

Machine Learning Enabled Structure-Based Drug Repurposing Approach to Identify Potential CYP1B1 Inhibitors

Baddipadige Raju, Gera Narendra, Himanshu Verma, Manoj Kumar, Bharti Sapra, Gurleen Kaur, Subheet Kumar Jain, and Om Silakari*



Cite This: *ACS Omega* 2022, 7, 31999–32013



Read Online

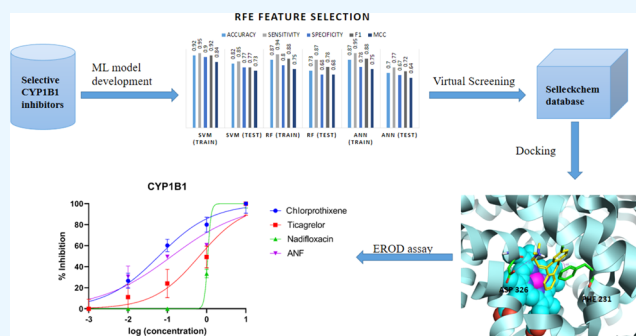
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Drug-metabolizing enzyme (DME)-mediated pharmacokinetic resistance of some clinically approved anticancer agents is one of the main reasons for cancer treatment failure. In particular, some commonly used anticancer medicines, including docetaxel, tamoxifen, imatinib, cisplatin, and paclitaxel, are inactivated by CYP1B1. Currently, no approved drugs are available to treat this CYP1B1-mediated inactivation, making the pharmaceutical industries strive to discover new anticancer agents. Because of the extreme complexity and high risk in drug discovery and development, it is worthwhile to come up with a drug repurposing strategy that may solve the resistance problem of existing chemotherapeutics. Therefore, in the current study, a drug repurposing strategy was implemented to find the possible CYP1B1 inhibitors using machine learning (ML) and structure-based virtual screening (SB-VS) approaches. Initially, three different ML models were developed such as support vector machines (SVMs), random forest (RF), and artificial neural network (ANN); subsequently, the best-selected ML model was employed for virtual screening of the selleckchem database to identify potential CYP1B1 inhibitors. The inhibition potency of the obtained hits was judged by analyzing the crucial active site amino acid interactions against CYP1B1. After a thorough assessment of docking scores, binding affinities, as well as binding modes, four compounds were selected and further subjected to *in vitro* analysis. From the *in vitro* analysis, it was observed that chlorprothixene, nadifloxacin, and ticagrelor showed promising inhibitory activity toward CYP1B1 in the IC₅₀ range of 0.07–3.00 μ M. These new chemical scaffolds can be explored as adjuvant therapies to address CYP1B1-mediated drug-resistance problems.



INTRODUCTION

Cytochrome P450 enzymes (CYPs) are heme-coupled monooxygenases involved in the metabolic biotransformation of a wide range of endogenous and exogenous compounds, including certain therapeutically approved anticancer medicines.¹ The human CYP superfamily is reported to be classified into 57 isoforms; among these, cytochrome P450 subfamily 1 (CYP1) enzymes including CYP1A1, CYP1A2, and CYP1B1 are of key interest to most researchers for their dominating involvement in the hydroxylation of procarcinogens such as PAHs and amines to cytotoxic and mutagenic compounds.² It is reported that CYP1B1 showed a 40% structural similarity with that of CYP1A1 and CYP1A2 isoforms.³ The specific overexpression of CYP1B1 was identified in extrahepatic tissues, such as ovarian, lung, brain, lymph, breast, and colon cancer tissues, whereas no detectable amount of CYP1B1 was found in nearby normal tissues.^{4–6} Unlike CYP1B1, overexpression of CYP1A1 was identified in the liver and some extrahepatic tissues including the pancreas, uterus, small intestine, and thymus, while CYP1A2 expression was found to be constitutive in human hepatic tissues.^{7,8} The expression

level of these three CYPs is highly influenced by interactions of certain planar aromatic compounds such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and 7,12-dimethylbenz[*a*]anthracene (DMBA) with the aryl hydrocarbon receptor (AhR) and a ligand-activated transcription factor.^{9–11} The CYP1B1 isoform is the most appealing therapeutic target among the three CYP1 members discussed above for the following reasons. (1) CYP1B1 overexpression has only been found in cancer tissues (ovarian, lung, brain, lymph, breast, and colon). (2) Besides its role in the bioactivation of a variety of procarcinogens, CYP1B1 may play a role in the carcinogenic action of estradiol by converting 17-estradiol (E2) into 4-hydroxyestradiol, which is a mutagenic compound capable of forming covalent interactions with DNA. This type of

Received: May 13, 2022

Accepted: August 23, 2022

Published: August 31, 2022



Table 1. Data Used in the Development of ML Models

target	ML algorithms	train set	test set	train set		test set	
CYP1B1 ^a	SVM RF ANN	192	80	selective inhibitors 96	nonselective inhibitors 96	selective inhibitors 40	nonselective inhibitors 40

^aCytochrome P450 1B1; drug-inactivating enzymes responsible for resistance to the Taxane class of anticancer drugs, and Imatinib.

conversion is not observed in the cases of CYP1A1 and CYP1A2 as these two isoforms mediate estrogen 17-hydroxylation primarily to 2-hydroxyestradiol, which is not a carcinogen. (3) It confers drug resistance by inactivating structurally diverse anticancer drugs such as docetaxel, imatinib, tamoxifen, cisplatin, and paclitaxel.^{12–14} As a result, a treatment strategy that promotes the selective inhibition of CYP1B1 as adjuvant therapy could be advantageous in avoiding CYP1B1-mediated drug resistance and cancer progression.

Currently, derivatives of various scaffolds such as α -naphthoflavones (ANF), flavonoids, trans-stilbenes, anthraquinones, coumarins, alkaloids, and chalcones have all been explored in the development of CYP1B1 inhibitors.^{3,5,10,15–18} However, none of these chemicals reached the market due to their limited solubility and selectivity problems.^{5,19,20} To identify potential CYP1B1 inhibitors with improved solubility and selectivity properties in the current study, the drug repurposing approach was utilized because it represents a cost-effective and appealing alternative strategy as existing licensed pharmacological compounds have undergone extensive safety and efficacy examinations.^{21–25} Drug repurposing accounts for roughly 30% of the new FDA-approved pharmaceuticals in the United States in recent years.²⁶ In the area of CYP1B1 research, numerous ligand-based computational techniques have already been forged;^{5,27} however, none of these models considered a broad chemical space (various chemical scaffolds) in the design process, and they are only applicable to certain chemical series inhibitors. Therefore, in the present study, the most diverse and potent CYP1B1 inhibitors were utilized in the development of machine learning (ML) models. The best-selected ML models were implemented in virtual screening of the FDA database (<https://www.selleckchem.com/screening-1>) in combination with structure-based virtual screening and *in vitro* analysis. Furthermore, molecular dynamics (MD) simulations were used to reveal the specific binding mechanisms between the hit molecule and its target. Thus, a novel CYP1B1 inhibitor with a well-established safety or toxicity profile can be anticipated from this study.

MATERIAL AND METHODS

Data Collection and Preparation. A total of 210 selective CYP1B1 inhibitors with IC_{50} ranges from 0.002 to 400 μ M were collected from ChEMBL, Pubchem, and literature sources.^{3–5,7,10,15,18,20,28–35} Some of the important considerations in data collection are as follows: (1) molecules tested with a similar assay technique, namely, the ethoxyresorufin-*O*-deethylase (EROD) assay, were considered for evaluating CYP1B1 and CYP1A1 enzyme inhibitory activity; and (2) the data of human recombinant cytochrome P450 1A enzymes (CYP1B1 and CYP1A1) were collected. Finally, one hundred thirty-six (136) inhibitors were marked as selective CYP1B1 inhibitors based on selectivity index values, i.e., IC_{50} of CYP1A1/ IC_{50} of CYP1B1 > 3-folds. The remaining 74 molecules were considered nonselective inhibitors, i.e., IC_{50} of CYP1A1/ IC_{50} of CYP1B1 < 3-folds. The details of the data with

biological activity values against CYP1B1 and CYP1A1 are given in the Supporting Information as Table S1.

Decoys Generation. The balance of data sets is critical for developing a solid predictive model. The performance of the ML model is highly influenced by data imbalance between classes.³⁶ To balance the data set, 62 decoy molecules were generated by computing their molecular properties based on the two diverse selective CYP1B1 structures using the DUDE online database (<http://dude.docking.org/>).³⁷ In the current study, the AM1-BCC method was used to compute partial charges of all molecules. The geometry of the whole data set was optimized using the AMBER force field by employing an implicit OBC2 solvent model with a minimum energy tolerance of 0.25 kcal/mol. Afterward, the data set was divided into training (192) and test sets (80) in the ratio of 7:3 using a random split algorithm in the CARET library³⁸ (Table 1).

Calculation of Molecular Fingerprints and Data Preprocessing. In predictive modeling, small molecules were represented in the form of a binary string, known as a molecular fingerprint. In particular, molecular fingerprints reveal details on the molecule's constitutive substructures. These molecular fingerprints are frequently supplied as a fixed-length string of integers 1 and 0, where 1 denotes the presence of a substructure in the given molecule and 0 denotes absence. The optimized structures of CYP1B1 inhibitors are imported into PaDEL software to compute different molecular fingerprints, namely, Estate (Est 79 bits), Molecular Access System (MACCS 166 bits), and PubChem fingerprint (PubChem 881 bits). A total of 1126 molecular fingerprints were calculated. To remove extraneous features from the data, some statistical filters were applied. The following methods were used to undertake data preprocessing: (1) deletion of molecular descriptors with zero variance values and (2) elimination of noise and bias among derived descriptors by the Pearson correlation analysis.

Feature Selection. To create a good prediction model, it is essential to choose an optimal number of descriptors. Feature selection methods have been extensively used to identify significant feature subsets for improving the prediction performance of the model. When developing a model, if molecular structures are represented by an excess number of descriptors than fixed samples, they may cause a problem of overfitting (unreliable predictions). Therefore, employing feature selection methods in predictive model development procedures is highly important. The details of the feature selection methods used in the current study are mentioned below.

Recursive Feature Elimination (RFE). The recursive feature elimination (RFE) method has recently gained a lot of popularity due to its high efficiency in discovering informative features or attributes in the classification of inhibitors associated with drug activity analysis in a variety of biochemical fields.^{39–42} The RFE algorithm assigns a ranking for features and recursively drops low-rank features and keeps important ones.⁴²

Boruta. The Boruta feature selection method is a wrapper strategy based on random forests that eliminates features that are less useful than random probes.⁴³ By considering the interactions between multiple features, the Boruta algorithm removes irrelevant inputs repeatedly. It creates shadow features by reversing the original feature's values. If a shadow feature's importance is greater than or equal to that of the original feature, it is classified as unimportant.

ML Classification Algorithms. In the ML approach, the mined properties of known compounds are used to predict the activity of unknown compounds.⁴⁴ In this work, different classification models such as SVM, RF, and ANN were developed in combination with RFE and Boruta feature selection methods to classify selective CYP1B1 inhibitors and noninhibitors. For developing ML models, Rstudio version 1.4.1717 was used in the current study.

Support Vector Machine (SVM). Vapnik created the support vector machine (SVM) model as a general data modeling methodology for pattern recognition.⁴⁵ The main idea behind the development of the SVM algorithm is to map the data into high dimensional space by constructing a hyperplane, which differentiates two separate classes of vectors with a maximum margin. The model can describe the position and direction of the hyperplane using a subset of training vectors. The initial training data in a space X is projected to a higher dimensional feature space in SVM using a Mercer kernel operator K . The following is a list of classifiers to be considered

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x)$$

Here, the maximal hyperplane is represented by α_i . When $f(x)$ is less than zero, x is classified as 1; when $f(x)$ is more than zero, x is classified as -1 . The radial basis kernel (RBF) function implemented with SVM was used to classify training and test set molecules.^{46–48} The SVM algorithm was performed with the help of the kernel parameters Cost (C) and Sigma (σ).³⁸ To optimize the performance of the developed SVM model, 10-fold cross-validation was performed.

Random Forest. A random forest is a collection of different decision trees. At each node, feature bagging finds the best data split for a random feature subset.⁴⁹ The final forecast is a consensus overall decision tree. The correlation between any two trees, as well as the strength of each tree, has an impact on the forest error rate. A tree with a low error rate is a strong classifier. The final prediction is based on a consensus among all decision trees $h(x;k)$, $k = 1$, where x is the observed input (covariate) vector of length p with the associated random vector X and θ_k is a set of independent and identically distributed random vectors.

$$h(x) = \left(\frac{1}{K}\right) \sum_{k=1}^k h(x; \theta_k)$$

There are a handful of parameters that can be adjusted (e.g., number of trees, number of features, and mtry), and these factors have a big impact on the algorithm's performance. In this study, random forest models were built using the 10-fold cross-validation method.

Artificial Neural Network (ANN). An artificial neural network (ANN) is made up of a large number of connected

artificial neurons.⁵⁰ Receiving neurons can process the input signal and then send it to downstream neurons connected to it. The weight assigned to the neurons can affect their learning rate. In the current study, the feed forward neural network was implemented with a single-layer logistic activation function to develop an ANN model. Furthermore, hyperparameter tuning parameters such as decay and cost were also utilized in combination with 10-fold cross-validation to optimize the developed models. The NNET package in Rstudio was used to create neural network models. The neuron's output signal O is given by the following equation

$$O = f(\text{net}) = f \sum_{j=1}^n (w_j x_j)$$

where w_j represents the weight vector and the variable (transfer) function is represented by $f(\text{net})$.

Model Validation. In this study, the robustness of the generated models was evaluated using 10-fold cross-validation of the training set. Data sets are frequently divided into two categories in binary classification models: selective and nonselective inhibitors. The following eqs 1–5 are used to calculate sensitivity (SE), specificity (SP), precision (P), accuracy (ACC), and Matthews's correlation coefficient (MCC) to measure the predictive quality of the developed ML models. The MCC of ML models generated by different algorithms can be compared to assess their quality. When a model's MCC score is near 1, it is deemed better, whereas when it reaches 0, it is considered a complete failure. Furthermore, the receiver operating characteristic (ROC) curve was plotted and the area under the ROC curve (AUC) was calculated. The values of AUC range from 0 to 1. An AUC value of 1 represents a flawless model, whereas 0.5 implies a random classifier, and 0.8 indicates a good model.^{51,52}

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

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

$$P = \frac{TP}{TP + FP} \quad (3)$$

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

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

Molecular Docking. Before performing the docking experiment conformation sampling method of the FLARE program was validated by the redocking experiment.⁵³ Initially, the cocrystal ligand ANF was extracted from its crystal structure of CYP1B1 (PDB ID: 3PM0); then, the extracted structure was redocked into protein 3PM0. Afterward, both cocrystallized and redocked ligands were aligned and the root-mean-square deviation (RMSD) of heavy atoms was calculated. When the RMSD is less than 2.0, the sampling is regarded as successful; otherwise, it is termed a failure. For the docking study, the PDB structures selected for CYP1B1 and CYP1A1 with PDB IDs 3PM0⁸ and 4I8V,⁵⁴ respectively, were

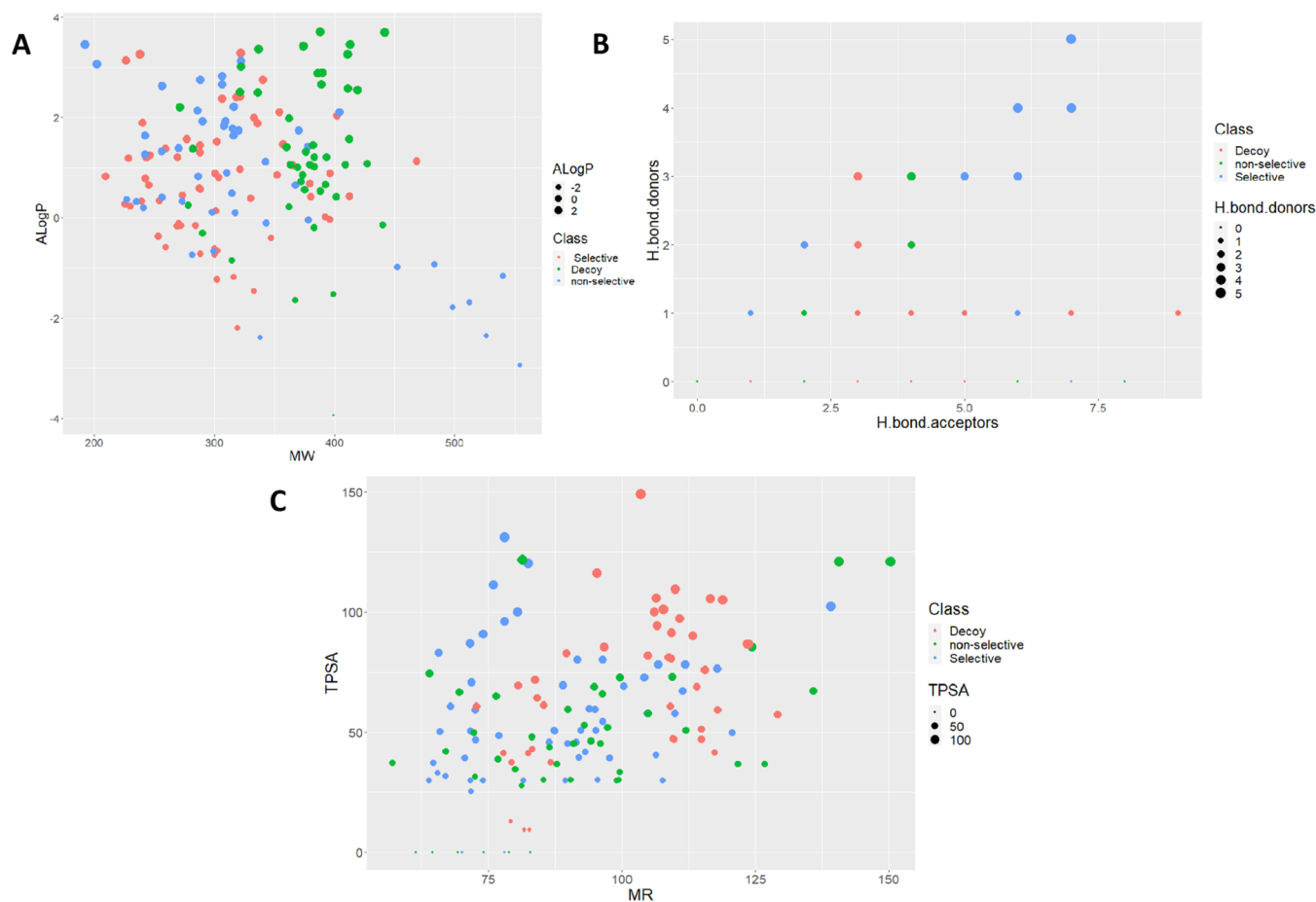


Figure 1. Chemical diversity analysis of training set molecules. (A) Molecular weight on the X-axis vs LogP on the Y-axis. (B) Hydrogen-bond acceptors on the X-axis vs hydrogen-bond donors on the Y-axis. (C) Molar refractivity on the X-axis vs topological polar surface area on the Y-axis.

based on the criteria reported in the literature. The proteins were prepared, which includes insertion of missing atoms in incomplete residues, modeling of missing loops, removal of cocrystallized external water, and protonation of the titratable residues using an XED force field at pH 7.4. A grid was generated by defining the binding site around the centroid of the cocrystallized ligand, i.e., α -naphthoflavone (ANF).

Molecular Dynamics. To examine the conformational changes and stability of the top docked ligands in association with CYP1B1, molecular dynamic (MD) simulations were performed for a period of 100 ns utilizing the Flare module of cresset working on the AMBER force field.⁵³ Molecules were solvated using the water-based TIP3P model to mimic the cellular environment. Following this, energy minimization of the system was performed to ensure that there were no steric conflicts and that the system had a good starting structure. The maximum number of iterations and minimum energy tolerance would be kept as 30 and 0.025, respectively. Flare's "Analysis" protocol is used to calculate the ligand-protein RMSD value and estimate the ligand's binding orientation. In this study, the remaining parameters have been set to default.

Enzymatic Assay. The recombinant human CYP1B1 and CYP1A1 enzymes were purchased from Sigma-Aldrich. 7-Ethoxyresorufin (7-ER), nicotinamide adenine dinucleotide phosphate (NADPH), glibenclamide, nadifloxacin, chlorprothixene, umbelliferone, propyl gallate, and α naphthoflavone (ANF) were also purchased from Sigma-Aldrich. Ticagrelor was received as a kind gift sample from Cipla Ltd, Mumbai.

The inhibitory activity of these repurposed drugs against the CYP1B1 and CYP1A1 enzymes was determined using the EROD assay.^{55,56} In general, a 200 μ L mixture containing various concentrations (final concentrations were 1, 10, 100, 1000, and 10000 nM) of tested compounds (except for positive and negative control wells), an enzyme source (5 μ L of 10 picomole CYP1B1 and 5 μ L of 2.5 picomole CYP1A1), 45 μ L of 0.1 M (pH = 7.4) phosphate buffer, and 50 μ L of 7-ER were incubated in a black 96-well flat-bottomed microplate at 37 $^{\circ}$ C for various durations of time (incubation times for CYP1B1 and CYP1A1 were 35 and 15 min, respectively). Finally, 50 μ L of 1.67 mM NADPH was added to the aforementioned mixture to start the reaction process. Later, using a Biotek Gen 5 microplate reader with 530 and 590 nm excitation and emission filters, the fluorescence intensity was measured for 10 min with a 2 min interval. Three replicates were used for each concentration, and each test was performed in triplicate. The IC_{50} value of each compound was calculated with the AAT Bioquest calculator <https://www.aatbio.com/tools>.

RESULTS AND DISCUSSION

Data Collection and Data Diversity. A total of 210 selective CYP1B1 inhibitors were collected from the different small-molecule databases and the literature. DUDE, a database for useful decoys, was used to generate 62 decoys. In chemometrics, the diversity of chemical compounds significantly influences the outcome and prediction of an ML model.

Table 2. List of Molecular Fingerprints Used in Model Development

feature selection method	no. of features	selected features
RFE	20	MACCSFP90 + MACCSFP126 + PubchemFP189 + PubchemFP798 + EStateFP11 + PubchemFP686 + MACCSFP150 + PubchemFP374 + MACCSFP49 + PubchemFP715 + PubchemFPS37 + PubchemFP667 + EStateFP29 + PubchemFP196 + MACCSFP144 + PubchemFP12 + PubchemFP695 + PubchemFP696 + MACCSFP69 + PubchemFP206
Boruta	19	EStateFP11 + EStateFP29 + MACCSFP49 + MACCSFP90 + MACCSFP126 + MACCSFP134 + MACCSFP137 + MACCSFP144 + MACCSFP150 + PubchemFP189 + PubchemFP196 + PubchemFP374 + PubchemFP537 + PubchemFP686 + PubchemFP695 + PubchemFP696 + PubchemFP715 + PubchemFP798

This can be conquered using a wide range of chemical diversity data sets. In the current study, the structural diversity of training set compounds was measured using AlogP, molecular weight (MW), hydrogen-bond donors (H.bond.donors), hydrogen-bond acceptors (H.bond.acceptors), topological polar surface area (TPSA), and molecular refractivity index (MR) descriptors. Figure 1A shows that the vast range of molecular weights ranging from approximately 144–600 and the AlogP ranging from -2.9 to 4 indicates that this training set has sufficient chemical space. Interestingly, it was also observed that nonselective inhibitors represented by the blue color showed a higher MW than selective inhibitors represented by the green color. Along with MW and AlogP, other drug-likeness descriptors including H.bond.donors and H.bond.acceptors were estimated, as seen in Figure 1B. This led to the discovery that H.bond.acceptors ranged from 0 to 7 , while H.bond.donors ranged from 0 to 5 . Likewise, TPSA and MR index values ranged from 0 to 150 , as illustrated in Figure 1C. Finally, from this entire analysis, it was concluded that the training set has sufficient chemical space and could be used for the development of ML models.

Data Preprocessing and Feature Selection. First of all, 1126 molecular fingerprints were calculated for CYP1B1 data using the freely available PaDEL tool. Later constant and highly correlated (>0.6) features were removed using the variable threshold and Pearson correlation method. Ultimately, after data preprocessing, 160 fingerprints were identified. These identified features were further subjected to feature selection methods using RFE and Boruta algorithms to identify important and relevant features. Finally, subsets of 20 and 19 features were selected based on the RFE and Boruta variable importance scores. The identified features using both algorithms are given in Table 2. Finally, different ML models, including SVM, RF, and ANN, were developed using each set of identified features.

Performance of Developed ML Models. A 10-fold cross-validation approach was used to test the performance of the developed predictive models (with optimized parameters), and the results of each model are represented in Tables 3 and 4. Additionally, the results are also illustrated in the form of bar charts in Figures 2 and 3. Initially, RFE-selected features were used to develop ML models. All of the developed models were optimized by various tuning parameters by implementing the grid search method. For example, the SVM model was created using the RBF and optimized using the caret package's Cost (0.1 – 2) and Sigma (1 – 3) parameters, whereas the RF model was optimized using the mtry parameter, i.e., 7 . On the other hand, the ANN model was optimized using size (1 – 3) and decay (0.01 – 0.1) parameters. The best model was chosen from the constructed models using statistical parameters such as SE, SP, P, ACC, and MCC. From Figure 2, it can be observed that all of the developed models showed predictive accuracy greater than 0.7 on both training and test sets. However, overall, the SVM model showed better results than RF and ANN in terms of ACC and MCC values on a 10-fold cross-validated test set. The results of each model developed using Boruta features are given in Figure 3. All models showed more or less similar results. However, in comparison, the SVM model showed better 10-fold cross-validated ACC and MCC values for the test set. Hence, the SVM models were selected as the best predictive models for both feature selection methods. Among these, the best model was selected by comparing their MCC values. According to Chicco et al., the MCC produces a

Table 3. Results of the Models Trained Using RFE-Selected Features

s. no.	method		accuracy	sensitivity	specificity	F1	MCC
1	SVM C = 2.6 Sigma = 0.1	train	0.92	0.95	0.9	0.92	0.84
		test	0.82	0.85	0.77	0.77	0.73
2	RF Mtry = 4	train	0.87	0.94	0.80	0.88	0.75
		test	0.73	0.87	0.68	0.78	0.68
3	ANN size = 1 decay = 0.01	train	0.87	0.95	0.78	0.88	0.75
		test	0.7	0.77	0.67	0.72	0.64

Table 4. Results of the Models Trained Using Boruta Selected Features

s. no.	method		accuracy	sensitivity	specificity	F1	MCC
1	SVM C = 2.1 Sigma = 0.1	train	0.91	0.94	0.86	0.91	0.80
		test	0.77	0.87	0.74	0.79	0.70
2	RF Mtry = 7	train	0.9	0.95	0.83	0.90	0.80
		test	0.72	0.87	0.71	0.76	0.67
3	ANN size = 1 decay = 0.01	train	0.90	0.96	0.83	0.90	0.80
		test	0.7	0.77	0.65	0.71	0.60

RFE Feature selection

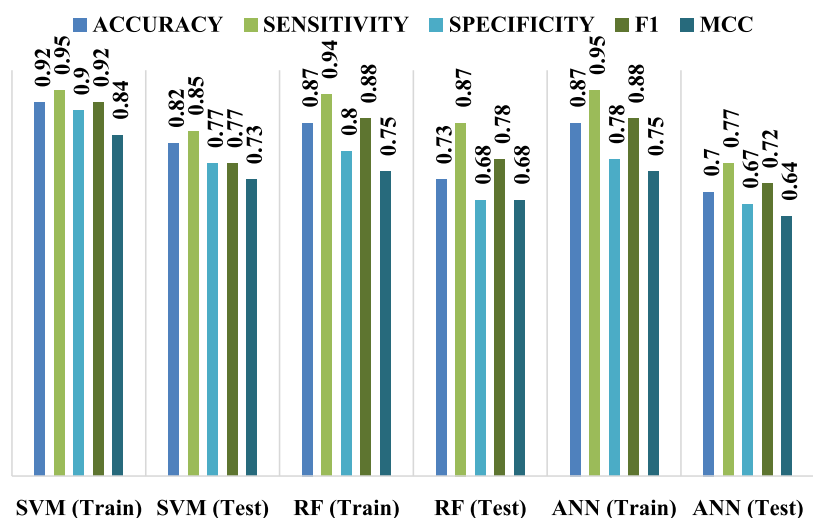


Figure 2. Performance of the 10-fold cross-validated models developed by the RFE features.

BORUTA FEATURE SELECTION

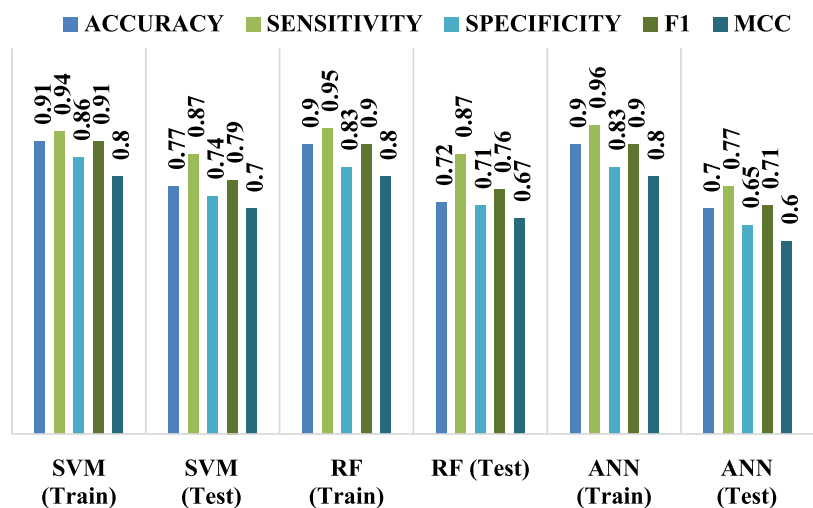


Figure 3. Performance of the 10-fold cross-validated models developed by the Boruta features.

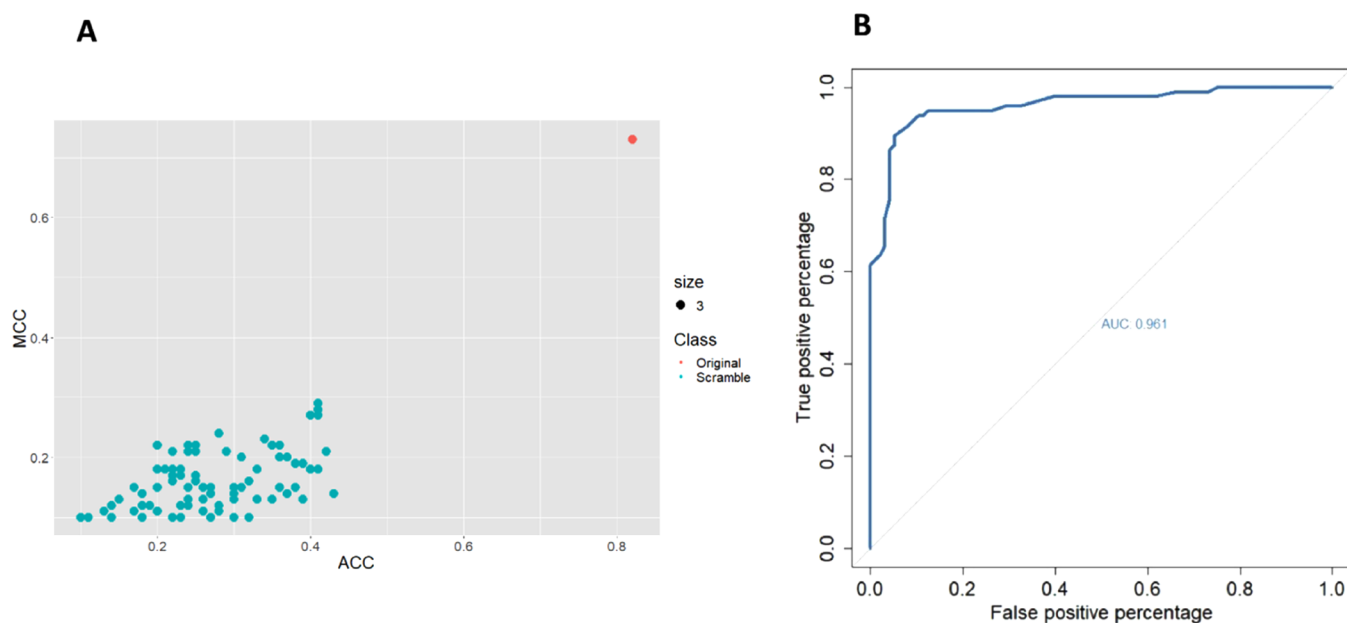


Figure 4. (A) Y-randomization results of the best-selected RFE+SVM model. Blue dots indicate the scrambled models and the red dot indicates the original model. (B) ROC curve of the best-selected RFE+SVM model on the test set.

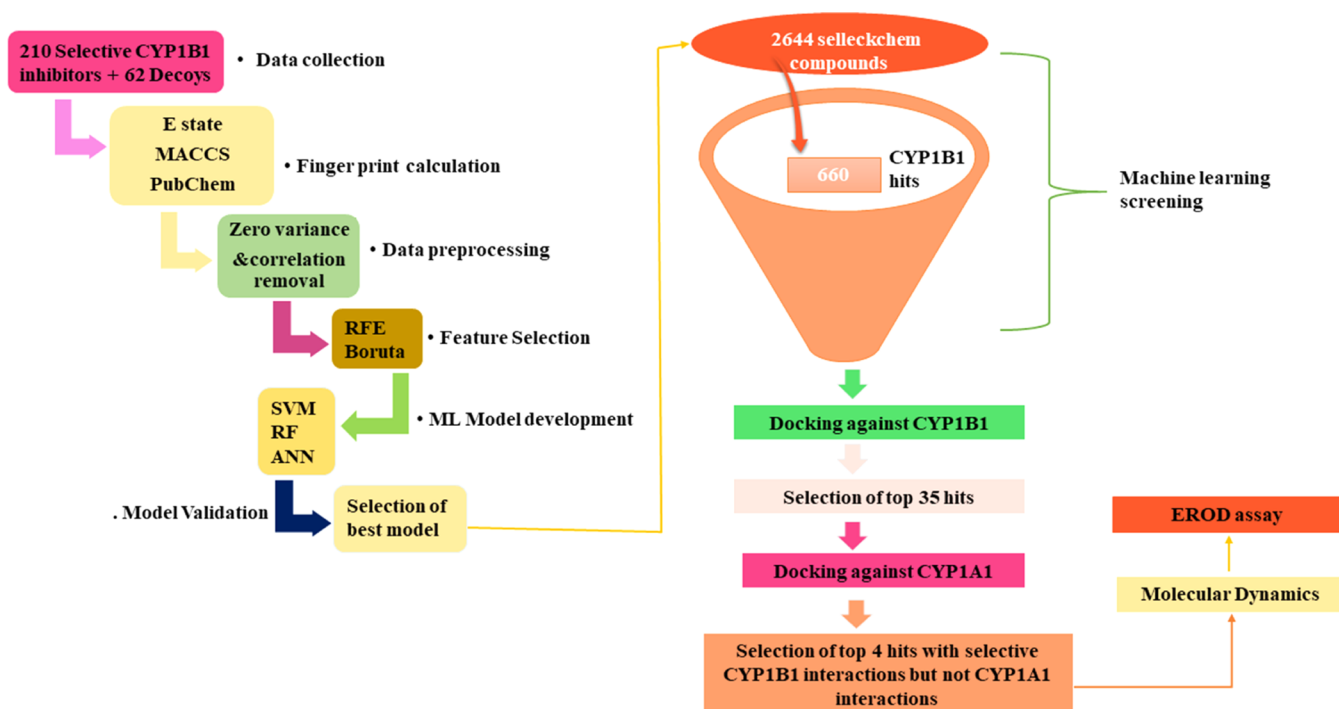


Figure 5. Schematic representation of the virtual screening pipeline.

more informative and truthful result in evaluating binary classification models than accuracy.⁵⁷ Hence, by considering the MCC values, the SVM model developed using RFE features (RFE+SVM) was selected as the best model. The best-selected SVM model was tuned using sigma (σ) 0.1 and Cost (C) 2.6 parameters. Furthermore, the performance of the best-selected model was assessed by an external validation method. An external data set of 37 selective inhibitors and 168 decoys was used in this method. The best-selected model provided 0.95% accuracy, 0.90% sensitivity, and 0.96% specificity for the external validation set. In addition, the MCC values showed

that the best-selected model (RFE+VM) is capable of classifying both negative and positive classes, as evidenced by MCC values of 0.73 and 0.82 for 10-fold CV and external validation, respectively. All of these findings show that the created SVM model is more effective at distinguishing highly selective CYP1B1 inhibitors from nonselective CYP1B1 inhibitors.

Y-Scrambling and ROC Curve Analysis. Random scrambling experiments were performed to prove that the best performance of the given SVM model was not a result of chance correlation. The classification labels (i.e., “S”, “N”) of

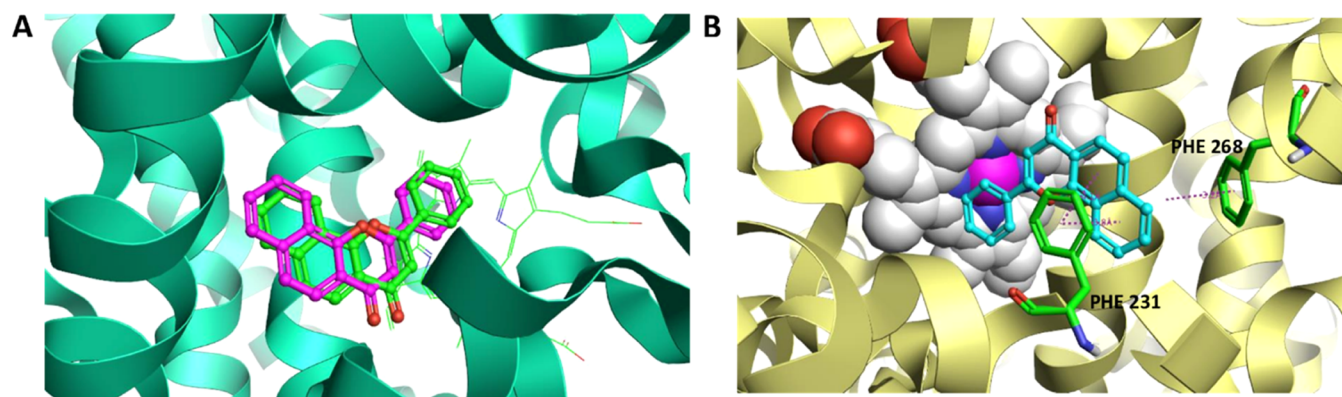


Figure 6. (A) Conformations of the docking ligand and extracted cocrystal ligand (ANF) are superimposed (the green ligand represents the cocrystal ligand pose, and the pink ligand represents the redocking ligand pose). (B) Binding mode of ANF inside the CYP1B1 active site.

Table 5. Top 35 Virtual Hits' Information toward CYP1B1

molecule	LF dG ^a	LF VS score ^b	LF Rank score ^c	selleckchem ID
ticagrelor	-11.9	-14.4	-11.1	878
bosutinib	-11.8	-13.7	-4.4	5
brinzolamide	-11.8	-11.5	-8.9	682
ceforanide	-11.8	-12.9	-6.2	2012
xipamide	-11.8	-12.1	-12.8	1961
dasatinib	-11.7	-14.3	-12.8	6
ezetimibe	-11.7	-13.9	-12.2	249
ozenoxacin	-11.7	-11.9	-10.7	2762
acotiamide- <i>d</i> ₆	-11.6	-12.5	-9.8	1981
trichlormethiazide	-11.6	-11.2	-10.7	750
vilazodone	-11.6	-13.5	-4.7	2269
darolutamide	-11.6	-12.3	-13.4	1689
donepezil	-11.6	-12.6	-5.5	2011
haloperidol	-11.6	-13.0	-12.1	108
dolutegravir	-11.5	-12.6	-12.7	583
cabergoline	-11.5	-12.8	-2.9	2267
isavuconazole	-11.5	-13.0	-10.7	1310
ceftazole	-11.5	-10.9	-8.9	2058
pantethine	-11.4	-13.3	-5.2	1989
lapatinib	-11.4	-15.2	-8.8	383
pyritinol	-11.4	-12.2	-11.7	1927
(-)-epigallocatechin gallate	-11.4	-13.1	0.53	445
<i>D</i> - α -tocopheryl acetate	-11.4	-12.8	-1.5	1919
pexidartinib	-11.4	-13.3	-13.2	1818
lenvatinib	-11.4	-12.8	-7.8	422
ozanimod	-11.4	-13.6	-14.1	1834
halofuginone	-11.3	-12.1	-12.4	1840
nebivolol	-11.3	-13.1	-12.1	2506
gefarnate	-11.3	-12.1	-1.5	2367
doripenem	-11.3	-12.3	-6.7	2339
glyceryl trioctanoate octanoic acid	-11.3	-12.4	0.434	2643
peimine	-11.3	-12.2	-2.6	2326
glibenclamide	-10.2	-12.4	-8.4	280
chlorprothixene	-9.9	-10.8	-12.3	303
nadifloxacin	-9.08	-10.3	-10.2	429

^aScore (kcal) used to rank the poses of docked compounds. ^bScore (kcal) used to rank the binding affinity of docked compounds. ^cScore (kcal) used to rank the docked compounds in the virtual screening experiment.

the training set (182) molecules were randomly shuffled. Afterward, attempts were made to build an RFE+SVM model using scrambled data. A total of 30 randomized models were created; all of these scrambled models were found to have lower accuracy and MCC scores than the unscrambled model (Original model), as shown in Figure 4A. The area under the ROC curve is an important measure for evaluating the model's performance. As a result, it was also plotted to demonstrate a binary classifier system's diagnostic ability to distinguish between two classes. From Figure 4B, it can be observed that the AUC value of the best-selected SVM models on the test set is 0.79, indicating that the generated models have good prediction ability and reliability.

ML-Based Virtual Screening. The best-selected ML model (i.e., RFE+SVM) was implemented in virtual screening (VS) of the Selleckchem database, which consists of 2644 FDA-approved chemical compounds. As a result of screening, 660 hits were identified to be selective CYP1B1 inhibitors, which are further subjected to docking studies to identify the novel chemical scaffolds as CYP1B1 inhibitors. The schematic representation of the VS protocol is shown in Figure 5.

Molecular Docking. Following the development of ML models, validation, and virtual screening, the best predicted FDA molecules were further evaluated for target binding affinity against CYP1B1 (PDB ID: 3PM0) and CYP1A1 (PDB ID: 4I8V) enzymes through molecular docking simulation studies. Redocking simulation was performed using the crystallographic structure of 3PM0. From Figure 6A, it was observed that both ligands (i.e., extracted crystal ligand and redocked ligand) were entirely overlapped in the same orientation with a lower RMSD (i.e., 0.5 Å), which indicates that the conformational sampling method of the Flare program appears to be reliable. The interaction patterns of ANF (cocystal ligand) have also been analyzed. As illustrated in Figure 6B, three pi–pi stacking interactions were observed between the naphthalene moiety of ANF and active site residues of PHE 231 and PHE 268, which are crucial for binding. From the literature, it is evident that PHE 231, PHE 268, and PHE 134 are crucial active residues involved in the development of pi–pi stacking interactions in CYP1B1 inhibition.⁵⁸ However, PHE 224 and 123 play an important role in selective inhibition of CYP1A1 by forming crucial pi–pi stacking interactions.⁵⁸ By considering this crucial interaction and its binding affinity scores, the selectivity of the identified hits was judged against 1B1 over 1A1.

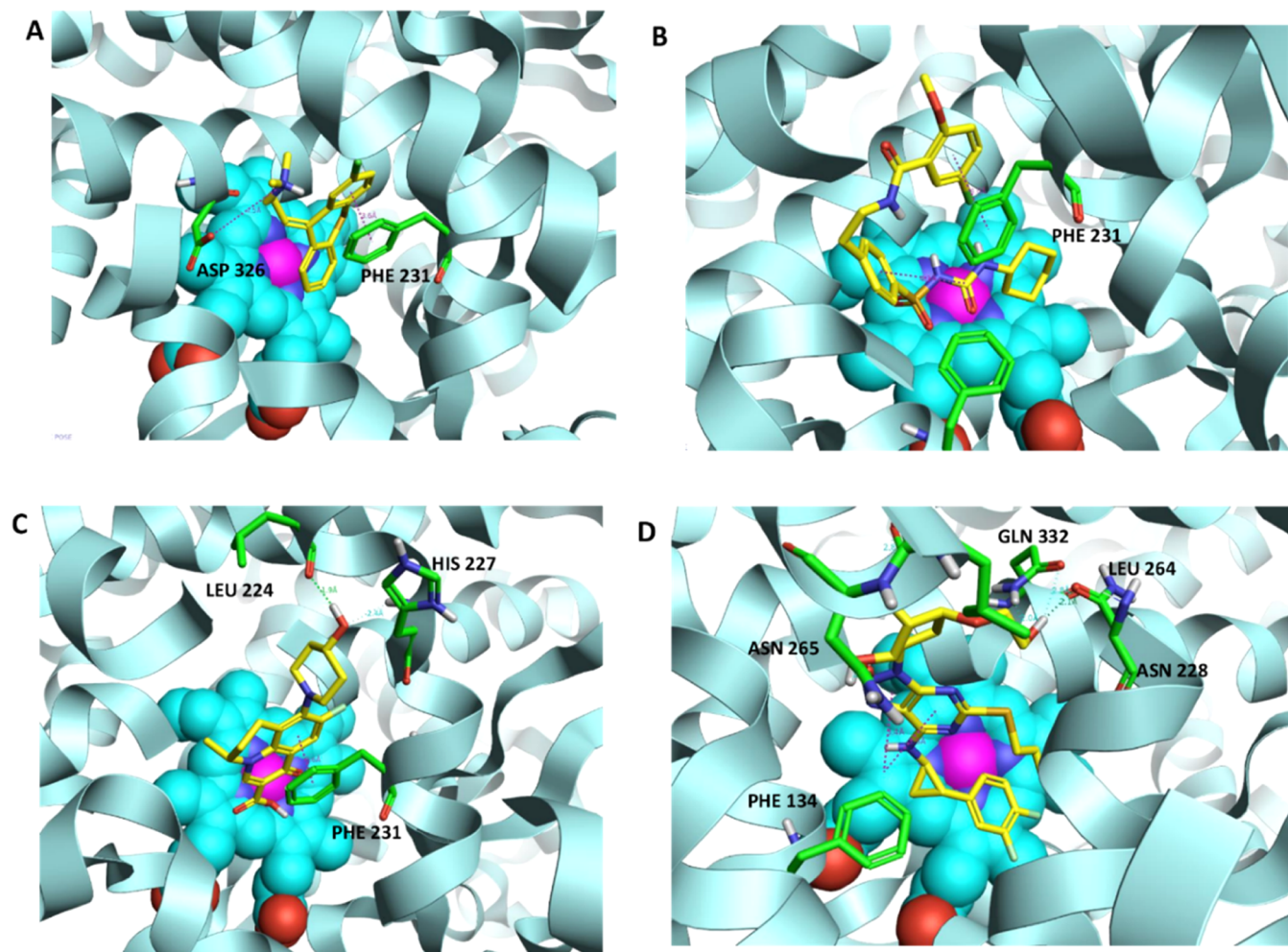


Figure 7. Binding interaction patterns of the top four compounds in the active site of CYP1B1: (A) chlorprothixene, (B) glibenclamide, (C) nadifloxacin, and (D) ticagrelor. Pink-colored-lined dots represent π - π stacking interactions, dark-green-colored-lined dots represent strong hydrogen bonds, and light-green color represents weak hydrogen-bond interactions.

Initially, all of the 660 hits were docked in the active site of CYP1B1, and their interaction profiles and docking scores were analyzed. Based on the docking scores, 35 molecules were selected, which showed high binding affinity scores (i.e., <-8.5 kcal/mol) than that of the cocrystal ligand (i.e., -8.5 kcal/mol). The identified top 35 molecules with their docking scores are given in Table 5. All of the 35 molecules showed crucial π - π stacking interactions with PHE 231, PHE 268, and PHE 134 of CYP1B1. Furthermore, to understand the selectivity of these identified hits over 1A1, all of the molecules were docked against CYP1A1, and their docking scores and interaction profiles were analyzed. After comprehensive consideration of docking scores, binding modes, and crucial interactions, four compounds, namely, chlorprothixene, glibenclamide, nadifloxacin, and ticagrelor, were considered selective CYP1B1 inhibitors because none of these molecules showed selective interaction against 1A1. Indeed, similar to the cocrystal ligand (ANF), these compounds display a high binding affinity with CYP1B1 by establishing the intramolecular interactions with its key amino acids such as PHE 231, PHE 134, ASP 326, LEU 224, HIS 227, GLN 332, ASN 228, LUE 264, and ASN 265. The docking interactions of the top four molecules are illustrated in Figure 7. To understand the stability of these interactions, the top four molecules were further subjected to molecular dynamics studies.

Molecular Dynamics. The stability and flexibility of each docking complex of CYP1B1 with chlorprothixene, glibenclamide, nadifloxacin, and ticagrelor were assessed as a function of time using a 100 ns MD simulation under an explicit hydration environment. MD simulation findings were assessed in terms of π - π stacking, hydrogen-bond dynamic stability, and structural flexibility of CYP1B1. The stability of the trajectory during MD simulation was determined by analyzing the RMSD of all complexes based on protein-heavy atoms. The RMSD plot for each ligand-protein complex is shown in Figure 8A–D, and it can be seen that the RMSD value of the ligand-protein fluctuates during an initial period of 35 ns and gradually stabilizes, and no significant fluctuation was observed in the four complexes. Following this, all complexes were equilibrated with average RMSD values of less than 1 Å. According to Gupta et al., stable protein backbone fluctuations and moderate RMSD values are good indications of molecular dynamics simulation stability.⁵⁹ The low RMSD values indicate that over the 100 ns MD simulation, the protein conformation did not change or changed only slightly, confirming the logic and stability of the initial docking complexes. RMSD for the complexes CYP1B1 with chlorprothixene showed that the structure of the systems equilibrated well after the 15 ns simulation (Figure 8A). A slight fluctuation was observed during the 80–85 ns simulation period with an RMSD of 0.5

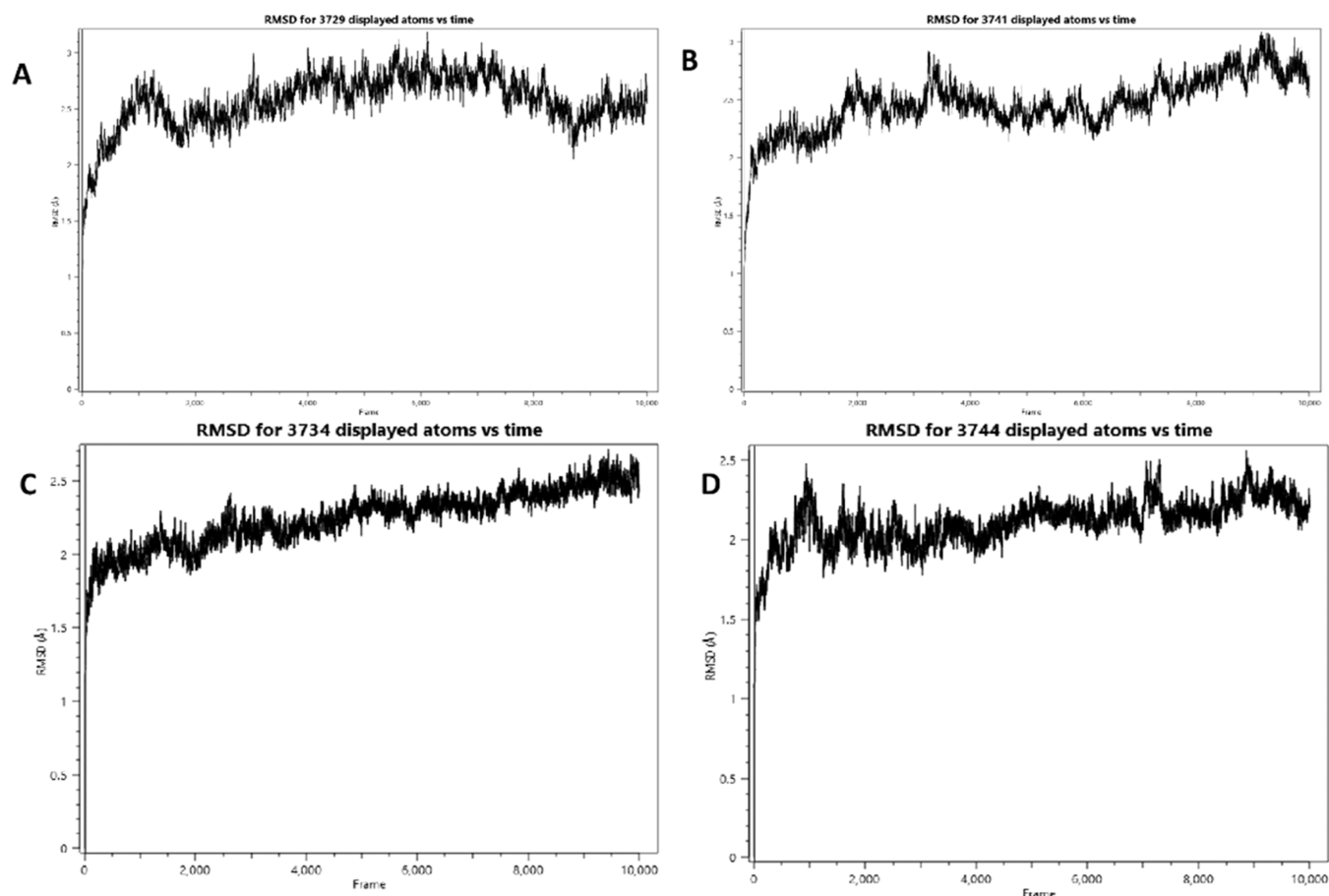


Figure 8. RMSD plot of top four molecules in complex with CYP1B1: (A) chlorprothixene, (B) glibenclamide, (C) nadifloxacin, and (D) ticagrelor.

Å; however, it showed that RMSD profiles were always less than 1 Å for chlorprothixene throughout 7–100 ns simulations. From Figure 9A, it was observed that two initial docking interactions such as PHE 231 and ASP 326 were retained in the case of the compound chlorprothixene, indicating the stability of the initial docking complex. The CYP1B1 complex with glibenclamide (Figure 8B) showed that the structure of the system is equilibrated well after 35 ns simulation with RMSD of less than 1 Å. As shown in Figure 9B, throughout the 100 ns simulation time, glibenclamide maintained two initial pi–pi stacking contacts with the residue PHE 231. In the case of nadifloxacin, slight fluctuations were observed with an RMSD of 0.5 Å during 20–40 ns simulations (Figure 8C); as a result, one new pi–pi stacking was observed with residue PHE 134, and two initial docking contacts were lost, i.e., LEU 224 and HIS 227 (Figure 9C). From Figure 8D, it was observed that after the 20 ns simulation, the CYP1B1-ticagrelor system reached equilibrium. The CYP1B1-ticagrelor complex slightly fluctuated (i.e., <1 Å) during the 65–75 ns simulation; as a result, one new pi–pi stacking interaction was observed with the residue PHE 231 shown in Figure 9D. Thus, the stability of CYP1B1 in chlorprothixene, glibenclamide, nadifloxacin, and ticagrelor-bound states was found to be suitable for postanalysis. The RMSD and crucial pi–pi stacking interactions of all complexes showed that they stably bounded to CYP1B1 throughout 100 ns simulations. The MD simulations revealed that key pi–pi stacking and hydrogen-bond interactions were conserved. The best four docked compounds

have essential contacts with the enzyme's active region, indicating that these compounds could potentially inhibit CYP1B1. Enzymatic test assays were performed utilizing human recombinant cytochrome P450 enzymes, namely, CYP1B1 and CYP1A1 enzymes, to see how well these four compounds can potentially inhibit CYP1B1.

In Vitro CYP Inhibition Assay. The inhibitory effects of the top four compounds on recombinant human CYP1B1 and CYP1A1 enzymes were evaluated using the standard EROD assay, which has been widely used as a method for the assessment of CYP1 inhibitory activity.^{19,60–62} ANF, a well-known CYP1 inhibitor, was used as a reference drug. The top selected drugs (chlorprothixene, nadifloxacin, glibenclamide, and ticagrelor) were examined at five concentrations as mentioned in the methodology section to investigate their CYP1B1 inhibitory activity. As shown in Figure 10, for each compound in the mixture, IC₅₀ was calculated using a dose–response curve plotted between percentage inhibition and log concentration (these observations were calculated in triplicate). According to general evaluation, three drugs, including chlorprothixene, nadifloxacin, and ticagrelor, exhibited potential inhibition effects on both CYP1 isoforms as mentioned in Table 6, whereas 50% inhibition was not observed at 10 μM in the case of glibenclamide. Among all compounds, chlorprothixene showed potent inhibitory activity against 1B1 (i.e., IC₅₀ = 0.072 ± 0.008 μM) compared to 1A1 (i.e., IC₅₀ = 0.49 ± 0.04 μM) with 7-fold selectivity. This high potency may be assumed to be due to their crucial pi–pi stacking and salt

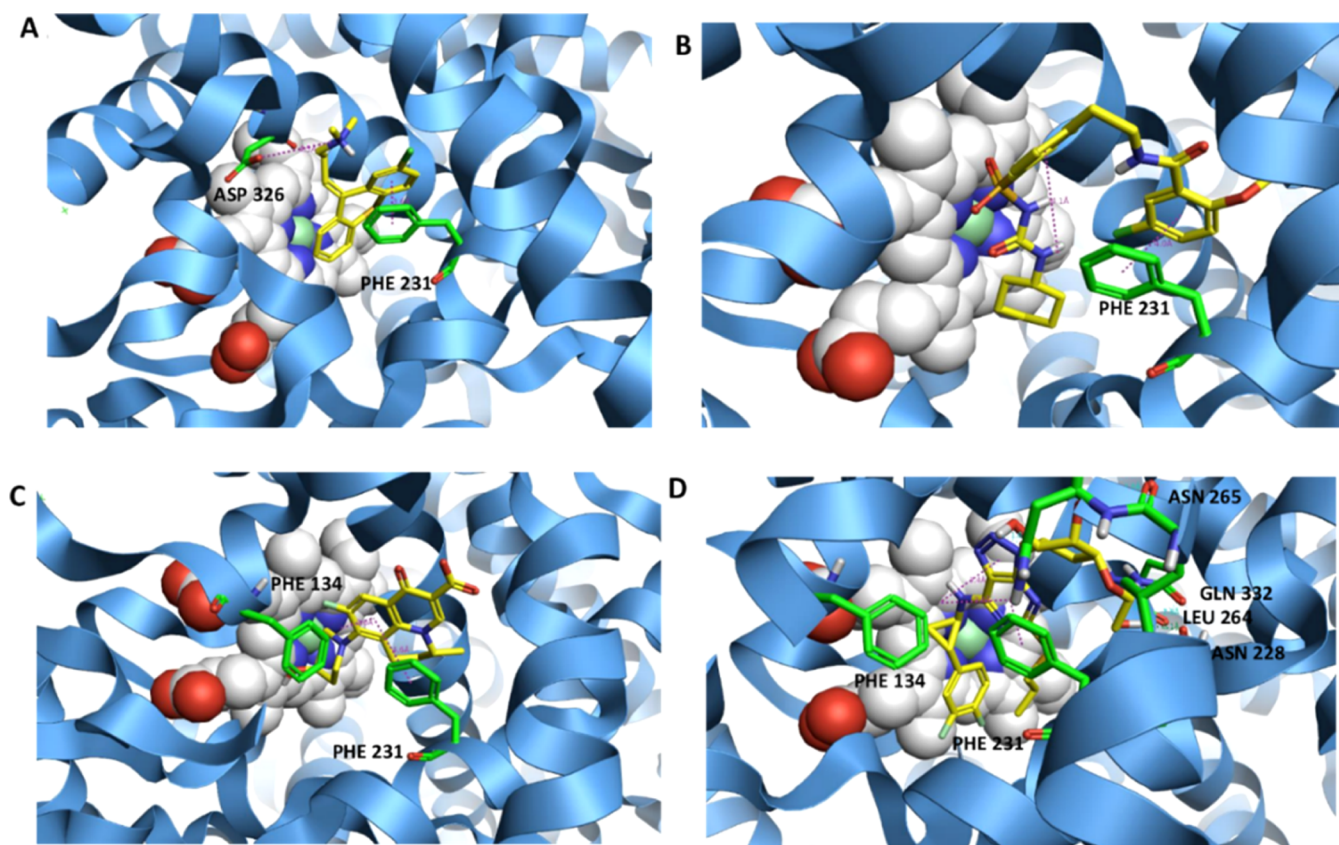


Figure 9. Interactions of top four ligands in complex with CYP1B1 (3PM0) after 100 ns molecular dynamics simulations: (A) chlorprothixene, (B) glibenclamide, (C) nadifloxacin, and (D) ticagrelor.

bridge interactions with PHE 231 and ASP 326 residues, as displayed in Figure 9A. From the literature, it was evident that salt bridge interactions play an important role in conformational specificity and the molecular recognition process. The binding energy of the salt bridge interactions is significantly higher than the hydrogen-bond and pi–pi stacking interactions.⁶³ In the current study, chlorprothixene showed one pi–pi stacking and one salt bridge interaction with residues PHE 231 and ASP 326, respectively (Figure 9A). However, no salt bridge interactions were observed in the case of compounds glibenclamide, nadifloxacin, and ticagrelor (Figure 9B–D). From these results, it was concluded that ASP 326 and PHE 231 are the only contributors to chlorprothixene potency. On the other hand, nadifloxacin and ticagrelor showed moderate potency against both CYP1 isoforms. However, slightly better activity was observed against 1B1 with IC_{50} values of 1.46 and 2.81 μ M, respectively. Finally, these findings led to the conclusion that pi–pi stacking interaction with PHE 231 is responsible for the potency of three compounds including chlorprothixene, nadifloxacin, and ticagrelor. A study conducted by Du et al. suggested that chlorprothixene, a dopamine receptor antagonist, can induce cell death by apoptosis in myeloid leukemia cell lines.^{64,65} This may direct oncologists to explore chlorprothixene as a potential adjuvant in the case of imatinib-resistant-acute myeloid leukemia cell lines. Additionally, the second hit Ticagrelor, an antiplatelet drug, has been reported to show breast cancer prevention activity in orthotopic 4T1 breast cancer models by inhibiting tumor-cell platelet aggregation.⁶³ From these findings, it was hypothesized that ticagrelor may be used as adjuvant therapy in docetaxel-resistant breast cancer. Nadifloxacin is a

fluoroquinolone class of antibacterial drug that has been reported to show promising anticancer activity in triple-negative breast cancer cell lines (i.e., MDA-MB-231) by apoptosis mechanism.⁶⁶ These results suggested that the combination therapy of docetaxel and nadifloxacin could improve the anticancer activity in docetaxel-resistant breast cancer cell lines. Herein, we report that chlorprothixene, ticagrelor, and nadifloxacin may ameliorate the anticancer activity in docetaxel-resistant cell lines by inhibiting CYP1B1.

CONCLUSIONS

Resistance against clinically authorized anticancer drugs is a major challenge in cancer treatment. Several commonly used anticancer medications, such as docetaxel, imatinib, and paclitaxel, are inactivated by CYP1B1, which ultimately results in drug resistance. A number of commonly used anticancer medications, such as docetaxel, tamoxifen, imatinib, and paclitaxel, are specifically quickly inactivated by CYP1B1, which ultimately results in drug resistance. Herein, efforts have been made to integrate *in silico* and *in vitro* approaches to identify possible selective CYP1B1 inhibitors. Based on the existing knowledge about already known selective and nonselective CYP1B1 inhibitors, three ML models such as SVM, RF, and ANN were developed. The best-selected SVM +RFE model was validated with known CYP1B1 inhibitors and Y-scrambling experiments. The best-selected model with an MCC value on the test set of 0.73 was used to screen the selleckchem database. To estimate the binding affinity, binding pose, and factors governing the binding process of the obtained hits, molecular docking and molecular dynamics (MD)

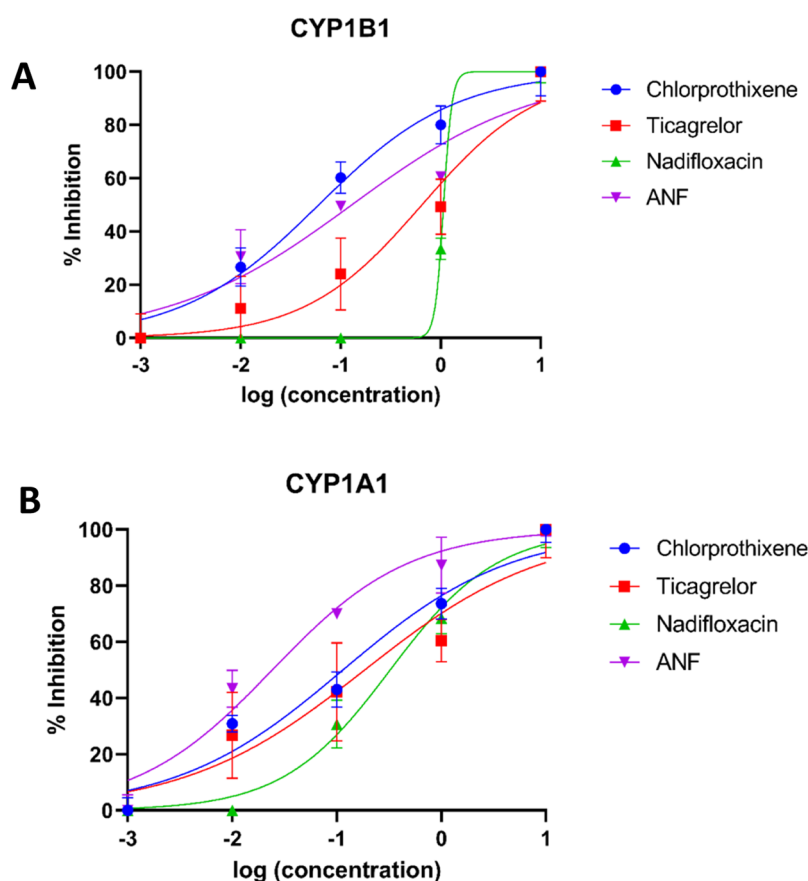


Figure 10. Inhibitory potencies of the top four molecules identified from the EROD assay. (A) CYP1B1 and (B) CYP1A1.

Table 6. Identified Inhibitory Potencies of the Top Four Molecules in Human Recombinant CYP-P450 Enzymes

molecule	recombinant human CYP-P450 enzymes/ IC_{50} μ M		selectivity CYP1A1/CYP1B1
	CYP1B1 \pm SD	CYP1A1 \pm SD	
chlorprothixene	0.07 \pm 0.01	0.49 \pm 0.04	6.8
glibenclamide	>10	>10	
nadifloxacin	1.46 \pm 0.06	2.36 \pm 0.20	1.61
ticagrelor	2.81 \pm 0.90	3.45 \pm 0.39	1.22
ANF	0.02 \pm 0.01	0.02 \pm 0.004	1

simulations have been performed. From interaction and RMSD analysis at the atomic level through the 100 ns MD simulation, it was found that four compounds such as chlorprothixene, glibenclamide, nadifloxacin, and ticagrelor maintain mandatory pi–pi stacking and hydrogen-bonding interactions and can be considered to be the best. The ligand-binding affinity was then assessed via an *in vitro* enzymatic assay. The 50% inhibition concentration (IC_{50}) of these four compounds to CYP1B1 was determined via an EROD assay. Three compounds including chlorprothixene, nadifloxacin, and ticagrelor were found to be highly potent inhibitors of CYP1B1 because they established IC_{50} values of 0.072, 1.463, and 2.81 μ M, respectively. Although it was expected to inhibit CYP1A1, to our surprise, chlorprothixene exhibited significant selectivity (7-fold) toward CYP1B1 over CYP1A1 than the cocrystal ligand (ANF). From molecular dynamics results, it was concluded that pi–pi stacking and salt bridge interaction with PHE 231 and ASP 326 play a crucial role in the inhibitory potency of chlorprothixene, whereas, on the other hand, the same kind

of salt bridge interaction, i.e., ASP 326, was absent in other compounds. These identified new leads can be considered for developing new therapeutics to treat CYP1B1-mediated drug-resistance problems in cancer patients. Overall, it can be concluded that the discovery of novel CYP1B1 inhibitors with the help of multiple ML and classical SB-VS methods may be beneficial to cope with the problem of chemoresistance.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c02983>.

List of small-molecule data used in the ML model development (PDF)

AUTHOR INFORMATION

Corresponding Author

Om Silakari – Molecular Modeling Lab (MML), Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab 147002, India; orcid.org/0000-0002-8314-7395; Phone: +91-9501542696; Email: omsilakari@gmail.com; Fax: +91-1752283075

Authors

Baddipadige Raju – Molecular Modeling Lab (MML), Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab 147002, India
Gera Narendra – Molecular Modeling Lab (MML), Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab 147002, India

Himanshu Verma – Molecular Modeling Lab (MML), Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab 147002, India
Manoj Kumar – Molecular Modeling Lab (MML), Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab 147002, India
Bharti Sapra – Molecular Modeling Lab (MML), Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab 147002, India
Gurleen Kaur – Center for Basic and Translational Research in Health Sciences, Guru Nanak Dev University, Amritsar 143005, India
Subheet Kumar Jain – Center for Basic and Translational Research in Health Sciences, Guru Nanak Dev University, Amritsar 143005, India

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c02983>

Funding

This work was financially supported by the Indian Council of Medical Research (ICMR), New Delhi [ISRM/12(10)/2019].

Notes

No potential conflict of interest was reported by the author(s). The authors declare no competing financial interest.

ACKNOWLEDGMENTS

B.R. would like to acknowledge the Indian Council of Medical Research (ICMR), New Delhi, for providing a Senior Research Fellowship (SRF). B.R. would like to acknowledge Shubham Thakur, ICMR senior research fellow, department of pharmaceutical sciences, Guru Nanak Dev University, Amritsar, for helping in *in vitro* studies. B.R. would like to acknowledge Cipla Pharmaceuticals, Mumbai, for providing a gift sample of the ticagrelor drug.

ABBREVIATIONS

DMEs, drug-metabolizing enzymes; CYPs, cytochrome P450s; ML, machine learning; VS, virtual screening; SB-VS, structure-based virtual screening; RFE, recursive feature elimination; RF, random forest; SVM, support vector machine; ANN, artificial neural network; MD, molecular dynamics; AML, acute myeloid leukemia; ANF, α -naphthoflavone; EROD, ethoxyresorufin-O-deethylase; TCDD, 2,3,7,8-tetrachlorodibenzo-p-dioxin; DMBA, 7,12-dimethylbenz[a]anthracene; AhR, aryl hydrocarbon receptor; DUDE, database of useful decoys; TP, true positive; TN, true negative; FP, false positive; FN, false negative; ACC, accuracy; SE, sensitivity; SP, specificity; P, precision; MCC, Matthews correlation coefficient; AUC, area under the curve; RMSD, root-mean-square deviation; ROC, receiver operating characteristic curve; NADPH, nicotinamide adenine dinucleotide phosphate; 7-ER, 7-ethoxyresorufin

REFERENCES

- (1) Lang, D.; Radtke, M.; Bairlein, M. Highly variable expression of CYP1A1 in human liver and impact on pharmacokinetics of Riociguat and Granisetron in humans. *Chem. Res. Toxicol.* **2019**, *32*, 1115–1122.
- (2) Dong, J.; Zhang, Q.; Cui, Q.; Huang, G.; Pan, X.; Li, S. Flavonoids and naphthoflavonoids: wider roles in the modulation of cytochrome P450 family 1 enzymes. *ChemMedChem* **2016**, *11*, 2102–2118.
- (3) Don, M.-J.; Lewis, D. F.; Wang, S.-Y.; Tsai, M.-W.; Ueng, Y.-F. Effect of structural modification on the inhibitory selectivity of

rutaecarpine derivatives on human CYP1A1, CYP1A2, and CYP1B1. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2535–2538.

- (4) Dutour, R.; Cortés-Benítez, F.; Roy, J.; Poirier, D. Structure-based design and synthesis of new estrane-pyridine derivatives as cytochrome P450 (CYP) 1B1 Inhibitors. *ACS Med. Chem. Lett.* **2017**, *8*, 1159–1164.

- (5) Dong, J.; Wang, Z.; Cui, J.; Meng, Q.; Li, S. Synthesis and structure-activity relationship studies of α -naphthoflavone derivatives as CYP1B1 inhibitors. *Eur. J. Med. Chem.* **2020**, *187*, No. 111938.

- (6) Murray, G. I.; Taylor, M. C.; McFadyen, M. C.; McKay, J. A.; Greenlee, W. F.; Burke, M. D.; Melvin, W. T. Tumor-specific expression of cytochrome P450 CYP1B1. *Cancer Res.* **1997**, *57*, 3026–3031.

- (7) Shimada, T.; Yamazaki, H.; Foroozesh, M.; Hopkins, N. E.; Alworth, W. L.; Guengerich, F. P. Selectivity of polycyclic inhibitors for human cytochrome P450s 1A1, 1A2, and 1B1. *Chem. Res. Toxicol.* **1998**, *11*, 1048–1056.

- (8) Wang, A.; Savas, U.; Stout, C. D.; Johnson, E. F. Structural characterization of the complex between α -naphthoflavone and human cytochrome P450 1B1. *J. Biol. Chem.* **2011**, *286*, 5736–5743.

- (9) Takemura, H.; Nagayoshi, H.; Matsuda, T.; Sakakibara, H.; Morita, M.; Matsui, A.; Ohura, T.; Shimoi, K. Inhibitory effects of chrysoeriol on DNA adduct formation with benzo [a] pyrene in MCF-7 breast cancer cells. *Toxicology* **2010**, *274*, 42–48.

- (10) Kim, S.; Ko, H.; Park, J. E.; Jung, S.; Lee, S. K.; Chun, Y.-J. Design, synthesis, and discovery of novel trans-stilbene analogues as potent and selective human cytochrome P450 1B1 inhibitors. *J. Med. Chem.* **2002**, *45*, 160–164.

- (11) Endo, K.; Uno, S.; Seki, T.; Ariga, T.; Kusumi, Y.; Mitsumata, M.; Yamada, S.; Makishima, M. Inhibition of aryl hydrocarbon receptor transactivation and DNA adduct formation by CYP1 isoform-selective metabolic deactivation of benzo [a] pyrene. *Toxicol. Appl. Pharmacol.* **2008**, *230*, 135–143.

- (12) McFadyen, M. C.; McLeod, H. L.; Jackson, F. C.; Melvin, W. T.; Doehmer, J.; Murray, G. I. Cytochrome P450 CYP1B1 protein expression: A novel mechanism of anticancer drug resistance. *Biochem. Pharmacol.* **2001**, *62*, 207–212.

- (13) Rochat, B.; Morsman, J. M.; Murray, G. I.; Figg, W. D.; McLeod, H. L. Human CYP1B1 and anticancer agent metabolism: mechanism for tumor-specific drug inactivation? *J. Pharmacol. Exp. Ther.* **2001**, *296*, 537–541.

- (14) McFadyen, M. C.; Murray, G. I. Cytochrome P450 1B1: a novel anticancer therapeutic target. *Future Oncol.* **2005**, *1*, 259–263.

- (15) Takemura, H.; Itoh, T.; Yamamoto, K.; Sakakibara, H.; Shimoi, K. Selective inhibition of methoxyflavonoids on human CYP1B1 activity. *Bioorg. Med. Chem.* **2010**, *18*, 6310–6315.

- (16) Sridhar, J.; Liu, J.; Foroozesh, M.; Klein Stevens, C. L. Inhibition of cytochrome p450 enzymes by quinones and anthraquinones. *Chem. Res. Toxicol.* **2012**, *25*, 357–365.

- (17) Mammen, J. S.; Kleiner, H. E.; DiGiovanni, J.; Sutter, T. R.; Strickland, P. T. Coumarins are competitive inhibitors of cytochrome P450 1B1, with equal potency for allelic variants. *Pharmacogenet. Genomics* **2005**, *15*, 183–188.

- (18) Horley, N. J.; Beresford, K. J.; Chawla, T.; McCann, G. J.; Ruparelia, K. C.; Gatchie, L.; Sonawane, V. R.; Williams, I. S.; Tan, H. L.; Joshi, P.; et al. Discovery and characterization of novel CYP1B1 inhibitors based on heterocyclic chalcones: Overcoming cisplatin resistance in CYP1B1-overexpressing lines. *Eur. J. Med. Chem.* **2017**, *129*, 159–174.

- (19) Dong, J.; Huang, G.; Cui, Q.; Meng, Q.; Li, S.; Cui, J. Discovery of heterocycle-containing α -naphthoflavone derivatives as water-soluble, highly potent and selective CYP1B1 inhibitors. *Eur. J. Med. Chem.* **2021**, *209*, No. 112895.

- (20) Kubo, M.; Yamamoto, K.; Itoh, T. Design and synthesis of selective CYP1B1 inhibitor via dearomatization of α -naphthoflavone. *Bioorg. Med. Chem.* **2019**, *27*, 285–304.

- (21) Pushpakom, S.; Iorio, F.; Meyers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; Doig, A.; Guilleams, T.; Latimer, J.; McNamee, C.; et al.

- Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discovery* **2019**, *18*, 41–58.
- (22) Catalano, R.; Rocca, R.; Juli, G.; Costa, G.; Maruca, A.; Artese, A.; Caracciolo, D.; Tagliaferri, P.; Alcaro, S.; Tassone, P.; Amodio, N. A drug repurposing screening reveals a novel epigenetic activity of hydroxychloroquine. *Eur. J. Med. Chem.* **2019**, *183*, No. 111715.
- (23) Hassan, A. H.; Yoo, S. Y.; Lee, K. W.; Yoon, Y. M.; Ryu, H. W.; Jeong, Y.; Shin, J.-S.; Kang, S.-Y.; Kim, S.-Y.; Lee, H.-H.; et al. Repurposing mosloflavone/S, 6, 7-trimethoxyflavone-resveratrol hybrids: Discovery of novel p38- α MAPK inhibitors as potent interceptors of macrophage-dependent production of proinflammatory mediators. *Eur. J. Med. Chem.* **2019**, *180*, 253–267.
- (24) Ramesh, R.; Shingare, R. D.; Kumar, V.; Anand, A.; Swetha, B.; Veeraraghavan, S.; Viswanadha, S.; Ummanni, R.; Gokhale, R.; Reddy, D. S. Repurposing of a drug scaffold: Identification of novel sila analogues of rimonabant as potent antitubercular agents. *Eur. J. Med. Chem.* **2016**, *122*, 723–730.
- (25) Wu, K.-J.; Zhong, H.-J.; Li, G.; Liu, C.; Wang, H.-M. D.; Ma, D.-L.; Leung, C.-H. Structure-based identification of a NEDD8-activating enzyme inhibitor via drug repurposing. *Eur. J. Med. Chem.* **2018**, *143*, 1021–1027.
- (26) Jin, G.; Wong, S. T. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discovery Today* **2014**, *19*, 637–644.
- (27) Tan, H. L. *Selective Inhibitors of the Cytochrome P450 Enzyme CYP1B1*; De Montfort University, 2006.
- (28) Androutsopoulos, V. P.; Spyrou, I.; Ploumidis, A.; Papalampros, A. E.; Kyriakakis, M.; Delakas, D.; Spandidos, D. A.; Tsatsakis, A. M. Expression profile of CYP1A1 and CYP1B1 enzymes in colon and bladder tumors. *PLoS One* **2013**, *8*, No. e82487.
- (29) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Motow, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954.
- (30) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; et al. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.
- (31) Mohd Siddique, M. U.; McCann, G. J.; Sonawane, V.; Horley, N.; Williams, I. S.; Joshi, P.; Bharate, S. B.; Jayaprakash, V.; Sinha, B. N.; Chaudhuri, B. Biphenyl urea derivatives as selective CYP1B1 inhibitors. *Org. Biomol. Chem.* **2016**, *14*, 8931–8936.
- (32) Mikstaka, R.; Rimando, A. M.; Dutkiewicz, Z.; Stefański, T.; Sobiak, S. Design, synthesis and evaluation of the inhibitory selectivity of novel trans-resveratrol analogues on human recombinant CYP1A1, CYP1A2 and CYP1B1. *Bioorg. Med. Chem.* **2012**, *20*, 5117–5126.
- (33) Mikstaka, R.; Baer-Dubowska, W.; Wiczorek, M.; Sobiak, S. Thiomethylstilbenes as inhibitors of CYP1A1, CYP1A2 and CYP1B1 activities. *Mol. Nutr. Food Res.* **2008**, *52*, S77–S83.
- (34) Mikstaka, R.; Przybylska, D.; Rimando, A. M.; Baer-Dubowska, W. Inhibition of human recombinant cytochromes P450 CYP1A1 and CYP1B1 by trans-resveratrol methyl ethers. *Mol. Nutr. Food Res.* **2007**, *51*, 517–524.
- (35) Doostdar, H.; Burke, M. D.; Mayer, R. T. Bioflavonoids: selective substrates and inhibitors for cytochrome P450 CYP1A and CYP1B1. *Toxicology* **2000**, *144*, 31–38.
- (36) Longadge, R.; Dongre, S. Class Imbalance Problem in Data Mining Review. 2013, arXiv:1305.1707. arXiv.org e-Print archive. <http://arxiv.org/abs/1305.1707>.
- (37) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594.
- (38) Kuhn, M. *Caret: Classification and Regression Training*; Astrophysics Source Code Library, 2015.
- (39) Lin, H.; Han, L.; Yap, C.; Xue, Y.; Liu, X.; Zhu, F.; Chen, Y. Prediction of factor Xa inhibitors by machine learning methods. *J. Mol. Graphics Modell.* **2007**, *26*, 505–518.
- (40) Xue, Y.; Li, H.; Ung, C.; Yap, C.; Chen, Y. Classification of a diverse set of Tetrahymena pyriformis toxicity chemical compounds from molecular descriptors by statistical learning methods. *Chem. Res. Toxicol.* **2006**, *19*, 1030–1039.
- (41) Zhang, L.; Ai, H.; Zhao, Q.; Zhu, J.; Chen, W.; Wu, X.; Huang, L.; Yin, Z.; Zhao, J.; Liu, H. In *Computational Prediction of Influenza Neuraminidase Inhibitors Using Machine Learning Algorithms and Recursive Feature Elimination Method*, International Symposium on Bioinformatics Research and Applications; Springer, 2017; pp 344–349.
- (42) Zhang, S.; Li, Y.; Qin, Z.; Tu, G.; Chen, G.; Yan, A. SAR study on inhibitors of GIIA secreted phospholipase A2 using machine learning methods. *Chem. Biol. Drug Des.* **2019**, *93*, 666–684.
- (43) Kursu, M. B.; Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **2010**, *36*, 1–13.
- (44) Wahi, D.; Jamal, S.; Goyal, S.; Singh, A.; Jain, R.; Rana, P.; Grover, A. Cheminformatics models based on machine learning approaches for design of USP1/UAF1 abrogators as anticancer agents. *Syst. Synth. Biol.* **2015**, *9*, 33–43.
- (45) Vishwanathan, S.; Murty, M. N. *SSVM: A Simple SVM Algorithm*, Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290); IEEE, 2002; pp 2393–2398.
- (46) Yuan, L.-F.; Ding, C.; Guo, S.-H.; Ding, H.; Chen, W.; Lin, H. Prediction of the types of ion channel-targeted conotoxins based on radial basis function network. *Toxicol. In Vitro* **2013**, *27*, 852–856.
- (47) Yao, X.; Panaye, A.; Doucet, J.; Chen, H.; Zhang, R.; Fan, B.; Liu, M.; Hu, Z. Comparative classification study of toxicity mechanisms using support vector machines and radial basis function neural networks. *Anal. Chim. Acta.* **2005**, *535*, 259–273.
- (48) Fernandez-Lozano, C.; Cuinas, R. F.; Seoane, J. A.; Fernandez-Blanco, E.; Dorado, J.; Munteanu, C. R. Classification of signaling proteins based on molecular star graph descriptors using Machine Learning models. *J. Theor. Biol.* **2015**, *384*, 50–58.
- (49) Wright, M. N.; Wager, S.; Probst, P. *Ranger: A Fast Implementation of Random Forests*, R package version 0.12 2020, 2020.
- (50) Bala, R.; Kumar, D. Classification using ANN: A review. *Int. J. Comput. Intell. Res.* **2017**, *13*, 1811–1820.
- (51) Niu, A.-q.; Xie, L.-j.; Wang, H.; Zhu, B.; Wang, S.-q. Prediction of selective estrogen receptor beta agonist using open data and machine learning approach. *Drug Des., Dev. Ther.* **2016**, *10*, No. 2323.
- (52) Xing, G.; Liang, L.; Deng, C.; Hua, Y.; Chen, X.; Yang, Y.; Liu, H.; Lu, T.; Chen, Y.; Zhang, Y. Activity prediction of small molecule inhibitors for anti-rheumatoid arthritis targets based on artificial intelligence. *ACS Comb. Sci.* **2020**, *22*, 873–886.
- (53) Bauer, M. R.; Mackey, M. D. Electrostatic complementarity as a fast and effective tool to optimize binding and selectivity of protein–ligand complexes. *J. Med. Chem.* **2019**, *62*, 3036–3050.
- (54) Walsh, A. A.; Szklarz, G. D.; Scott, E. E. Human cytochrome P450 1A1 structure and utility in understanding drug and xenobiotic metabolism. *J. Biol. Chem.* **2013**, *288*, 12932–12943.
- (55) Chaudhary, A.; Willett, K. L. Inhibition of human cytochrome CYP1 enzymes by flavonoids of St. John's wort. *Toxicology* **2006**, *217*, 194–205.
- (56) Wincent, E.; Le Bihan, F.; Dreij, K. Induction and inhibition of human cytochrome P4501 by oxygenated polycyclic aromatic hydrocarbons. *Toxicol. Res.* **2016**, *5*, 788–799.
- (57) Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **2020**, *21*, 1–13.
- (58) Wang, Y.; Hu, B.; Zhang, Y.; Wang, D.; Luo, Z.; Wang, J.; Zhang, F. Perspective of structural flexibility on selective inhibition towards CYP1B1 over CYP1A1 by α -naphthoflavone analogs. *Phys. Chem. Chem. Phys.* **2021**, *23*, 20230–20246.
- (59) Gupta, S.; Jadaun, A.; Kumar, H.; Raj, U.; Varadwaj, P. K.; Rao, A. Exploration of new drug-like inhibitors for serine/threonine protein phosphatase 5 of Plasmodium falciparum: a docking and simulation study. *J. Biomol. Struct. Dyn.* **2015**, *33*, 2421–2441.
- (60) Chen, P.; Wang, S.; Cao, C.; Ye, W.; Wang, M.; Zhou, C.; Chen, W.; Zhang, X.; Zhang, K.; Zhou, W. α -naphthoflavone-derived cytochrome P450 (CYP) 1B1 degraders specific for sensitizing

CYP1B1-mediated drug resistance to prostate cancer DU145: Structure activity relationship. *Bioorg. Chem.* **2021**, *116*, No. 105295.

(61) Chen, D.; Fan, Q.; Xu, T.; Dong, J.; Cui, J.; Wang, Z.; Wang, J.; Meng, Q.; Li, S. Design, Synthesis and Binding Affinity Evaluation of Cytochrome P450 1B1 Targeted Chelators. *Anti-Cancer Agents Med. Chem.* **2022**, *22*, 261–269.

(62) Shoieb, S. M.; Dakarapu, R.; Falck, J. R.; El-Kadi, A. O. Novel Synthetic Analogues of 19 (S/R)-Hydroxyeicosatetraenoic Acid Exhibit Noncompetitive Inhibitory Effect on the Activity of Cytochrome P450 1A1 and 1B1. *Eur. J. Drug Metab. Pharmacokinet.* **2021**, *46*, 613–624.

(63) Donald, J. E.; Kulp, D. W.; DeGrado, W. F. Salt bridges: geometrically specific, designable interactions. *Proteins* **2011**, *79*, 898–915.

(64) Du, Y.; Li, K.; Wang, X.; Kaushik, A. C.; Junaid, M.; Wei, D. Identification of chlorprothixene as a potential drug that induces apoptosis and autophagic cell death in acute myeloid leukemia cells. *FEBS J.* **2020**, *287*, 1645–1665.

(65) Rochat, B.; Zoete, V.; Grosdidier, A.; von Grünigen, S.; Marull, M.; Michielin, O. In vitro biotransformation of imatinib by the tumor expressed CYP1A1 and CYP1B1. *Biopharm. Drug Dispos.* **2008**, *29*, 103–118.

(66) Beberok, A.; Wrześniok, D.; Rok, J.; Rzepka, Z.; Rzepka, Z.; Respondek, M.; Respondek, M.; Buszman, E. Ciprofloxacin triggers the apoptosis of human triple-negative breast cancer MDA-MB-231 cells via the p53/Bax/Bcl-2 signaling pathway. *Int. J. Oncol.* **2018**, *52*, 1727–1737.