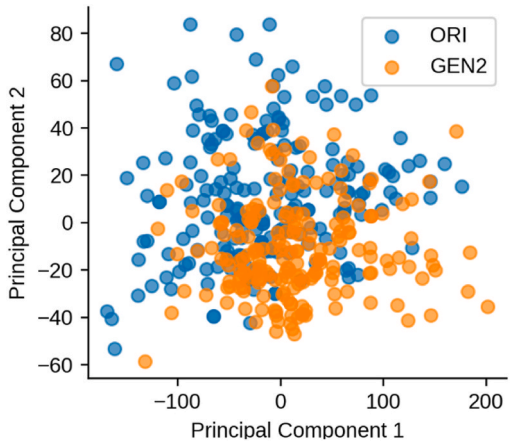


Fig. 1. The results of a principal component analysis performed on the datasets consisting of 200 randomly selected molecules from the second-generation ligands (GEN2) and known CDK1 inhibitors (ORI), based on their physiochemical features. Additionally, further statistical analysis reveals significant differences in the mean values among various groups within the first two principal components (all p-values <0.001).

(RMSD) and residue root mean square fluctuations (RMSF) of each snapshot were calculated to determine the equilibrium time and stability of the protein backbone atoms. The ligand-protein conformational analysis was performed using the Pymol graphical software. The interactions between receptor and ligands were analyzed by using PLIP server [37]. Furthermore, Radius of gyration (Rg) and inter-molecular hydrogen bond were calculated using the gyrate module and the h-bond modules in GROMACS, respectively. The protein-ligand interaction energy was calculated using the energy module of GROMACS to extract the short-range Coulombic (Coul-SR) and Lennard-Jones (LJ-SR) terms. The total interaction energy, after propagating the error according to the standard formula for the addition of two quantities, was then determined. By performing these calculations, we can gain valuable insights into the dynamics, stability, and interactions of the protein-ligand complex throughout the molecular dynamics simulation.

3. Results and discussions

3.1. Generated ligands

The LSTM_Chem network underwent a fine-tuning process to generate new molecular compounds. Each round of fine-tuning yielded 5000 molecules, which were subsequently checked for validity, novelty and uniqueness. After careful examination, a total of 1349 molecules were deemed eligible for further analysis (Supplementary materials: Table S2).

To assess the similarity between the generated molecules and the original molecules, we conducted a principal component analysis (PCA) of the physiochemical features (Fig. 1). The results indicate that there are significant differences between the two datasets as revealed by the first two principal components. Further statistical analysis demonstrates significant differences in the mean values between different groups within the first two principal components (p-values all <0.001).

We applied a clustering analysis method to categorize the generated ligands into five distinct groups based on their Tanimoto similarity scores. From each group, we selected one molecule and calculated its structural similarity to known CDK1 inhibitors (Fig. 2). The findings indicate that the chosen compounds exhibit relatively low structural similarity to CDK1 inhibitors, ranging from a minimum of 0.35 to a maximum of 0.66. The results suggest that after two rounds of fine-tuning, the generated ligands underwent significant alterations in their chemical space compared to the original molecules.

3.2. Properties of generated ligands

3.2.1. QED and Lipinski's rule

QED (quantitative estimate of drug-likeness) is a measure of drug-likeness based on the concept of desirability. The QED values span from zero, indicating unfavorable properties in all aspects, to one, denoting favorable properties in all aspects. By utilizing the

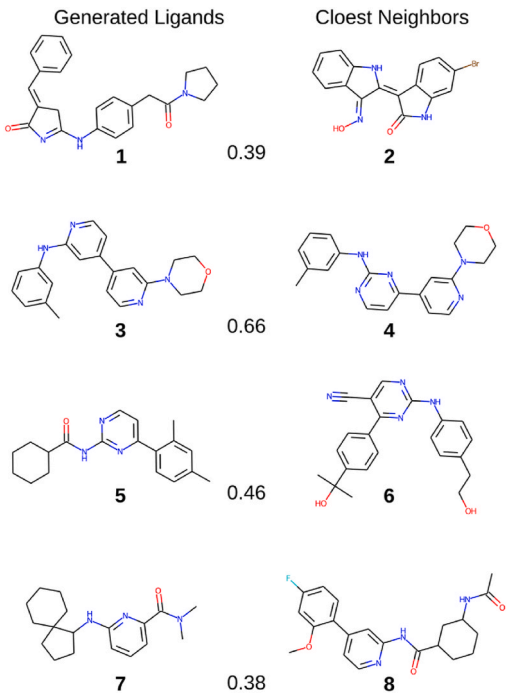


Fig. 2. The generated ligands were divided into four groups based on Tanimoto similarity using Butina cluster method. From each group, one molecule (1, 3, 5, 7) was randomly selected and listed on the left side (Generated ligands). The corresponding molecules with the highest Tanimoto similarity in CDK1 inhibitors were listed on the right side of each selected compound.

desirability approach, it is possible to generate functions that characterize various sets of compounds according to specific requirements [31]. After each round of fine-tuning, the generated molecules demonstrated a substantial enhancement in terms of QED (Quantitative Estimate of Drug-likeness) as shown in Fig. 3. The average QED values for the second-generation ligands and the original CDK1 inhibitors were 0.733 ± 0.10 and 0.547 ± 0.15, respectively. This improvement was statistically significant ($p < 0.001$, *t*-test), suggesting that the second generation of molecules exhibits notably improved drug-like properties.

Lipinski's Rule of 5 is a widely used guideline in the field of drug discovery and medicinal chemistry to assess the drug-likeness of a compound. It was developed by Christopher Lipinski, a scientist at Pfizer, and has been adopted as a rule of thumb in the pharmaceutical industry. According to Lipinski's Rule of 5 [38], a compound is more likely to possess favorable drug-like properties if it meets the following criteria: a molecular weight (MW) less than or equal to 500, a partition coefficient (logP) less than or equal to 5, no more than 5 hydrogen bond donors (HBD), no more than 10 hydrogen bond acceptors (HBA), and no more than 10 rotatable bonds (NRB). It is important to note that while Lipinski's Rule of 5 provides a useful framework for evaluating drug-likeness, it is not an absolute rule and there are exceptions.

For CDK1 inhibitors, the first round of fine-tuned data, and second round of fine-tuned data, the proportions of compounds that pass Lipinski's Rule are 84.9 %, 89.1 %, and 88.58 % respectively. This represents a significant improvement in the number of molecules meeting Lipinski's Rule. It can be observed that the average molecular weight (MW) for CDK1 inhibitors gradually decreases from 485.3 to 368.8 g/mol in the three datasets (Fig. 4A), while the optimal values for MW should be between 150 and 500 g/mol [39].

In the three datasets, none of them has more than 5 hydrogen bond donors (Fig. 4B), while 92.5 %, 99.9 %, and 100.0 % have no more than 10 hydrogen bond acceptors respectively. The octanol-water partition coefficient (logP) is a measure of molecular lipophilicity, which is crucial for drug absorption, bioavailability, hydrophobic drug-receptor interactions, as well as metabolism and molecule toxicity [40]. Among the three groups of ligands, the average logP values were all below 5.0, with the first-generation ligands and second-generation ligands showing a more significant decrease compared to CDK1 inhibitors (p-value <0.001).

3.2.2. Docking analysis

The ligands generated in our study exhibited a more robust binding affinity towards the target CDK1 complex (4Y72), as demonstrated by the results of molecular docking illustrated in Fig. 5. On average, the generated ligands displayed a binding affinity of −10.65 ± 0.877 kcal/mol towards CDK1. The ligand with the highest binding affinity reached −13.1 kcal/mol, while the lowest was −6.5 kcal/mol. Among the 821 CDK1 inhibitors, the average binding affinity was −10.16 ± 2.18 kcal/mol, where Entrectinib exhibited the lowest binding affinity at −13.5 kcal/mol and Temsirolimus had the highest at −6.0 kcal/mol.

In the selection process, we identified four molecules out of the 1209 generated molecules that exhibited binding affinity values lower than −12.0 kcal/mol and QED values higher than 6.5. These four ligands were further investigated for their interactions with the CDK1 complex. Fig. 6 provides a visual representation of the interaction between these ligands and CDK1, with the amino acid residues involved in the ligand-receptor interaction clearly indicated. Among the ligands, it was observed that most of them interacted with specific receptor residues, namely Ala 31, Glu 12, Glu 51, and Lys 33, through hydrogen bonding. An interesting finding is that Ligand1 displayed the lowest binding affinity value of −12.7 kcal/mol along with a QED value of 0.51. In addition to hydrogen bonding interactions with Lys33 and Glu51, this ligand also demonstrated hydrophobic interactions with Leu 135, Ala 31, Val 64, and Gln 132.

3.2.3. Evaluation of pharmacokinetic properties

Pharmacokinetic properties refer to the processes by which a drug is absorbed, distributed, metabolized, and excreted by the body. These properties help determine how the drug behaves within the body and how it is eventually eliminated. Understanding the

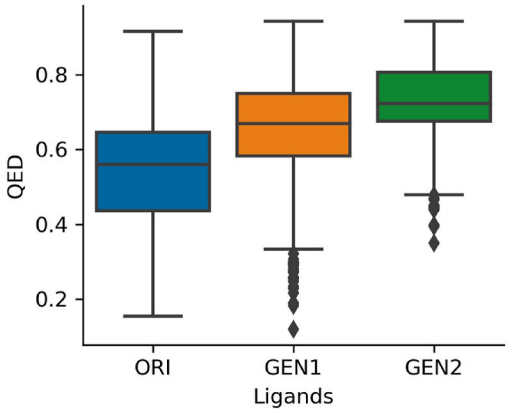


Fig. 3. The graphical representation of the QED (Quantitative Estimate of Drug-likeness) statistics distribution for CDK1 inhibitors (ORI), first-generation ligands (GEN1), and second-generation ligands (GEN2) is presented through box-plots, with the medians indicated by solid lines in the center of each box. The average QED values for these three sets of data are as follows: 0.54 for ORI, 0.66 for GEN1, and 0.73 for GEN2. Notably, the results demonstrate a significant enhancement in QED values between the ORI and GEN1 groups, as well as between the GEN1 and GEN2 groups (ORI vs. GEN1 and GEN1 vs. GEN2, with all p-values <0.001 according to *t*-test analysis).

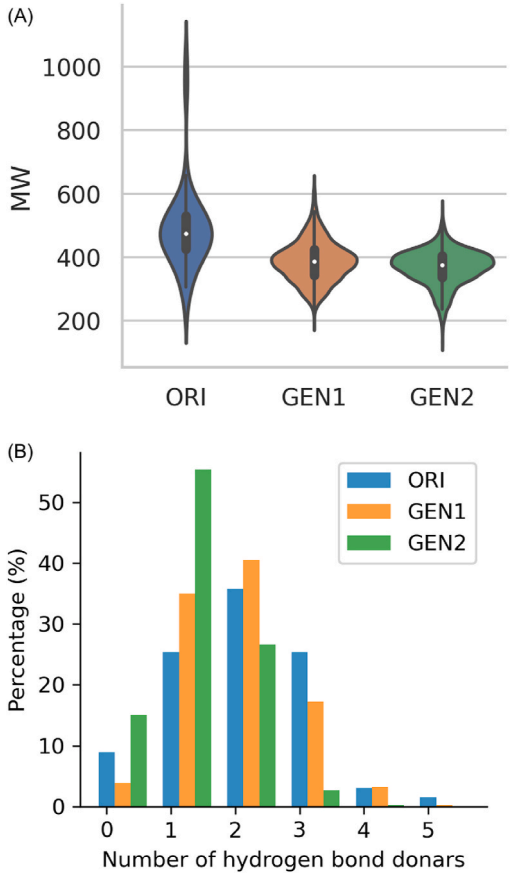


Fig. 4. The distribution of molecular weight (MW) and the number of hydrogen bonds in the reference and generated ligands. (A) The graph illustrates a gradual decrease in MW across the three datasets. The average MW values are reported as 486.7, 387.5, and 368.8, respectively. Significantly different MWs are observed between the ORI and GEN1 datasets, as well as between the GEN1 and GEN2 datasets ($p < 0.001$, *t*-test). (B) The bar chart reveals that the number of hydrogen bond donors of the second-generation ligands tends to concentrate around one after two rounds of selection (>50 % of compounds), and no compound has more than four hydrogen bond donors.

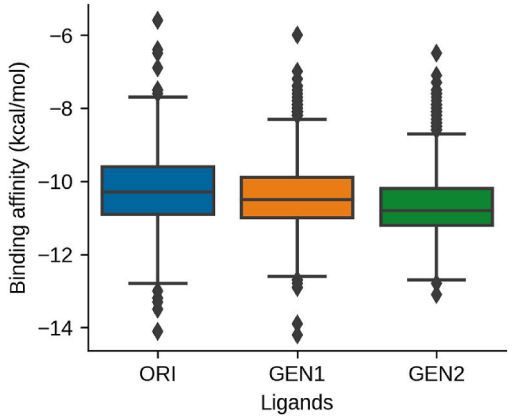


Fig. 5. The distribution of binding affinity towards CDK1 is examined among reference ligands (ORI), first-generation (GEN1) ligands, and second-generation (GEN2) ligands. The average binding affinities are reported as −10.2, −10.4, and −10.7 kcal/mol, respectively. Statistical analysis using the Kruskal-Wallis test reveals significant differences in binding affinity among the three datasets ($p < 0.0001$).

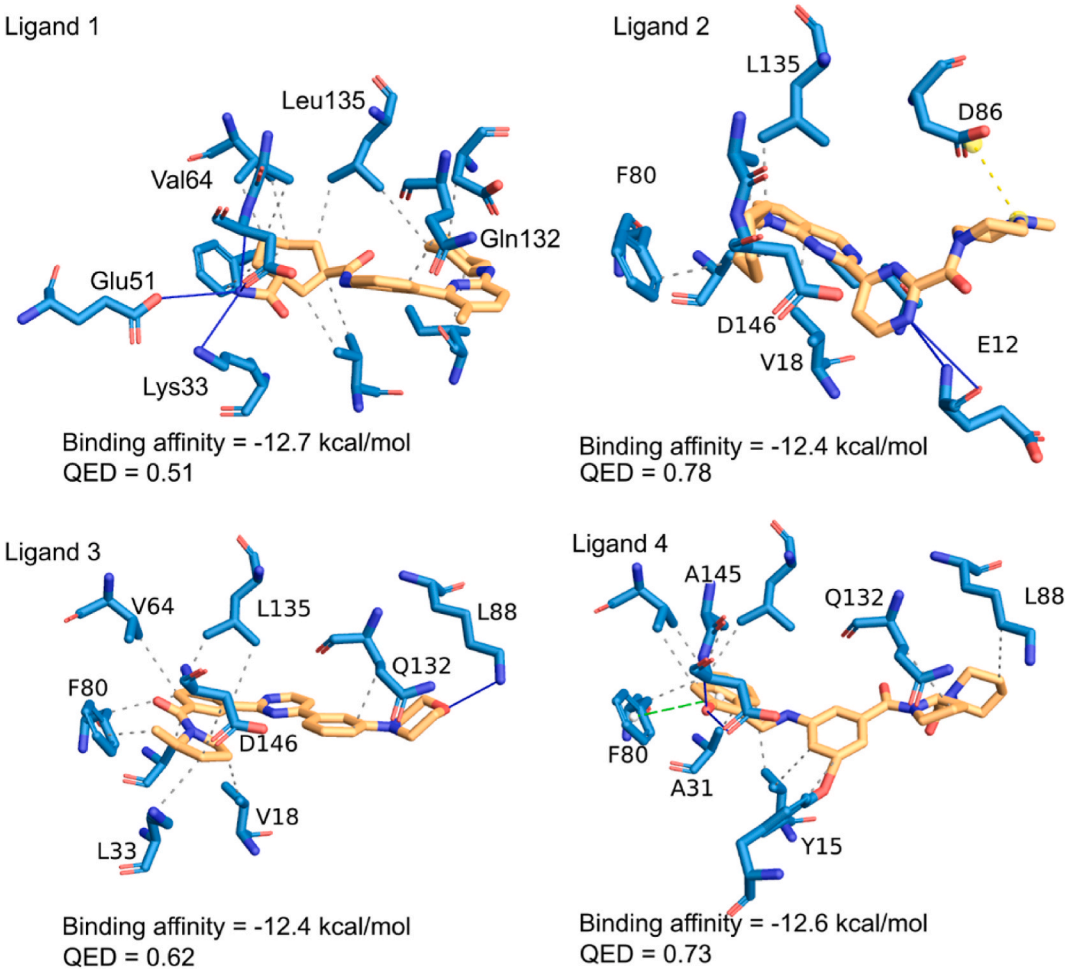


Fig. 6. 3D visualization of docked ligands interacting with CDK1, where hydrogen bonds are represented by blue solid lines, hydrophobic interactions by gray dashed lines, and π stacking interactions by green dashed lines. The figures were generated using the PLIP server [37].

pharmacokinetic properties of a drug is vital for optimizing its therapeutic use, predicting potential drug-drug interactions, and ensuring patient safety [41,42].

A total of 200 molecules were randomly selected from both the reference and generated ligands for the assessment of pharmacokinetic properties using the SwissADME server [39,43]. The proportion of compounds with high gastrointestinal (GI) absorption was found to be 81 % for CDK1 inhibitors, 0.88 for GEN1 inhibitors, and 0.99 for GEN2 inhibitors. The blood-brain barrier (BBB) is a selectively permeable membrane that separates the circulating blood from the brain tissue. BBB permeability refers to the ability of a compound to cross this barrier and enter the brain. Assessing BBB permeability provides insights into the efficacy of inhibitors for

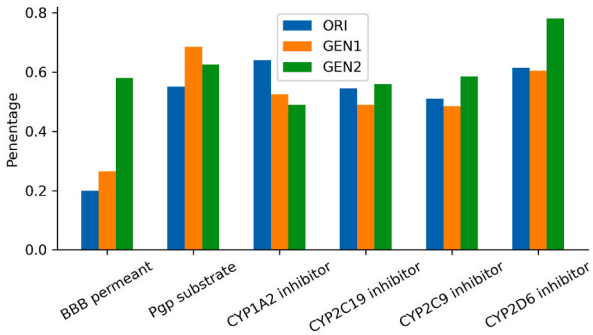


Fig. 7. Prediction of the pharmacokinetics of known CDK1 inhibitors (ORI), first-generation ligands (GEN1), and second-generation ligands (GEN2).

central nervous system (CNS) conditions. For the three sets of inhibitors analyzed (CDK1, GEN1, and GEN2 inhibitors), the results indicate that GEN2 inhibitors have the highest proportion (58 %) of compounds that can permeate the BBB, followed by GEN1 inhibitors (26.5 %) and CDK1 inhibitors (20 %). Additionally, other properties such as P-glycoprotein substrate activity and inhibitory effects on cytochrome P450 enzymes (CYP1A2, CYP2C19, CYP2C9, and CYP2D6) were also evaluated (Fig. 7).

4. Molecular dynamics simulation

To determine the stability and intermolecular interaction profiles over time, molecular dynamics simulations were employed to investigate docked complexes. Various results such as Root-Mean-Square Deviation (RMSD), Root-Mean-Square Fluctuation (RMSF), number of hydrogen bonds and radius of gyration (Rg) were obtained from the molecular dynamics simulations.

4.1. RMSD

The root mean square deviation (RMSD) was calculated between the backbone of protein atoms and the ligand for each complex in order to assess system stability and quantify conformational changes in both the ligands and protein of the docked complexes. The RMSD plot for all simulation systems is shown in Fig. 8.

Based on the analysis, the RMSD values of the complexes for all systems ranged between 0.3 and 0.5 nm. After approximately 30 ns into the simulation, the RMSD values of all systems reached a plateau with minimal fluctuation. This indicates that the complex remained stable throughout the simulation. In contrast, the average root mean square deviation (RMSD) value for Ligand1, Ligand2 and Ligand3 ranges from 0.15 to 0.20 nm. The root-mean-square deviation (RMSD) of Ligand4 exhibited a periodic fluctuation trend, which gradually stabilized after 80 ns. This indicates that the selected compounds except Ligand4 were successfully stabilized and properly fit into the active site of the protein.

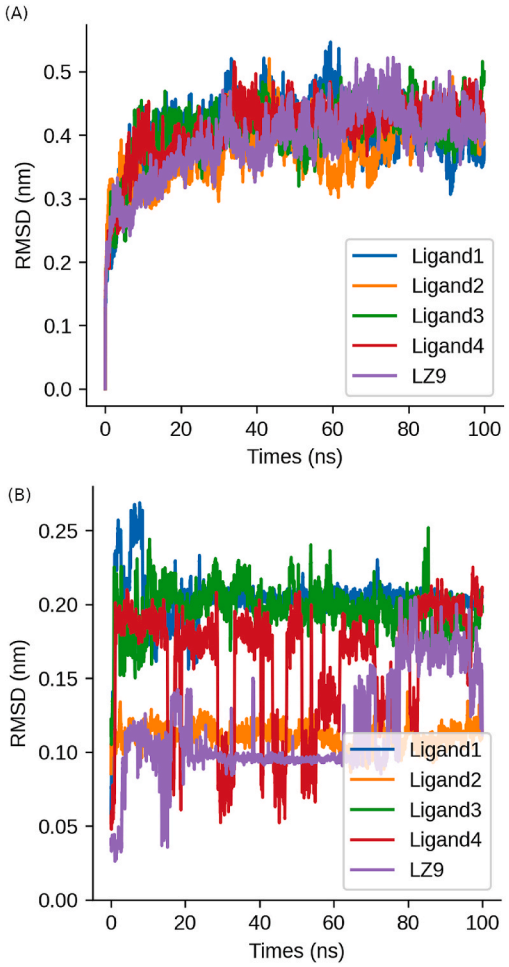


Fig. 8. Analysis of MD trajectories from simulations of CDK1 complex. In each case, MD simulation of 100 ns was performed. (A) RMSD of the protein backbone in complex with the four selected ligands and the known CDK1 inhibitors LZ9. (B) RMSD of the five ligands.

4.2. RMSF

The root mean square fluctuation (RMSF) is commonly employed to evaluate the dynamical behavior of amino acid residues within a protein. Analysis of RMSF reveals that the greatest fluctuations primarily occur in the loop regions. The RMSF of Cα atoms was computed for all complexes, with average values ranging from 0.10 to 0.30 nm (excluding the loop region). The CDK1 complex comprises three proteins (CDK1/CyclinB1/CKS2), wherein the RMSF values of CyclinB1 and CKS2 display minor alterations across the various complexes. CDK1 serves as the target for the inhibitor, with its interaction sites predominantly located near amino acid residues 31, 81, and 135. Within the CDK1-Ligand3 complex, residues in the vicinity of 31, 81, and 135 exhibit notably higher RMSF values compared to other complexes. These specific residues represent the main binding region with the ligand. The graphical representation highlights substantial fluctuations in the RMSF values at these sites (Fig. 9). Overall, the relatively low RMSF values indicate that the protein-ligand complexes are relatively stable while displaying a moderate level of flexibility. The proximity of the complexes to the apoprotein further emphasizes their stability, suggesting that the ligands can bind to the protein without inducing significant conformational changes.

4.3. Radius of gyration

The compactness and stability of protein structures were measured using the radius of gyration (Rg), which represents the mass-weighted root mean square distance of the atomic distribution from their mutual center of mass [44]. Smaller Rg values indicate a more compact and stable protein structure.

In this case, the Rg values of the protein-ligand complexes were analyzed, and it was found that all the complexes displayed similar Rg values ranging from 2.90 to 3.08 nm throughout the simulations (Fig. 10). This consistency suggests that the complexes are stable and exhibit appropriate interactions between the protein and ligand. These observations suggest a stronger binding between the protein and ligand, as they maintain a stable conformation throughout the simulation. Ligands3 complex exhibited the most stable compact folding, with an average radius of gyration (Rg) value of 2.96 nm (ranging from 1.99 to 3.08 nm) and a standard deviation (SD) of 0.02 nm. In comparison, the Ligand1 complex had an average Rg value of 3.02 nm (ranging from 2.98 to 3.12 nm) with an SD value of 0.02 nm.

4.4. Hydrogen bonds and interaction energy

Based on the analysis of hydrogen bonding patterns during the 100 ns MD simulations, it was observed that the complex formed with Ligand1 exhibited a higher number of hydrogen bonds compared to the other complexes. In the interaction between Ligand1 and the CDK1 protein, a single hydrogen bond was consistently observed, occurring 82.5 % of the time out of the analyzed instances (Fig. 11). This suggests that there is a stable and specific interaction between the ligand and the protein [45].

In evaluating the binding interactions between the generated ligands and CDK1, it is crucial to consider not only hydrogen bonding but also other non-covalent interactions such as hydrophobic interactions, van der Waals forces, and electrostatic interactions, which collectively contribute to the overall binding strength [46]. Ligand LZ9 (a potent CDK1/CDK2 ATP-competitive inhibitor) exhibited the highest number of hydrogen bonds during molecular dynamics simulations, suggesting strong and stable interactions. To gain a more comprehensive understanding, we analyzed the total interaction energies for each ligand, which integrate various types of non-covalent interactions. The results (Fig. 11) showed Ligand 3 had the highest interaction energy (-234.52 ± 6.93 kJ/mol), indicating strong binding despite having an average number of hydrogen bonds less than 1 during the molecular dynamics simulations. This suggests that other non-covalent interactions, such as hydrophobic and van der Waals forces, play a significant role in its binding stability and affinity. Ligand 1, despite having a relative stable hydrogen bond, showed a lower interaction energy compared to Ligand 3, highlighting the importance of considering a broader range of intermolecular forces.

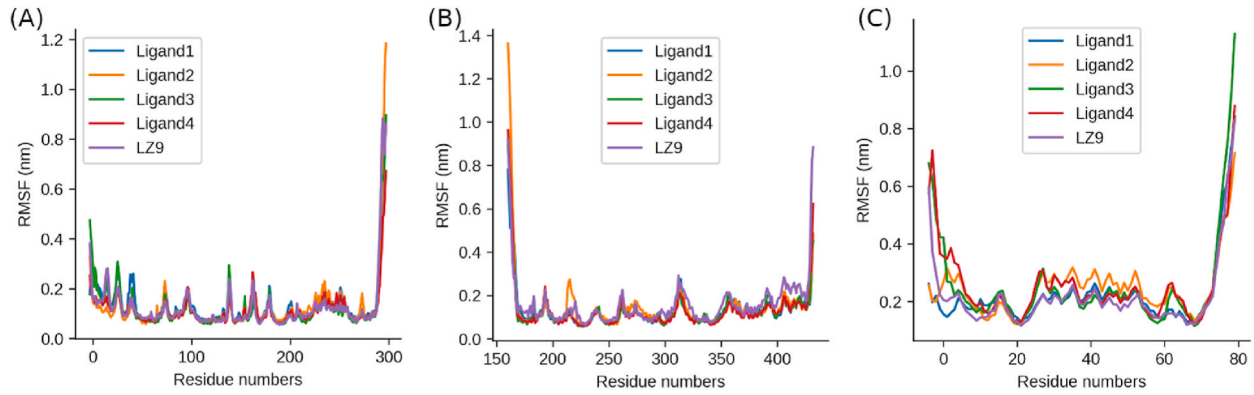


Fig. 9. Root-mean-square fluctuation (RMSF) analysis was performed on the Cα atoms of residues belonging to the three chains (A: CDK1, B: CyclinB1, C: CKS2) in the CDK1 complex.

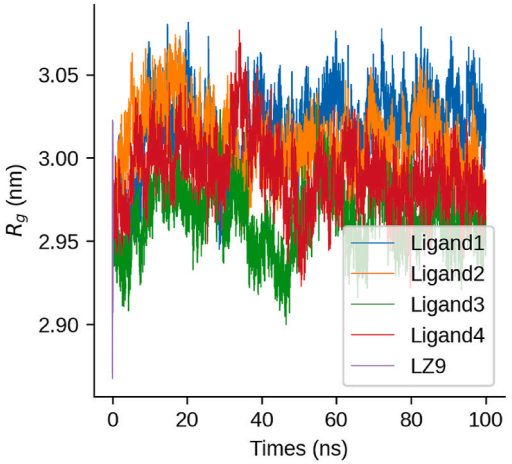


Fig. 10. Radius of gyration (R_g)of the CDK1 complex structure.

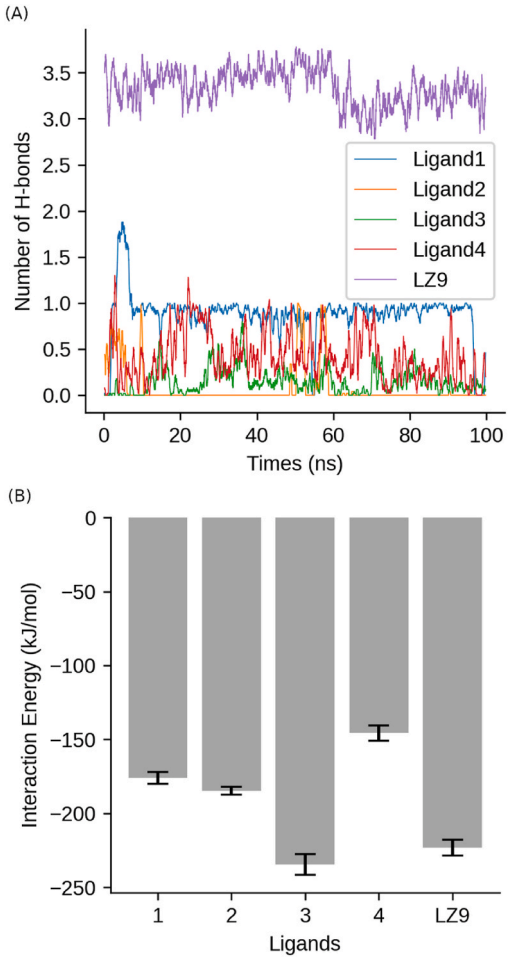


Fig. 11. Protein-Ligand Hydrogen Bonds and Interaction Energy Analysis. (A) The number of hydrogen bonds between CDK1 and the selected ligands is depicted, with a 50-ps running average plotted for improved clarity. In the 100 ns trajectory simulation of the Ligand1-CDK1 complex, the number of hydrogen bonds remains relatively stable, with the majority of the time exhibiting one hydrogen bond. Notably, Ligand LZ9 maintains an average hydrogen bond count of 3.3 throughout the simulation period. (B)The Ligand3-CDK1 complex demonstrates the highest stability throughout the simulation despite having an average number of hydrogen bonds less than 1. This suggests that other non-covalent interactions, such as hydrophobic and van der Waals forces, contribute significantly to the overall binding stability and affinity.

5. Conclusion

This study presents a novel computational framework combining deep learning with molecular docking and dynamics to generate potential CDK1 inhibitors. The most significant finding is the identification of CDK1 inhibitors that not only demonstrate enhanced target affinity but also possess favorable drug-likeness profiles. Among the four molecules subjected to MD simulations, Ligand 3 stands out with the highest interaction energy, despite forming fewer hydrogen bonds compared to traditional inhibitors like ligand LZ9. This underscores the potential to identify CDK1 inhibitors that rely on a broader spectrum of non-covalent interactions, thereby offering a new direction for drug development in cancer therapy. Future work should involve experimental validation to confirm the efficacy and safety of these compounds in biological systems.

CRediT authorship contribution statement

Zuokun Lu: Writing – review & editing, Writing – original draft, Visualization, Investigation, Funding acquisition, Formal analysis. **Jiayuan Han:** Investigation, Formal analysis. **Yibo Ji:** Investigation, Formal analysis. **Bingrui Li:** Investigation, Formal analysis. **Aili Zhang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of generative AI in the writing process

During the preparation of this work the authors used GPT-3 from openAI in order to ensure that the work is written in correct English. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work was supported by Natural Science Foundation of Henan (grant numbers 222300420278), General research projects of Xuchang University (grant numbers 2024YB017).

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e40345.

References

[1] M. Malumbres, M. Barbacid, Cell cycle, CDKs and cancer: a changing paradigm, Nat. Rev. Cancer 9 (2009) 153–166, https://doi.org/10.1038/nrc2602.
[2] H. Hochegger, S. Takeda, T. Hunt, Cyclin-dependent kinases and cell-cycle transitions: does one fit all? Nat. Rev. Mol. Cell Biol. 9 (2008) 910–916, https://doi.org/10.1038/nrm2510.
[3] Q. Wang, A.M. Bode, T. Zhang, Targeting CDK1 in cancer: mechanisms and implications, npj Precis. Oncol. 7 (2023) 58, https://doi.org/10.1038/s41698-023-00407-7.
[4] R.S. Finn, M. Martin, H.S. Rugo, S. Jones, S. Im, K. Gelmon, N. Harbeck, O.N. Lipatov, J.M. Walshe, S. Moulder, E. Gauthier, D.R. Lu, S. Randolph, V. Diéras, D. J. Slamon, Palbociclib and letrozole in advanced breast cancer, N. Engl. J. Med. 375 (2016) 1925–1936, https://doi.org/10.1056/NEJMoa1607303.
[5] E. Panagiotou, G. Gomatou, I.P. Trontzas, N. Syrigos, E. Kotteas, Cyclin-dependent kinase (CDK) inhibitors in solid tumors: a review of clinical trials, Clin. Transl. Oncol. Off. Publ. Fed. Spanish Oncol. Soc. Natl. Cancer Inst. Mex. 24 (2022) 161–192, https://doi.org/10.1007/s12094-021-02688-5.
[6] B. O'Leary, R.S. Finn, N.C. Turner, Treating cancer with selective CDK4/6 inhibitors, Nat. Rev. Clin. Oncol. 13 (2016) 417–430, https://doi.org/10.1038/nrclinonc.2016.26.
[7] A.B. Heptinstall, I. Adiyasa, C. Cano, I.R. Hardcastle, Recent advances in CDK inhibitors for cancer therapy, Future Med. Chem. 10 (2018) 1369–1388, https://doi.org/10.4155/fmc-2017-0246.
[8] B. Goel, N. Tripathi, N. Bhardwaj, S.K. Jain, Small molecule CDK inhibitors for the therapeutic management of cancer, Curr. Top. Med. Chem. 20 (2020) 1535–1563, https://doi.org/10.2174/1568026620666200516152756.
[9] P. Mounika, B. Gurupadayya, H.Y. Kumar, B. Namitha, An overview of CDK enzyme inhibitors in cancer therapy, Curr. Cancer Drug Targets 23 (2023) 603–619, https://doi.org/10.2174/1568009623666230320144713.
[10] L. Wang, W. Qin, Y. Huo, X. Li, Q. Shi, J.E.J. Rasko, A. Janin, W. Zhao, Advances in targeted therapy for malignant lymphoma, Signal Transduct. Targeted Ther. 5 (2020) 15, https://doi.org/10.1038/s41392-020-0113-2.
[11] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674, https://doi.org/10.1016/j.cell.2011.02.013.
[12] M. Malumbres, M. Barbacid, Cell cycle kinases in cancer, Curr. Opin. Genet. Dev. 17 (2007) 60–65, https://doi.org/10.1016/j.gde.2006.12.008.
[13] A. V Sadybekov, V. Katritch, Computational approaches streamlining drug discovery, Nature 616 (2023) 673–685, https://doi.org/10.1038/s41586-023-05905-z.
[14] X. Tong, X. Liu, X. Tan, X. Li, J. Jiang, Z. Xiong, T. Xu, H. Jiang, N. Qiao, M. Zheng, Generative models for de novo drug design, J. Med. Chem. 64 (2021) 14011–14027, https://doi.org/10.1021/acs.jmedchem.1c00927.
[15] W.P. Walters, R. Barzilay, Applications of deep learning in molecule generation and molecular property prediction, Acc. Chem. Res. 54 (2021) 263–270, https://doi.org/10.1021/acs.accounts.0c00699.

[16] R. Gómez-Bombarelli, J.N. Wei, D. Duvenaud, J.M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T.D. Hirzel, R.P. Adams, A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules, ACS Cent. Sci. 4 (2018) 268–276, https://doi.org/10.1021/acscentsci.7b00572.
[17] S. He, D. Zhao, Y. Ling, H. Cai, Y. Cai, J. Zhang, L. Wang, Machine learning enables accurate and rapid prediction of active molecules against breast cancer cells, Front. Pharmacol. 12 (2021) 796534, https://doi.org/10.3389/fphar.2021.796534.
[18] X. Pang, Y. Zhao, G. Li, J. Liu, A. Yan, A SAR and QSAR study on cyclin dependent kinase 4 inhibitors using machine learning methods, Dig. Dis. 2 (2023) 1026–1041, https://doi.org/10.1039/d2dd00143h.
[19] H. Zhang, J. Huang, R. Chen, H. Cai, Y. Chen, S. He, J. Xu, J. Zhang, L. Wang, Ligand- and structure-based identification of novel CDK9 inhibitors for the potential treatment of leukemia, Bioorg. Med. Chem. 72 (2022) 116994, https://doi.org/10.1016/j.bmc.2022.116994.
[20] S. Kaveh, A. Mani-Varnosfaderani, M.S. Neiband, Deriving general structure-activity/selectivity relationship patterns for different subfamilies of cyclin-dependent kinase inhibitors using machine learning methods, Sci. Rep. 14 (2024) 15315, https://doi.org/10.1038/s41598-024-66173-z.
[21] J. Liang, Y. Zheng, X. Tong, N. Yang, S. Dai, In silico identification of anti-SARS-CoV-2 medicinal plants using cheminformatics and machine learning, Molecules 28 (2022), https://doi.org/10.3390/molecules28010208.
[22] D. Merk, L. Friedrich, F. Grisoni, G. Schneider, De novo design of bioactive small molecules by artificial intelligence, Mol. Inform. 37 (2018) 1700153, https://doi.org/10.1002/minf.201700153.
[23] G. Choi, D. Kim, J. Oh, AI-based drug discovery of TKIs targeting L858R/T790M/C797S-mutant EGFR in non-small cell lung cancer, Front. Pharmacol. 12 (2021) 660313, https://doi.org/10.3389/fphar.2021.660313.
[24] P. Prabhakaran, A.V. Hebbani, S. V Menon, B. Paital, S. Murmu, S. Kumar, M.K. Singh, D.K. Sahoo, P.P.D. Desai, Insilico generation of novel ligands for the inhibition of SARS-CoV-2 main protease (3CL(pro)) using deep learning, Front. Microbiol. 14 (2023) 1194794, https://doi.org/10.3389/fmicb.2023.1194794.
[25] L. Salas-Estrada, D. Provasi, X. Qiu, H.Ü. Kaniskan, X.-P. Huang, J.F. DiBerto, J.M. Lamim Ribeiro, J. Jin, B.L. Roth, M. Filizola, De novo design of κ-opioid receptor antagonists using a generative deep-learning framework, J. Chem. Inf. Model. 63 (2023) 5056–5065, https://doi.org/10.1021/acs.jcim.3c00651.
[26] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[27] A.H. Arshia, S. Shadravan, A. Solhjoo, A. Sakhteman, A. Sami, De novo design of novel protease inhibitor candidates in the treatment of SARS-CoV-2 using deep learning, docking, and molecular dynamic simulations, Comput. Biol. Med. 139 (2021) 104967, https://doi.org/10.1016/j.compbiomed.2021.104967.
[28] Y. Bian, X. Xie, Artificial intelligent deep learning molecular generative modeling of scaffold-focused and cannabinoid CB2 target-specific small-molecule sublibraries, Cells 11 (2022), https://doi.org/10.3390/cells11050915.
[29] A. Gupta, A.T. Müller, B.J.H. Huisman, J.A. Fuchs, P. Schneider, G. Schneider, Generative recurrent networks for de novo drug design, Mol. Inform. 37 (2018) 1700111, https://doi.org/10.1002/minf.201700111.
[30] A. Gaulton, A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magariños, J.P. Overington, G. Papadatos, I. Smit, A.R. Leach, The ChEMBL database in 2017, Nucleic Acids Res. 45 (2017) D945–D954, https://doi.org/10.1093/nar/gkw1074.
[31] G.R. Bickerton, G.V. Paolini, J. Besnard, S. Muresan, A.L. Hopkins, Quantifying the chemical beauty of drugs, Nat. Chem. 4 (2012) 90–98, https://doi.org/10.1038/nchem.1243.
[32] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 42717, https://doi.org/10.1038/srep42717.
[33] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings, J. Chem. Inf. Model. 61 (2021) 3891–3898, https://doi.org/10.1021/acs.jcim.1c00203.
[34] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461, https://doi.org/10.1002/jcc.21334.
[35] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminf. 3 (2011) 33, https://doi.org/10.1186/1758-2946-3-33.
[36] A.W. Sousa da Silva, W.F. Vranken, Acpype - AnteChamber PYthon parser interfacE, BMC Res. Notes 5 (2012) 367, https://doi.org/10.1186/1756-0500-5-367.
[37] M.F. Adasme, K.L. Linnemann, S.N. Bolz, F. Kaiser, S. Salentin, V.J. Haupt, M. Schroeder, Plip 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA, Nucleic Acids Res. 49 (2021) W530–W534, https://doi.org/10.1093/nar/gkab294.
[38] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25, https://doi.org/10.1016/S0169-409X(96)00423-1.
[39] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 1–13, https://doi.org/10.1038/srep42717.
[40] A. Daina, O. Michielin, V. Zoete, iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach, J. Chem. Inf. Model. 54 (2014) 3284–3301, https://doi.org/10.1021/ci500467k.
[41] K.M. Honório, T.L. Moda, A.D. Andricopulo, Pharmacokinetic properties and in silico ADME modeling in drug discovery, Med. Chem. 9 (2013) 163–176, https://doi.org/10.2174/1573406411309020002.
[42] C.A. Lipinski, Drug-like properties and the causes of poor solubility and poor permeability, J. Pharmacol. Toxicol. Methods 44 (2000) 235–249, https://doi.org/10.1016/s1056-8719(00)00107-6.
[43] A. Daina, V. Zoete, A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules, ChemMedChem 11 (2016) 1117–1121, https://doi.org/10.1002/cmdc.201600182.
[44] M.Y. Lobanov, N.S. Bogatyreva, O. V Galzitskaya, Radius of gyration as an indicator of protein structure compactness, Mol. Biol. 42 (2008) 623–628, https://doi.org/10.1134/S0026893308040195.
[45] J.E.M. Pereira, J. Eckert, S. Rudic, D. Yu, R. Mole, N. Tsapatsaris, H.N. Bordallo, Hydrogen bond dynamics and conformational flexibility in antipsychotics, Phys. Chem. Chem. Phys. 21 (2019) 15463–15470, https://doi.org/10.1039/c9cp02456e.
[46] X. Du, Y. Li, Y. Xia, S. Ai, J. Liang, P. Sang, X. Ji, S. Liu, Insights into protein-ligand interactions: mechanisms, models, and methods, Int. J. Mol. Sci. 17 (2016), https://doi.org/10.3390/ijms17020144.