

Virtual Screening of Antitumor Inhibitors Targeting BRD4 Based on Machine Learning Methods

Keliang Wu

Nanyang Institute of Technology

Chenghua Zhang

School of Basic Medical Sciences, North Sichuan Medical College

Bing He

Chengdu Normal University

Huanxin Li

Nanyang Institute of Technology

Shan Tang

Chengdu Normal University

Tao Han

Chengdu Normal University

Bingke Li (✉ libingke86@nyist.edu.cn)

Nanyang Institute of Technology <https://orcid.org/0000-0002-4324-6073>

Research Article

Keywords: machine learning methods, virtual screening, BRD4, antitumor inhibitors, molecular dynamics simulations.

Posted Date: September 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-768419/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at ChemistrySelect on February 1st, 2022. See the published version at <https://doi.org/10.1002/slct.202104054>.

Abstract

BRD4 is a hot antitumor target. In this study, three kinds of machine learning methods were used to establish classification models of BRD4 inhibitors, achieving satisfactory prediction performance. Through comparison, random forest model worked best, the parameters of which were also optimized. Then, the best random forest model was applied to perform virtual screening against ZINC database and a total of 89 potential compounds with BRD4 inhibitory activity were eventually identified. Further, seven molecules were chosen from the hits, and a docking calculation was carried out for each molecule, showing a strong interaction between ligand and BRD4. Subsequently, these molecules were evaluated by molecular dynamics simulations, all having certain binding stability. The results have proved the effectiveness of the developed models based on machine learning methods and the molecules filtered by virtual screening not only have a significant guiding in practice for the molecular design and synthesis, but also can provide great possibility for the discoveries and final approvals of anti-cancer drugs targeting BRD4.

Introduction

The bromodomain and extra-terminal domain (BET) family of proteins acts as the “readers” of acetylated lysines (KAc) on chromatin and plays a crucial role in regulating some cellular activities including transcription, DNA replication, cell cycle progression, cell proliferation, differentiation, migration and apoptosis [1, 2]. This family which contains four isoforms, namely bromodomain 2 (BRD2), BRD3, BRD4, and bromodomain testis-specific, shows high sequence similarity, and BRD4 is the most abundant and extensively investigated protein [3, 4]. BRD4 has a highly conserved modular architecture, consisting of an extra C-terminal domain and two tandem N-terminal bromodomains (BD1, BD2) that can recognize KAc residues [5, 6].

Strongly associated with several types of malignancies, such as melanoma [7], lymphoma [8], leukaemia [9], glioblastoma [10], neuroblastoma [11], breast cancer [12], prostate cancer [13], multiple myeloma [14], thyroid carcinoma [15], hepatocellular carcinoma [16], non-small cell lung cancer [17], coupled with other diseases including inflammation [5] and HIV [18], BRD4 is emerging as a novel promising target for lead identification, and its KAc binding pocket affords a druggable site fit for developing small molecule inhibitors [19]. Many structurally diverse BRD4 inhibitors (BRD4is) have been reported over the last few years, which can be mainly divided into eight categories [20]: (1) pyridine derivatives [21], (2) isoxazole derivatives [22], (3) 4-acyl pyrrole derivatives [23], (4) triazoloazepine derivatives [24], (5) triazolopyrazine derivatives [25], (6) 2-thiazolidinone derivatives [26], (7) tetrahydroquinoline derivatives [27], (8) others [28, 29]. A majority of these molecules mimic the acetyl moiety, occupy the central hydrophobic binding cavity and form a direct hydrogen bond with Asn140 at the top of the pocket and indirect hydrogen bonds with Tyr97 via well-ordered water molecules [30, 31]. Although more than a dozen of BRD4 drugs are currently or have been successfully progressed into different phases of clinical trials, none of them have yet been granted an FDA-approval thus far [32].

Hence, search for novel BRD4is with high efficacy and low toxicity is still a highly active research field. More recently, a number of computational drug design approaches, e.g. virtual screening (VS), computer-guided optimization of lead molecules, quantitative structure–activity relationship (QSAR), and machine learning (ML), have been devoted to the discovery and analysis of BRD4is, receiving increasing attention.

Tumdam et al. [33] performed VS by Glide module of Schrödinger suite, carried out molecular dynamics (MD) simulations and molecular mechanics/Poisson–Boltzmann surface area calculations, and finally produced two novel promising inhibitors towards BRD4-BD1. Meanwhile, Yan et al. [34] made an attempt to find BRD4is using pharmacophore based VS by Discovery Studio (DS) 3.1, molecular docking, and MD simulations, obtaining two molecules with potent anticancer activity. Additionally, Tahir et al. [35] conducted 3D-QSAR studies employing comparative molecular field analysis and comparative molecular similarity indices analysis approaches by SYBYL-X 2.1.1 molecular modeling package, who identified prominent structural features and designed ten new BRD4is based on the best predictive models. Moreover, Xing et al. [36] demonstrated a novel, structure-based VS approach utilizing ML algorithms on the basis of binding patterns, established a target-specific scoring function BRD4LGR, and generated 17 new compounds, which exhibited significant inhibitory potencies.

In this work, three kinds of ML methods: random forest (RF) [37], k-nearest neighbor (k-NN) [38] and C4.5 decision tree (DT) [39] were employed to develop classification models. By tuning parameters, performing VS, carrying out docking calculations and molecular dynamics simulations, we gained one promising BRD4 inhibitor.

Materials And Methods

For comparison, a total of 757 BRD4is and 757 BRD4 non-inhibitors (non-BRD4is) were collected. All of BRD4is were download from Thomson Reuters Integrity database (<https://integrity.thomson-pharma.com>), while the non-BRD4is were abstracted from the MDL Drug Data Report (<http://www.mdli.com>, MDDR) database by k-means clustering [40].

The data sets were further separated into training sets (973 molecules, 477 BRD4is and 496 non-BRD4is) and testing sets (541 molecules, 280 BRD4is and 261 non-BRD4is) in relation to their distribution in the chemical space [41].

The calculation of 189 molecular descriptors in the present study was depended on the 3D structure of each compound, and the calculating program was developed by our laboratory. More information about the 189 molecular descriptors, can be obtained in Table S1 from Supporting Information.

The RF, k-NN and C4.5 DT models were generated by open source Fortran codes on the network. More information about the three methods can be easily found in these literatures [37–39].

The 3D X-ray crystal structure of the first bromodomain of human BRD4 in complex with a 3,5-dimethylisoxazol ligand (pdb id: 3SVG, resolution = 1.68 Å) was downloaded from the protein data bank (<https://www.rcsb.org/>), and some tasks such as water deletion, impurity atom removal and hydrogenation were performed for 3SVG.

The 89 hits from virtual screening were docked to BRD4 protein by Autodock Vina software [42], and a grid box of size 16 × 16 × 16 Å with coordinates X = 30.313, Y = 15.738 and Z = 0.0 was selected to ensure ligand molecules were completely encapsulated.

Conventional MD simulations were performed on the docked complexes using the Gromacs 2019.5 software package [43]. The Amber14SB force field [44] was applied to describe the protein, while the generalized amber force field (GAFF) [45] to the inhibitors, and the TIP3P water model was used. The topology files of small molecules were generated by ACPYPE [46], and the charge was calculated as the restrained electro static potential (RESP) charge [47] at HF/6-31G(d) level. First, the complex was placed in a box with a boundary of 15 Å, and then water molecules were added. Afterwards, when the energy was minimized, all non-hydrogen atoms of the protein and the inhibitor were restricted to a force constant of 1000 kJ•mol⁻¹•nm⁻¹ with a maximum drop rate of 5000 steps and a conjugate gradient, respectively. Only Alpha C atoms were restricted to optimization without restriction. In the first 100 ps of NVT simulation, the system was heated from 0 K to 300 K, and remained at 300 K for the next 500 ps. In the next 2.5 ns NPT simulation, the position limiting force constant was successively reduced from 1000 to 0 kJ•mol⁻¹•nm⁻¹, and the equilibrium system was simulated for 100 ns. The Particle Mesh Ewald (PME) method [48] was used to describe long range electrostatic interactions. All H related keys are limited by linear constraint solver (LCS) algorithm, with a step size of 2 fs and data saved every 10 ps.

For the binding free energy calculations, MM/PBSA method [49] was chosen, which is expressed as:

$$\Delta G_{\text{bind}} = \Delta G_{\text{MM}} + \Delta G_{\text{PB}} + \Delta G_{\text{SA}} - T\Delta S \quad (1)$$

Here, ΔG_{MM} is the molecular mechanics interaction (sum of electrostatic and van der Waals interaction), ΔG_{PB} and ΔG_{SA} correspond to polar and nonpolar solvation energies, respectively. $T\Delta S$ is the entropic contribution. Due to the high computational time, the entropy contribution was not considered. The binding free energy calculations were made using the `g_mmpbsa` [50] and the final 30 ns data of MD simulations were used.

Results And Discussion

The Comparison of the Prediction Accuracies

Table 1 The comparison on the prediction accuracies of BRD4is and non-BRD4is from RF, k-NN and C4.5 DT in the testing set

Method	Parameter	BRD4is			non-BRD4is			Q (%)	MCC
		TP	FN	SE (%)	TN	FP	SP (%)		
RF	$M_{try}=13$	274	6	97.86	252	9	96.55	97.23	0.9445
k-NN	$k = 6$ eq	271	9	96.79	232	29	88.89	92.98	0.8614
C4.5 DT	/	259	21	92.50	226	35	86.59	89.65	0.7934

Three methods, RF, k-NN and C4.5 DT, were utilized to build classification prediction models for the same training set, and the performance of the models were evaluated by the same testing set. The data are summarized in Table 1.

Among them, M_{try} is the parameter of RF method, and its default value is 13, which is approximately equal to the square root of the number of descriptors (189) [51]. k is the parameter of k-NN method, whose value was optimized by the internal parameter selection program. TP (True positive) denotes the number of positive samples predicted correctly, TN (True negative) denotes the number of negative samples predicted correctly, FP (False positive) denotes the number of negative samples mispredicted as positive, and FN (False negatives) denotes the number of positive samples mispredicted as negative. For the predicting outcomes in this article, TP means the correct number of BRD4is, TN means the correct number of non-BRD4is, FN means the wrong number of BRD4is, and FP means the wrong number of non-BRD4is. At the same time, there are several precision functions to measure the prediction performance, including SE (sensitivity), SP (specificity), Q (total prediction accuracy) and MCC (Matthews correlation coefficient) [52]. These functions have the following relationship with the previous variables:

$$SE = \frac{TP}{TP+FN} \quad (2)$$

$$SP = \frac{TN}{TN+FP} \quad (3)$$

$$Q = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(FN+FP)}} \quad (5)$$

From Table 1, it can be seen that RF model possesses the highest Q value (97.23%) and the highest MCC value (0.9445). That is to say, the prediction accuracy of the RF model is the best of all. Compared with the models established by k-NN and C4.5 DT, the RF model has greater advantages.

Optimization of Parameters for RF method

Different RF models were generated based on different values of parameters M_{try} and N_{tree} . The parameters of the best-performing RF model were chosen by comparing prediction error rates (ERs) of testing set for all the models. The effect of the different values of M_{try} (1-189) and N_{tree} (100-3000) on the prediction ERs of the testing set for different RF models are illustrated in Figure 1(A) and 1(B), respectively.

As can be seen from Figure 1(A), for different prediction models established by RF method with different parameters, when M_{try} value is 9 (prediction ER of training set is 6.06%) and M_{try} value is 18 (prediction ER of training set is 6.37%), the prediction ER of the testing set reaches the lowest 2.40%. However, when the value is taken as 9, the prediction ER of the corresponding training set is lower than 18, so 9 was selected as the optimal solution of parameter M_{try} .

After determining the parameter M_{try} , the M_{try} value was fixed at 9 and the N_{tree} value was constantly changed. The results are displayed in Figure 1(B), which illustrate that the prediction ER of the testing set is the lowest 2.22% when N_{tree} equals to 300 or 400. Nevertheless, when N_{tree} is 400, the prediction ER of the corresponding training set is lower (300: 6.17% VS 400: 5.96%), therefore, 400 is more suitable as the optimal parameter.

Altogether, when the parameters of RF method are set to $M_{try} = 9$ and $N_{tree} = 400$, the associated model achieves the best prediction performance, and, by extension, the Q of the training set is 94.04%, while the Q of the testing set is 97.60%. Compared with the model before the parameters optimization, the prediction accuracy has been improved.

Evaluation of the Optimal RF Model

According to the special feature selection procedure of the RF method, the model with the optimum parameters was processed further. 25 descriptors (listed in Table 2) most relevant to the properties of BRD4is were screened out from the 189 descriptors, which can serve as a theoretical basis for structural modifications. Each of these descriptors has its corresponding contribution rate, arranged in Figure 2(A) on the basis of its relative importance.

Table 2 The most relevant 25 descriptors identified by the best RF model for the prediction of BRD4is

Descriptor	Description
S(16)	Atom-type Estate sum for -CH ₃
Q _{N, Max}	Most positive charge on N atoms
Q _{C, Min}	Most negative charge on C atoms
⁴ χ _{PC}	Simple molecular connectivity Chi indices for path/cluster
Hlb	Hydrophilic-Hydrophobic balance
Q _{H, Max}	Most positive charge on H atoms
³ χ _C	Simple molecular connectivity Chi indices for cluster
S(26)	Atom-type Estate sum for :C:-
S(22)	Atom-type Estate sum for >CH-
S(39)	Atom-type Estate sum for -OH
Tcent	Centric Index
S(34)	Atom-type Estate sum for =N-
S(12)	Atom-type H Estate sum for CH _n (Saturated)
Q _{H, Min}	Most negative charge on H atoms
Capy	Capacity factor
Q _{O, Max}	Most positive charge on O atoms
³ χ _C ^v	Valence molecular connectivity Chi indices for cluster
⁴ χ _{PC} ^v	Valence molecular connectivity Chi indices for path/cluster
Shpl	Hydrophilic region
S(10)	Atom-type H Estate sum for :CH: (sp ² , aromatic)
S(1)	Atom-type H Estate sum for -OH
S(25)	Atom-type Estate sum for =C<
S(35)	Atom-type Estate sum for :N:
Shpb	Hydrophobic region
Q _{O, SS}	Sum of squares of charges on O atoms

As shown in Table 2 and Figure 2(A), the importance of these 25 descriptors decreases in turn, and the top three are $S(16)$, $Q_{N, \text{Max}}$ and $Q_{C, \text{Min}}$. It is quite clear that these three characteristics have very important reference value in predicting the BRD4is, which sequentially stand for the sum of the electric topological states of $-\text{CH}_3$ atom type, the largest positive charge on N atom and the smallest negative charge on C atom.

Figure 2(B) displays the distribution of 541 molecules in the testing set from the established RF optimal model. As can be seen in the figure, the classification boundary line of the model can separate BRD4is from non-BRD4is very well.

By plotting the receiver operating characteristic (ROC) curve [53], we can further analyze and evaluate the discriminant effect of binary classification model. ROC curve combines both SE and SP together. With the change of prediction probability threshold, many pairs of SE and "1-SP" will be produced. If SE is taken as the ordinate and "1-SP" as the abscissa, the ROC curve can be drawn by connecting each point, and the points on the curve represent the compromise between SE and SP when the prediction probability threshold is constantly changed. There is also a very important index to evaluate the prediction ability of classification model: the area under the ROC curve (AUC). The range for the AUC value is from 0.5 to 1, and the larger the value, the better the classification performance of the model is. The ROC curves of the training set and the testing set for RF optimal model in this paper are emerged in Figure 3. The curve fitting exhibits that the AUC value in the training set and the testing set is 0.981 and 0.993, separately, all reflecting excellent prediction performance of the RF model.

Virtual Screening of BRD4is

The above optimal RF model was applied to screen virtually about 100,000 compounds from the "drug-like" subset of ZINC database. Finally, 89 promising drug molecules were obtained, the details of which are given in Table S2 of Supporting Information. It can be found that some compounds share the same skeleton structure, for instance, ZINC00126622 and ZINC00126628, as exemplified in Figure 4.

The results indicated that the optimal RF model has filtered a number of highly useful structures from the database, such as sulfonyl, triazole and isoxazole, which have better potential inhibitory activity.

Molecular docking calculations

Using Autodock Vina software, the 89 molecules were docked to BRD4 protein afterwards. Seven molecules with binding energies less than or equal to -8 kcal/mol were selected in order of ascending energy. At the same time, the binding energy of the original ligand was also calculated, which is -7.4 kcal/mol, higher than that of the seven molecules above-mentioned. Their structures and binding energies scores are detailed in Table S3 from Supporting Information.

The Table S3 shows that the molecules chosen by us have stronger affinity with BRD4 protein than the original ligand and therefore can be used for further structural modification, chemical synthesis and

biological testing. Among them, compound ZINC59239754 has isoxazole structure, which belongs to one category of known BRD4 inhibitors. The binding energy between ZINC59239754 and protein ranks third, so the complex is relatively stable.

Molecular dynamics simulations

In order to verify the binding stability of the above seven molecules to BRD4, molecular dynamics simulations were conducted respectively at 100 ns, not only on the complexes between the seven molecules and BRD4 after docking, but also on the complex of the original ligand with BRD4. The simulation results (Figure 5) show that the root mean square deviations (RMSD) for backbones of all systems hold steady after 60 ns, and the radius of gyration (Rg) is also stable at about 15 Å. The binding free energies obtained by molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) calculations (Table 3) indicate that compound ZINC59239754 has a lower binding free energy than the other six compounds, which is close to the original ligand. Therefore, compound ZINC59239754 is more likely to be a potential BRD4 inhibitor.

In addition, the ranking of binding free energies by molecular dynamics simulations is very different from the Molecular docking calculations

Table 3 Average RMSD, Rg and binding free energy of the seven top-ranked molecules-BRD4 complexes and the original ligand 3,5-dimethylisoxazol-BRD4 complex

Complex	RMSD(Å)	Radius of gyration(Å)	Binding free energy (kJ/mol)
ZINC67473070-BRD4	1.33±0.28	15.20±0.08	-81.59±0.68
ZINC00481768-BRD4	1.25±0.27	15.24±0.08	-67.60±0.76
ZINC59239754-BRD4	1.05±0.19	15.24±0.07	-100.54±2.56
ZINC04487544-BRD4	1.31±0.31	15.21±0.08	-82.51±0.73
ZINC22055514-BRD4	1.26±0.19	15.19±0.08	-90.93±0.60
ZINC71783667-BRD4	1.38±0.27	15.29±0.08	-75.53±0.70
ZINC71782051-BRD4	1.24±0.21	15.24±0.07	-79.38±0.56
The original ligand-BRD4	1.21±0.18	15.21±0.06	-99.75±1.06

Conclusion

Research on small molecule inhibitors targeting BRD4 protein is a hot topic in cancer therapy.

In this paper, three ML methods, RF, k-NN and C4.5 DT, were applied for developing the classification models of BRD4is, achieving satisfactory results. Through comparative analysis, we found that the RF model had better prediction performance, and the parameters of the model were further debugged to

obtain the optimal RF model followed with the corresponding 25 highest ranked molecular descriptors most relevant to inhibition of BRD4.

Then, the best RF predictive model was used to screen virtually the ZINC database, and 89 potential compounds with BRD4 inhibitory activity were collected. One of the screened compounds, with typical characteristic structure, was selected as the mother nucleus for structural optimization and molecular design. Subsequently, we designed three molecules and carried out docking calculations. The binding free energies gave a good answer that each of the three molecules modified according to the VS hit had strong interaction with BRD4 protein.

In summary, the above facts have proved that the models established by the ML methods and the molecules screened virtually have a strong guiding role in the practice of molecular design and synthesis, raising the possibility for the discoveries and final approvals of anti-cancer drugs targeting BRD4.

Declarations

Funding: This study was supported by National College Students Innovation and Entrepreneurship Training Program (No.201914389012), the Interdisciplinary Sciences Project, Nanyang Institute of Technology, and the Doctoral Scientific Research Foundation for Returned Scholars, Nanyang Institute of Technology.

Conflicts of interest/Competing interests: The authors declare that they have no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors. Informed consent was obtained from all individual participants included in the study.

Availability of data and material: All material used and data generated or analysed during this study are included in this published article and its supplementary information files.

Code availability: The software or custom code used or analysed during the current study are available from the corresponding authors on reasonable request.

Authors' contributions: Keliang Wu, Writing – original draft, Data curation; Chenghua Zhang, Writing – original draft, Validation; Bing He, Methodology, Software; Huanxin Li, Project administration, Data curation; Shan Tang, Funding acquisition, Visualization; Tao Han, Writing – review & editing, Resources; Bingke Li: Writing – review & editing, Conceptualization, Funding acquisition.

References

1. Fujisawa T, Filippakopoulos P (2017) Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat Rev Mol Cell Biol* 18:246-262.
2. Shi J, Vakoc CR (2014) The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol Cell* 54:728-736.

3. Mahesh B, Anup D, Keiko O (2016) Chromatin landscape of the IRF genes and role of the epigenetic reader BRD4. *J Interferon Cytokine Res* 36:470-475.
4. Filippakopoulos P, Picaud S, Mangos M, et al (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* 149:214-231.
5. Liu Z, Tian B, Chen H, et al (2018) Discovery of potent and selective BRD4 inhibitors capable of blocking TLR3-induced acute airway inflammation. *Eur J Med Chem* 151:450-461.
6. Wu SY, Chiang CM (2007) The double bromodomain-containing chromatin adaptor BRD4 and transcriptional regulation. *J Biol Chem* 282:13141-13145.
7. Segura MF, Fontanals CB, Gaziel SA, et al (2013) BRD4 sustains melanoma proliferation and represents a new target for epigenetic therapy. *Cancer Res* 73:6264-6276.
8. Gopalakrishnan R, Matta H, Tolani B, et al (2015) Immunomodulatory drugs target IKZF1-IRF4-MYC axis in primary effusion lymphoma in a cereblon-dependent manner and display synergistic cytotoxicity with BRD4 inhibitors. *Oncogene* 35:1797-1810.
9. Zuber J, Shi J, Wang E, et al (2011) RNAi screen identifies BRD4 as a therapeutic target in acute myeloid leukaemia. *Nature* 478:524-528.
10. Pastori C, Kapranov P, Penas C, et al (2015) The bromodomain protein BRD4 controls HOTAIR, a long noncoding RNA essential for glioblastoma proliferation. *Proc Natl Acad Sci* 112:8326-8331.
11. Puissant A, Frumm SM, Alexe G, et al (2013) Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Discov* 3:308-323.
12. Andrieu G, Tran AH, Strissel KJ, et al (2016) BRD4 regulates breast cancer dissemination through Jagged1/Notch1 signaling. *Cancer Res* 76:6555-6567.
13. Dai X, Gan W, Li X, et al (2017) Prostate cancer-associated SPOP mutations confer resistance to BET inhibitors through stabilization of BRD4. *Nat Med* 23:1063-1071.
14. Delmore JE, Issa GC, Lemieux ME, et al (2011) BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 146:904-917.
15. Xu K, Chen D, Qian D, et al (2019) a novel BRD4 inhibitor, suppresses human thyroid carcinoma cell growth in vitro and in vivo. *Biochem Biophys Res Commun* 499:531-537.
16. Li GQ, Guo WZ, Zhang Y, et al (2015) Suppression of BRD4 inhibits human hepatocellular carcinoma by repressing MYC and enhancing BIM expression. *Oncotarget* 7:2462-2474.
17. Liao YF, Wu YB, Long X, et al (2016) High level of BRD4 promotes non-small cell lung cancer progression. *Oncotarget* 7:9491-9500.
18. Abner E, Stoszko M, Zeng L, et al (2018) A new quinoline BRD4 inhibitor targets a distinct latent HIV-1 reservoir for reactivation from other "shock" drugs. *J Virol* 92:e02056-17.
19. Feng Y, Xiao S, Chen Y, et al (2018) Design, synthesis and biological evaluation of benzo[cd]indol-2(1H)-ones derivatives as BRD4 inhibitors. *Eur J Med Chem* 152:264-273.
20. Duan Y, Guan Y, Qin W, et al (2018) Targeting BRD4 for cancer therapy: inhibitors and degraders. *Med Chem Comm* 9:1779-1802.

21. Wang L, Pratt JK, Soltwedel T, et al (2017) Fragment-based, structure-enabled discovery of novel pyridones and pyridone macrocycles as potent bromodomain and extra-terminal domain (BET) family bromodomain inhibitors. *J Med Chem* 60:3828-3850.
22. Zhang M, Zhang Y, Song M, et al (2018) Structure-based discovery and optimization of benzo[d]isoxazole derivatives as potent and selective BET inhibitors for potential treatment of castration-resistant prostate cancer (CRPC). *J Med Chem* 61:3037-3058.
23. Hügler M, Lucas X, Weitzel G, et al (2016) 4-acyl pyrrole derivatives yield novel vectors for designing inhibitors of the acetyl-lysine recognition site of BRD4(1). *J Med Chem* 59:1518-1530.
24. Filippakopoulos P, Qi J, Picaud S, et al (2010) Selective inhibition of BET bromodomains. *Nature* 468:1067-1073.
25. Abdel Magid, AF (2016) Inhibitors of BRD4 as potential cancer therapy. *ACS Med Chem Lett* 7:728-729.
26. Zhao L, Wang Y, Cao D, et al (2015) Fragment-based drug discovery of 2-thiazolidinones as BRD4 inhibitors: 2. Structure-based optimization. *J Med Chem* 58:1281-1297.
27. Gosmini R, Nguyen VL, Toum J, et al (2014) The discovery of I-BET726 (GSK1324726A), a potent tetrahydroquinoline ApoA1 up-regulator and selective BET bromodomain inhibitor. *J Med Chem* 57:8111-8131.
28. Qin C, Hu Y, Zhou B, et al (2018) Discovery of QCA570 as an exceptionally potent and efficacious proteolysis targeting chimera (PROTAC) degrader of the bromodomain and extra-terminal (BET) proteins capable of inducing complete and durable tumor regression. *J Med Chem* 61:6685-6704.
29. Zhong HJ, Lu L, Leung KH, et al (2015) An iridium(iii)-based irreversible protein–protein interaction inhibitor of BRD4 as a potent anticancer agent. *Chem Sci* 6:5400-5408.
30. Donati B, Lorenzini E, Ciarrocchi A (2018) BRD4 and cancer: going beyond transcriptional regulation. *Mol Cancer* 17:164.
31. Nicodeme E, Jeffrey KL, Schaefer U, et al (2010) Suppression of inflammation by a synthetic histone mimic. *Nature* 468:1119-1123.
32. Pervaiz M, Mishra P, Günther S (2018) Bromodomain drug discovery - the past, the present, and the future. *Chem Rec* 18:1808-1817.
33. Tumdam R, Kumar A, Subbarao N, et al (2018) In silico study directed towards identification of novel high-affinity inhibitors targeting an oncogenic protein: BRD4-BD1. *SAR QSAR Environ Res* 29:975-996.
34. Yan G, Hou M, Luo J, et al (2018) Pharmacophore-based virtual screening, molecular docking, molecular dynamics simulation, and biological evaluation for the discovery of novel BRD4 inhibitors. *Chem Biol Drug Des* 91:478-490.
35. Tahir A, Alharthy DR, Naseem S, et al (2018) Investigations of structural requirements for BRD4 inhibitors through ligand- and structure-based 3D QSAR approaches. *Molecules* 23:1527.

36. Xing J, Lu W, Liu R, et al (2017) Machine-learning-assisted approach for discovering novel inhibitors targeting bromodomain-containing protein 4. *J Chem Inf Model* 57:1677-1690.
37. Breiman L (2001) Random forests. *Mach Learn* 45:5-32.
38. Lavecchia A (2015) Machine-learning approaches in drug discovery: methods and applications. *Drug Discov Today* 20:318-331.
39. Ruggieri S (2002) Efficient C4.5. *IEEE Trans Knowl Data Eng* 14:438-444.
40. Jain AK (2010) Data clustering: 50 years beyond K-means. *Pattern Recogn Lett* 31:651-666.
41. Xue Y, Yap CW, Sun LZ, et al (2004) Prediction of P-glycoprotein substrates by a support vector machine approach. *J Chem Inf Comp Sci* 44:1497-1505.
42. Morris GM, Huey R, Lindstrom W, et al (2009) Autodock4 and Autodocktools4: automated docking with selective receptor flexibility. *J Comput Chem* 30(16):2785-2791.
43. Pronk S, Pli S, Schulz R, et al (2013) A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845-854.
44. Maier JA, Martinez C, Kasavajhala K, et al (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11(8):3696-3713.
45. Dickson CJ, Rosso L, Betz RM, et al (2012) GAFFlipid: a general amber force field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* 8:9617-9627.
46. Da Silva AWS, Vranken WF (2012) ACPYPE-Antechamber python parser interface. *BMC Res Notes* 5:185.
47. Dupradeau FY, Pigache A, Zaffran T, et al (2010) The R.E.d. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* 12:7821-7839.
48. Essmann U, Perera L, Berkowitz ML, et al (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577-8593.
49. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Dis* 10(5):449-461.
50. Kumari R, Kumar R, Lynn A (2014) g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* 54(7):1951-1962.
51. Li BK, He B, Tian ZY et al (2015) Modeling, predicting and virtual screening of selective inhibitors of MMP-3 and MMP-9 over MMP-1 using random forest classification. *Chemometr Intell Lab* 147:30-40.
52. Li BK, Cong Y, Yang XG et al (2013) In silico prediction of spleen tyrosine kinase inhibitors using machine learning approaches and an optimized molecular descriptor subset generated by recursive feature elimination method. *Comput Biol Med* 43:395-404.
53. Yang XG, Lv W, Chen YZ et al (2010) In silico prediction and screening of gamma-secretase inhibitors by molecular descriptors and machine learning methods. *J Comput Chem* 31:1249-1258.

Figures

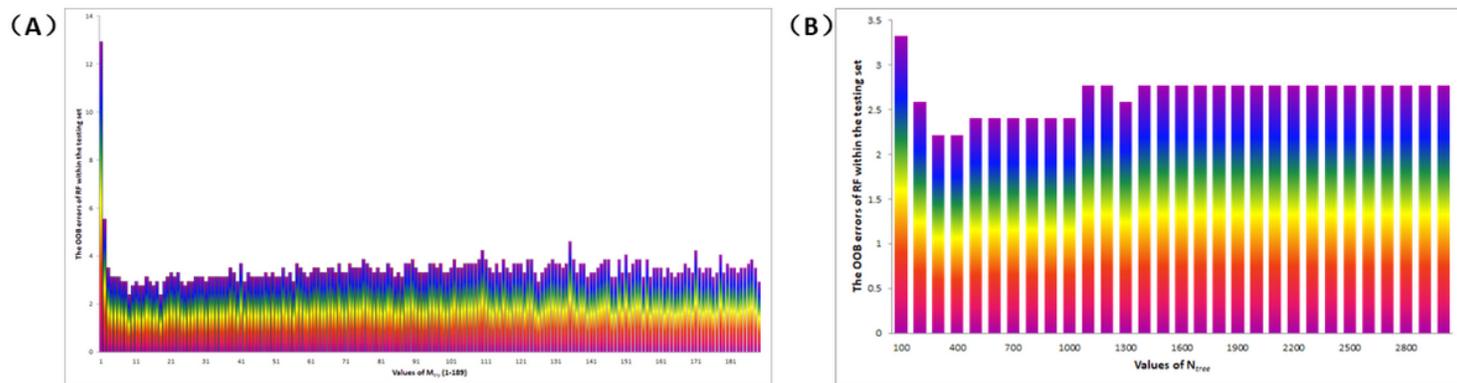


Figure 1

The effect of the different values of (A) Mtry (1-189), (B) Ntree (100-3000) on the prediction ERs of the testing set for different RF models

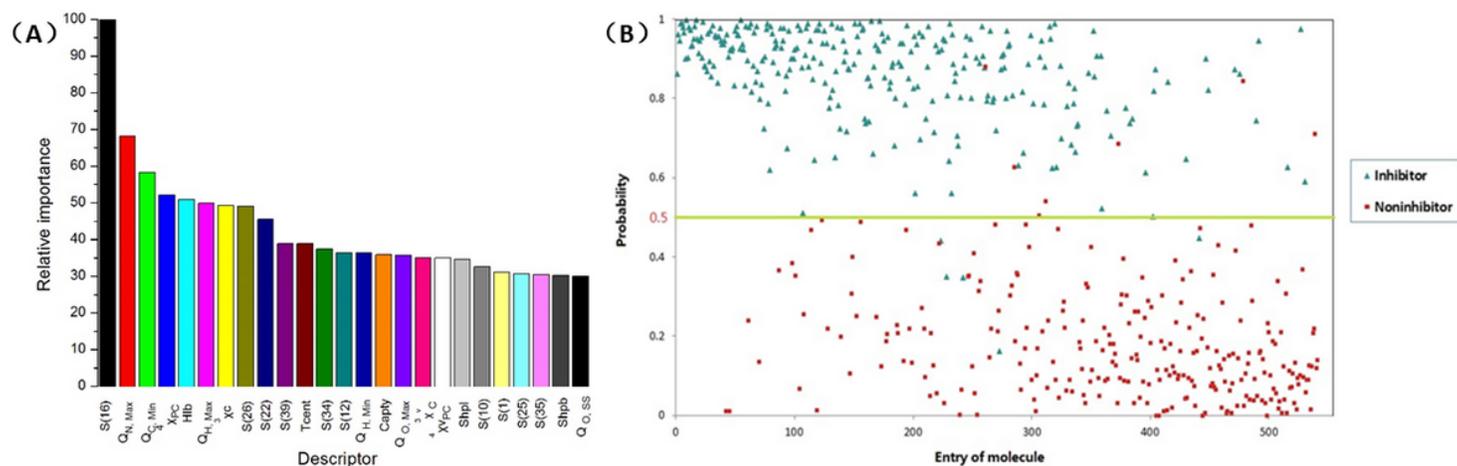


Figure 2

(A) The 25 descriptors with the highest ranking of relative importance in the best RF model for the prediction of BRD4is; (B) The visualized distributions of 541 molecules in the testing set from the best developed RF model

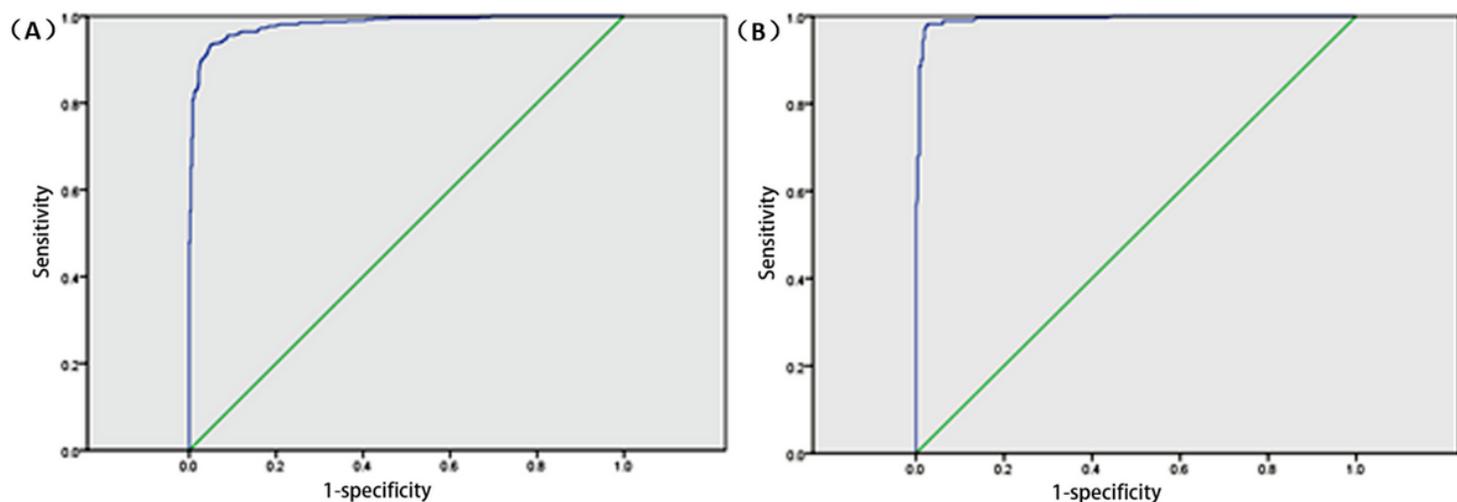


Figure 3

The ROC curves of the best RF model: (A) for the training set, (B) for the testing set

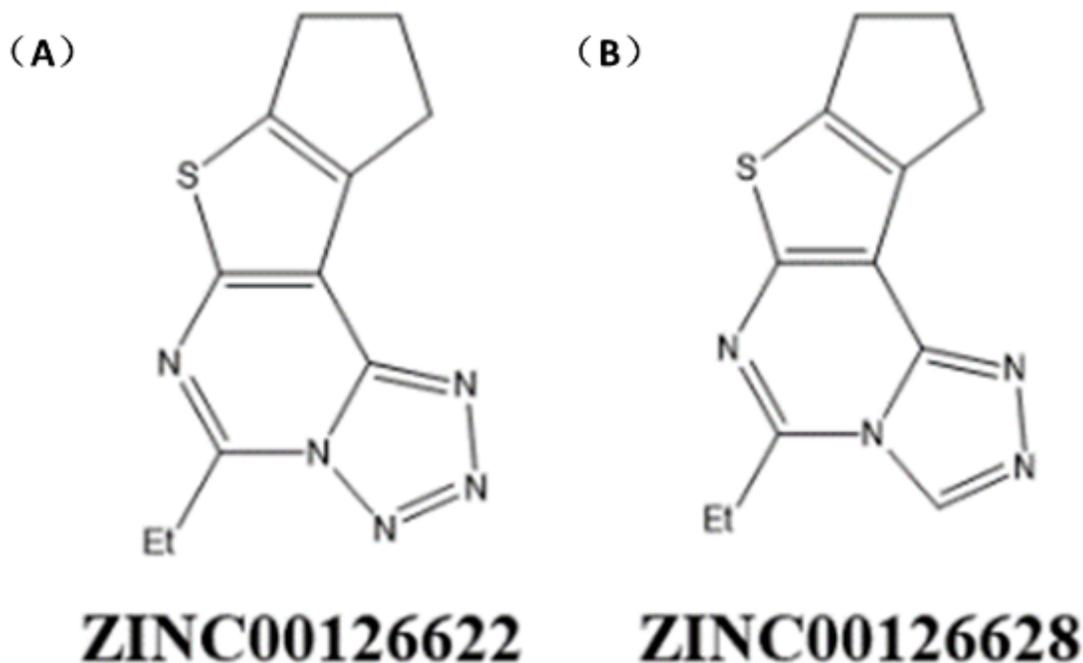


Figure 4

The selected molecules (Excerpts): (A) ZINC00126622, (B) ZINC00126628

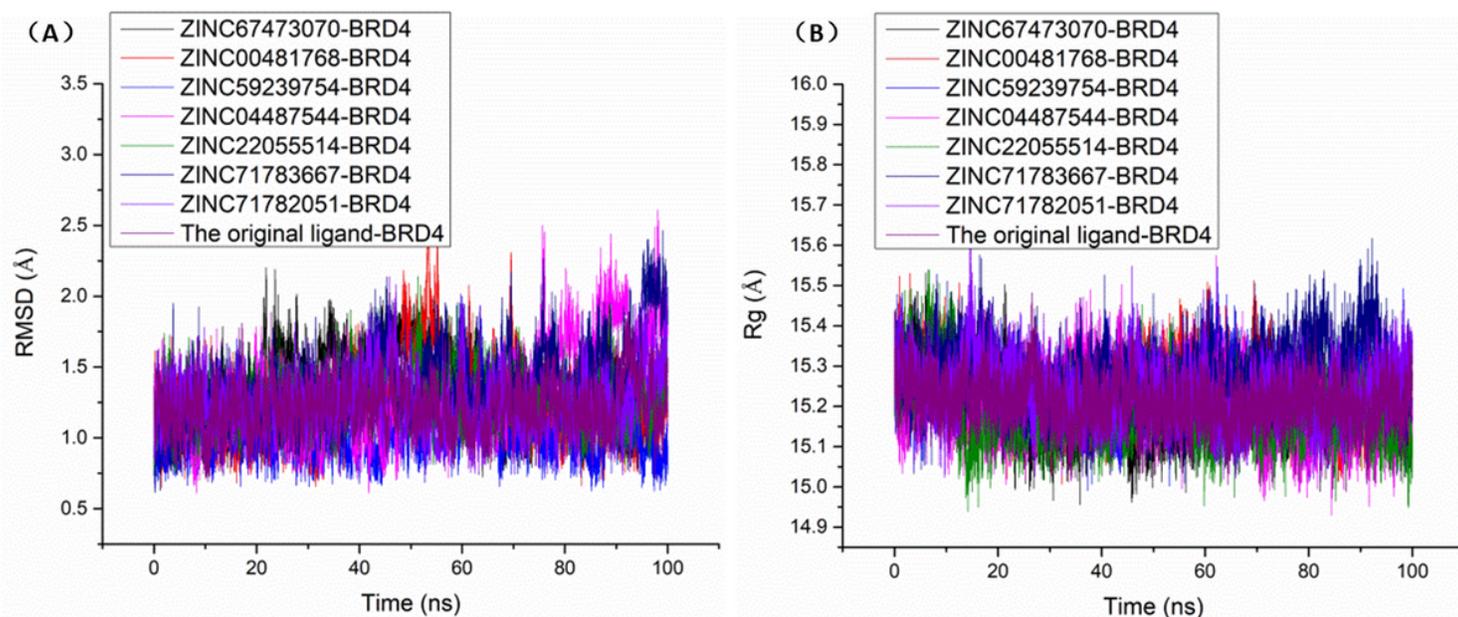


Figure 5

Analysis of RMSD and Rg of the seven top-ranked molecules-BRD4 complexes and the original ligand 3,5-dimethylisoxazol-BRD4 complex at 100 ns: (A) Root mean square deviation RMSD (Å) of the backbone atoms over the phase of 100 ns, (B) Radius of gyration (Rg) over the entire simulation