



Identification of a Novel Bcl-2 Inhibitor by Ligand-Based Screening and Investigation of Its Anti-cancer Effect on Human Breast Cancer Cells

Mei Wen^{1†}, Zhen-ke Deng^{1†}, Shi-long Jiang¹, Yi-di Guan¹, Hai-zhou Wu¹, Xin-luan Wang², Song-shu Xiao³, Yi Zhang⁴, Jin-ming Yang⁵, Dong-sheng Cao^{1,6*} and Yan Cheng^{1,6*}

¹ Xiangya School of Pharmaceutical Sciences, Central South University, Changsha, China, ² Translational Medicine R&D Center, Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, ³ Department of Gynecology and Obstetrics, The Third Xiangya Hospital, Central South University, Changsha, China, ⁴ Department of Pharmacology, College of Pharmaceutical Sciences, Soochow University, Suzhou, China, ⁵ Department of Pharmacology, The Penn State Hershey Cancer Institute, The Pennsylvania State University College of Medicine and Milton S Hershey Medical Center, Hershey, PA, United States, ⁶ Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China

OPEN ACCESS

Edited by:

Tingjun Hou,
Zhejiang University, China

Reviewed by:

Bo Liu,
Sichuan University, China
Katarina Nikolic,
University of Belgrade, Serbia

*Correspondence:

Dong-sheng Cao
oriental-cds@163.com
Yan Cheng
chengyan0677@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Experimental Pharmacology and Drug
Discovery,
a section of the journal
Frontiers in Pharmacology

Received: 02 November 2018

Accepted: 29 March 2019

Published: 17 April 2019

Citation:

Wen M, Deng Z, Jiang S, Guan Y, Wu H, Wang X, Xiao S, Zhang Y, Yang J, Cao D and Cheng Y (2019) Identification of a Novel Bcl-2 Inhibitor by Ligand-Based Screening and Investigation of Its Anti-cancer Effect on Human Breast Cancer Cells. *Front. Pharmacol.* 10:391. doi: 10.3389/fphar.2019.00391

Bcl-2 family protein is an important factor in regulating apoptosis and is associated with cancer. The anti-apoptotic proteins of Bcl-2 family, such as Bcl-2, are overexpression in numerous tumors, and contribute to cancer formation, development, and therapy resistance. Therefore, Bcl-2 is a promising target for drug development, and several Bcl-2 inhibitors are currently undergoing clinical trials. In this study, we carried out a QSAR-based virtual screening approach to develop potential Bcl-2 inhibitors from the SPECS database. Surface plasmon resonance (SPR) binding assay was performed to examine the interaction between Bcl-2 protein and the screened inhibitors. After that, we measured the anti-tumor activities of the 8 candidate compounds, and found that compound M1 has significant cytotoxic effect on breast cancer cells. We further proved that compound M1 downregulated Bcl-2 expression and activated apoptosis by inducing mitochondrial dysfunction. In conclusion, we identified a novel Bcl-2 inhibitor by QSAR screening, which exerted significant cytotoxic activity in breast cancer cells through inducing mitochondria-mediated apoptosis.

Keywords: Bcl-2, small molecule inhibitors, breast cancer cell, virtual screening, QSAR

INTRODUCTION

Apoptosis is a process of cellular suicide through which unwanted or unhealthy cells are eliminated during organism development or cellular stress (Reed, 2002; Fernald and Kurokawa, 2013). Deregulated apoptosis is a characteristic of cancers (Doi et al., 2012; Roizen, 2012). The B-cell lymphoma-2 (Bcl-2) family is composed of pro- and anti-apoptotic proteins, plays pivotal role in the mitochondrial pathway of apoptosis by promoting the release of cytochrome *c* and Smac (Second mitochondrial-derived activator of caspases) into the cytosol, resulting in caspase-dependent cell death (Day et al., 2008; Bai and Wang, 2013). The pro-apoptotic proteins of Bcl-2 family include Bak and Bax, the multi-domain proteins, and Bim, Bid, Puma, and Noxa, the BH3 domain only proteins (Fulda, 2013). The most prominent anti-apoptotic proteins contain Bcl-xl, Bcl-2, and Mcl-1 (Oltersdorf et al., 2005). The BH3-only proteins stimulate

apoptosis either by interacting with Bak and Bax or by binding to the anti-apoptotic proteins to release Bax and Bak (Martin and Dowsett, 2013). Therefore, the cell survival or death is largely determined by the ratio of Bcl-2 family anti- and pro-apoptotic proteins. In most tumor tissues and cancer cell lines, the anti-apoptotic Bcl-2 proteins are frequently highly expressed (Roberts and Huang, 2017). For example, Bcl-2 is overexpressed in prostate cancer, breast cancer, B-cell lymphomas, colorectal cancer (Buolamwini, 1999; Tóthová et al., 2002; Kirkin et al., 2004). The overexpression of anti-apoptotic proteins of Bcl-2 family promote cell proliferation and survival, and lead to therapy resistance due to the evasion of apoptosis (Buolamwini, 1999; Green and Evan, 2002). Therefore, they are well-validated drug targets for cancer treatment. The development of Bcl-2 anti-apoptotic proteins inhibitors has broad prospects in clinical application. Recently, a number of compounds have been demonstrated to inhibit anti-apoptotic Bcl-2 proteins, and some of them have entered clinical trials as potential cancer treatments. ABT-737 was designed using a fragment-based strategy and has a high affinity with Bcl-2 and Bcl-xl. Preclinical researches have demonstrated that ABT-737 and ABT-263, an ABT-737 derivative, have significant cytotoxicity against a variety of hematologic malignancies. However, the study presented that ABT-737 or ABT-263 can cause thrombocytopenia *in vivo*, which limits the dosage and may impair the full antitumor effect (Schoenwaelder et al., 2011). Structure-guided reverse engineering of ABT-263 led to the discovery of ABT-199, a novel selective inhibitor of Bcl-2 (Souers et al., 2013). In phase I trials, ABT-199 showed greater activity than ABT-263 in patients with lymphoma, and had no significant effects on platelet counts, but tumor lysis syndrome was caused by ABT-199 in patients with chronic lymphocytic leukemia (Seymour et al., 2013; Vandenberg and Cory, 2013). AT-101 is an oral pan Bcl-2 inhibitor, which exhibits affinity for Bcl-2, Mcl-1, and Bcl-xl at submicromolar concentrations. AT-101 was well-tolerated in patients (Liu et al., 2009). However, early phase II trials show that AT-101 failed to exert significant clinical activity either used as monotherapy or in combination with conventional chemotherapies in lung and prostate cancers (Ready et al., 2011). Therefore, it is urgent to develop novel Bcl-2 anti-apoptotic proteins inhibitors with excellent anti-tumor activity and little side effects.

In the present study, we design a ligand-based screening workflow for identification of new Bcl-2 inhibitors using a virtual screening that includes a series of QSAR classification and regression models (Figure 1). In the virtual screening framework, both classification and regression models based on random forest were firstly constructed and strictly evaluated, and then used for step-by-step screening the SPECS database filtered by Lipinski's rule of five. The molecules obtained were subsequently subjected to ADMET evaluation, and finally several compounds were manually selected by scaffold analysis from self-organizing map. A series of experimental validation and investigation of their anti-cancer effect on human breast cancer cells were performed. Finally, compound M1 was identified as a novel Bcl-2 inhibitor, which interacts with Bcl-2 and downregulates its expression. M1 exerts significant anti-cancer effects, and induces apoptosis by mitochondrial pathway in breast cancer cells.

MATERIALS AND METHODS

Data Set

To obtain a comprehensive and diverse dataset, we collected the Bcl-2 data from three public chemogenomics resources: ChEMBL, BindingDB, and PubChem databases (Liu et al., 2007; Gaulton et al., 2012; Kim et al., 2016). Combined with these data sets, a Bcl-2 dataset with various active values was obtained. In order to further improve the reliability and quality of data, several preprocessing steps are adopted: (1) removing compounds without explicit description for Bcl-2 and explicit molecular structures, and only retaining compounds with IC₅₀, K_i, K_D, and EC₅₀ values; (2) if we have two or more active data for the same molecule, we will take their arithmetic average to reduce the random error. All molecules were preprocessed by "wash" function in MOE software (Molecular Operating Environment software) (MOE, 2016) to disconnect group metals in simple salts, remove minor components, deprotonate strong acids, protonate strong bases, and add explicit hydrogens. After that, the structure of all molecules were optimized to structure of minimize energy by Merck Molecular Force Field 94 (Halgren, 2015), and the gradient-threshold of potential energy was set to 0.001 kcal⁻¹.mol⁻¹. After above-mentioned pretreatments, we finally obtained a Bcl-2 data set consisting of 1,259 compounds for further study. For the identification of more active compounds, all compounds were divided into "highly active" or "lowly active" sets based on the activity value threshold of 100 nM in our study. Thus, the dataset used for classification consists of 766 positive samples and 493 negative samples. For regression task, all 1,259 compounds with their activity values are subjected to construct the regression prediction model. It should be noted that the activity values in regression should be converted to their logarithmic form. Moreover, the molecules used for virtual screening were obtained from SPECS database, consisting of 230,673 compounds. These molecules are prepared and processed by MOE software in the similar way.

According to the principles of the Organization for Economic Co-operation and Development (OECD), in the process of validating the reliability and predictive capabilities of the model, not only internal validation but also external validation is needed (Roy, 2007; Tropsha, 2010; Ojha et al., 2011; Wang et al., 2016). In this study, Bcl-2 bioactivity data containing 1,259 compounds were divided into a training set and a test set according to the distribution of their chemical space by MOE. And then, we obtained a training set containing 1,007 compounds (80%) and a test set containing 252 compounds (20%). For training set, their activity values range from 0.001 to 9,900 nM, with an average value of 800.82 nM. For test set, their activity values range from 0.01 to 9,490 nM, with an average value of 906.73 nM. We used training set to build prediction models, while test set was used to further evaluate the predictive ability of the model. Additionally, to further validate the prediction ability of our constructed regression model, we additionally collected 128 compounds with Bcl-2 bioactivity values as external validation set from the ChEMBL database.

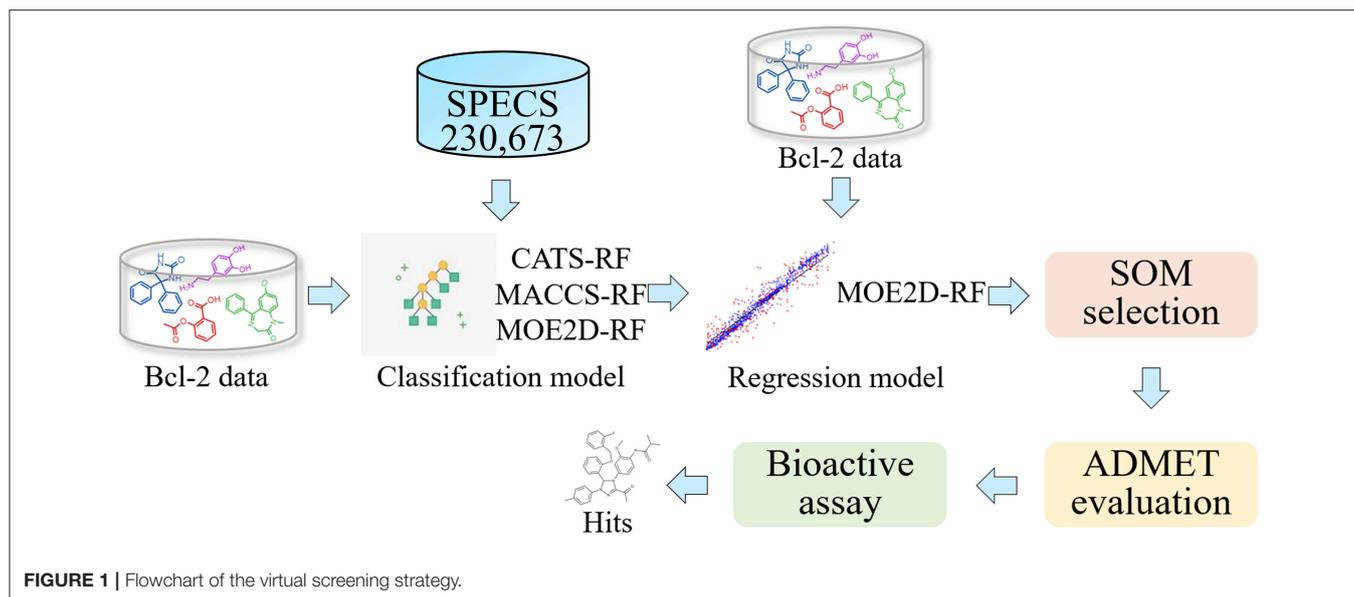


FIGURE 1 | Flowchart of the virtual screening strategy.

Molecular Representation

In the process of establishing a robust and convincing SAR/QSAR model, the informative descriptors are very important. Different types of descriptors represent different molecular structural contents, and thus help to capture diverse molecular scaffolds for screening Bcl-2 inhibitors. In this study, three kinds of molecular descriptors were used for molecular representation, including 206 two-dimensional (MOE2D) molecular properties calculated from MOE software, 166 MACCS fragments, 210 chemically advanced template search (CATS) pharmacophore descriptors calculated by ChemDes (Jie et al., 2015), ChemPy (Cao et al., 2013), and PyBioMed (Dong et al., 2018b), developed by our group. The CATS topological pharmacophore descriptor, it contains six atom types, namely, hydrogen-bond donor (D), hydrogen-bond acceptor (A), positively charged (+), negatively charged (-), Hydrophobe (H), and Aromatic atom (R). The numbers of all 21 possible pairs of generalized atom types are determined; distances of up to ten bonds were considered which led to a 210-dimensional vector representation of a molecular compound. The MACCS fingerprint uses a dictionary of MDL keys, which contains a set of 166 mostly common substructure features. 206 MOE two-dimensional molecular descriptors, including 20 physical properties, 14 Hückel theory descriptors, 18 subdivided surface areas, 42 atom counts and bond counts, 16 Kier & Hall connectivity and Kappa shape indices, 33 adjacency and distance matrix descriptors, 13 pharmacophore feature descriptors, and 50 partial charge descriptors. For classification tasks, three types of representations were all used to individually build classification models. For regression tasks, MOE2D molecular descriptors covering different molecular structural contents were applied to build regression models.

Drug-Likeness Filtering

Lipinski's five rules were used to remove chemical structures that have too many undesired properties in the screening database

(Lipinski et al., 2012). The parameters of Lipinski's rule are calculated by MOE program and ADMETlab (Dong et al., 2018a). In this work, we have removed compounds that violated more than two criteria. The 212,749 compounds filtered with Lipinski's rule of five were subjected to the subsequent virtual screening by SAR/QSAR models.

RF Model Generation and Validation

As a more flexible modeling technology, Random Forest (RF) algorithms can naturally deal with the classification and regression of complex non-linear systems (Breiman, 2001; Svetnik et al., 2003). Compared to some traditional approaches, RF offers several advantages, including simplicity, excellent prediction accuracy, and high ability to handle large or severely unbalanced datasets (Breiman, 2001; Cao et al., 2010). Several reports have introduced the effective prediction ability of RF in a wide range of classification and regression problems for SAR/QSAR studies (Wang et al., 2017; Ye et al., 2019). RF was chosen for constructing the screening model in our study. Herein, we intended to construct different models represented by three kinds of descriptors to screened more diverse compounds. We made full use of known Bcl-2 bioactivity data and combined RF classification and regression models. Firstly, the database was pre-screened by RF classification model, and then the screening results are further screened by RF regression model. Finally, the potential Bcl-2 inhibitors were obtained for experimental validation. The combined application of RF classification and regression methods could be considered as reliable strategy in the exploration of the activity of Bcl-2 inhibitors. This strategy has been generally performed in other activity prediction studies (Dong et al., 2009). All above-mentioned models are constructed by RF node in KNIME¹ analytics platform.

¹KNIME. Available online at: <https://www.knime.com/> (accessed May 18, 2018).

To ensure the established SAR/QSAR model has good generalization ability, independent test sets, and 5-fold cross-validation were applied for this purpose. For classification tasks, the statistical parameters of these models such as specificity (SP), sensitivity (SE), accuracy (Q) are calculated as follows:

$$SE = \frac{TP}{TP + FN} \quad (1)$$

$$SP = \frac{TN}{TN + FP} \quad (2)$$

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

where TP, FN, TN, and FP denote the number of true positives, false negatives, true negatives and false positives, respectively. In addition, the receiver operating characteristic (ROC) curve that shows the separation ability of a binary classifier was also plotted. The ROC curve was used to graphically present the model behavior in a visual way.

For the regression model, we used six main statistic parameters to evaluate the predictive models: the square correlation coefficients of training (R_F^2); the root mean squared error of training (RMSE_F); the square correlation coefficients of test (R_{T^2}); the root mean squared error of test (RMSE_T); the square correlation coefficients of cross-validation (Q^2); the root mean squared error of cross validation (RMSE_{CV}); the square correlation coefficients of external dataset (R_{Ex^2}); the root mean squared error of external dataset (RMSE_{Ex}). They are calculated as follows:

$$R_F^2 = 1 - \frac{\sum (\hat{y}_i - y_i)^2}{\sum (y_i - \bar{y})^2} \quad (4)$$

$$RMSE_F = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

$$Q^2 = 1 - \frac{\sum (\hat{y}_{(v)i} - y_i)^2}{\sum (y_i - \bar{y})^2} \quad (6)$$

$$RMSE_{CV} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_{(v)i})^2} \quad (7)$$

where \hat{y}_i and y_i are the predicted and experimental values of the i th sample in the training set; \bar{y} is the mean value of all the experimental values in the training set; $\hat{y}_{(v)i}$ is the predicted value of i th sample for cross validation; N is the number of samples in the training set. When independent test set and external dataset are applied, R_{T^2} , RMSE_T, R_{Ex^2} , and RMSE_{Ex} are calculated in the similar way.

ADMET Evaluation

As we know, the failure of many drug candidates in clinical trials may be due to poor ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties (Prentis et al., 1988; Kennedy, 1997). However, it is time-consuming and laborious to evaluate the pharmacokinetics and toxicological properties of compounds by experimental methods. Therefore,

during the early stages of drug discovery, computational techniques that can predict pharmacokinetics and toxicity profiles have become alternatives (Wishart, 2007; Cheng et al., 2013). Nowadays, researchers can use computational approaches to evaluate the ADMET properties of screened compounds to filter the compounds with undesired properties. Here, we evaluated five ADMET properties for hit compounds including octanol/water partition coefficient (logP), aqueous solubility (logS), Caco-2 cell permeability (Caco-2), percent human oral absorption (PHOA), and IC₅₀ value for blockage of HERG K⁺ channels (HERG) by QikProp module of Schrödinger (Quick Prop, version 3.5, Schrödinger, LLC, New York, NY, 2016). QikProp provided 95% of known drug's ADMET properties distribution as reference values.

Surface Plasmon Resonance Assay

The interaction between virtual screening candidates and Bcl-2 was determined using a PlexArray HT biological molecular interaction analyzer (PLEXERA LLC, USA). Candidates were covalently immobilized to the surface of sensor chip following the standard coupling reaction guide. Candidates were dissolved in DMSO to a final concentration of 10 μM. To analyze binding signal, a set of concentration gradient (700, 350, 175, 87.5 nM) of Bcl-2 protein were, respectively, diluted with PBST (1x PBS + 0.1% Tween-20, pH 7.4). PBST was also used as running buffer and samples were injected into the flow cell chamber of the sensor chip one by another with rate of 2 μL/s for 300 s. After the injection stage, running buffer was flowed over the chip surface for 300 s at 2 μL/s to allow the bound analytes to dissociate from the Bcl-2. The response unit (RU) of sample dots on the chip surface was recorded by the interaction analyzer during the whole process. In this way, dissociation curves obtained. HBS-P buffer was used as vehicle controls. Lastly, the changes in RU and the binding curve were measured by the PlexeraDE software (PLEXERA LLC, USA).

Cell Lines and Culture

The human breast cancer cell line MDA-MB-231, T47D, and BT549, were purchased from Cell Bank of Chinese Academy of Sciences. MDA-MB-231 were cultured in L-15 medium with 10% FBS at 37°C with 100% air, and T47D were cultured in DMEM/High glucose medium with 10% FBS at 37°C with 5% CO₂. BT549 was cultured in 1,640 medium with 10% FBS at 37°C with 5% CO₂.

Reagents and Antibodies

Compound M1 were purchased from SPECS database. Anti-Bcl-2, anti-Cyt c, anti-COX-IV, anti-PARP, and anti-cleaved caspase-3 were purchased from Cell Signaling Technology (Danvers, MA, USA). Anti-β-actin was purchased from Proteintech. Mitochondria Isolation Kit for Cultured Cells (CAT. NO. 89874) was purchased from Thermo. JC-1 Assay Kit (CAT. NO. C2006) and Reactive Oxygen Species Assay Kit (CAT. NO. S0033) were purchased from Beyotime. The enhanced chemiluminescence (ECL) kit was purchased from Beijing Com Win Biotech Co, Ltd (Cwbio, China). CCK-8 reagent was purchased from Bimake (shanghai, China).

Cell Viability Assay

Cell viability was measured by Cell Counting Kit-8 (CCK-8) assay following the manufacturer's protocol. Cells were plated at 5×10^3 cells per well in 96-well plates. After compound M1 treatment, cells were incubated with 10 μ L CCK-8 for 2 h at 37°C with 5% CO₂. The results were determined at 450 nm wave length.

EdU Assay

Cell Proliferation was measured by Cell Light EdU DNA Cell Proliferation Kit. Cells were plated in 96-well culture plates. Cells were treated with 4 or 8 μ M compound M1 for 48 h. At the end of treatment, cells were incubated with EdU reagent for 2 h at 37°C. Four percent of paraformaldehyde was used to fix the cells for 15 min. Then, cells were incubated with 0.5% Triton X-100 for 30 min. Washed cells with PBS three times. After that, cells were dyed with reaction cocktail for 30 min, and stained cell nucleus with 5 μ g/mL of Hoechst 33,342 for 30 min. Images were observed by fluorescent microscope.

Clonogenic Assays

The cells were plated at 200 cells per well in six-well culture plates followed by compound M1 for 48 h. At the end of treatment, the cells were cultured with fresh medium, and were grown at 37°C for 10 days. After colony formation, the cells were washed with PBS and immobilized by 4% formaldehyde at room temperature for 30 min. Finally, the cells were dyed with 0.5% crystal violet.

Western Blot Analysis

After compound M1 treatment, cells were lysed with RIPA buffer supplemented with protease inhibitor (Selleck) on ice for 30 min following by centrifugation at $14,000 \times g$ for 15 min. Protein concentrations of the lysates were determined by BCA assay kit. Equivalent amounts of cellular protein were separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes. And then, the membranes were blocked with 5% skim milk in TBST. The membranes were incubated with primary antibodies and peroxidase-conjugated secondary antibodies. Finally, the membranes were visualized with an enhanced chemiluminescent detection kit.

Assessment of Mitochondrial Membrane Potential

The MMP was measured using JC-1 assay kit according to the manufacturer's protocol. Cells were treated with 2 or 4 μ M compound M1 for 48 h. The cells were cultured with JC-1 staining reagent (1x) at 37°C for 20 min. Cells were washed twice by JC-1 staining buffer. Next, the fluorescence intensity of JC-1 aggregate or monomer was observed by fluorescent microscopic. The ratio of aggregates to monomers was estimated as a symbol of mitochondrial membrane potential.

Detection of Reactive Oxygen Species

Cells were treated with 2 or 4 μ M compound M1 for 48 h. After that, cells were rinsed twice with PBS and incubated with DCFH-DA (10 μ mol/L) at 37°C in a darkroom. The fluorescence

intensity of DCF was detected at an emission wavelength of 525 nm and excitation wavelength of 488 nm.

Apoptosis Assay

Cell apoptosis was measured by Annexin V-FITC/PI double-staining Kit following the manufacturer's protocol. After treatment with 8 or 16 μ M compound M1 for 48 h, cells were washed twice with cold phosphate buffer saline and then, suspended in 100 μ L binding buffer (BD CAT. NO.556547) with 5 μ L Annexin-V (BD CAT. NO.556420) and 5 μ L PI (BD CAT. NO.556463). And cells were gently vortexed and incubated at room temperature for 15 min in darkness for flow cytometric (Becton Dickison) analysis.

Data Analysis

The differences between the samples with compound M1 treatment and control, were analyzed by *t*-test. Differences between groups were considered statistically significant at $P < 0.05$. The RF classification and regression models are constructed by the Konstanz Information Miner (KNIME) software (KNIME, version 3.4.1). The molecular preparation and processing are finished by MOE software. The ADMET evaluation and SOM are performed by the Qikprop module and Canvas module of Schrodinger software, and scaffold analysis are performed by RDKit² cheminformatics package (RDKit, version 2016).

RESULTS AND DISCUSSION

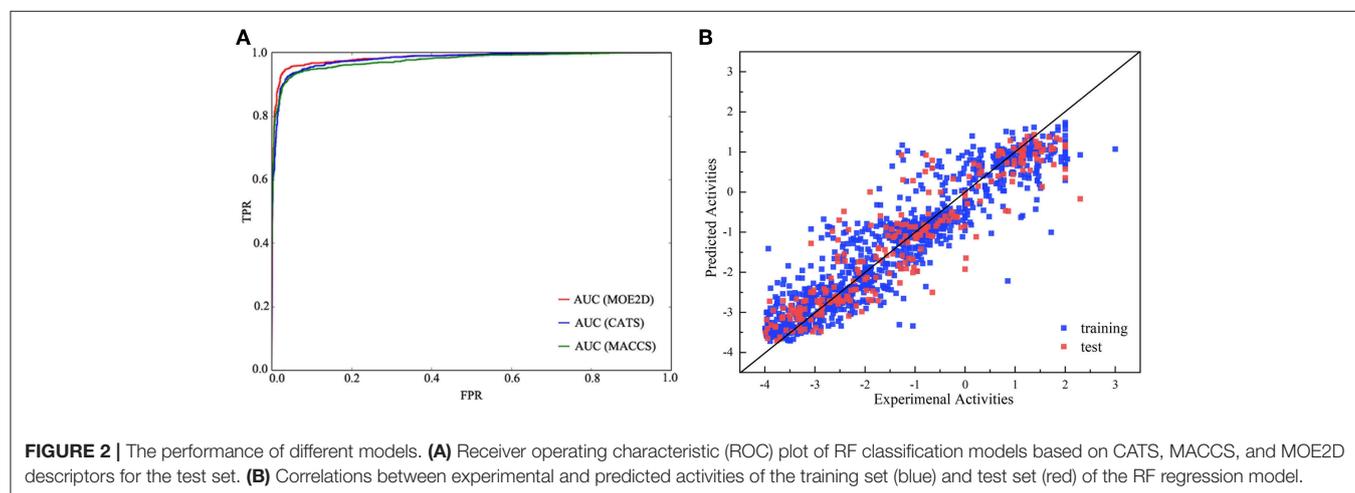
RF Classification and Regression Models

In the present study, we adopted a ligand-based virtual screening strategy to screen the potential Bcl-2 inhibitors. The ligand-based screening strategy combined the classification and regression RF models. The combined use of both classification and regression models has been performed as useful strategy in the exploration of various activities (Dong et al., 2009; Michielan et al., 2009). Before applying these models, we strictly evaluated the model performance for classification and regression tasks. Firstly, the RF classification models were generated with three different molecular representations, which were CATS, MACCS and MOE2D, respectively. **Table 1** lists the prediction statistics of RF classification models with three representations for distinguishing the Bcl-2 highly active inhibitors and lowly active inhibitors. Clearly, three models with different representations all yield high prediction accuracy, and the best performance was achieved by MACCS-RF classification model, with values of prediction accuracy of 91.3 ± 1.6 , sensitivity of 93.1 ± 2.1 , and specificity of 88.4 ± 3.2 , respectively. However, from the statistical point of view, three models had no statistically significant difference, indicating that three models constructed by three different representations could all be used for the subsequent screening process. As shown in **Table 1**, the AUC values of MOE2D-RF, MACCS-RF, and CATS-RF models was 96.5 ± 1.0 , 96.5 ± 1.1 , and 96.5 ± 0.9 , respectively. The ROC plot of three models constructed by three different representations

²RDKit: *Open-Source Cheminformatics*. Available online at: <http://www.rdkit.org> (accessed November 28, 2016).

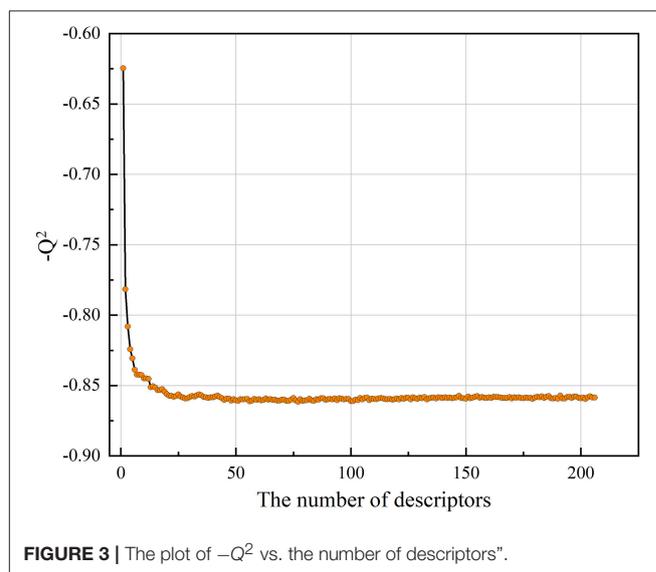
TABLE 1 | Performance of the RF classification models by using different molecular descriptors.

Models	TP ^a	FP ^b	TN ^c	FN ^d	SE (%) ^e	SP (%) ^f	Q (%) ^g	AUC (%) ^h
MOE2D-RF	141.2 ± 8.3	10.5 ± 2.6	87.6 ± 8.4	12.7 ± 3.6	91.7 ± 2.3	89.1 ± 2.6	90.8 ± 1.5	96.5 ± 1.0
MACCS-RF	144.6 ± 6.9	11.2 ± 3.0	85.5 ± 7.4	10.7 ± 3.3	93.1 ± 2.1	88.4 ± 3.2	91.3 ± 1.6	96.5 ± 1.1
CATS-RF	140.6 ± 5.7	12.6 ± 3.1	87.4 ± 5.3	11.5 ± 3.4	92.5 ± 2.2	87.4 ± 2.9	90.5 ± 1.5	96.5 ± 0.9

^aTP, true positives.^bFP, false positives.^cTN, true negatives.^dFN, false negatives.^eSE, sensitivity, $SE = TP/(TP + FN)$.^fSP, specificity, $SP = TN/(TN + FP)$.^gQ, accuracy, $Q = (TP + TN)/(TP + FP + TN + FN)$.^hAUC, the area of under the receiver operating characteristic curve.

is shown in **Figure 2A**. AUC scores from three models further indicated that there is no statistically significant difference in model performances. Overall, the three kinds of RF classification models were found to perform well and capable to accurately predict the Bcl-2 inhibitors.

As described above, the collected Bcl-2 bioactivity molecules represented by the MOE2D descriptor was used to construct a predictive model using an RF regression algorithm. In the modeling stage, the feature selection was carried out by recursive feature elimination method of random forest. The influence of the number of descriptors to the performance of model (Q^2) can be seen from **Figure 3**. Clearly, when the number of features reaches 50, the model can achieve the highest accuracy, and the Q^2 reached 0.852. Thus, we applied 50 informative MOE2D descriptors to construct the RF regression model. We used 5-fold cross-validation and independent test set to evaluate the predictive ability of regression model. For the training set, $R_F^2 = 0.980$ and $RMSE_F = 0.238$; for the cross-validation, $Q^2 = 0.852$ and $RMSE_{cv} = 0.690$. For the test set, $R_T^2 = 0.847$ and $RMSE_T = 0.695$. From the statistic results, we can see that the regression models have good performance. The relationship between the experimental values and predicted values in training set and test set for the model was shown in **Figure 2B**. From this point, we can conclude that the model has good predictive ability for



most compounds in training and test sets. At the same time, the regression model was evaluated using the external dataset. For the external dataset, the $R_{Ex}^2 = 0.803$ and $RMSE_{Ex} = 0.684$. Although the R^2 value of the external dataset were a little smaller than

those for the training set, test set and cross-validation, we still deemed that the predictive ability of this model is satisfactory due to the decline is within acceptable range. These prediction results of external validation process proved the validity and reliability of final model again. Furthermore, we used the tool of applicability domain by standardization approach developed by Kunal Roy's group to evaluate the applicability domain of our QSAR model (Kunal and Supratik, 2015; Roy et al., 2015). From the result, we found that about two-thirds of the test set molecules fall within the applicability domain. Thus, we can use it to evaluate the reliability of the RF regression model for the prediction of compounds.

QSAR-Based Virtual Screening

In this research, QSAR-based methods were adopted to screening potent Bcl-2 inhibitors from the SPECS database. After the validation of RF classification and regression models, we constructed a virtual screening process. SPECS database was used to screening new Bcl-2 inhibitors. In the RF-classification stage, three RF models with different kinds of descriptors were used to screen the entire database. Combining the results of three models screened resulted in 10,530 unique hits. After that, 358 compounds were further screened by RF-regression model represented by MOE2D descriptors in the RF-regression stage.

Selection of Hit Compounds Based on Scaffold Analysis and SOM

To assess the diversity of the molecular structure of the hit compounds, we first performed a scaffold analysis for them. The RDKit package is used to calculate and analyze the scaffold of 358 hit molecules. The RDKit package decomposes the molecules into scaffolds and carbon skeletons according to their two-dimensional structure. The definition of scaffold decomposition was proposed and established by Bemis and murcko, and it is the most widely used scaffold definition (Bemis and Murcko, 1996). In this definition, the scaffold is defined by removing all R-groups of compounds but retaining the linkers between the

ring systems in the compounds. On the basis of scaffold, Xu and Johnson defined the carbon skeleton in 2002 (Xu, 2002). The carbon skeleton is derived from scaffolds by converting each heteroatom into a carbon atom and converting all bond orders into single bonds. Therefore, different carbon skeletons represent topologically distinct scaffolds. From 358 screened molecules, we obtained 177 unique scaffolds. There are five scaffolds containing more than ten molecules, and most of the other scaffolds contain only one molecule. The scaffold analysis indicated that the hit compounds screened by ligand-based screening strategy have diverse molecular structures to a certain extent. Subsequently, the prediction results are mapped to the SOM space along with the Bcl-2 highly active molecules mentioned above, by using

TABLE 2 | Assessment of ADMET properties of the hits by Qikprop.

ID	K_D (M) ^a	logPo/w ^b	logS ^c	Caco-2 ^d	PHOA ^e	logHERG ^f
M1	3.17×10^{-7}	7.03	-7.78	2099.01	100.00	-5.94
M2	4.06×10^{-7}	6.19	-6.14	766.34	88.91	-6.15
M3	8.56×10^{-7}	6.54	-7.44	256.59	82.44	-9.36
M4	4.76×10^{-6}	8.80	-11.76	2729.26	100.00	-9.08
M5	5.70×10^{-7}	5.06	-6.61	373.92	76.73	-7.30
M6	6.85×10^{-8}	7.60	-8.45	2672.14	100.00	-6.98
M7	1.59×10^{-6}	8.99	-9.42	4460.15	100.00	-6.08
M8	1.08×10^{-5}	7.92	-7.61	6557.40	100.00	-5.50
M9	-	8.70	-9.02	2320.00	100.00	-6.93
M10	-	7.31	-10.35	571.15	93.16	-8.19
M11	-	7.51	-9.78	1057.80	100.00	-7.75
M12	-	8.49	-11.00	1300.25	100.00	-7.84

^a K_D , dissociation equilibrium constant. $K_D = K_d/K_a$. K_a , K_d values are association- and dissociation-rate constants, respectively.

^bPredicted octanol/water partition coefficient, logP (recommended range: -2.0 to 6.5).

^cPredicted aqueous solubility, logS (recommended range: -6.5 to 0.5).

^dPredicted apparent Caco-2 cell permeability (<25 is poor and >500 is great).

^ePercentage of human oral absorption (<25% is poor and >80% is high).

^fPredicted IC_{50} value for blockage of HERG K^+ channels (concern below -5).

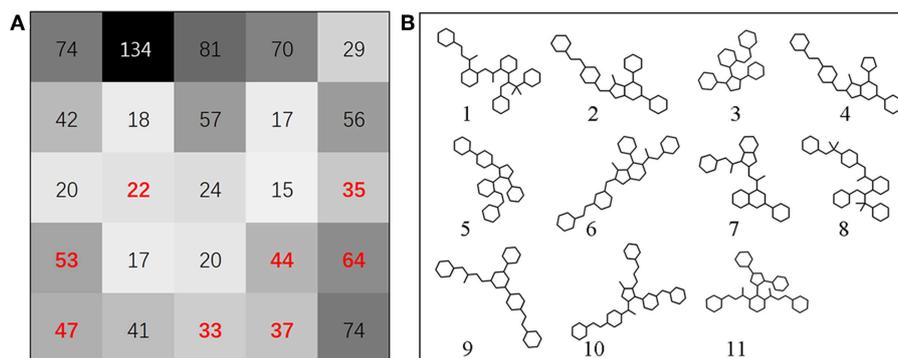
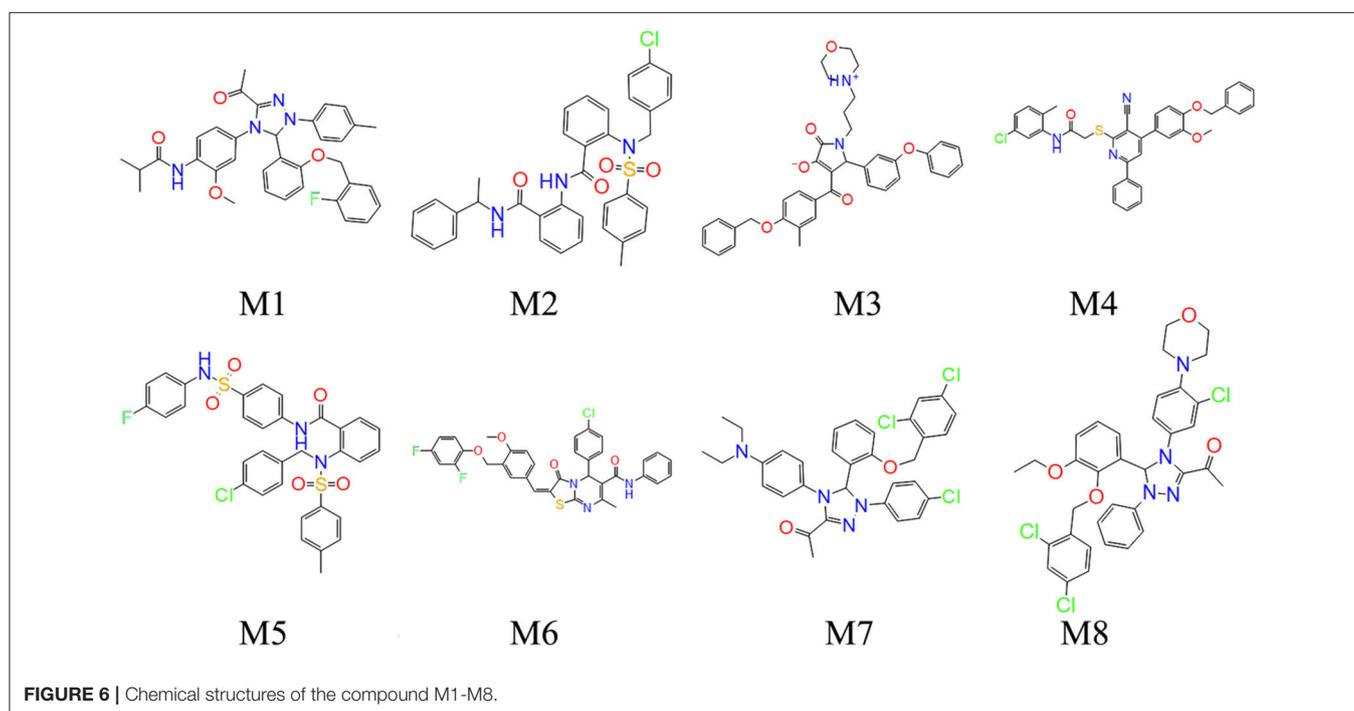
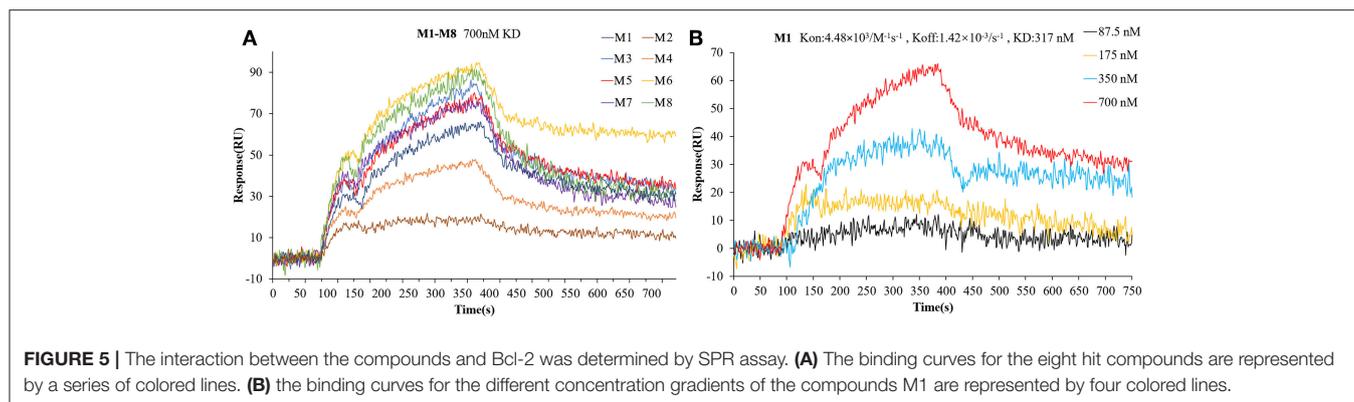


FIGURE 4 | (A) SOM projection of positive set and screened compounds. Black shading indicates compound density. Compounds were represented by CATS descriptors. The number in the grid represents the number of compounds. The position of red number in grid represents the source of the hit compounds we finally selected. In this research, the SOM technique can be used to identify preferred hit compounds. **(B)** The scaffold structure of selected molecules. Scaffold 1, 3, and 5 are located in cluster (1/4), scaffold 2 is located in cluster (1/1), scaffold 4, 6 are located in cluster (2/1), scaffold 7 is located in cluster (2/5), scaffold 8 is located in cluster (1/3), scaffold 9 is located in cluster (3/5), scaffold 10 is located in cluster (3/2), scaffold 11 is located in cluster (2/4).



SOM method based on CATS descriptors (Robinson et al., 1999; Schneider et al., 2009). **Figure 4** shows the distribution of positive set and screened compounds in the SOM space. Black shading indicates compound density. The number in the grid represents the number of compounds. Then the representative molecules that are distributed together with highly active molecules in the different grid were selected as the final screening hits. At the same time, we considered the structural diversity of hit molecules and to select molecules with different scaffolds as far as possible. Finally, we selected 12 Bcl-2 potential hits from the grid where the red numbers are located.

Evaluation of ADMET Properties

In this study, the Qikprop module of Schrodinger software was used to calculate the pharmacokinetic properties of twelve

TABLE 3 | The IC_{50} of eight small molecule compounds on breast cancer cells.

ID	IC_{50} (μ M)
M1	5
M2	>50
M3	>50
M4	>50
M5	>50
M6	>50
M7	>50
M8	>50

MDA-MB-231 cells were treated with eight small molecule compounds for 48 h. At the end of treatment, cell viability was measured by CCK-8 reagent. And then the IC_{50} of each compound was calculated according to the CCK-8 assay results.

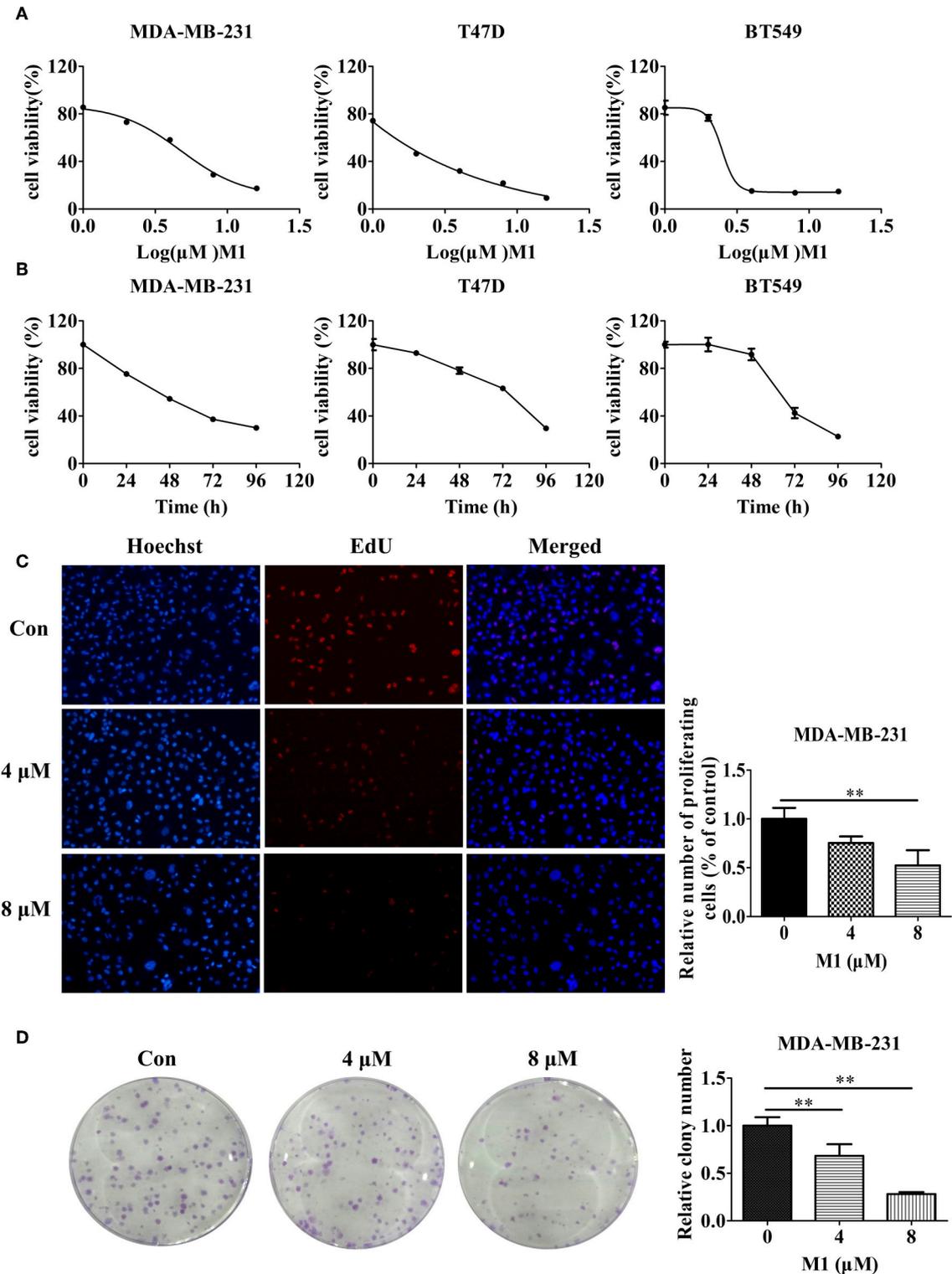


FIGURE 7 | Compound M1 inhibits cell proliferation in breast cancer cells. **(A)** MDA-MB-231, T47D, and BT549 cells were treated with a series of doses of compound M1 for 72 h. **(B)** MDA-MB-231, T47D, or BT549 cells were treated with 5, 2, or 3 μ M compound M1 for difference time periods. After treatment, cell viability was estimated by CCK-8 assay. Compound M1 vs. control, $*p < 0.05$, $**p < 0.01$, student *t*-test, $n = 3$, means \pm SD. **(C)** MDA-MB-231 cells were treated with 4 or 8 μ M compound M1 for 48 h. The EdU-marked replicating cells were observed by fluorescent microscope. Compound M1 vs. control, $*p < 0.05$, $**p < 0.01$, student *t*-test, $n = 3$, means \pm SE. **(D)** MDA-MB-231 cells were treated with 4 or 8 μ M compound M1 for 48 h, and the cell colonies were fixed and stained for camerawork. Compound M1 vs. control, $*p < 0.05$, $**p < 0.01$, student *t*-test, $n = 3$, means \pm SE.

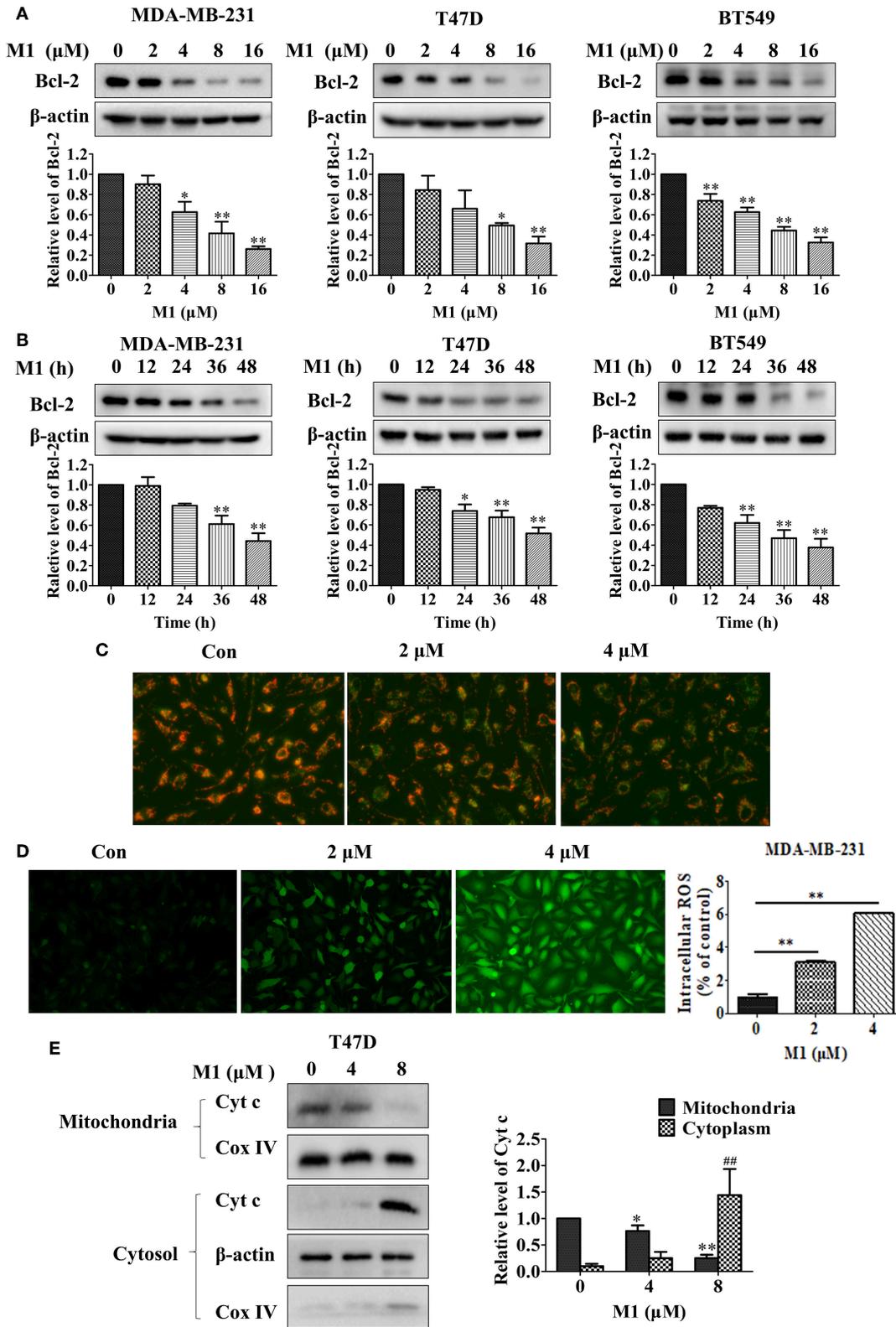
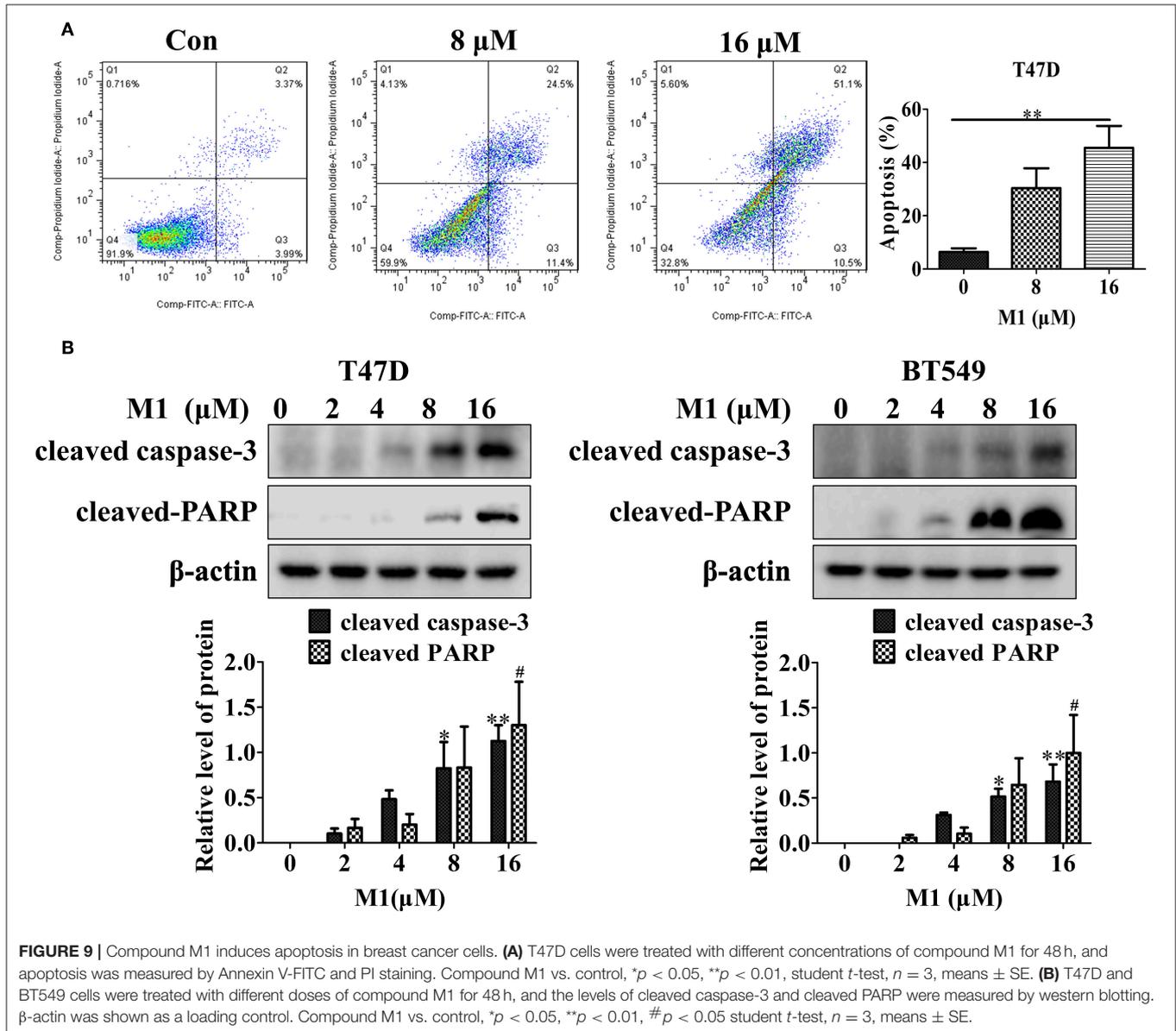


FIGURE 8 | Compound M1 downregulates Bcl-2 expression and induces mitochondrial dysfunction in breast cancer cells. MDA-MB-231, T47D, and BT549 cells were treated with different doses of compound M1 for 48 h (A), or were treated with 8 μM M1 for difference time periods (B), and the expressions of Bcl-2 were (Continued)

FIGURE 8 | estimated by western blotting. β -actin was shown as a loading control. Compound M1 vs. control, $*p < 0.05$, $**p < 0.01$, student *t*-test, $n = 3$, means \pm SE. **(C)** MDA-MB-231 cells were treated with 2 or 4 μ M compound M1 for 48 h. The MMP was observed by JC-1 staining by fluorescent microscopic. **(D)** MDA-MB-231 cells were treated with 2 or 4 μ M compound M1 for 48 h. The levels of ROS were measured by staining with DCF-DA and the fluorescent intensity observed by fluorescent microscopic. The levels of ROS were measured by TECAN SPARK 10M. Compound M1 vs. control, $*p < 0.05$, $**p < 0.01$, student *t*-test, $n = 3$, means \pm SD. **(E)** T47D cells were treated with 4 or 8 μ M compound M1 for 48 h. The levels of Cytochrome c in mitochondrial or cytosol fraction were measured by western blotting. COX-IV and β -actin were shown as loading controls for the mitochondria and cytosol, respectively. Compound M1 vs. control, $*p < 0.05$, $**p < 0.01$, $##p < 0.01$ student *t*-test, $n = 3$, means \pm SE.



potential inhibitors. This would be extremely beneficial if information on the ADMET properties of the studied molecules could be obtained at an early stage of the drug discovery process, which might help chemists improve the pharmacokinetic profile of these compounds (Wang et al., 2015). Once obtained, this information is expected to help chemists to ameliorate the pharmacokinetic profile of the compounds.

The ADMET evaluation results are shown in **Table 2**. The predicted logPo/w ranges from 5.06 to 8.99, logS ranges from -11.76 to -6.14 , logHERG ranges from -9.36 to -5.5 , Caco-2 ranges from 256.59 to 6557.4, and PHOA ranges from 76.73 to 100 in this study. The predictions suggest that candidates may have good Caco-2 cell permeability, and have highly human oral absorption, the HERG was performed well. The properties of

most hit compounds are within acceptable range available for human use, which indicates that they have potential as drug-like molecules.

Measurement of Direct Binding Between Screened Compounds With Bcl-2 by SPR Assay

To verify the interactions between Bcl-2 and the screened twelve hit compounds, we performed SPR-based binding assays. We found that eight compounds (M1-8) could directly bind to the recombinant human Bcl-2 protein (Figure 5A). The structures of eight compounds were shown in Figure 6. When the concentration of Bcl-2 is 700 nM, they achieved the strongest binding with each other. The average equilibrium dissociation constants (K_D) for eight candidates at a concentration of 700 nM are range from 10^{-8} to 10^{-5} M, and compound M1 have the lowest K_D (370 nM) among these compounds. The binding signal of M1 became stronger as the concentration increased, and the K_{On} and K_{Off} value of M1 are $4.48 \times 10^3/M^{-1} s^{-1}$ and $1.42 \times 10^{-3}/s^{-1}$ respectively (Figure 5B).

M1 Shows Significant Cytotoxic Effects in Breast Cancer Cells

We further measured the effects of eight small molecule compounds on breast cancer cell viability by the CCK-8 assay (Figure S1). We found compound M1 showed significant cytotoxicity, IC_{50} value (48 h) of M1 in MDA-MB-231 cells is $5 \mu M$, whereas the IC_{50} (48 h) of the other seven compounds were more than $50 \mu M$ (Table 3). Furthermore, we found that compound M1 inhibits cell proliferation in concentration- and time- dependent in three breast cancer cell lines (Figures 7A,B). The EdU incorporation assay further demonstrated that compound M1 has an inhibitory effect on the proliferation of breast cancer cells, as evidenced by the reduced number of EdU-positive cells treated with compound M1 (Figure 7C). Next, we performed clonogenic assay to measure the long-term anti-proliferative activity of compound M1. After treatment with compound M1, the clonogenicity of breast cancer cells was reduced in a dose-dependent manner (Figure 7D). And then, we measured that the compound M1 were negative in tests for known pain assay interference (PAINS) substructures (Baell and Holloway, 2010) at the FAF-Drugs4 webserver (Lagorce et al., 2015). These results demonstrate that compound M1 has significant cytotoxic effects in breast cancer cells and has passed current PAINS removal filters.

Compound M1 Downregulates Bcl-2 Expression and Induces Mitochondrial Dysfunction in Breast Cancer Cells

Next, we want to investigate the relationship between the compound M1 and the level of Bcl-2 in breast cancer cells. Compound M1 decreased the protein expression of Bcl-2 in breast cancer cells in a dose- and time-dependent manner (Figures 8A,B). The downregulation of Bcl-2 can cause mitochondrial dysfunction. Therefore, we explored whether

mitochondrial function can be damaged by this compound in breast cancer cells. As shown in Figure 8C, compound M1 caused the disturbance of MMP (mitochondrial membrane potential), as determined by the shift of JC-1-produced fluorescence from red to green. Reactive oxygen species (ROS) are mainly produced in mitochondria, when the cells are stimulated by external stimuli, the increased level of ROS will cause further mitochondrial damage (Chong et al., 2014). We found that there was an elevation in the level of generated ROS in the M1-treated breast cancer cells compared to untreated cells (Figure 8D). We further found that compound M1 increased the release of cytochrome c from mitochondria to cytosol in a concentration-dependent manner in breast cancer cells (Figure 8E). These results confirmed that compound M1 could induce mitochondrial dysfunction by downregulating the expression of Bcl-2.

Compound M1 Induces Apoptosis in Breast Cancer Cells

The loss of MMP is an early but critical step in the mitochondrial pathways of apoptosis. Mitochondrial pro-apoptotic proteins such as cytochrome c release from mitochondria to cytoplasm are one of the crucial events in the intrinsic apoptosis pathway. Therefore, we further measured the effect of M1 on apoptosis induction, and found that this compound caused a dose-dependent increase in apoptosis, as indicated by an increase in Annexin V staining (Figure 9A), and in cleaved PARP and cleaved caspase-3 (Figure 9B). These results show that apoptosis can be induced by compound M1 in breast cancer cells.

CONCLUSIONS

In this study, a virtual screening process including RF classification and regression models was applied to identify the novel Bcl-2 inhibitors. We finally screened out the compound M1 as a novel Bcl-2 small molecule inhibitor, which binds to and down-regulates Bcl-2, has significant cytotoxic effect by inducing mitochondria-mediated apoptosis. Compound M1 could be a novel and promising hit or lead compound for further structural optimization in drug discovery.

ETHICS STATEMENT

We have explored the anti-tumor effect of Compound M1 *in vitro*. The human breast cancer cell lines MDA-MB-231, T47D, and BT549, were purchased from Cell Bank of Chinese Academy of Sciences.

AUTHOR CONTRIBUTIONS

YC and DC designed, conceived the study, and revised the manuscript. MW, ZD, SJ, YG, and HW performed the

experiments and analyzed data. MW and ZD drafted the manuscript. XW, SX, YZ, and JY provided technical support in the experiments.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China No. 81422051, 81402853 and 81472593, the National Key Basic Research Program (2015CB910700), the Project of Innovation-driven Plan in Central South University (502211812), and the Hunan Natural Science Foundation of China No. 2016JJ1020.

REFERENCES

- Baell, J. B., and Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53, 2719–2740. doi: 10.1021/jm901137j
- Bai, L., and Wang, S. (2013). Targeting apoptosis pathways for new cancer therapeutics. *Annu. Rev. Med.* 65, 139–155. doi: 10.1146/annurev-med-010713-141310
- Bemis, G. W., and Murcko, M. A. (1996). The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* 39, 2887–2893. doi: 10.1021/jm9602928
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Buolamwini, J. K. (1999). Novel anticancer drug discovery. *Curr. Opin. Chem. Biol.* 3, 500–509. doi: 10.1016/S1367-5931(99)80073-8
- Cao, D.-S., Xu, Q.-S., Liang, Y.-Z., Zhang, L.-X., and Li, H.-D. (2010). The boosting: a new idea of building models. *Chemometr. Intell. Lab. Syst.* 100, 1–11. doi: 10.1016/j.chemolab.2009.09.002
- Cao, D. S., Xu, Q. S., Hu, Q. N., and Liang, Y. Z. (2013). ChemoPy: freely available python package for computational biology and chemoinformatics. *Bioinformatics* 29, 1092–1094. doi: 10.1093/bioinformatics/btt105
- Cheng, F., Li, W., Liu, G., and Tang, Y. (2013). *In silico* ADMET prediction: recent advances, current challenges and future trends. *Curr. Top. Med. Chem.* 13, 1273–1289. doi: 10.2174/15680266113139990033
- Chong, S. J. F., Low, I. C. C., and Pervaiz, S. (2014). Mitochondrial ROS and involvement of Bcl-2 as a mitochondrial ROS regulator. *Mitochondrion* 19, 39–48. doi: 10.1016/j.mito.2014.06.002
- Day, C. L., Smits, C., Fan, F. C., Lee, E. F., Fairlie, W. D., and Hinds, M. G. (2008). Structure of the BH3 domains from the p53-inducible BH3-only proteins noxa and puma in complex with Mcl-1. *J. Mol. Biol.* 380, 958–971. doi: 10.1016/j.jmb.2008.05.071
- Doi, K., Li, R., Sung, S. S., Wu, H., Liu, Y., Manieri, W., et al. (2012). Discovery of marinopyrrole A (Maritoclax) as a selective Mcl-1 antagonist that overcomes ABT-737 resistance by binding to and targeting Mcl-1 for proteasomal degradation. *J. Biol. Chem.* 287, 10224–10235. doi: 10.1074/jbc.M111.334532
- Dong, J., Wang, N. N., Yao, Z. J., Zhang, L., Cheng, Y., Ouyang, D., et al. (2018a). ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminform.* 10:29. doi: 10.1186/s13321-018-0283-x
- Dong, J., Yao, Z.-J., Zhang, L., Luo, F., Lin, Q., Lu, A.-P., et al. (2018b). PyBioMed: a python library for various molecular representations of chemicals, proteins and DNAs and their interactions. *J. Cheminform.* 10, 16–16. doi: 10.1186/s13321-018-0270-2
- Dong, X., Jiang, C., Hu, H., Yan, J., Chen, J., and Hu, Y. (2009). QSAR study of Akt/protein kinase B (PKB) inhibitors using support vector machine. *Eur. J. Med. Chem.* 44, 4090–4097. doi: 10.1016/j.ejmech.2009.04.050
- Fernald, K., and Kurokawa, M. (2013). Evading apoptosis in cancer. *Trends Cell Biol.* 23, 620–633. doi: 10.1016/j.tcb.2013.07.006
- Fulda, S. (2013). Modulation of mitochondrial apoptosis by PI3K inhibitors. *Mitochondrion* 13, 195–198. doi: 10.1016/j.mito.2012.05.001
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, 1100–1107. doi: 10.1093/nar/gkr777
- Green, D. R., and Evan, I. G. (2002). A matter of life and death. *Cancer Cell* 1, 19–30. doi: 10.1016/S1535-6108(02)00024-7
- Halgren, T. A. (2015). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comp. Chem.* 17, 490–519. doi: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P
- Jie, D., Cao, D. S., Miao, H. Y., Shao, L., Deng, B. C., Yun, Y. H., et al. (2015). ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminform.* 7:60. doi: 10.1186/s13321-015-0109-z
- Kennedy, T. (1997). Managing the drug discovery/development interface. *Drug Discov. Today* 2, 436–444. doi: 10.1016/S1359-6446(97)01099-4
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951
- Kirkin, V., Joos, S., and Zörnig, M. (2004). The role of Bcl-2 family members in tumorigenesis. *BBA Mol. Cell Res.* 1644, 229–249. doi: 10.1016/j.bbamcr.2003.08.009
- Kunal, R., and Supratik, K. (2015). “Importance of applicability domain of QSAR Models,” in *Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment*, ed K. Roy (Hershey, PA: IGI Global), 180–211
- Lagorce, D., Sperandio, O., Baell, J. B., Miteva, M. A., and Villoutreix, B. O. (2015). FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res.* 43, 200–207. doi: 10.1093/nar/gkv353
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 64, 4–17. doi: 10.1016/j.addr.2012.09.019
- Liu, G., Kelly, W. K., Wilding, G., Leopold, L., Brill, K., and Somer, B. (2009). An open-label, multicenter, phase I/II study of single-agent AT-101 in men with Castrate-Resistant Prostate Cancer (CRPC). *Clini. Cancer Res.* 15, 3172–3176. doi: 10.1158/1078-0432.CCR-08-2985
- Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K. (2007). BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 35, 198–201. doi: 10.1093/nar/gkl999
- Martin, L. A., and Dowsett, M. (2013). BCL-2: a new therapeutic target in estrogen receptor-positive breast cancer? *Cancer Cell* 24, 7–9. doi: 10.1016/j.ccr.2013.06.006
- Michielan, L., Bolcato, C., Federico, S., Cacciari, B., Bacilieri, M., Klotz, K. N., et al. (2009). Combining selectivity and affinity predictions using an integrated Support Vector Machine (SVM) approach: an alternative tool to discriminate between the human adenosine A(2A) and A(3) receptor pyrazolo-triazolo-pyrimidine antagonists binding sites. *Bioorg. Med. Chem.* 17, 5259–5274. doi: 10.1016/j.bmc.2009.05.038
- Molecular Operating Environment (MOE) (2016). *Molecular Operating Environment*. Chemical Computing Group Inc., ULC.

The studies meet with the approval of the university's review board.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00391/full#supplementary-material>

Figure S1 | The cytotoxic effects of eight small molecule compounds on breast cancer cells. MDA-MB-231 cells were treated with eight small molecule compounds for 48h. At the end of treatment, cell viability was measured by CCK-8 reagent.

- Ojha, P. K., Mitra, I., Das, R. N., and Roy, K. (2011). Further exploring rm2 metrics for validation of QSPR models. *Chemometr. Intell. Lab. Syst.* 107, 194–205. doi: 10.1016/j.chemolab.2011.03.011
- Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., et al. (2005). An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435, 677–681. doi: 10.1038/nature03579
- Prentis, R. A., Lis, Y., and Walker, S. R. (1988). Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964–1985). *Br. J. Clin. Pharmacol.* 25, 387–396. doi: 10.1111/j.1365-2125.1988.tb03318.x
- Ready, N., Karaseva, N. A., Orlov, S. V., Luft, A. V., Popovych, O., Holmlund, J. T., et al. (2011). Double-blind, placebo-controlled, randomized phase 2 study of the proapoptotic agent AT-101 plus docetaxel, in second-line non-small cell lung cancer. *J. Thor. Oncol.* 6, 781–785. doi: 10.1097/JTO.0b013e31820a0ea6
- Reed, J. C. (2002). Apoptosis-based therapies. *Nat. Rev. Drug Dis.* 1, 111–121. doi: 10.1038/nrd726
- Roberts, A. W., and Huang, D. (2017). Targeting BCL2 with BH3 mimetics: basic science and clinical application of venetoclax in chronic Lymphocytic leukemia and related B cell malignancies. *Clin. Pharmacol. Ther.* 101, 89–98. doi: 10.1002/cpt.553
- Robinson, D. D., Winn, P. J., Lyne, P. D., and Richards, W. G. (1999). Self-organizing molecular field analysis: a tool for structure-activity studies. *J. Med. Chem.* 42, 573–583. doi: 10.1021/jm9810607
- Roizen, M. F. (2012). Hallmarks of cancer: the next generation. *Yearbook Anesthesiol. Pain Manage.* 2012:13. doi: 10.1016/j.yane.2012.02.046
- Roy, K. (2007). On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opin. Drug Discov.* 2, 1567–1577. doi: 10.1517/17460441.2.12.1567
- Roy, K., Kar, S., and Ambure, P. (2015). On a simple approach for determining applicability domain of QSAR models. *Chemometr. Intell. Lab. Syst.* 145, 22–29. doi: 10.1016/j.chemolab.2015.04.013
- Schneider, P., Tanrikulu, Y., and Schneider, G. (2009). Self-organizing maps in drug discovery: compound library design, scaffold-hopping, repurposing. *Curr. Med. Chem.* 16, 258–266. doi: 10.2174/092986709787002655
- Schoenwaelder, S. M., Jarman, K. E., Gardiner, E. E., Hua, M., Qiao, J., White, M. J., et al. (2011). Bcl-xL-inhibitory BH3 mimetics can induce a transient thrombocytopenia that undermines the hemostatic function of platelets. *Blood* 118, 1663–1674. doi: 10.1182/blood-2011-04-347849
- Seymour, J., Davids, M., and Pagel, P. (2013). Updated results of a phase I first-in-human study of the BCL-2 inhibitor ABT-199 (Gdc-0199) in patients with relapsed/refractory (R/R) chronic lymphocytic leukaemia (CLL) [ICML abstract 057]. *J. Clin. Oncol.* 31, 7018–7018. doi: 10.1200/jco.2013.31.15_suppl.7018
- Souers, A. J., Levenson, J. D., Boghaert, E. R., Ackler, S. L., Catron, N. D., Chen, J., et al. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* 19, 202–208. doi: 10.1038/nm.3048
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., and Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958. doi: 10.1021/ci034160g
- Tóthová, E., Fricova, M., Stecová, N., Kafková, A., and Elbertová, A. (2002). High expression of Bcl-2 protein in acute myeloid leukemia cells is associated with poor response to chemotherapy. *Blood* 81, 3091–3096. doi: 10.1038/sj.neo.7900211
- Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* 29, 476–488. doi: 10.1002/minf.2010.00061
- Vandenberg, C. J., and Cory, S. (2013). ABT-199, a new Bcl-2-specific BH3 mimetic, has *in vivo* efficacy against aggressive Myc-driven mouse lymphomas without provoking thrombocytopenia. *Blood* 121, 2285–2288. doi: 10.1182/blood-2013-01-475855
- Wang, N.-N., Huang, C., Dong, J., Yao, Z.-J., Zhu, M.-F., Deng, Z.-K., et al. (2017). Predicting human intestinal absorption with modified random forest approach: a comprehensive evaluation of molecular representation, unbalanced data, and applicability domain issues. *RSC Adv.* 7, 19007–19018. doi: 10.1039/C6RA28442F
- Wang, N. N., Deng, Y. H., Zhu, M. F., Wang, J. B., Wen, M., Yao, Z. J., et al. (2016). ADME properties evaluation in drug discovery: prediction of Caco-2 cell permeability using a combination of NSGA-II and boosting. *J. Chem. Inf. Model.* 56, 763–773. doi: 10.1021/acs.jcim.5b00642
- Wang, Y., Xing, J., Xu, Y., Zhou, N., Peng, J., Xiong, Z., et al. (2015). *In silico* ADME/T modelling for rational drug design. *Q. Rev. Biophys.* 48, 488–515. doi: 10.1017/S0033583515000190
- Wishart, D. D. S. (2007). Improving early drug discovery through ADME modelling. *Drugs R D* 8, 349–362. doi: 10.2165/00126839-200708060-00003
- Xu, J. (2002). A new approach to finding natural chemical structure classes. *J. Med. Chem.* 45, 5311–5320. doi: 10.1021/jm010520k
- Ye, W.-L., Yang, S.-Q., Zhang, L.-X., Deng, Z.-K., Li, W.-Q., Zhang, J.-W., et al. (2019). Multistep virtual screening for rapid identification of G Protein-Coupled Receptors Kinase 2 inhibitors for heart failure treatment. *Chemometr. Intell. Lab. Syst.* 185, 32–40. doi: 10.1016/j.chemolab.2018.12.015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wen, Deng, Jiang, Guan, Wu, Wang, Xiao, Zhang, Yang, Cao and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.