molecular pharmaceutics

Article

Cite This: *Mol. Pharmaceutics* 2018, 15, 4336−4345

Prediction of Human Cytochrome P450 Inhibition Using a Multitask Deep Autoencoder Neural Network

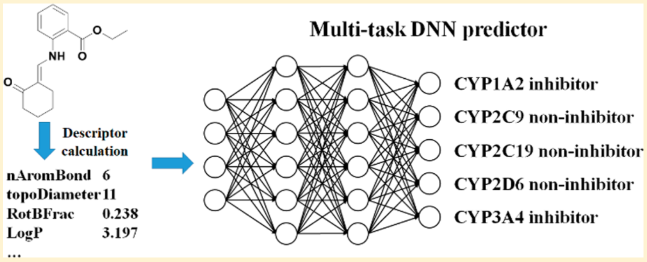
Xiang Li,[†,‡] Youjun Xu,[§] Luhua Lai,[†,‡,§] and Jianfeng Pei*[,§]

[†] Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

[‡] BNLMS, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

[§] Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

Supporting Information



ABSTRACT: Adverse side effects of drug−drug interactions induced by human cytochrome P450 (CYP450) inhibition is an important consideration in drug discovery. It is highly desirable to develop computational models that can predict the inhibitive effect of a compound against a specific CYP450 isoform. In this study, we developed a multitask model for concurrent inhibition prediction of five major CYP450 isoforms, namely, 1A2, 2C9, 2C19, 2D6, and 3A4. The model was built by training a multitask autoencoder deep neural network (DNN) on a large dataset containing more than 13 000 compounds, extracted from the PubChem BioAssay Database. We demonstrate that the multitask model gave better prediction results than that of single-task models, previous reported classifiers, and traditional machine learning methods on an average of five prediction tasks. Our multitask DNN model gave average prediction accuracies of 86.4% for the 10-fold cross-validation and 88.7% for the external test datasets. In addition, we built linear regression models to quantify how the other tasks contributed to the prediction difference of a given task between single-task and multitask models, and we explained under what conditions the multitask model will outperform the single-task model, which suggested how to use multitask DNN models more effectively. We applied sensitivity analysis to extract useful knowledge about CYP450 inhibition, which may shed light on the structural features of these isoforms and give hints about how to avoid side effects during drug development. Our models are freely available at http://repharma.pku.edu.cn/deepcyp/home.php or http://www.pkumdl.cn/deepcyp/home.php.

KEYWORDS: *multitask deep neural network, quantitative structure−activity relationship, cytochrome P450, drug−drug interaction, sensitivity analysis*

INTRODUCTION

Human cytochrome P450 (CYP450), a hemoprotein superfamily containing 57 isoforms, is responsible for catalyzing the metabolism of many endogenous substrates and exogenous compounds. They can oxidize, peroxidize, and reduce small molecules of a variety of chemicals.[1] Inhibition of CYP450 isoforms may cause drug−drug or food−drug interactions, and they result in severe side effects. In the last few decades, several commercial drugs were withdrawn from the market due to adverse inhibition to CYP450.[2,3]

Five CYP450 isoforms (1A2, 2C9, 2C19, 2D6, and 3A4) are in charge of nearly 90% of metabolic reactions.[4] The crystal structures of these five isozymes[5−10] indicate the ability of CYP450 isozymes to change their conformations to accommodate different small molecules.[11] The flexibility of CYP450 conformations makes it difficult to predict inhibitors of CYP450 using structure-based methods, such as molecular docking and pharmacophore mapping.[12] Alternative approaches, like quantitative structure−activity relationship (QSAR) modeling, especially those using machine learning methods, have been widely used to predict CYP450 inhibitors.

In 2009, Auld et al. used an *in vitro* bioluminescent assay to measure EC_{50} values of 17 143 compounds against these five CYP isozymes,[13] which provided a large dataset for developing QSAR models. In 2011, Cheng et al. used an artificial neural network to combine the prediction results from five models developed using k-nearest neighbors (k-NN), support vector machine (SVM), C4.5 decision tree (C4.5 DT), and naïve Bayes (NB) classifiers.[14] The predictive accuracies of the test sets were 73.1%, 86.7%, 81.0%, 87.8%, and 76.0% for 1A2, 2C9, 2C19, 2D6, and 3A4, respectively. Sun et al. developed a SVM classification model for these five CYP450 isozymes, and the

Special Issue: Deep Learning for Drug Discovery and Biomarker Development

Received: January 31, 2018
Revised: May 3, 2018
Accepted: May 18, 2018
Published: May 18, 2018

DOI: 10.1021/acs.molpharmaceut.8b00110

areas under the receiver operating characteristic curves (AUC) for the test sets was 0.93, 0.89, 0.89, 0.85, and 0.87 for 1A2, 2C9, 2C19, 2D6, and 3A4, respectively.[12] In 2015, Su et al. utilized a rule-based C5.0 DT algorithm to construct prediction models for these five isozymes.[15] When the training sets and test sets were split randomly, the predictive accuracies were 79.5%, 76.8%, 86.0%, 89.8%, and 73.3% for 1A2, 2C9, 2C19, 2D6, and 3A4, respectively.

All previous work used five independent models for five isozymes inhibitor prediction with different prediction accuracies. Given the similarity between different CYP450 isoforms and the similarity between inhibitors, one can suppose a multitask model that can predict inhibitors for five isozymes at the same times to give better predictive power. Deep learning, as a novel technique, are thought to be especially useful for this kind of problem.

The deep learning technique, which has been widely used in many areas, such as computer vision and natural language processing, has also shown great performance in QSAR tasks in the past several years and been applied in drug discovery researches. In 2012, Merck hosted a Kaggle competition involving several QSAR problems. A multitask DNN model developed by Dahl et al. won this competition, which achieved a relative accuracy improvement of nearly 15% over Merck's internal baseline.[16] A multitask DNN model developed by Mayr et al. also outperformed other methods in the 12 toxicity prediction tasks of the Tox21 data challenge.[17] Besides chemical data, deep learning technique has also been applied to analyze genomic data, transcriptome data and proteomic data, to deal with problems related to drug discovery, such as virtual screening,[18−22] ADMET properties,[23−27] protein structure,[28−33] function,[34,35] and interaction.[36]

Ramsundar et al. investigated several aspects of the multitask DNN and showed that multitask networks can achieve higher predictive accuracy than single-task methods. Adding tasks and data can both help improve model performance.[37] Furthermore, training several tasks in one model is computational efficient. In the present study, we developed a multitask autoencoder-based (AE-based) DNN model for CYP450 isozyme inhibitor prediction. Single-task AE-based DNN, multitask DNN, logistic regression, SVM, C4.5 DT, and k-NN models were also constructed for comparison. Our multitask AE-based DNN model outperformed all other models and previously reported classifiers. In addition, we qualitatively explained the difference between single-task and multitask models using linear regression analysis. We also identified features of compounds for their differences in CYP450 isoform inhibition via sensitivity analysis, and some of our discoveries are in agreement with known knowledge of CYP450 isoform inhibitors.

■ MATERIALS AND METHODS

Dataset Preparation. All six datasets were downloaded from the PubChem BioAssay Database. The first dataset (AID: 1851) contained 17 143 compounds. In this dataset, the experimental data were assayed using CYP1A2, 2C9, 2C19, 2D6, and 3A4 to measure the dealkylation of various proluciferin substrates to luciferin. After addition of a luciferase detection reagent, the luciferin was measured by luminescence. Compounds were added at several concentrations to test how much the luminescence decreased and thereby determined the potency values of these compounds against five isoforms.[13] The other five assays used the same protocol, but each examined one isozyme (AID: 410 for 1A2, AID: 883 for 2C9, AID: 899 for 2C19, AID: 891 for 2D6, and AID: 884 for 3A4). Each dataset contained compound activity score, potency, curve description, fitted log EC_{50}, fitted R-square, and activity at 0.00368, 0.018, 0.091, 0.457, 2.286, 11.43, and 57.14 μM for each isozyme. Figure 1 showed the correlation among different assays of dataset AID 1851. The activities were correlated given these isoforms are all from the CYP450 superfamily and share sequence identity.

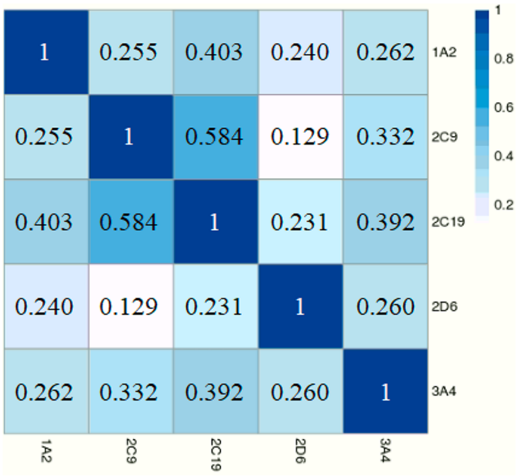


Figure 1. Correlation among different assays of dataset AID 1851. Each assay was represented by a 17 143-dimensional vector, whose element equals to 1, 0, or 0.5 if corresponding compound is an inhibitor, noninhibitor, or inconclusive for this assay. The correlation of two assays was calculated by the Pearson correlation coefficient of two vectors.

We used dataset AID 1851 as training set, AID 410, AID 883, AID 899, AID 891, and AID 884 as test sets. Compounds contained in both training set and test sets were excluded from the test sets. For each dataset, entries containing mixtures, noncovalent complexes, inorganic compounds, and atoms other than C, H, O, N, P, S, F, Cl, Br, and I were excluded using the KNIME software.[38] OpenBabel software was used to exclude duplicate compounds based on the inchi fingerprint consistency.[39] The remaining compounds were processed using the ChemAxon Standardizer to add explicit hydrogens and aromatize for descriptor calculation.[40]

Descriptor Calculation. PaDEL-1D&2D descriptors and PubChem fingerprints of all compounds were calculated using PaDEL-Descriptor software.[41] PaDEL-1D&2D contained 1444 descriptors, including atom type electrotopological state (E-state) descriptors, Crippen's logP, and molecular linear free energy. PubChem fingerprints had 881 binary fingerprints counting chemical substructures, such as element counts, ring counts, simple atom pairs, atom nearest neighbors, and SMARTS patterns. Descriptors that were not calculable for all compounds, or took same value for all compounds were removed. In total, 1253 PaDEL-1D&2D descriptors and 688 PubChem fingerprints remained for model training. The values of each descriptor were normalized to range between 0 and 1 by subtracting the minimum value of the descriptor and dividing by the range.

Labeling of Inhibitors and Noninhibitors. In the previous work related to these datasets, three criteria were used to classify inhibitors and noninhibitors (Table 1). They

Table 1. Criteria for Classification of Inhibitors and Noninhibitors

criteria	inhibitor	noninhibitor
AC50	≤10 μM	>57 μM
score	≥40	0
curve class	−1.1, −1.2, −2.1	4

were fitted EC_{50} value,[14] PubChem activity score,[14] and concentration−response curves (see in Supplementary Table S1).[12,42] The EC_{50} was fitted from concentration−response curve, so it may be significantly influenced by outliers. Therefore, the other two criteria were used together, which are highly consistent (see in Supplementary Table S2). Only compounds identified as inhibitors or noninhibitors by these two criteria at the same time were included in the datasets, and the others were assigned as inconclusive and removed. The information on processed datasets is shown in Table 2. The overall workflow of dataset processing is shown in Figure 2.

Table 2. Detailed Information of the Datasets

	AID	isoform	total	number of inhibitors	number of noninhibitors
training set	1851	1A2	9488	3559	5929
		2C9	9385	2552	6833
		2C19	10058	4450	5608
		2D6	10103	1351	8752
		3A4	9055	3070	5985
test set	410	1A2	584	107	477
	883	2C9	665	69	596
	899	2C19	711	142	569
	891	2D6	748	77	671
	884	3A4	2382	537	1845

Modeling. Autoencoder-based multitask and single-task DNN models were developed. The architectures of the networks are shown in Figure 3. The only difference between the multitask model and the single-task model was the number of output layers. Both network models contained an input layer, three fully connected encoded layers, one fully connected hidden layer, and two-node softmax output layers. Models were first initialized by layer-wise unsupervised pretraining and then fine-tuned using a supervised criterion.[43] In the pretraining step, a three-layer network was constructed initially. The first two layers of the network were the input and the first encoded layer of original network, and the training output was the same as the input. In other words, the pretraining step was to train an identify function, and the encoder layer would learn an overcomplete (hidden code has dimension greater than the input) or undercomplete (hidden code has dimension less than the input) representation of input data. After training the first autoencoder, the weights between input layer and the first encoder layer were assigned to the original network. Then the first and second encoder layers were trained using the raw input data as the target. Its input values were the output of the first encoder layer of the previous three-layer network. The pretraining was repeated until all encoder layer learned a representation. In the fine-tuning step, the entire classifier was trained with label information.

The hyperparameter setting is shown in Table 3. The greedy algorithm was applied to determine the optimizer, L2 normalization term, dropout rate, and batch size. These hyperparameters were determined one by one, according to this order, by finding the values yielded best accuracy on the validation set while other hyperparameters were fixed.

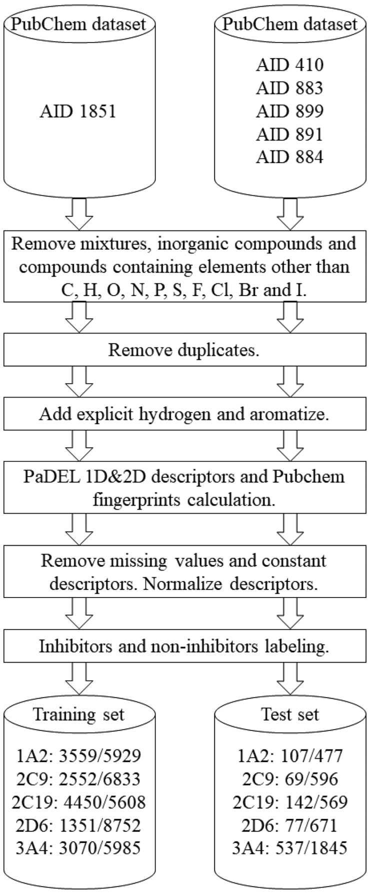

PubChem dataset
AID 1851
PubChem dataset
AID 410
AID 883
AID 899
AID 891
AID 884
Remove mixtures, inorganic compounds and compounds containing elements other than C, H, O, N, P, S, F, Cl, Br and I.
Remove duplicates.
Add explicit hydrogen and aromatize.
PaDEL 1D&2D descriptors and Pubchem fingerprints calculation.
Remove missing values and constant descriptors. Normalize descriptors.
Inhibitors and non-inhibitors labeling.
Training set
1A2: 3559/5929
2C9: 2552/6833
2C19: 4450/5608
2D6: 1351/8752
3A4: 3070/5985
Test set
1A2: 107/477
2C9: 69/596
2C19: 142/569
2D6: 77/671
3A4: 537/1845

Figure 2. Overall workflow of dataset processing.

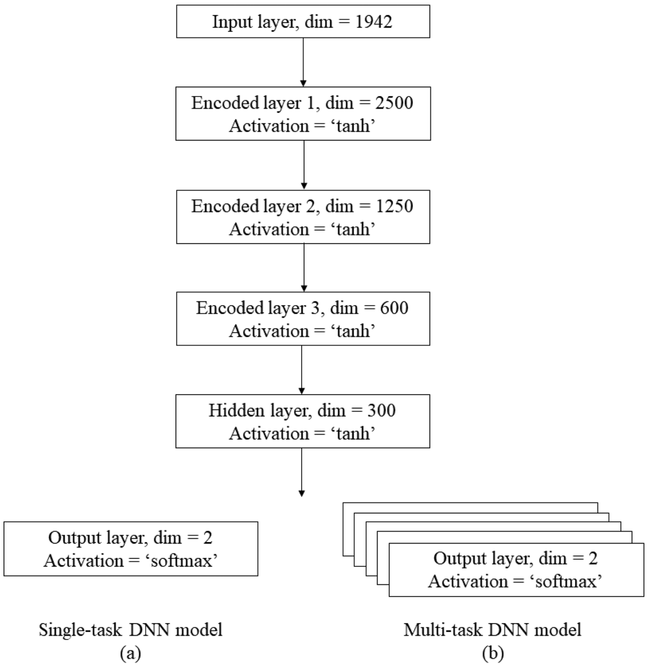


Figure 3. Architecture of DNN models. (a) The architecture of single-task DNN models. (b) The architecture of multitask DNN models.

Table 3. Hyperparameter Setting

hyperparameter	setting
optimizer	Adam(lr = 1e-4) (sgd(lr = 0.1, 0.01, 0.001), rmsprop(lr = 1 e-3, 5e-4, 1e-4), Adam(lr = 1e-3, 5e-4))
L2 normalization term	2e-6 (1e-6, 3e-6, 5e-6, 1e-5)[a]
dropout rate	0.3 (0.2, 0.4, 0.5)
batch size	1024 (128, 256, 512)
initializer	RandomUniform(−0.05, 0.05)
loss	mse for AE, binary crossentropy for classifier
monitor	mae for AE, acc for classifier

[a] The numbers written in the brackets are other values tried for this hyperparameter.

Cross entropy was usually used as the loss function for the classification task. In the ith single-task binary classification problem, it took form:

$$L_i = -\sum_{k=1}^{n} m_{ik}(r_i y_{ik}(\log a_{ik}) + (1 - y_{ik})\log(1 - a_{ik})) \tag{1}$$

where y_{ik} represents the k^{th} compound label in the i^{th} task. y_{ik} = 1 or 0, means the k^{th} compound was the inhibitor or noninhibitor of the i^{th} isozyme, respectively. $a_{ik}=P(y_{ik} = 1|x_k)$ was the probability of k^{th} compound is the inhibitor of i^{th} isozyme calculated by model. $m_{ik} = 0$ indicates that the k^{th} compound were excluded in the i^{th} task, otherwise it equaled to 1. n is the total number of the compounds. Given the imbalance of the training data, especially for the CYP2D6 dataset of which the number of noninhibitors is 6 times more than the number of inhibitors, we weighted a coefficient r_i before the loss for inhibitors, r_i = #noninhibitors/#inhibitors in the i^{th} task, such that the ratio of inhibitors to noninhibitor would be balanced at 1:1.

The loss function of the multitask model is the sum of loss of all single tasks:

$$L = -\sum_{i=1}^{5}\sum_{k=1}^{n} m_{ik}(r_i y_{ik}(\log a_{ik}) + (1 - y_{ik})\log(1 - a_{ik})) \tag{2}$$

All models were evaluated by five indexes. The accuracy (ACC), sensitivity (SE), and specificity (SP) were defined as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{3}$$

$$\text{SE} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{SP} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{5}$$

where TP is the number of inhibitors correctly classified, TN is the number of noninhibitors correctly classified, FN is the number of inhibitor wrongly classified, and FP is the number of noninhibitor wrongly classified. These three indexes represented the proportion of correctly classified samples, inhibitors, and noninhibitors, respectively. We also calculated the Matthews correlation coefficient (MCC) to evaluate the model without the influence of the imbalanced dataset. MCC is defined as

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{FP} + \text{TN})(\text{FP} + \text{TP})(\text{FN} + \text{TN})(\text{FN} + \text{TP})}} \tag{6}$$

The area under the receiver operating characteristic curve (AUC) was also used to evaluate classifiers. Each model was evaluated by 10-fold cross-validation.

Model Interpretation. We used the sensitivity analysis (SA) method, developed by Cortez et al., to interpret our models and find the sensitive variables that are important for model prediction.[44] Sensitive variables were defined as the variables whose change would significantly influence the output of the model. The first step of SA is to generate a baseline vector $b = (b_1,b_2,\cdots,b_p)$, where p is the number of variables. Here we set b_i as the median of the i^{th} variable in the training set. Then, an arithmetic sequence $l = \{l_1,\cdots l_L\}$ ranging from 0 to 1 was defined, where $l_1 = 0$ and $l_L = 1$, L is the length of the sequence.

To calculate the sensitivity of the j^{th} variable, we built L input examples by changing the j^{th} element of b to each term of sequence l (each variable was normalized to the [0,1] range as previously mentioned). These L examples were input to the DNN model, and denoted their probability as an inhibitor predicted by the model as $y_j=(y_{1j},\cdots,y_{Lj})$. The sensitivity can be measured in many ways, such as the range of y_j, the variance of y_j or the average square deviation from the median. In fact, for our problem, the ranking of the most sensitive variables remained the same no matter what metric of sensitivity was used.

■ RESULTS AND DISCUSSION

Prediction Results. The performances of single-task and multitask DNN are shown in Tables 4 and 5, respectively. The multitask model outperformed the single-task model in the training processes. The 1A2 model performed best among all single-task models for predicting the test sets. The multitask model performed slightly worse than the single-task model in predicting 1A2 isoform, but it outperforms all other single-task models in predicting the inhibitors of 2C9, 2C19, 2D6, and 3A4 isoforms.

Both multitask and single-task models were compared with previously reported CYP450 inhibitor classifiers developed by Cheng et al.,[15] Sun et al.[12] and Su et al.[16] All these classifiers

Table 4. Performance on the Training Set

	single-task models					multitask model				
isoform	ACC	SE	SP	MCC	AUC	ACC	SE	SP	MCC	AUC
1A2	0.856	0.809	0.884	0.693	0.915	**0.892**	**0.871**	**0.905**	**0.772**	**0.956**
2C9	0.849	0.794	0.869	0.638	0.912	**0.857**	**0.808**	**0.876**	**0.658**	**0.922**
2C19	0.827	0.830	0.824	0.651	0.900	**0.838**	**0.846**	**0.832**	**0.674**	**0.908**
2D6	0.899	0.587	0.947	0.550	0.880	0.885	**0.665**	0.919	0.544	**0.886**
3A4	0.845	0.803	0.866	0.660	0.920	**0.850**	**0.823**	0.864	**0.673**	**0.926**

Table 5. Performance on the Test Sets

	single-task models					multitask model				
isoform	ACC	SE	SP	MCC	AUC	ACC	SE	SP	MCC	AUC
1A2	**0.971**	**0.944**	**0.977**	**0.905**	**0.983**	0.968	0.925	**0.977**	0.893	0.982
2C9	0.842	0.493	0.883	0.315	0.771	**0.860**	**0.522**	**0.899**	**0.365**	**0.799**
2C19	0.793	0.570	0.849	0.395	0.810	**0.809**	**0.599**	**0.861**	**0.436**	**0.832**
2D6	0.893	0.481	0.940	0.421	0.829	**0.893**	**0.533**	0.934	**0.447**	**0.878**
3A4	0.884	0.652	0.951	0.649	0.921	**0.896**	**0.708**	0.951	**0.692**	**0.929**

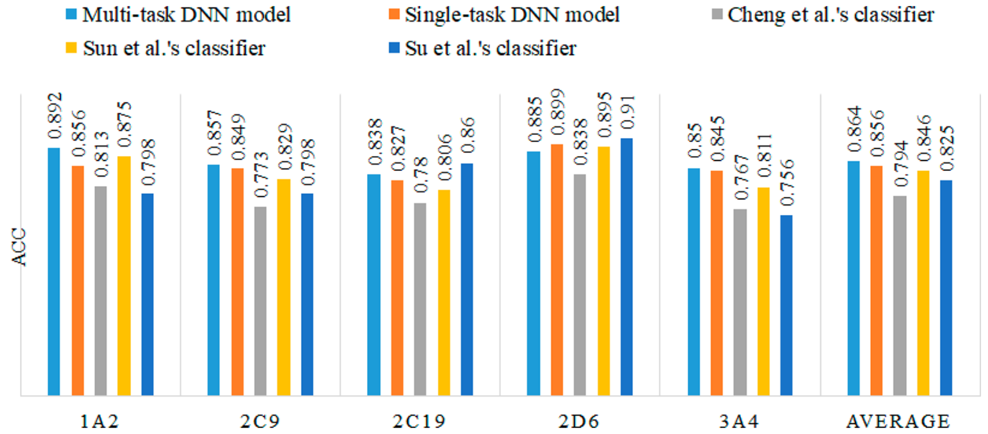


Figure 4. Cross-validation accuracies of different classifiers on AID 1851 dataset.

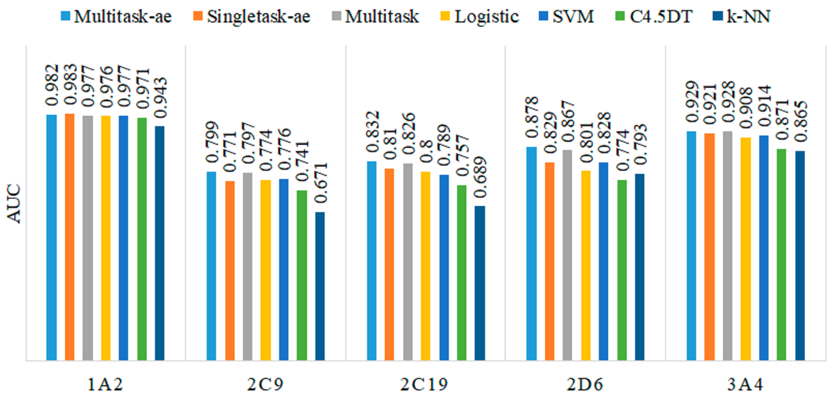


Figure 5. AUC values of different machine learning models on five test sets. Multitask-ae means multitask model with unsupervised pretraining. Multitask means multitask model without unsupervised pretraining.

were trained on AID 1851 dataset. Figure 4 showed the generalization ability of different classifiers, evaluated by either cross validation or testing on a subset of AID 1851 dataset. Multitask DNN model performed best on 1A2, 2C9, and 3A4 isoforms, first runner up on 2C19 isoforms, and achieved the best average ACC across all tasks.

Ma et al. suggested that unsupervised pretraining was slightly harmful to their chemical activity prediction problems.[17] To verify the effect of autoencoder-based unsupervised pretraining, another multitask model was trained without pretraining. Logistic, SVM, C4.5 DT and k-NN (k = 5) models were also built based on the same datasets. The AUC values of all models on five test sets are shown in Figure 5. The multitask DNN model with pretraining showed better predictive power than the model without pretraining. Moreover, all DNN models showed better predictive power than traditional machine learning models, suggesting that DNN has an advantage in solving this cheminformatics problem.

Quantification of the Contributions of Multi Tasks. In this study, multitask model outperformed the single-task models, as with many others, but not all QSAR tasks. Xu et al. made an explanation of why multitask DNNs makes a difference in predictive performance in QSAR tasks.[45] They found that the multitask DNNs will perform better than single-task DNNs if the molecular activities of these tasks are positively or negatively correlated. On the other hand, the multitask DNNs perform worse than single-tasks DNNs if the activities of these tasks are uncorrelated. However, their explanation was just qualitative. Here we built linear regression models for each task to quantify the contributions to the prediction of all single tasks, and we showed how these models can be used to explain the prediction difference between single-task and multitask models.

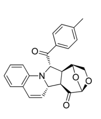
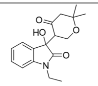
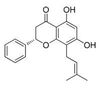
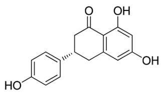
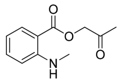
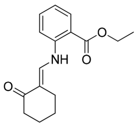
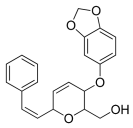
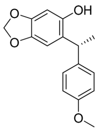
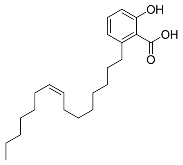
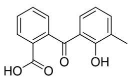
For each test set compound, we supposed that its classification was mainly dependent on the most similar inhibitor and noninhibitor in the training set. Therefore, we measured the similarity between each test compound and the

DOI: 10.1021/acs.molpharmaceut.8b00110

Table 6. Coefficients and R^2 of Five Linear Regression Models Fitted Δlogit(p) by Maximum Similarity to Inhibitors and Non-Inhibitors of Five Isoforms in Training Sets

	isoform				
parameter	1A2	2C9	2C19	2D6	3A4
intercept	4.249	5.732	3.727	3.614	1.875
max similarity to 1A2 inhibitors	−10.982	0.196	1.106	−13.096	0.178
max similarity to 2C9 inhibitors	0.443	−18.980	3.238	9.275	−0.778
max similarity to 2C19 inhibitors	3.626	6.023	−12.229	7.753	1.117
max similarity to 2D6 inhibitors	1.213	5.248	2.549	−19.057	4.345
max similarity to 3A4 inhibitors	2.612	4.682	0.950	13.671	−8.547
max similarity to 1A2 noninhibitors	8.521	−1.270	−0.107	−0.503	−2.089
max similarity to 2C9 noninhibitors	1.844	10.612	−1.942	17.014	−4.073
max similarity to 2C19 noninhibitors	−4.364	−14.777	5.179	−20.637	2.540
max similarity to 2D6 noninhibitors	−4.461	−2.900	1.002	−20.067	−5.481
max similarity to 3A4 noninhibitors	−2.064	5.302	−2.353	20.631	12.811
R^2	0.6487	0.3584	0.3427	0.3679	0.2947

Table 7. Some Examples of Test Set Compounds and Their Most Similar Compounds in Training Set

Test set compound	Most similar compound in training set	Tanimoto similarity	p^i	p^m	Δlogit(p)
 2C9 inhibitor	 1A2 non-inhibitor 2D6 non-inhibitor 3A4 non-inhibitor	0.7166	0.7980	0.1282	-3.291
 2C9 inhibitor	 1A2 inhibitor 2C9 non-inhibitor 2D6 non-inhibitor 3A4 inhibitor	0.9632	0.0817	0.5320	2.547
 2C19 non-inhibitor	 1A2 inhibitor 2C19 inhibitor 3A4 non-inhibitor	0.9353	0.7974	0.2380	-2.534
 2D6 inhibitor	 2C9 inhibitor 2C19 inhibitor	0.8273	0.1534	0.6006	2.116
 3A4 non-inhibitor	 1A2 non-inhibitor 2C19 non-inhibitor 2D6 non-inhibitor	0.8537	0.8457	0.1889	-3.159

most similar inhibitor and noninhibitor of each isoform in the training set, and we obtained a total of 10 variables as explanatory variables. The similarity between two compounds was defined by the Tanimoto coefficient based on PubChem fingerprints. The response variable was

$$\Delta\text{logit}(p_i) = \text{logit}(p_i^m) - \text{logit}(p_i^s) \quad (7)$$

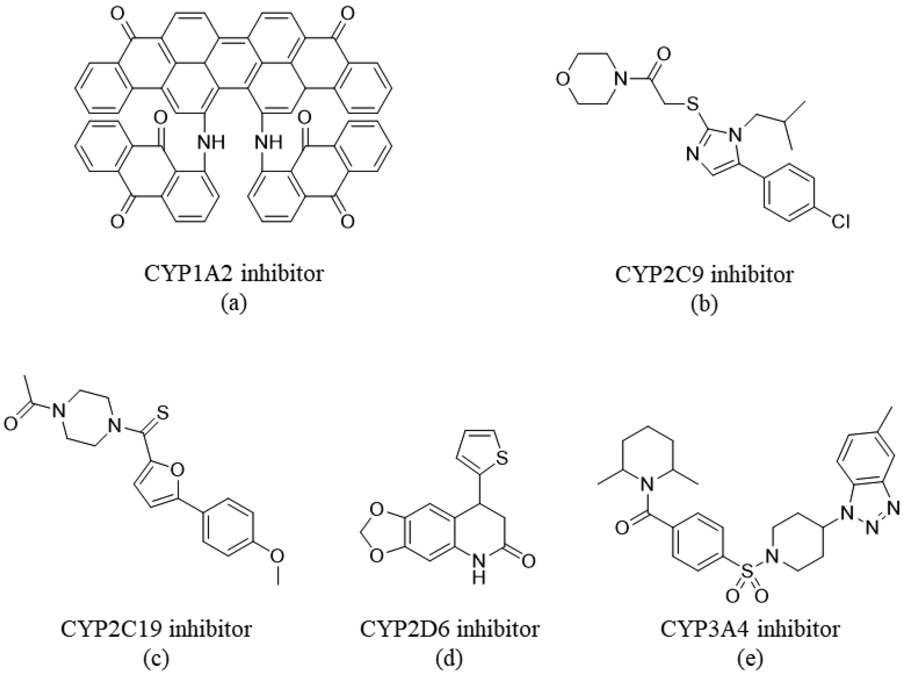
where p_i^m is the probability the i^{th} compound of the test set as an inhibitor predicted by multitask model, and p_i^s was the probability of the i^{th} compound of the test set as an inhibitor predicted by single-task model. Logit is the logistic function, which takes the form

$$\text{logit}(x) = \log\frac{x}{1-x} \quad (8)$$

Thus, Δlogit(p) can be a measure of the difference between multitask and single-task models. The regression models were built for five test sets, respectively, and for compounds for which the prediction significantly changed from single-task model to multitask model results ($|\Delta\text{logit}(p)| \geq 2$).

Table 6 shows the coefficients and R^2 of five regression models. For each task, Δlogit(p_i) was negatively correlated with the maximum similarity to inhibitor of its training set and generally positively correlated with the maximum similarity to the inhibitor of other tasks. The correlations were opposite for noninhibitors. These results met our expectations. The more similar a test set compound is to a corresponding training set inhibitor, the higher probability it would be predicted as an inhibitor by the single-task model. As a result, the probability of being predicted as an inhibitor by the multitask model increased less. In contrast, the similarity of a test set compound to the training set inhibitors of other tasks would increase its probability to become an inhibitor predicted by the multitask model. The effect of training set noninhibitors was just the opposite. Here, the coefficients can be used to quantify the contributions of all tasks to a specific prediction. Based on the R^2, more than 30% of the deviance of Δlogit(p) can be explained by the similarity to training set inhibitors and noninhibitors.

These regression models could to some extent explain the prediction difference between single-task and multitask models. Some examples of test set compounds whose predictions significantly changed from single-task model to multitask model and their most similar compounds in the training set are shown in Table 7. In these examples, the prediction differences are associated with the labels of their most similar compounds in the training set. If the most similar compound is an inhibitor of another isoform, then the probability as inhibitor predicted by multitask model would increase, and vice versa. Some of the


CYP1A2 inhibitor
(a)

CYP2C9 inhibitor
(b)

CYP2C19 inhibitor
(c)

CYP2D6 inhibitor
(d)

CYP3A4 inhibitor
(e)

Figure 6. Some typical inhibitors of CYP450 isozymes. (a) A CYP1A2 inhibitor. (b) A CYP2C9 inhibitor. (c) A CYP2C19 inhibitor. (d) A CYP2D6 inhibitor. (e) A CYP3A4 inhibitor.

examples are noteworthy where the effects of the most similar compound in other isoforms are inconsistent, such as the second and the third examples. In these two examples, their most similar compounds were labeled in their own tasks, and the prediction results of single-task models are the same as their most similar compounds. As for the predictions of multitask model, the influences from other tasks can be judged from Table 6. In the second example, the most similar compound is also the inhibitor of 1A2 and 3A4 and noninhibitor of 2D6. For the 2C9 model, the coefficients of maximum similarity to 1A2 inhibitor, 3A4 inhibitor, and 2D6 noninhibitor were 0.196, 4.682, and −2.900, respectively; therefore, the multitask model tends to predict higher probability as inhibitor. In likely manner, for 2C19 model, the coefficients of maximum similarity to 1A2 inhibitor and 3A4 noninhibitor were 1.106 and −2.353, respectively, and the prediction of multitask model changed toward the label of task 3A4. This rule is correct for 81.8% of 1A2 test set compounds, 60.0% of 2C9 test set compounds, 76.0% of 2C19 test set compounds, 53.5% of 2D6 test set compounds, and 70.5% of 3A4 test set compounds. The correctness of this rule is associated with the Tanimoto similarity between the test set compound and its most similar compound in the training set.

On the basis of the above analysis, we might draw the following conclusions. In a single-task model, the prediction result of the test set compound is likely to be the same as its most similar compound in the training set. In a multitask model, the prediction will change on the basis of the sum of the contributions of other tasks. The contributions of all tasks to a given task can be quantified by using a linear regression model. Whether the multitask model outperforms the single-task models on test set depends on whether the sum of the contributions of other tasks changes the prediction in the right way, in other words, closer to the truly classification of test set compounds. These conclusions can guide how to improve the predictive power using multitask model.

Sensitive Variable Analysis. By using the SA method, we could quantify the influence of each descriptor to the probability a molecule is predicted as a CYP450 isozyme inhibitor. For each sensitive variable, we distinguished it as positively correlated if the probability as an inhibitor predicted by the model monotonous increased with its increase, or negatively correlated if the probability monotonous decreased with its increase. Both positively correlated and negatively correlated variables were sorted according to the absolute values of their sensitivity, and the most sensitive variables were analyzed in detail. Here we demonstrated that some of our discoveries are in agreement with existing knowledge of CYP450 isozyme inhibitors.

The negatively correlated variables for both single-task model and multitask model are almost the same for CYP1A2, including the number of basic groups (nBase), the number of acidic groups (nAcid) and the E-state of quaternary nitrogen (ssssNp). These descriptors are associated with water solubility. The positively correlated variables included Pubchem FP385 [C(:C)(:C)(:C), ':' means aromatic bond], number of 10-membered fused rings and the E-state of ::C:. All these results are consistent with what we have known about CYP1A2. Sansen's study on the structure of CYP1A2 revealed that CYP1A2 contained a compact and closed active site cavity that is suitable for accommodating lipophilic planar polyaromatic or polyheteroaromatic small molecules.[7,46] A typical example was shown in Figure 6a, where each of these positive correlated variables achieved its maximum value.

CYP2C9 and CYP2C19 belong to the same subfamily, and they share 91% amino acid sequence identity.[46] Their highly identity is reflected in the high correlation between these two classification tasks (Figure 1), and the sensitive variables of these two tasks are quite similar. The negatively correlated variables of CYP2C9 and CYP2C19 are similar to those of CYP1A2, which also included nBase, nAcid, and E-state of ssssNp, indicating that CYP2C9 and CYP2C19 also favor

DOI: 10.1021/acs.molpharmaceut.8b00110

hydrophobic ligands. The positively correlated variables of both isoforms included the E-state of sulfur (forms two single bonds or a double bond) and number failures of the Lipinski's rule of five. The inhibitors of CYP2C9 and CYP2C19 are more inclined to have sulfur atoms; that is, 18.69% of CYP2C9 inhibitors and 17.16% of CYP2C19 inhibitors contain at least one sulfur atom to form two single bonds, whereas only 7.83% of CYP2C9 noninhibitors and 7.17% of CYP2C19 noninhibitors contain at least one sulfur atom to form two single bonds. In many of the inhibitors containing sulfur atoms, the sulfur atoms or thiocarbonyl groups directly bond or conjugate to an aromatic heterocycle (Figure 6b,c). Another interesting feature is that, the inhibitors of CYP2C9 and CYP2C19 are more inclined to break one of Lipinski's rule of five (Table 8).

Table 8. Proposition of CYP450 Inhibitors and Noninhibitors Break Lipinski's Rule of Five

	number of broken Lipinski's rule of five					
isoform	0	1	2	3	4	5
1A2 inhibitor	70.7%	29.0%	0.3%	0.0%	0.0%	0.0%
1A2 noninhibitor	71.7%	22.5%	5.2%	0.5%	0.0%	0.0%
2C9 inhibitor	52.1%	**42.0%**	5.9%	0.0%	0.0%	0.0%
2C9 noninhibitor	78.1%	18.7%	2.6%	0.5%	0.0%	0.0%
2C19 inhibitor	59.6%	**36.7%**	3.6%	0.0%	0.0%	0.0%
2C19 noninhibitor	77.7%	18.6%	3.1%	0.5%	0.0%	0.0%
2D6 inhibitor	61.3%	34.9%	3.6%	0.2%	0.0%	0.0%
2D6 noninhibitor	72.5%	23.8%	3.4%	0.3%	0.0%	0.0%
3A4 inhibitor	58.1%	34.0%	7.7%	0.2%	0.0%	0.0%
3A4 noninhibitor	77.8%	19.6%	2.1%	0.5%	0.0%	0.0%

Nearly 70% of these compounds have lipid−water partition coefficient (LogP) ≥ 5, which means CYP2C9 and CYP2C19 have more affinity toward lipophilic ligands than other isoforms.

The CYP2D6 dataset is the most imbalanced one. Only 13.4% of compounds in the dataset are inhibitors. There was no meaningful information accessible about the negatively correlated variables because the noninhibitors of CYP2D6 are highly diverse. The positively correlated variables of CYP2D6 inhibitors include Pubchem FP367 (C(~H)(~O)(~O)), Pubchem FP661 (O−C−O−C−C) and the number of 9-membered fused rings, which corresponding to the structure of the piperonyl group (Figure 6d). 7.4% of CYP2D6 inhibitors contained this substructure, while only 2.3% of CYP2D6 noninhibitors contain this substructure.

As with CYP1A2, 2C9, and 2C19, negatively correlated variables of CYP3A4 include nBase, nAcid, and E-state of ssssNp. A significantly positively correlated variable is the number of saturated or aromatic carbon-only six-membered ring. More than two-thirds (67.8%) of CYP3A4 inhibitors contain more than 2 carbon-only six-membered ring, whereas this proposition is 34.3% in noninhibitors. A typical example is shown in Figure 6e. A previous study on the structure of CYP3A4 indicates that CYP3A4 has a highly flexible cavity, with an estimated volume ranging from 1173 Å^3 to 2682 Å^3.[10] In addition, CYP3A4 has the ability to bind and metabolize multiple substrates simultaneously.[47] Our discovery indicates that the existence of two or more six-membered rings may be favorable for the interaction with CYP3A4 and may shed light on the structural features of this isoform.

■ CONCLUSIONS

In this study, a multitask deep autoencoder neural network model was built to predict the inhibitors of five major isozymes of CYP450 (1A2, 2C9, 2C19, 2D6, and 3A4) based on a large dataset containing more than 13 000 compounds. The prediction accuracies of the training set validated by 10-CV were 89.2%, 85.7%, 83.8%, 88.5%, and 85.0% for 1A2, 2C9, 2C19, 2D6, and 3A4, respectively. The prediction accuracies of the external test sets was 96.8%, 86.0%, 80.9%, 89.3%, and 89.6% for 1A2, 2C9, 2C19, 2D6, and 3A4, respectively. The prediction accuracies of the multitask model is on average 0.8% and 0.5% higher than that of single-task model on training set and test set, respectively. The cross-validation accuracy of our multitask DNN model is 1.8% to 7% higher than previously reported classifiers. The predictive power of multitask model also outperformed traditional machine learning methods. Moreover, we used linear regression models to explain how all the other tasks influence the prediction of a given task in the multitask model, which can help to use multitask model more effectively. Sensitivity analysis was applied to find the sensitive variables that significantly influenced the prediction, which can help to acquire useful knowledge about CYP450 inhibitors, and give hints on how to prevent side effects during drug development.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.molpharmaceut.8b00110.

> Table S1: Curve classes and corresponding description; Table S2: The number of inhibitors and noninhibitors, under different criteria combinations; and Table S3−Table S12: Top 10 sensitive variables of all isoforms in multitask and single-task DNN models (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail for J.P.: jfpei@pku.edu.cn. Fax: (+86)10-62759595.

ORCID

Luhua Lai: 0000-0002-8343-7587
Jianfeng Pei: 0000-0002-8482-1185

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been supported in part by the Ministry of Science and Technology of China (grant numbers: 2016YFA05023032, 2015CB910302) and the National Natural Science Foundation of China (grant numbers: 21673010, 21633001).

■ ABBREVIATIONS

CYP450, cytochromes P450; ACC, accuracy; QSAR, quantitative structure−activity relationship; k-NN, k-nearest neighbors; SVM, support vector machine; DT, decision tree; NB, naïve Bayes; AUC, range of areas under the receiver operating characteristic curve; DNN, deep neural network; E-state, electrotopological state; AE, autoencoder; SE, sensitivity; SP, specificity; MCC, Matthews correlation coefficient; 10-CV, 10-fold cross-validation.

■ REFERENCES

(1) Nebert, D. W.; Russell, D. W. Clinical importance of the cytochromes P450. *Lancet* **2002**, *360* (9340), 1155−1162.

(2) Lasser, K. E.; Allen, P. D.; Woolhandler, S. J.; Himmelstein, D. U.; Wolfe, S. M.; Bor, D. H. Timing of new black box warnings and withdrawals for prescription medications. *JAMA* **2002**, *287* (17), 2215−20.

(3) Wienkers, L. C.; Heath, T. G. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat. Rev. Drug Discovery* **2005**, *4* (10), 825−33.

(4) Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S.; Goosen, T. C.; Peterkin, V.; Koup, J. R.; Ball, S. E. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.* **2004**, *32* (11), 1201−1208.

(5) Reynald, R. L.; Sansen, S.; Stout, C. D.; Johnson, E. F. Structural characterization of human cytochrome P450 2C19: active site differences between P450s 2C8, 2C9, and 2C19. *J. Biol. Chem.* **2012**, *287* (53), 44581−91.

(6) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **2006**, *281* (11), 7614−22.

(7) Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J. Biol. Chem.* **2007**, *282* (19), 14348−55.

(8) Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305* (5684), 683−686.

(9) Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D.; Jhoti, H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424* (6947), 464−8.

(10) Gay, S. C.; Roberts, A. G.; Halpert, J. R. Structural features of cytochromes P450 and ligands that affect drug metabolism as revealed by X-ray crystallography and NMR. *Future Med. Chem.* **2010**, *2* (9), 1451−68.

(11) Pochapsky, T. C.; Kazanis, S.; Dang, M. Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxid. Redox Signaling* **2010**, *13* (8), 1273−96.

(12) Sun, H.; Veith, H.; Xia, M.; Austin, C. P.; Huang, R. Predictive models for cytochrome p450 isozymes based on quantitative high throughput screening data. *J. Chem. Inf. Model.* **2011**, *51* (10), 2474−81.

(13) Veith, H.; Southall, N.; Huang, R.; James, T.; Fayne, D.; Artemenko, N.; Shen, M.; Inglese, J.; Austin, C. P.; Lloyd, D. G.; Auld, D. S. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat. Biotechnol.* **2009**, *27* (11), 1050−5.

(14) Cheng, F.; Yu, Y.; Shen, J.; Yang, L.; Li, W.; Liu, G.; Lee, P. W.; Tang, Y. Classification of cytochrome P450 inhibitors and noninhibitors using combined classifiers. *J. Chem. Inf. Model.* **2011**, *51* (5), 996−1011.

(15) Su, B. H.; Tu, Y. S.; Lin, C.; Shao, C. Y.; Lin, O. A.; Tseng, Y. J. Rule-Based Prediction Models of Cytochrome P450 Inhibition. *J. Chem. Inf. Model.* **2015**, *55* (7), 1426−34.

(16) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **2015**, *55* (2), 263−74.

(17) Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* **2016**, *3*, 80.

(18) Erhan, D.; L'Heureux, P. J.; Yue, S. Y.; Bengio, Y. Collaborative filtering on a family of biological targets. *J. Chem. Inf. Model.* **2006**, *46* (2), 626−635.

(19) Subramanian, G.; Ramsundar, B.; Pande, V.; Denny, R. A. Computational Modeling of beta-Secretase 1 (BACE-1) Inhibitors Using Ligand Based Approaches. *J. Chem. Inf. Model.* **2016**, *56* (10), 1936−1949.

(20) Aliper, A.; Belikov, A. V.; Garazha, A.; Jellen, L.; Artemov, A.; Suntsova, M.; Ivanova, A.; Venkova, L.; Borisov, N.; Buzdin, A.; Mamoshina, P.; Putin, E.; Swick, A. G.; Moskalev, A.; Zhavoronkov, A. In search for geroprotectors: in silico screening and in vitro validation of signalome-level mimetics of young healthy state. *Aging* **2016**, *8* (9), 2127−2152.

(21) Artemov, A. V.; Putin, E.; Vanhaelen, Q.; Aliper, A.; Ozerov, I. V.; Zhavoronkov, A. Integrated deep learned transcriptomic and structure-based predictor of clinical trials outcomes. *bioRxiv* **2016**, 095653.

(22) Kadurin, A.; Aliper, A.; Kazennov, A.; Mamoshina, P.; Vanhaelen, Q.; Khrabrov, K.; Zhavoronkov, A. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* **2017**, *8* (7), 10883−10890.

(23) Xu, Y. J.; Dai, Z. W.; Chen, F. J.; Gao, S. S.; Pei, J. F.; Lai, L. H. Deep Learning for Drug-Induced Liver Injury. *J. Chem. Inf. Model.* **2015**, *55* (10), 2085−2093.

(24) Lusci, A.; Pollastri, G.; Baldi, P. Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules. *J. Chem. Inf. Model.* **2013**, *53* (7), 1563−1575.

(25) Hughes, T. B.; Miller, G. P.; Swamidass, S. J. Modeling Epoxidation of Drug-like Molecules with a Deep Machine Learning Network. *ACS Cent. Sci.* **2015**, *1* (4), 168−180.

(26) Kew, W.; Mitchell, J. B. O. Greedy and Linear Ensembles of Machine Learning Methods Outperform Single Approaches for QSPR Regression Problems. *Mol. Inf.* **2015**, *34* (9), 634−647.

(27) Xu, Y. J.; Pei, J. F.; Lai, L. H. Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *J. Chem. Inf. Model.* **2017**, *57* (11), 2672−2685.

(28) Di Lena, P.; Nagata, K.; Baldi, P. Deep architectures for protein contact map prediction. *Bioinformatics* **2012**, *28* (19), 2449−2457.

(29) Lyons, J.; Dehzangi, A.; Heffernan, R.; Sharma, A.; Paliwal, K.; Sattar, A.; Zhou, Y. Q.; Yang, Y. D. Predicting Backbone C alpha Angles and Dihedrals from Protein Sequences by Stacked Sparse Auto-Encoder Deep Neural Network. *J. Comput. Chem.* **2014**, *35* (28), 2040−2046.

(30) Eickholt, J.; Cheng, J. L. Predicting protein residue-residue contacts using deep networks and boosting. *Bioinformatics* **2012**, *28* (23), 3066−3072.

(31) Eickholt, J.; Cheng, J. L. A study and benchmark of DNcon: a method for protein residue-residue contact prediction using deep networks. *BMC Bioinformatics* **2013**, *14*, S12 DOI: 10.1186/1471-2105-14-S14-S12.

(32) Lena, P. D.; Nagata, K.; Baldi, P. F. Deep Spatio-Temporal Architectures and Learning for Protein Structure Prediction. In *Advances in Neural Information Processing Systems*; Pereira, F., Burges, C. J. C., Bottou, L., Weinberger, K. Q., Eds.; Neural Information Processing Systems Foundation, Inc., 2012; pp 512−520.

(33) Skwark, M. J.; Raimondi, D.; Michel, M.; Elofsson, A. Improved Contact Predictions Using the Recognition of Protein Like Contact Patterns. *PLoS Comput. Biol.* **2014**, *10* (11), e1003889.

(34) Alipanahi, B.; Delong, A.; Weirauch, M. T.; Frey, B. J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **2015**, *33* (8), 831.

(35) Koch, C. P.; Perna, A. M.; Weissmuller, S.; Bauer, S.; Pillong, M.; Baleeiro, R. B.; Reutlinger, M.; Folkers, G.; Walden, P.; Wrede, P.; Hiss, J. A.; Waibler, Z.; Schneider, G. Exhaustive Proteome Mining for Functional MHC-I Ligands. *ACS Chem. Biol.* **2013**, *8* (9), 1876−1881.

(36) Sun, T. L.; Zhou, B.; Lai, L. H.; Pei, J. F. Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC Bioinformatics* **2017**, *18*, 277 DOI: 10.1186/s12859-017-1700-2.

(37) Ramsundar, B.; Kearnes, S.; Riley, P.; Webster, D.; Konerding, D.; Pande, V. Massively multitask networks for drug discovery. **2015**, *arXiv preprint arXiv:1502.02072*. https://arxiv.org/abs/1502.02072.

(38) Beisken, S.; Meinl, T.; Wiswedel, B.; de Figueiredo, L. F.; Berthold, M.; Steinbeck, C. KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinf.* **2013**, *14*, 257.

(39) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.

(40) *ChemAxon Standardizer, version 5.4. 4.1.*; ChemAxon: Budapest, Hungary, 2010.

(41) Yap, C. W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32* (7), 1466−74.

(42) Inglese, J.; Auld, D. S.; Jadhav, A.; Johnson, R. L.; Simeonov, A.; Yasgar, A.; Zheng, W.; Austin, C. P. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (31), 11473−8.

(43) Hinton, G. E.; Osindero, S.; Teh, Y. W. A fast learning algorithm for deep belief nets. *Neural Comput* **2006**, *18* (7), 1527−1554.

(44) Cortez, P.; Embrechts, M. J. Using sensitivity analysis and visualization techniques to open black box data mining models. *Inf. Sci.* **2013**, *225*, 1−17.

(45) Xu, Y.; Ma, J.; Liaw, A.; Sheridan, R. P.; Svetnik, V. Demystifying Multitask Deep Neural Networks for Quantitative Structure−Activity Relationships. *J. Chem. Inf. Model.* **2017**, *57* (10), 2490−2504.

(46) Lewis, D. F.; Ito, Y. Human P450s involved in drug metabolism and the use of structural modelling for understanding substrate selectivity and binding affinity. *Xenobiotica* **2009**, *39* (8), 625−35.

(47) Shou, M.; Grogan, J.; Mancewicz, J. A.; Krausz, K. W.; Gonzalez, F. J.; Gelboin, H. V.; Korzekwa, K. R. Activation of CYP3A4: evidence for the simultaneous binding of two substrates in a cytochrome P450 active site. *Biochemistry* **1994**, *33* (21), 6450−5.