Bioorg. Med. Chem. 72 (2022) 116994

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

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

Ligand- and structure-based identification of novel CDK9 inhibitors for the potential treatment of leukemia

Huimin Zhang [a,1], Jindi Huang [a,1], Rui Chen [b], Hanxuan Cai [a], Yihao Chen [a], Shuyun He [a], Jianrong Xu [c,d], Jiquan Zhang [b], Ling Wang [a,*]

[a] Guangdong Provincial Key Laboratory of Fermentation and Enzyme Engineering, Joint International Research Laboratory of Synthetic Biology and Medicine, Guangdong Provincial Engineering and Technology Research Center of Biopharmaceuticals, School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China

[b] State Key Laboratory of Functions and Applications of Medicinal Plants & College of Pharmacy, Guizhou Provincial Engineering Technology Research Center for Chemical Drug R&D, Guizhou Medical University, Guiyang 550004, China

[c] Department of Pharmacology and Chemical Biology, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[d] Academy of Integrative Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

ARTICLE INFO

Keywords:
CDK9 inhibitor
Machine learning
Deep learning
Molecular docking

ABSTRACT

Cyclin-dependent kinase 9 (CDK9) plays a vital role in controlling cell transcription and has been an attractive target for cancer treatment. Herein, ten predictive models derived from 1330 unique molecules against CDK9 were constructed based on molecular fingerprints and graphs using two conventional machine learning and four deep learning methods. The evaluation results showed that FP-GNN deep learning architecture performed best for CDK9 inhibitors prediction with the highest BA and F1 values of 0.681 and 0.912 for testing set. We then performed virtual screening to identify new CDK9 inhibitors by incorporating the optimal established predictive model and molecular docking. Five compounds were identified to show broad anticancer activity against various cancer cell lines through bioassays. For example, **C9** exhibited antiproliferative activities against HeLa, MOLM-13 and MDA-MB-231 with IC_{50} values of 2.53, 3.92 and 11.65 μM. Kinase inhibition assay results demonstrated that these compounds displayed submicromolar (214 ~ 504 nM) inhibitory activities against CDK9. Further cellular mechanism evaluation revealed that **C9** suppressed the activity of CDK9 and interfered with the expression of Mcl-1 and cleaved PARP in MOLM-13 cells, resulting in the induction of cellular apoptosis. In addition, **C9** displayed a good stability in rat liver microsomes, artificial gastrointestinal fluid and plasm. An online platform (called DEEPCDK9Pred) was developed based on the FP-GNN models to predict or design new CDK9 inhibitors. Collectively, our findings demonstrated that FP-GNN algorithm can achieve accurate prediction of CDK9 inhibitors and the subsequent discovery of **C9** as a new potential CDK9 inhibitor deserves further structural modification for the treatment of leukemia.

1. Introduction

Cyclin-dependent kinases (CDKs) as members of serine/threonine protein kinases participate in the control of cell cycle progression (CDKs 1–6, 11 and 14–18) and cell transcription (CDKs 7–10, 12–13, 19 and 20) [1–3]. The abnormal expression of CDKs can cause various malignant tumors, many efforts are exerted to discovery new drugs targeted CDKs [4,5], especially for CDK9 subtype. CDK9 is a component of positive transcription elongation factor b (P-TEFb), which functions to phosphorylate the serine-2 at the C-terminal domain of RNA polymerase II to stimulate the transcription elongation of cellular proteins [6,7]. Hyperactivation of CDK9 is involved in the occurrence of many cancers, such as acute myeloid leukemia (AML) [8], hepatocellular carcinoma (HCC) [9], triple negative breast cancer (TNBC) [10], and non-small cell lung cancer (NSCLC) [11]. CDK9-mediated transcription of anti-apoptotic protein, such as Mcl-1, is closely related to the occurrence of various cancers [12], and the inhibition of CDK9 to indirectly block the transcription of apoptotic-associated protein would be a promising cancer therapy. All of these make CDK9 an attractive target for the treatment of cancers.

Many inhibitors target CDK9 have been developed for the treatment

* Corresponding author.
E-mail address: lingwang@scut.edu.cn (L. Wang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bmc.2022.116994
Received 5 July 2022; Received in revised form 21 August 2022; Accepted 29 August 2022
Available online 5 September 2022

of cancers including solid tumor and hematological malignancies (Fig. 1). These inhibitors are divided into pan-CDKs inhibitors and selective CDK9 inhibitors, pan-CDKs inhibitors target CDK9 and multiple CDK subtypes, such as Flavopiridol [13], Dinaciclib [14], and AT7519 [15]. Flavopiridol (Alvocidib), the first pan-inhibitor of CDKs inhibiting CDK1, CDK2, CDK4 and CDK9 has entered clinical investigation for the treatment of AML, chronic lymphocytic leukemia (CLL), breast cancer, and NSCLC [16–19]. However, pan-CDKs inhibitors usually lack selectivity to cause side effects, thereby limiting their entry into clinical research and/or their clinical use. A number of selective inhibitors targeting CDK9 have been developed to overcome the clinical limitations of lack of selectivity and potential toxicity of pan-CDKs inhibitors, such as AZD4573 [20], BAY-1143572 [21] and BAY-1251152 [22]. To date, there are no CDK9 inhibitors approved for the treatment of cancers, and only four selective CDK4/6 inhibitors (Palbociclib, Ribociclib, Abemaciclib and Trilaciclib) are approved by Food and Drug Administration (FDA) in CDKs family [23,24]. There is no doubt that CDK9 as clinically validated target is worthy of continued development of potential inhibitors for the treatment of various cancers. Hence, the identification of novel and diverse scaffolds targeting CDK9 is still required.

Currently, the discovery of small-molecule inhibitors targeting CDK9 (i.e. ATP-competitive CDK9 inhibitors) is mainly through *in vitro* or *in vivo* screening of compounds from various chemical databases and/or rational design and synthesis [25,26]. However, these biochemical experimental assays are time-consuming, expensive and laborious, particularly in large-scale applications. As an alternative and complementary choice for bioassays, computational approaches play a critical role in the early stages of hit or lead compounds identification, design and optimization, as they can provide fast and efficient screening models for the discovery of new actives for different targets or diseases [27–34]. Actually, the application of computational methods to discover CDK9 inhibitors has been reported in recent years. For example, in 2014, molecular dynamics (MD) simulations, molecular docking, pharmacophore modeling and binding free energy calculation methods were employed to identify the small molecule inhibitors against CDK9/cyclin T1 [35]. In 2019, Hussain et al. [36] reported 11 potential hit compounds against CDK9 by using the integration of pharmacophores modeling and molecular docking. Wu et al. [37] identified six hits with weak inhibitory activity against CDK9 through a hybrid virtual screening (VS) strategy that consists of ADMET prediction, 3D-QSAR pharmacophore, molecular docking and MD simulations. In 2021, Shao et al. applied molecular docking to design selective CDK9 inhibitors by analyzing the different docking poses of compounds with CDK2 and CDK9 [38].

Although the aforementioned computational models have been reported to accelerate the discovery of inhibitors targeting CDK9, there is still no specific platform that provides highly accurate predictive models based on diversity scaffolds for CDK9.

In this study, with the aim to construct reliable classification models for CDK9 inhibitors prediction and identify new inhibitors towards CDK9 based established optimal models, ten models were first established based on two different types of molecular representations (i.e. molecular fingerprints and graphs) using two conventional machine learning (CML) algorithms including support vector machine (SVM) [39] and extreme gradient boosting (XGBoost) [40], and four advanced deep learning (DL) methods including graph convolutional network (GCN) [41], graph attention network (GAT) [42], Attentive FP [43], and FP-GNN [44]. An online webserver called DEEPCDK9Pred (https://deepcdk9pred.idruglab.cn/) was then developed based on FP-GNN models with high predictive accuracy. In addition, an integrated VS strategy combining FP-GNN models and molecular docking was proposed to identify new ATP-competitive CDK9 inhibitors (Fig. 2). *In vitro* CDK9 inhibition and antiproliferative assays demonstrated that five compounds were confirmed as novel CDK9 inhibitors and displayed antitumor activity against eight tumor cell lines. Among these active compounds, **C9** displayed potential inhibition activity against CDK9 with an IC_{50} of 295 nM and considerable anticancer activity against HeLa, MOLM-13 and MDA-MB-231. The mechanisms of cell death induced by compound **C9** were also investigated in MOLM-13 using the quantification of apoptosis and western blot analyses.

2. Results and discussion

2.1. Dataset analysis and model construction

After dataset preprocessing according to the predefined principles in "Computational methods" section, we collected a total of 1330 unique compounds against CDK9 and their enzymatic inhibitory activity data. 1087 compounds were labeled inhibitors (actives) and 243 compounds

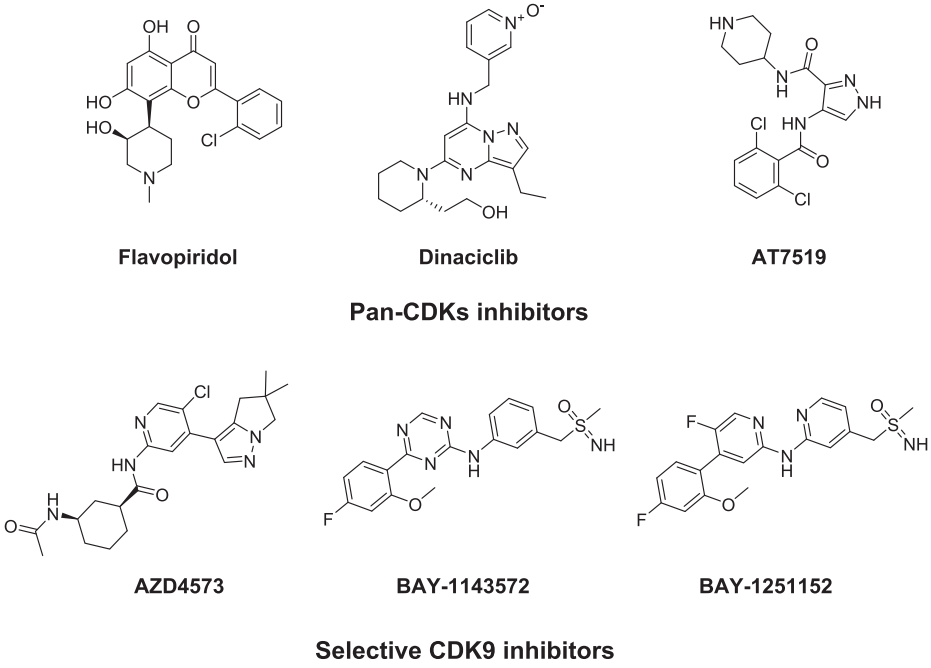

Fig. 1. Representatives of the clinical CDK9 inhibitors.

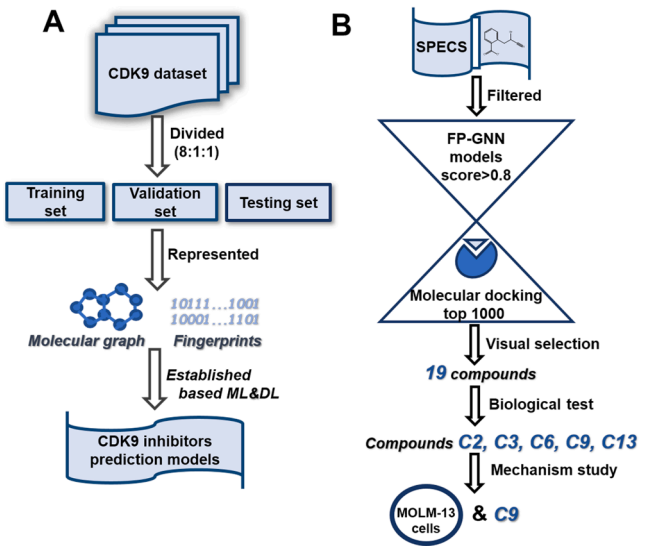

Fig. 2. Flowcharts of models construction and VS. The workflow of constructing predictive models for CDK9 inhibitors (A) and the integrated VS strategy for discovering new CDK9 inhibitors (B). VS: Virtual screening.

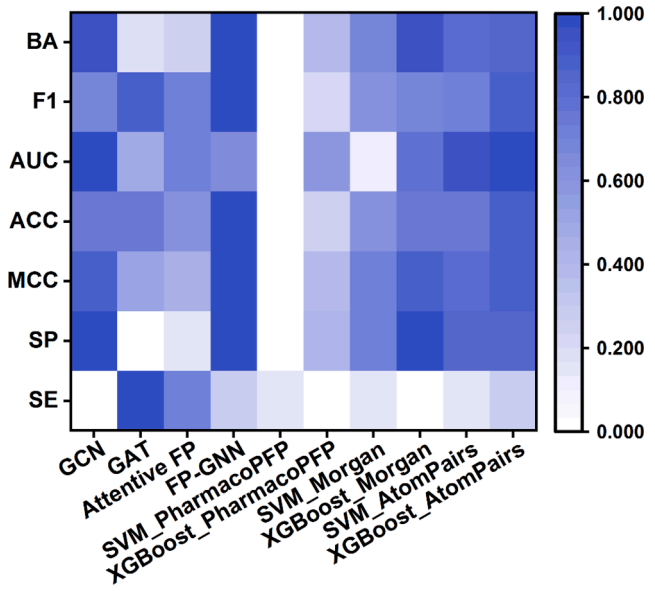

Fig. 3. Performance of established CDK9 inhibitors prediction models for the testing sets. The data given were processed by Min-Max Normalization.

were labeled non-inhibitors (inactives), the proportion of actives in CDK9 dataset achieved 81.7 %.

The structural diversity and chemical space of the compounds in the modeling dataset have a significant impact on the prediction accuracy and stability of the classification models [45–48]. We then used the Bemis–Murcko scaffold analysis method [49] to calculate the structural diversity of the molecules in CDK9 inhibitors dataset. The Bemis–Murcko scaffold for each molecule was generated using open-source RDKit software (https://www.rdkit.org/). The results showed that the proportion of the scaffolds for inhibitors of CDK9 was 21.43 %, indicating that the chemical structures of the CDK9 dataset displayed a slightly scaffold diversity. The molecular weight (MW) and AlogP values of molecules in the dataset were calculated and then displayed on a two-dimensional space to analyze the chemical space. The results indicated that the compounds in training, validation and testing sets distributed over a wide range of MW (185.190 ~ 916.855 Da) and AlogP (-0.724 ~ 7.813) (Fig. S1), suggesting that the molecules in the CDK9 modeling dataset showed a broad chemical space. Based on two types of molecular feature representations (molecular fingerprints and graphs), ten classification models were developed using two CML (XGBoost and SVM) and four DL (GCN, GAT, Attentive FP and FP-GNN) algorithms. The detailed performance of these models is described as follows.

2.2. Performance of fingerprints- and graph-based predictive models

For each CML or DL model, hyperparameters were optimized based on the validation set, and the best set of hyperparameters for each model are summarized in Table S1. The details of evaluation results for the fingerprints- and graph-based models are summarized in Table S2. The evaluation metrics (BA, F1, AUC, SE, SP, MCC, ACC) from the testing set are used to evaluate the predictive and generalization capabilities of the established models. The BA metric is suitable for evaluating the models based on the unbalanced dataset. Since the ratio of active and inactive molecules in the CDK9 dataset is approximately 4.5:1, the BA metric is mainly applied to evaluate the performance of the established models.

Six predictive models were built based on three types of molecular fingerprints (Morgan, Atompairs, and PharmacoPFP) using XGBoost and SVM methods. As shown in Fig. 3 and Table S3, compared to SVM algorithm against different fingerprints, XGBoost algorithm achieved comparable or slightly better performance with average ACC, AUC, F1 and BA values of 0.827, 0.850, 0.898 and 0.640 for the testing set. Among these three different fingerprints, predictive models based on AtomPairs fingerprint performed best with the highest average BA value of 0.658 for the testing set, indicating that AtomPairs represents the best fingerprint-based featurization for CDK9-associated modeling dataset (Fig. 3 and Table 1).

Recently, DL (especially graph neural network, GNN) methods have been reported to yield state-of-the-art (SOTA) performance in molecular properties prediction tasks [50–53], including physicochemical, biological activity and ADME/T properties. Therefore, four graph-based predictive models (Table S2) were developed using the four advanced DL algorithms (i.e. GCN, GAT, Attentive FP and FP-GNN). FP-GNN as a novel DL algorithm developed in our Lab (https://github.com/idrugLab/FP-GNN), can operate over a hybrid molecular feature representation that combines molecular graph and fingerprints. FP-GNN architecture can achieve SOTA performance for molecular properties prediction tasks (data not shown). Taking the BA score (Fig. 3 and Table 1) of the testing set into consideration, FP-GNN method displayed the overall best performance compared with other GNN methods, with the highest BA value of 0.681.

Comparisons of the established fingerprints-, and graph-based predictive models (Fig. 3 and Table 1) indicated that FP-GNN model exhibited the best prediction abilities because it was associated with the highest BA and F1 values of the testing set for CDK9 modeling dataset. In addition, the performance of the testing set in fingerprints-based models (average BA = 0.623) was comparable or superior with other three graph-based DL (i.e. GCN, GAT and Attentive FP, average BA = 0.600) models, suggesting that these commonly used graph-based DL algorithms do not exhibit SOTA performance than the CML (i.e. XGBoost and SVM) algorithms for CDK9 inhibitors dataset. The FP-GNN models for CDK9 are implemented in a freely accessible web server (DEEPCDK9Pred: https://deepcdk9pred.idruglab.cn/, Fig. 4), and the source code is available on https://github.com/idrugLab/DEEPCDK9Pred. DEEPCDK9Pred would be a useful tool for predicting or designing new inhibitors against CDK9.

2.3. Virtual screening and bioassays for the identification of new CDK9 inhibitors

To verify whether the established FP-GNN models can identify new CDK9 inhibitors in real-world drug discovery scenarios, we designed an integrated VS strategy using the combination of the FP-GNN models

Table 1
The performance of established CDK9 inhibitors prediction models for the training set and testing set.

Molecular features	Algorithms	Training set							Testing set						
		SE[a]	SP[b]	MCC[c]	ACC[d]	AUC[e]	F1[f]	BA[g]	SE	SP	MCC	ACC	AUC	F1	BA
PharmacoPFP	SVM	0.992	0.579	0.699	0.916	0.966	0.951	0.786	0.936	0.125	0.089	0.789	0.747	0.879	0.530
	XGBoost	0.970	0.456	0.531	0.876	0.915	0.927	0.713	0.927	0.250	0.221	0.805	0.823	0.886	0.588
	Mean	0.981	0.518	0.615	0.896	0.940	0.939	0.749	0.931	0.188	0.155	0.797	0.785	0.883	0.559
Morgan	SVM	0.992	0.815	0.861	0.960	0.990	0.976	0.904	0.936	0.333	0.327	0.827	0.763	0.899	0.635
	XGBoost	0.987	0.621	0.713	0.920	0.969	0.953	0.804	0.927	0.417	0.386	0.835	0.848	0.902	0.672
	Mean	0.990	0.718	0.787	0.940	0.980	0.964	0.854	0.931	0.375	0.357	0.831	0.806	0.900	0.653
AtomPairs	SVM	0.993	0.749	0.820	0.948	0.992	0.969	0.871	0.936	0.375	0.367	0.835	0.870	0.903	0.655
	XGBoost	0.988	0.841	0.868	0.961	0.994	0.977	0.915	0.945	0.375	0.389	0.842	0.878	0.907	0.660
	Mean	0.991	0.795	0.844	0.955	0.993	0.973	0.893	0.940	0.375	0.378	0.838	0.874	0.905	0.658
Molecular graph	GCN	0.994	0.964	0.962	0.989	1.000	0.993	0.979	0.927	0.417	0.386	0.835	0.875	0.902	0.672
	GAT	0.990	0.236	0.394	0.852	0.869	0.916	0.613	0.991	0.125	0.261	0.835	0.812	0.908	0.558
	Attentive FP	0.985	0.272	0.412	0.854	0.893	0.917	0.628	0.972	0.167	0.240	0.827	0.843	0.902	0.570
	FP-GNN	0.958	0.721	0.707	0.914	0.959	0.948	0.840	0.945	0.417	0.427	0.850	0.830	0.912	0.681
	Mean	0.972	0.496	0.559	0.884	0.926	0.933	0.734	0.959	0.292	0.334	0.838	0.836	0.907	0.625

[a] SE: Sensitivity.
[b] SP: Specificity.
[c] MCC: Matthews correlation coefficient.
[d] ACC: Accuracy.
[e] AUC: The area under receiver operating characteristic.
[f] F1: F1-measure.
[g] BA: Balanced accuracy.

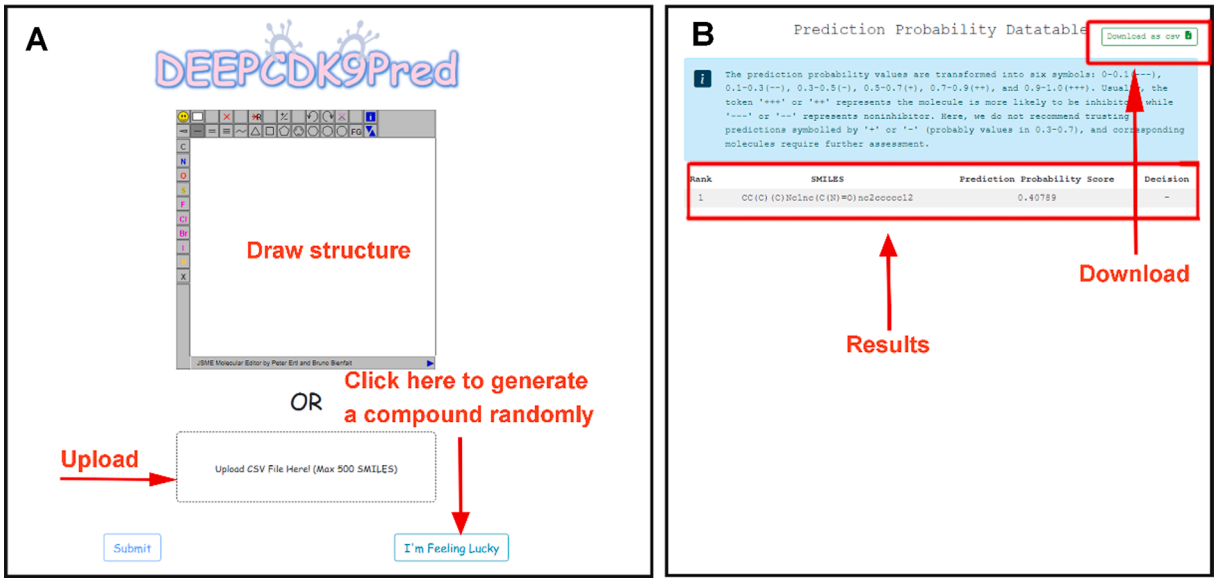

Fig. 4. Website schematic diagram of DEEPCDK9Pred for CDK9 inhibitors prediction. (A) represents the page of submission of compounds. (B) represents the prediction results of the compounds against CDK9.

with molecular docking to discover new ATP-competitive CDK9 inhibitors (Fig. 2). As shown in Fig. 2, 208,670 compounds in SPECS library were filtered by Lipinski's Rule of Five and the filtered database (194,916 compounds) was then predicted using the FP-GNN-based CDK9 model, resulting in a CDK9 inhibitor-like library (51,457 compounds with predictive scores > 0.8). The CDK9 inhibitor-like library was subsequently used for Glide docking-based VS. After this docking procedure, 1000 compounds with top Glide SP scores were used for visual analysis. These molecules were inspected to check whether they interact with the ATP-binding pocket of CDK9, especially including hydrogen bond interactions with Cys106 at the hinge region. Finally, 19 compounds were purchased for bioassays (Table S4). These compounds were initially measured for their inhibitory activities against MDA-MB-231 and HeLa cells at 100 μM. As shown in Fig. 5A, six compounds displayed a minimum of 70 % inhibitory rates against MDA-MB-231 and HeLa cells. The IC_{50} values of these six compounds were then determined against eight human tumor cell lines from different tissues, including MDA-MB-231, HeLa, CNE2, A549, HepG2, MCF-7, HCT116 and MOLM-13. Flavopiridol was used as a positive control drug. The detailed IC_{50} values are summarized in Table 2. Table 2 illustrated that six compounds (**C2**, **C3**, **C6**, **C9**, **C13** and **C19**, Fig. 5B) exhibited anticancer effects on cells of solid tumors (MDA-MB-231, MCF-7, HepG2, HCT116, CNE2 and A549) and hematological malignancies (MOLM-13). For example, **C9** showed considerable antitumor activities against MOLM-13, HeLa and MDA-MB-231 with IC_{50} values of 3.92, 2.53 and 11.65 μM, respectively.

Next, we performed *in vitro* CDK9 kinase inhibition assays to verify whether six compounds could target CDK9. As shown in Table 2, five compounds (**C2**, **C3**, **C6**, **C9** and **C13**) displayed > 70 % inhibitory activities against CDK9 at 20 μM. While **C19** did not show inhibitory activity against CDK9 due to its lower inhibition rate (~22.7 %), suggesting that **C19** was off-target for CDK9. The IC_{50} values of compounds **C2**, **C3**, **C6**, **C9** and **C13** against CDK9 were further determined and listed in Table 2. Apparently, these compounds showed

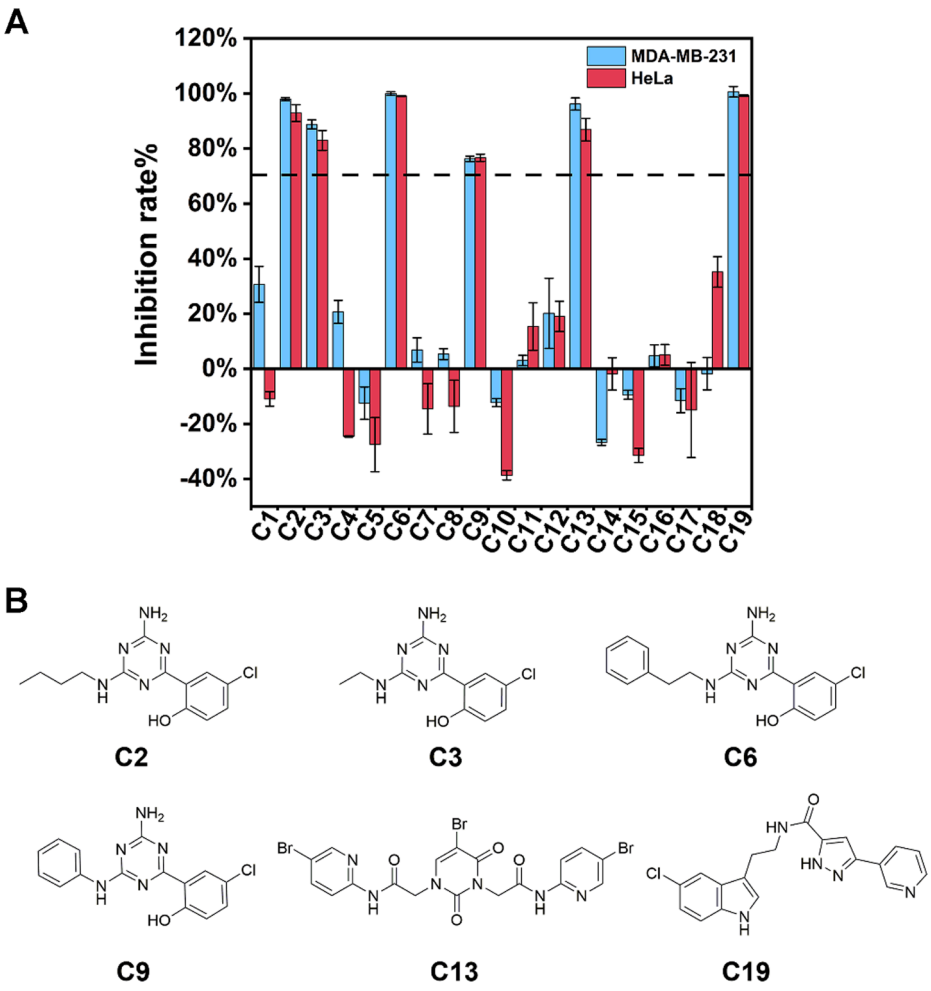

Fig. 5. (A) The inhibitory rates of 19 compounds against HeLa and MDA-MB-231 cells at 100 μM. (B) Chemical structures of compounds **C2**, **C3**, **C6**, **C9**, **C13** and **C19**.

Table 2
Inhibitory activities of selected compounds against CDK9 and eight cancer cell lines.

Compd.	CDK9 Inhibition % at 20 μM[a]	CDK9 IC_{50} (nM)[b]	IC_{50} (μM) ± SD[c]							
			MDA-MB-231	HeLa	CNE2	A549	MCF-7	HepG2	HCT116	MOLM-13
C2	93.6 ± 0.76	214	16.01 ± 0.68	13.74 ± 2.51	16.99 ± 1.08	19.21 ± 2.93	11.29 ± 0.50	40.92 ± 1.94	27.21 ± 4.18	13.48 ± 0.75
C3	93.7 ± 2.43	335	21.19 ± 6.26	7.38 ± 0.81	30.42 ± 4.80	>50	18.95 ± 0.33	49.21 ± 6.35	26.63 ± 7.91	15.21 ± 0.49
C6	92.2 ± 0.76	504	14.52 ± 0.46	25.05 ± 12.15	> 50	28.16 ± 0.04	28.99 ± 1.42	28.97 ± 2.54	> 50	4.58 ± 0.08
C9	88.4 ± 1.82	295	11.65 ± 1.52	2.53 ± 0.43	34.54 ± 11.82	> 50	23.52 ± 6.12	> 50	22.22 ± 8.91	3.92 ± 1.59
C13	76.4 ± 0.99	219	8.34 ± 1.81	6.61 ± 0.66	28.32 ± 3.36	15.35 ± 1.91	6.14 ± 0.08	> 50	3.22 ± 2.13	11.79 ± 0.51
C19	22.7 ± 1.59	NT[e]	28.15 ± 3.64	29.45 ± 3.03	17.06 ± 1.38	> 50	20.86 ± 1.33	19.79 ± 2.11	15.63 ± 0.73	2.52 ± 0.53
Flavopiridol[d]	97.2 ± 1.36	4.5	0.06 ± 0.02	0.15 ± 0.03	0.12 ± 0.01	0.1 ± 0.00	0.71 ± 0.00	0.56 ± 0.29	0.03 ± 0.00	0.07 ± 0.02

[a,b] Data are the mean of n ≥ 2.
[c] Data are presented as the mean values ± SD from experiments conducted in triplicate at three independent times.
[d] Positive control drug.
[e] *NT*: not tested.

submicromolar (214 ~ 504 nM) activities against CDK9, revealing that compounds **C2**, **C3**, **C6**, **C9** and **C13** can directly target CDK9.

To assess the structural novelty of these five CDK9 inhibitors with respect to the known CDK9 inhibitors, Tanimoto similarity scores between these five molecules and active molecules from the CDK9 inhibitors modeling dataset were calculated based on the ECFP_6

fingerprint using Discovery Studio software. As shown in Fig. S2, five molecules had low Tanimoto similarity scores (0.132 ~ 0.265) with the known CDK9 inhibitors, demonstrating that the five CDK9 inhibitors identified in this study are structurally novel.

2.4. Structure activity relationships of active compounds

We analyzed the structure activity relationships of compounds **C2**, **C3**, **C6**, and **C9** in MOLM-13 cells, since these compounds were sensitive to MOLM-13 cells with IC_{50} values lower than 20 μM. As shown in Fig. 6, compounds **C2**, **C3**, **C6**, and **C9** contained the 1,3,5-triazine-2,4-diamine scaffold. Compound **C2** with R_1 group of *n*-butyl (IC_{50} = 13.48 μM) displayed better antitumor activity than **C3** with R_1 group of ethyl (IC_{50} = 15.21 μM) in MOLM-13 cells. Compound **C9** with R_1 group of a benzene ring (IC_{50} = 3.92 μM) showed better antitumor activity than compound **C6** with R_1 group of an indirectly linked benzene ring (IC_{50} = 4.58 μM). Collectively, compounds **C6** and **C9** with larger R_1 substitutions will show better activity against MOLM-13 cells.

2.5. *C9* inhibits CDK9 activity in cells

In recent years, some CDK9 inhibitors (e.g. AZD4573 [20], BAY-1143572 [54], Dinaciclib [55]) have been tested in clinical trials for the treatment of AML. Accordingly, compound **C9** was selected to verify whether this compound could act on CDK9 in MOLM-13 cells (AML) since **C9** showed potential inhibitory activities against CDK9 and MOLM-13. The intracellular CDK9 inhibitory activities were measured by determining the phosphorylation level of its target (RNAP II CTD) in the MOLM-13 cells through western blotting (WB) analysis. As shown in Figs. 7A-B, the phosphorylation of p-RNAP II CTD (Ser2) was suppressed with the treatment of **C9** in a concentration-dependent manner, demonstrating that **C9** can inhibit the activity of CDK9 in cells.

2.6. *C9* induces cellular apoptosis to cause tumor cell death

The inhibition of CDK9 has been reported to cause the induction of cellular apoptosis [11,56,57]. Annexin V-FITC and PI staining were used to verify whether cell proliferation was affected by **C9** through apoptosis. As shown in Figs. 7E-F, the MOLM-13 cells were treated with **C9** at concentrations of 2.5, 5 and 10 μM for 24 h. The total number of early and late apoptosis cells were 7.08 %, 7.93 % and 11.56 %, respectively. Such results confirmed that **C9** induced cell apoptosis in a concentration-dependent manner, contributing to its anticancer effects on tumor cells.

To further investigate the cellular mechanism of the observed apoptosis, WB analysis was employed to detect apoptosis-associated proteins (i.e. Mcl-1 and cleaved PARP) of CDK9. As shown in Figs. 7A, 7C-D, compared to the control, **C9** not only decreased the expression of Mcl-1 in a dose-dependent manner, but increased the level of the cleaved PARP. These results indicated that the effects of **C9** on cell apoptosis are caused by interfering with the expression of both Mcl-1 and cleaved PARP apoptosis-associated proteins, which is consistent with previous studies [38,58].

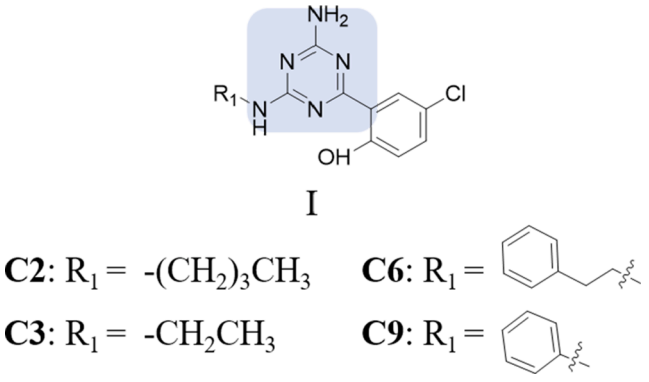

Fig. 6. Structure activity relationships of compounds **C2**, **C3**, **C6**, **C9**.

2.7. Stability analysis of *C6, C9* and *C13*

Metabolic stability of a given molecule is crucial for further lead optimization and subsequently clinical development [59]. Thus, *in vitro* stabilities of compounds **C6**, **C9** and **C13** were evaluated by co-incubating these compounds in a standard artificial gastric, intestinal fluids, liver microsomes and the blood plasma of SD rats [60,61]. Flavopiridol was used as a positive control. As shown in Figs. 8A-D, the remaining percentages of these tested compounds in the liver microsomes, blood plasm and artificial gastrointestinal fluid were no less than 90 %, illustrating that **C6**, **C9** and **C13** displayed a good metabolic stability *in vitro*.

2.8. Binding model analysis of *C9* and CDK9

The binding mode of compound **C9** with CDK9 was explored using molecular docking. As shown in Fig. 9, **C9** occupies well in the ATP-binding pocket of CDK9. The aminotriazine core of **C9** forms a hydrogen bond with Cys106 in the hinge region of CDK9 kinase domain, which plays a key role in maintaining its inhibitory activity against CDK9. The hydroxyl group on the benzene ring of **C9** forms an additional hydrogen bond with Ala153. Meanwhile, the phenyl ring substituent of **C9** also forms hydrophobic interactions with Phe103. In addition, the predicted docking scores of **C9** with CDK2, CDK5 (representative cell cycle associated CDKs), and CDK7 (representative cell transcription associated CDK) suggest that **C9** may exhibit selectivity for these representative CDKs (Table S5). The binding mode analysis also revealed that the substituents on the aminotriazine core of **C9** still have room for further structural modification to enhance the inhibitory activity against CDK9.

3. Conclusion

In this study, we collected comprehensive small-molecule inhibitors dataset for CDK9 and ten classification models were developed using two CML and four DL methods to distinguish inhibitors and non-inhibitors. Among these established predictive models, FP-GNN models exhibited the best performance. An online platform called DEEPCDK9Pred was developed based on the optimal FP-GNN models to predict and/or design new CDK9 inhibitors. We designed an integrated VS protocol, including FP-GNN model prediction, molecular docking and a series of biological assays, we finally identified five novel CDK9 inhibitors (**C2**, **C3**, **C6**, **C9** and **C13**) with sub-micromolar inhibitory activities against CDK9 and micromolar inhibitory activities against eight tumor cell lines. Further cellular studies illustrated that **C9** inhibited CDK9 activity in MOLM-13 cells, and **C9** could induce cell apoptosis to cause MOM-13 cell death through the regulation of apoptosis-related proteins (Mcl-1 and cleaved PARP). More importantly, compound **C9** showed a good metabolic stability. Collectively, **C9** as a new CDK9 inhibitor deserves further lead optimization to achieve better inhibitory activities against both CDK9 and cancer cells. Structural optimization of **C9** is currently underway in our Lab.

4. Computational and experimental methods and materials

4.1. Computational methods

4.1.1. Dataset collection and processing

CDK9 associated-compounds dataset was collected from ChEMBL database (version 28) [62], and then processed by the following procedure: (1) Compounds with only CDK9 inhibition assay data (assay type = B)

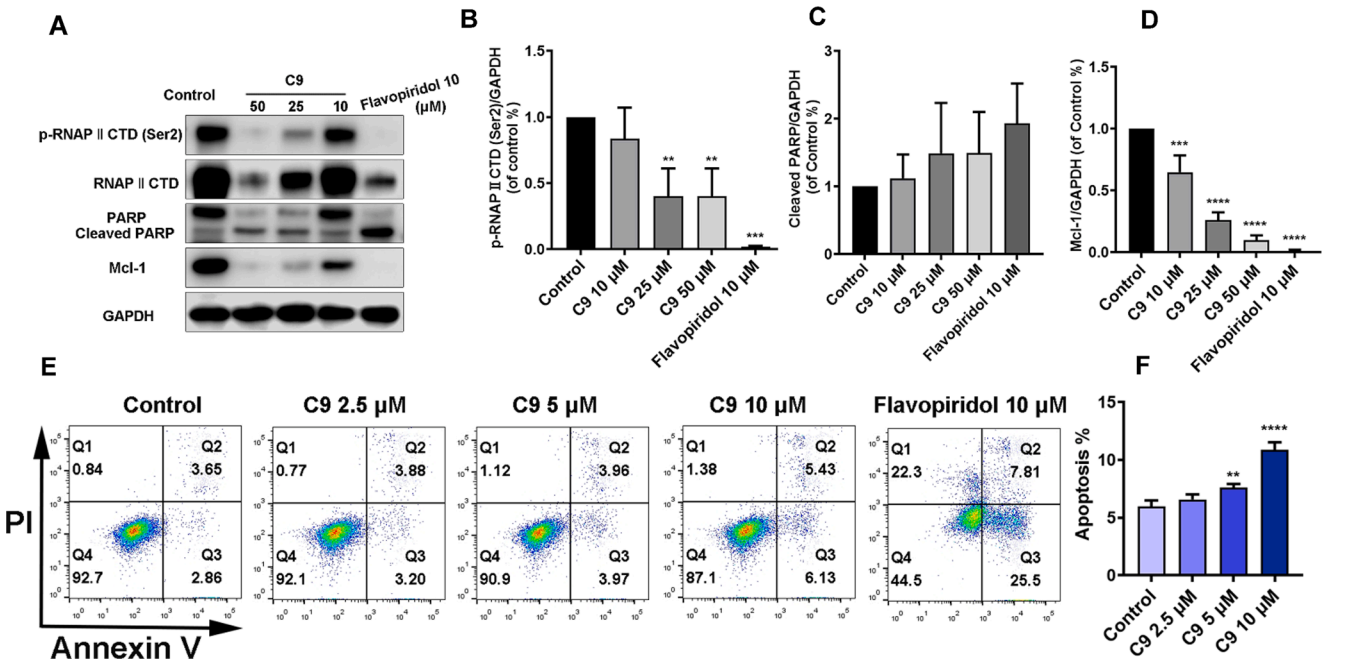

Fig. 7. (A) The effects of **C9** on CDK9 activity and apoptosis-related proteins (Mcl-1 and cleaved PARP) in MOLM-13 cells. Quantitative western blotting analysis for p-RNAP II CTD (Ser2) (B), cleaved PARP (C) and Mcl-1 (D). (E) Cellular apoptosis analysis of **C9** in MOLM-13 cells and quantitative analysis of cellular apoptosis. Three independent experiments were conducted for each data analysis. Flavopiridol was used as a positive control drug.

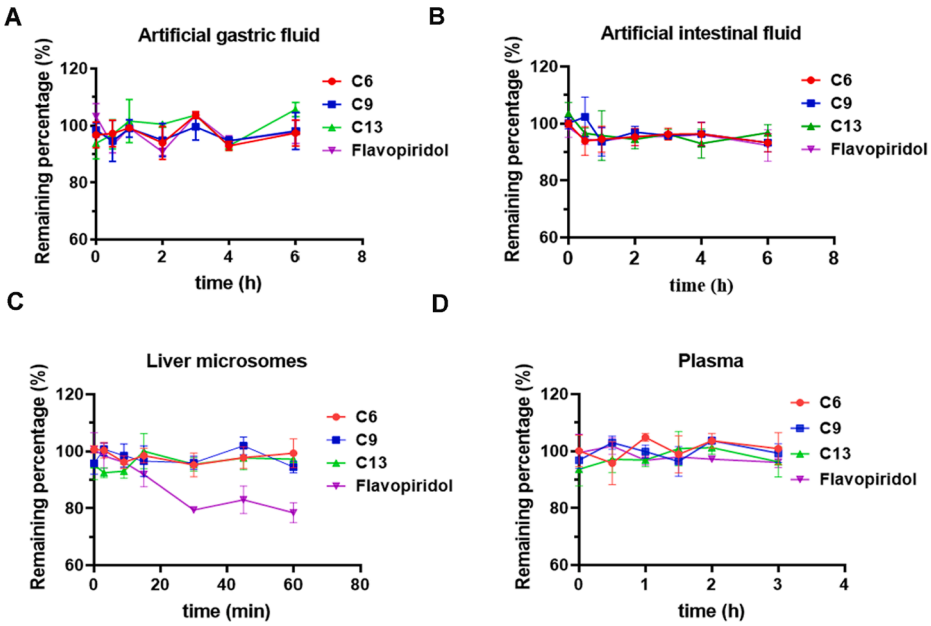

Fig. 8. The stability analysis of **C6**, **C9** and **C13** in artificial gastric fluid (A), artificial intestinal fluid (B), liver microsome (C) and plasm (D). The data given are mean values from three independent replicate experiments.

were kept; (2) Molecules with detailed assay values (i.e. IC_{50}, EC_{50}, K_i, and K_d) were retained, if a compound has many test values, computing the average value of its reported activity values as the final assay value; (3) Duplicates were then removed; (4) Compounds in CDK9 dataset were divided into inhibitors (≤10 μM) and non-inhibitors (>10 μM). Finally, 1330 molecules for CDK9 were obtained. CDK9 dataset was randomly divided into three sub-datasets: training set (80 %), validation set (10 %), and testing set (10 %). The training set was used to build a predictive model. The validation set was employed to optimize hyperparameters, and the testing set was used to test the predictive ability of the established model.

4.1.2. *Molecular representation calculation*

Herein, two types of molecular representations (molecular fingerprints and graphs) were applied to construct predictive models. Three molecular fingerprints (Morgan, AtomPairs and PharmacoPFP) were calculated using RDKit software (version: 2020.03.1), while molecular graph was generated using Deepchem software (https://deepchem.io/, version: 2.5.0.).

4.1.3. *In silico model construction and evaluation*

Two CML algorithms (XGBoost and SVM) and four DL algorithms (GCN, GAT, Attentive FP and FP-GNN) were applied to build classifi-

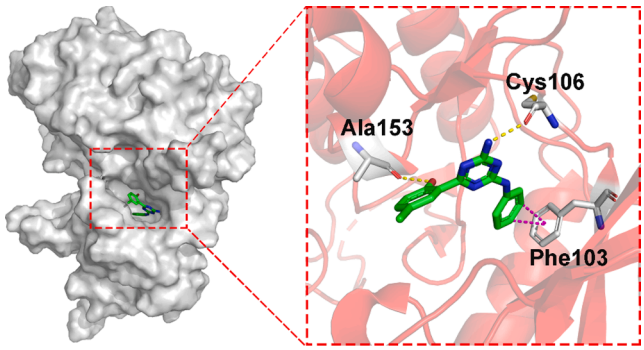

Fig. 9. Predicted binding mode of **C9** with CDK9. The binding mode was predicted based on Glide SP docking algorithm and the figure was generated using PyMOL software (https://pymol.org/2/). Hydrogen bonds are depicted by yellow dashes and arene-H and arene-arene interactions are depicted by magenta dashes.

cation models to predict the inhibitory activity of CDK9 inhibitors. SVM models were established using the Scikit-learn software (https://github.com/scikit-learn/scikit-learn, version: 0.24.1). Based on XGBoost python package (https://github.com/dmlc/xgboost, version: 1.3.3), XGBoost models were established. The FP-GNN models were constructed using FP-GNN DL architecture developed in our Lab (https://github.com/idrugLab/FP-GNN). FP-GNN algorithm can combine and simultaneously learn the information from molecular graph and mixed fingerprints (MACCS, Pharmacophore ErG and PubChem FPs). Other DL-based (GCN, GAT and Attentive FP) models were developed by using DeepChem python package (https://deepchem.io/). In addition, the grid search was employed to optimize hyperparameters for each predictive model. TP, FN, TN and FP represent true positives, false negatives, true negatives and false positives. Seven standard metrics including specificity (SP), sensitivity (SE), matthews correlation coefficient (MCC), accuracy (ACC), the area under the receiver operating characteristic curve (AUC), F1-measure (F1 score) and balanced accuracy (BA) were used to evaluate the performance of classification models, and were calculated as follows.

$$SP = \frac{TN}{TN + FP} \tag{1}$$

$$SE = \frac{TP}{TP + FN} \tag{2}$$

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN) \times (TP + FP) \times (TN + FN) \times (TN + FP)}} \tag{3}$$

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{4}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{5}$$

$$BA = \frac{SE + SP}{2} \tag{6}$$

4.1.4. Virtual screening

SPECS library (containing 208,670 compounds, https://www.specs.net/) was elected for VS to discover CDK9 inhibitors. All molecules were processed by the following steps: (1) Deleting duplicates, washing salt atoms, adding hydrogen atoms, and energy minimization using MOE software (version 2018); (2) Then filtering based on Lipinski's Rule of Five through Discovery Studio 3.5. The established FP-GNN model of CDK9 inhibitors was used to predict the inhibitory activity of the optimized SPECS library (approximately 194,916 compounds). The crystal structure of CDK9 with T6Q (PDB ID: 4BCF) [63] was employed for docking-based VS. The protein structure was refined using the "Protein Preparation Wizard" module in Maestro software, including removal of all water molecules, protonation, and optimization based on the OPLS_2005 force field. A grid file was generated based on the native ligand (T6Q) with default parameters. Glide-SP was used to dock the CDK9 inhibitors-like library into the ATP-binding pocket of CDK9. Top 1000 molecules were selected for further analysis according to the docking score. Finally, 19 compounds were elected through visual analysis of protein–ligand interactions. All 19 compounds passed PAINS filtration and then purchased from SPECS for biological assays. The purity of all 19 compounds from SPECS has been verified by SPECS vendor via liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) spectra, ensuring that the purity of each compound was no less than 95 %.

4.2. Biology

4.2.1. Cell cytotoxicity assay

Briefly, cells (MDA-MB-231, HeLa, MCF-7, MOLM-13, CNE2, HCT116, HepG2, A549) were seeded at 96-well plates with DMEM or 1640 medium. After 24 h cultured, cells were treated with different concentrations of each testing compound for 48 h. The viability of the cells was quantified using CCK8 reagent kit (Glpbio, USA). The OD values were measured by the microplate reader (Tecan, Switzerland) and then analyzed using GraphPad Prism 8.0 software.

4.2.2. In vitro CDK9 inhibition assay

In vitro CDK9 inhibition assay for the selected compounds was performed by Shanghai ChemPartner Co. ltd. (China). Briefly, 1 × kinase base buffer (20 mM HEPES, pH 7.5, 0.01 % Triton X-100) and stop buffer (100 mM HEPES, pH 7.5, 0.015 % Brij-35, 0.2 % Coating Reagent #3, 50 M EDTA) were prepared. Compounds were diluted to 50 × of the final desired highest inhibitor concentration in reaction by 100 % DMSO, and a total of 100 μL compounds dilution was transferred into a 96-well plate (marked as a source plate). Two empty wells were filled by 100 μL 100 % DMSO as a blank control. An intermediate plate (96-well) was prepared by filling 10 compounds from the source plate, and 90 μL 1 × kinase buffer was then added to each well. 5 μL of each well from the intermediate plate was transferred to a 384-well assay plate. 10 μL 2.5 × enzyme solution was added to each well and incubated with compounds for 10 min at room temperature. 20 μL 2.5 × peptide solution was added into each well of 384-well plate and incubated at 28 °C, and 25 μL stop buffer was then added into the well to stop reaction. The plate was read by the microplate reader and the IC_{50} values were calculated using GraphPad Prism 5.

4.2.3. Western blotting

MOLM-13 cells were treated with various concentrations of compound **C9** for 24 h, and then harvested and collected the total protein with RIPA lysis Buffer (Beyotime, China) and 1 mM PMSF at 12000 rpm for 15 min at 4 °C. After centrifugation, the supernatant was collected and the protein concentration was determined using BCA kit (Pierce, USA). The protein (20 μg) was loaded into the 8 % SDS-PAGE gel and separated at 80 V for 20 min and 120 V for 80 min, and the separated protein in the gel was then transferred into the PVDF membrane (Millipore, USA). The PVDF membranes were blocked with 5 % BSA, washed with PBS, incubated with primary antibodies overnight at 4 °C and subsequently washed for incubating with secondary antibodies (anti-mouse IgG, anti-mouse IgG) for 90 min. Finally, the membrane was detected by ECL reagents (Millipore, USA) and photographed by Amersham Image 600 instrument (GE, USA). Antibody information: anti-p-RNAP II CTD (Ser2) (CST; #13499S), anti-RNA polymerases II CTD (CST; #2629S), anti-Mcl-1 (CST; #94296S), anti-PARP (CST; #9532S), anti-GAPDH (CST; #5174S).

4.2.4. Cell apoptosis assay

Cell apoptosis assay was performed using an Annexin V-FITC apoptosis detection kit (Keygen, China) and analyzed using a BD Accuri C6 plus flow cytometer. Briefly, MOLM-13 cells were seeded into 6-well plates, and treated with different concentrations of compound **C9** for 24 h. The cells were stained with Annexin V-FITC and propidium iodide for analysis using FlowJo X10.0.7 software.

4.2.5. Stability assays

The stability of **C3**, **C9** and **C13** was performed in rat plasma. Flavopiridol was used as a positive control drug. Briefly, compounds **C3**, **C9**, C**13** and Flavopiridol (1 mg/mL) were dissolved in DMSO, and the reaction (10 μg/mL, n = 3) was prepared by mixing 198 μL plasma with 2 μL compounds solution, then incubated in 37 °C and stopped by adding 400 μL methanol at time of 0, 1, 2 and 3 h, followed by centrifugation at 13,000*g* at 4 °C for 10 min. The supernatant was dried and redissolved in 100 μL methanol. After centrifugation, the supernatant was analyzed by HPLC. The stability data was presented as remaining percentage *vs* time.

According to the guideline of Chinese Pharmacopeia (version 2015), compounds reagents (1 mg/mL, n = 3) were incubated with artificial gastric fluid and artificial intestinal fluid at 37 °C, and 400 μL acetonitrile was then added into the reagent to stop reaction at time of 0, 0.5, 1, 2, 3, 4 and 6 h. Supernatant was collected by centrifugation at 20,000*g* for 10 min and analyzed by HPLC. The data were processed by GraphPad Prism software, and the stability of compounds in liver microsomes referenced our previous work.[61]

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.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (81973241) and the Natural Science Foundation of Guangdong Province (2020A1515010548).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bmc.2022.116994.

References

1 Malumbres M. Cyclin-dependent kinases. *Genome Biol.* 2014;15:122.
2 Cheng W, Yang Z, Wang S, et al. Recent development of CDK inhibitors: An overview of CDK/inhibitor co-crystal structures. *Eur J Med Chem.* 2019;164:615–639.
3 Ding L, Cao J, Lin W, et al. The Roles of Cyclin-Dependent Kinases in Cell-Cycle Progression and Therapeutic Strategies in Human Breast Cancer. *Int J Mol Sci.* 2020;21:1960.
4 Shi Z, Tian L, Qiang T, et al. From Structure Modification to Drug Launch: A Systematic Review of the Ongoing Development of Cyclin-Dependent Kinase Inhibitors for Multiple Cancer Therapy. *J Med Chem.* 2022;65:6390–6418.
5 Xie Z, Hou S, Yang X, et al. Lessons Learned from Past Cyclin-Dependent Kinase Drug Discovery Efforts. *J Med Chem.* 2022;65:6356–6389.
6 Sonawane YA, Taylor MA, Napoleon JV, Rana S, Contreras JI, Natarajan A. Cyclin Dependent Kinase 9 Inhibitors for Cancer Therapy. *J Med Chem.* 2016;59:8667–8684.
7 Wu T, Qin Z, Tian Y, et al. Recent Developments in the Biology and Medicinal Chemistry of CDK9 Inhibitors: An Update. *J Med Chem.* 2020;63:13228–13257.
8 Boffo S, Damato A, Alfano L, Giordano A. CDK9 inhibitors in acute myeloid leukemia. *J Exp Clin Cancer Res.* 2018;37:36.
9 Borowczak J, Szczerbowski K, Stec E, Grzanka D, Szylberg L. CDK9: Therapeutic Perspective in HCC Therapy. *Curr Cancer Drug Targets.* 2020;20:318–324.
10 Wei D, Wang H, Zeng Q, et al. Discovery of Potent and Selective CDK9 Degraders for Targeting Transcription Regulation in Triple-Negative Breast Cancer. *J Med Chem.* 2021;64:14822–14847.
11 Wang X, Yu C, Wang C, et al. Novel cyclin-dependent kinase 9 (CDK9) inhibitor with suppression of cancer stemness activity against non-small-cell lung cancer. *Eur J Med Chem.* 2019;181, 111535.
12 Zhang B, Gojo I, Fenton RG. Myeloid cell factor-1 is a critical survival factor for multiple myeloma. *Blood.* 2002;99:1885–1893.
13 Chao SH, Price DH. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *J Biol Chem.* 2001;276:31793–31799.
14 Johnson AJ, Yeh YY, Smith LL, et al. The novel cyclin-dependent kinase inhibitor dinaciclib (SCH727965) promotes apoptosis and abrogates microenvironmental cytokine protection in chronic lymphocytic leukemia cells. *Leukemia.* 2012;26:2554–2557.
15 Wyat PG, Woodhead AJ, Berdini V, et al. Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J Med Chem.* 2008;51:4986–4999.
16 Lin TS, Ruppert AS, Johnson AJ, et al. Phase II study of flavopiridol in relapsed chronic lymphocytic leukemia demonstrating high response rates in genetically high-risk disease. *J Clin Oncol.* 2009;27:6012–6018.
17 Tan AR, Yang X, Berman A, et al. Phase I trial of the cyclin-dependent kinase inhibitor flavopiridol in combination with docetaxel in patients with metastatic breast cancer. *Clin Cancer Res.* 2004;10:5038–5047.
18 George S, Kasimis BS, Cogswell J, et al. Phase I study of flavopiridol in combination with Paclitaxel and Carboplatin in patients with non-small-cell lung cancer. *Clin Lung Cancer.* 2008;9:160–165.
19 Nelson DM, Joseph B, Hillion J, Segal J, Karp JE, Resar LM. Flavopiridol induces BCL-2 expression and represses oncogenic transcription factors in leukemic blasts from adults with refractory acute myeloid leukemia. *Leuk Lymphoma.* 2011;52:1999–2006.
20 Barlaam B, Casella R, Cidado J, et al. Discovery of AZD4573, a Potent and Selective Inhibitor of CDK9 That Enables Short Duration of Target Engagement for the Treatment of Hematological Malignancies. *J Med Chem.* 2020;63:15564–15590.
21 Lucking U, Scholz A, Lienau P, et al. Identification of Atuveciclib (BAY 1143572), the First Highly Selective, Clinical PTEFb/CDK9 Inhibitor for the Treatment of Cancer. *ChemMedChem.* 2017;12:1776–1793.
22 Luecking UT, Scholz A, Kosemund D, Bohlmann R, Brands M. Abstract 984: Identification of potent and highly selective PTEFb inhibitor BAY 1251152 for the treatment of cancer: from p.o. to i.v. application via scaffold hops. *Cancer Res.* 2017;77:984.
23 Braal CL, Jongbloed EM, Wilting SM, Mathijssen RHJ, Koolen SLW, Jager A. Inhibiting CDK4/6 in Breast Cancer with Palbociclib, Ribociclib, and Abemaciclib: Similarities and Differences. *Drugs.* 2021;81:317–331.
24 Trilaciclib DS. First Approval. *Drugs.* 2021;81:867–874.
25 Nemeth G, Greff Z, Sipos A, et al. Synthesis and evaluation of phosphorus containing, specific CDK9/CycT1 inhibitors. *J Med Chem.* 2014;57:3939–3965.
26 Ghanem NM, Farouk F, George RF, Abbas SES, El-Badry OM. Design and synthesis of novel imidazo[4,5-b]pyridine based compounds as potent anticancer agents with CDK9 inhibitory activity. *Bioorg Chem.* 2018;80:565–576.
27 Guo Q, Zhang H, Deng Y, et al. Ligand- and structural-based discovery of potential small molecules that target the colchicine site of tubulin for cancer treatment. *Eur J Med Chem.* 2020;196, 112328.
28 Wang L, Gu Q, Zheng X, et al. Discovery of new selective human aldose reductase inhibitors through virtual screening multiple binding pocket conformations. *J Chem Inf Model.* 2013;53:2409–2422.
29 Wang L, Pang X, Li Y, Zhang Z, Tan W. RADER: a RApid DEcoy Retriever to facilitate decoy based assessment of virtual screening. *Bioinformatics.* 2017;33:1235–1237.
30 Shang J, Dai X, Li Y, Pistolozzi M, Wang L. HybridSim-VS: a web server for large-scale ligand-based virtual screening using hybrid similarity recognition techniques. *Bioinformatics.* 2017;33:3480–3481.
31 Jiang Z, Xu J, Yan A, Wang L. A comprehensive comparative assessment of 3D molecular similarity tools in ligand-based virtual screening. *Brief Bioinform.* 2021;22:bbab231.
32 Ye Q, Chai X, Jiang D, et al. Identification of active molecules against Mycobacterium tuberculosis through machine learning. *Brief Bioinform.* 2021;22.
33 Wang L, Chen L, Yu M, et al. Discovering new mTOR inhibitors for cancer treatment through virtual screening methods and in vitro assays. *Sci Rep.* 2016;6:18987.
34 Guo Q, Luo Y, Zhai S, et al. Discovery, biological evaluation, structure-activity relationships and mechanism of action of pyrazolo[3,4-b]pyridin-6-one derivatives as a new class of anticancer agents. *Org Biomol Chem.* 2019;17:6201–6214.
35 Randjelovic J, Eric S, Savic V. In silico design of small molecule inhibitors of CDK9/cyclin T1 interaction. *J Mol Graph Model.* 2014;50:100–112.
36 Hussain A, Verma CK. Ligand- and structure-based pharmacophore modeling, docking study reveals 2-[[4-[6-(isopropylamino) pyrimidin-4-yl]-1H-pyrrolo[2,3-b] pyridin-6-yl] amino] ethanol as a potential anticancer agent of CDK9/cyclin T1 kinase. *J Cancer Res Ther.* 2019;15:1131–1140.
37 Wu M, Han J, Liu Z, et al. Identification of novel CDK 9 inhibitors based on virtual screening, molecular dynamics simulation, and biological evaluation. *Life Sci.* 2020;258, 118228.
38 Shao H, Foley DW, Huang S, et al. Structure-based design of highly selective 2,4,5-trisubstituted pyrimidine CDK9 inhibitors as anti-cancer agents. *Eur J Med Chem.* 2021;214, 113244.
39 Zernov VV, Balakin KV, Ivaschenko AA, Savchuk NP, Pletnev IV. Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-

likeness, and enzyme inhibition predictions. *J Chem Inf Comput Sci*. 2003;43:2048–2056.
40 Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *ACM*. 2016. https://doi.org/10.48550/arXiv.1603.02754.
41 Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. 2016. https://doi.org/10.48550/arXiv.1609.02907.
42 Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. *Graph Attention Networks*; 2017. https://doi.org/10.48550/arXiv.1710.10903.
43 Xiong Z, Wang D, Liu X, et al. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J Med Chem*. 2020;63:8749–8760.
44 Cai H, Zhang H. Zhao D, Wu J, Wang L, FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. 2022. https://doi.org/10.48550/arXiv.2205.03834.
45 Luo Y, Zeng R, Guo Q, Xu J, Sun X, Wang L. Identifying a novel anticancer agent with microtubule-stabilizing effects through computational cell-based bioactivity prediction models and bioassays. *Org Biomol Chem*. 2019;17:1519–1530.
46 He S, Zhao D, Ling Y, et al. Machine Learning Enables Accurate and Rapid Prediction of Active Molecules Against Breast Cancer Cells. *Front Pharmacol*. 2021;12, 796534.
47 Wang L, Le X, Li L, et al. Discovering new agents active against methicillin-resistant Staphylococcus aureus with ligand-based approaches. *J Chem Inf Model*. 2014;54:3186–3197.
48 Wang L, Chen L, Liu Z, Zheng M, Gu Q, Xu J. Predicting mTOR inhibitors with a classifier using recursive partitioning and Naive Bayesian approaches. *PLoS ONE*. 2014;9, e95221.
49 Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem*. 1996;39:2887–2893.
50 Wu Z, Ramsundar B, Feinberg EN, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. 2018;9:513–530.
51 Rifaioglu AS, Nalbat E, Atalay V, Martin MJ, Cetin-Atalay R, Dogan T, et al. DEEPScreen: high performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem Sci*. 2020;11:2531–2557.
52 Wu Z, Jiang D, Hsieh CY, Chen G, Liao B, Cao D, Hou T. Hyperbolic relational graph convolution networks plus: a simple but highly efficient QSAR-modeling method. *Brief Bioinform*. 2021;22:bbab112.
53 Wu Z, Ramsundar B, Feinberg EN, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. 2017;9:513–530.
54 Scholz A, Oellerich T, Hussain A, et al. BAY 1143572, a first-in-class, highly selective, potent and orally available inhibitor of PTEFb/CDK9 currently in Phase I, shows convincing anti-tumor activity in preclinical models of acute myeloid leukemia (AML). *Cancer Res*. 2016;76.
55 Kumar SK, LaPlant B, Chng WJ, et al. Dinaciclib, a novel CDK inhibitor, demonstrates encouraging single-agent activity in patients with relapsed multiple myeloma. *Blood*. 2015;125:443–448.
56 Han X, Song N, Saidahmatov A, et al. Rational Design and Development of Novel CDK9 Inhibitors for the Treatment of Acute Myeloid Leukemia. *J Med Chem*. 2021;64:14647–14663.
57 Xu J, Li H, Wang X, et al. Discovery of coumarin derivatives as potent and selective cyclin-dependent kinase 9 (CDK9) inhibitors with high antitumour activity. *Eur J Med Chem*. 2020;200, 112424.
58 Wang J, Li T, Zhao T, et al. Design of wogonin-inspired selective cyclin-dependent kinase 9 (CDK9) inhibitors with potent in vitro and in vivo antitumor activity. *Eur J Med Chem*. 2019;178:782–801.
59 Sloczynska K, Gunia-Krzyzak A, Koczurkiewicz P, et al. Metabolic stability and its role in the discovery of new chemical entities. *Acta Pharm*. 2019;69:345–361.
60 Wu CF, Wang QC, Chen R, et al. Synthesis and bioevaluation of diaryl urea derivatives as potential antitumor agents for the treatment of human colorectal cancer. *Eur J Med Chem*. 2022;229, 114055.
61 Zhai S, Zhang H, Chen R, et al. Design, synthesis and biological evaluation of novel hybrids targeting mTOR and HDACs for potential treatment of hepatocellular carcinoma. *Eur J Med Chem*. 2021;225, 113824.
62 Mendez D, Gaulton A, Bento AP, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res*. 2019;47:D930–D940.
63 Hole AJ, Baumli S, Shao H, et al. Comparative structural and functional studies of 4-(thiazol-5-yl)-2-(phenylamino)pyrimidine-5-carbonitrile CDK9 inhibitors suggest the basis for isotype selectivity. *J Med Chem*. 2013;56:660–670.