Trainable In-Silico Screening Filter for Various Human Cytochrome P450 Isoforms Inhibition Liability

Pranas Japertas[1], Remigijus Didziapetris[1], Justas Dapkunas[1,2], Andrius Sazonovas[1]

[1] ACD/Labs, Inc., A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania,
[2] Department of Biochemistry and Biophysics, Vilnius University, M.K.Ciurlionio g. 21/27, LT-03101 Vilnius, Lithuania.

Advanced Chemistry Development

INTRODUCTION

Metabolism related drug-drug interactions caused by the inhibition of cytochrome P450 enzymes are among the main problems in modern drug discovery. Inhibition of CYP450s can lead to undesired accumulation of their substrates in the organism, potentially resulting in toxic side effects. To date, a number of drugs (mibefradil, terfenadine, astemizole) have been excluded from the market due to the induction of drug-drug interactions. As a result, testing of novel compounds for cytochrome P450 inhibition has become a common practice in the pharmaceutical industry. In this work we present predictive models for CYP450 Inhibition covering five major isoforms (3A4, 2D6, 1A2, 2C9, and 2C19). These models provide the probabilities that the compound of interest will inhibit a certain CYP450 isoform with IC_{50} below selected threshold. "General inhibition" models estimate whether the compound will exhibit any inhibition at all (IC_{50} < 50 μM), while "Efficient inhibition" models predict the probability that the compound will inhibit a selected enzyme with clinically significant IC_{50} < 10 μM.

DATA SET

Datasets of up towards 10,000 compounds have been used in the development of presented models. These have been collected from both original scientific publications (mainly considering the inhibition of the metabolism of probe CYP450 substrates) and the PubChem project (AIDs 410, 883, 884, 891, and 899) [1]. For main characteristics of the employed dataset see Table 1 and Figure 1.

TABLE 1. Datasets used for modeling.

Isoform	*N*	Literature	PubChem
3A4	9,844	976	8,868
2D6	8,329	699	7,630
2C9	8,201	425	7,776
2C19	7,958	302	7,656
1A2	6,009	295	5,714

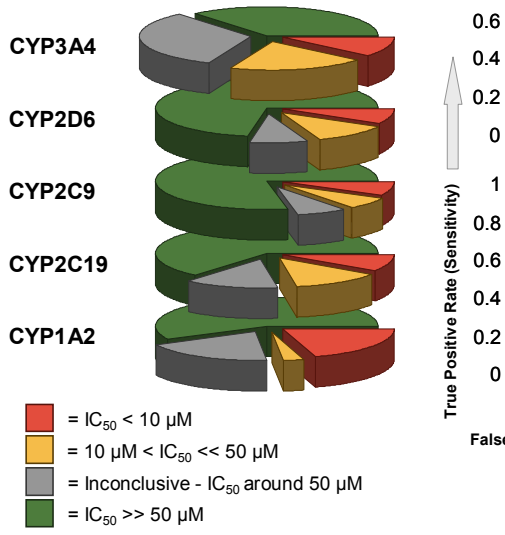


FIGURE 1. Distribution of the compounds according to the experimentally observed CYP450 inhibition IC_{50} values.

MODEL DEVELOPMENT

CYP450 inhibition models have been derived using a novel GALAS (**G**lobal, **A**djusted **L**ocally **A**ccording to **S**imilarity) modeling methodology. Each GALAS model consists of the following parts:

- A structure based QSAR for the prediction of the property of interest (i.e., baseline model)
- A user defined data set with the experimental values for the property of interest (i.e., Self-training Library)
- A similarity based routine that identifies the most similar compounds in the Self-training Library and calculates systematic deviations produced by the baseline model (i.e., training engine)

The result is a prediction that is corrected according to the experimental values of the most similar compounds in the user-defined Self-training Library and supported by the calculated Reliability Index (RI) value providing the basis for prediction quality assessment. More details about the GALAS modeling method can be found in our recent articles [2,3].

INTERNAL MODEL VALIDATION

The baseline model for general CYP3A4 inhibition produces relevant predictions for a test set with accuracy, specificity, and sensitivity close to 85%. Further improvements are introduced by local corrections using experimental data for similar compounds. Overall accuracy of the GALAS model exceeds 90% if the compound belongs to the Model Applicability Domain (RI>0.3), which constitutes 86% of the test set. For compounds with high reliability predictions (RI>0.5), accuracy exceeds 95%.

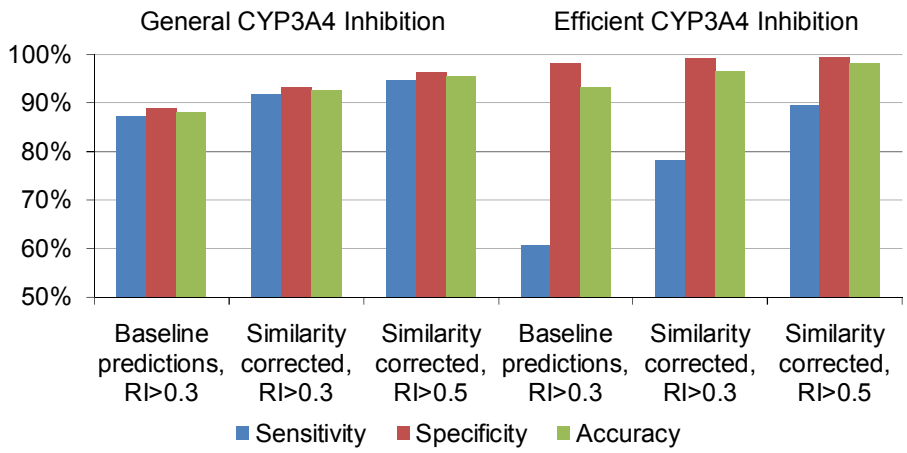


FIGURE 2. The detailed results of validation of the general and efficient CYP3A4 inhibition models on the test sets constituting 20% of the initial sets.

Despite high overall accuracy (92%), the sensitivity of the baseline model for the identification of effective CYP3A4 inhibitors (IC_{50} < 10 μM) is only 61%. After the similarity corrections, sensitivity increases to 78% for test set compounds belonging to the Model Applicability Domain (RI>0.3), and 89% for compounds with high reliability predictions (RI>0.5).

EXTERNAL MODEL VALIDATION

The same GALAS modeling methodology has been applied in modeling the inhibition of the remaining four P450 isoforms (2D6, 1A2, 2C9, and 2C19) with analogous outcome results on internal test sets. Subsequently, a more sophisticated validation has been performed for every model utilizing an external validation set. New screening data from the PubChem project (AID 1851) [4] classified according to a IC_{50} <10 μM threshold were used in this evaluation.

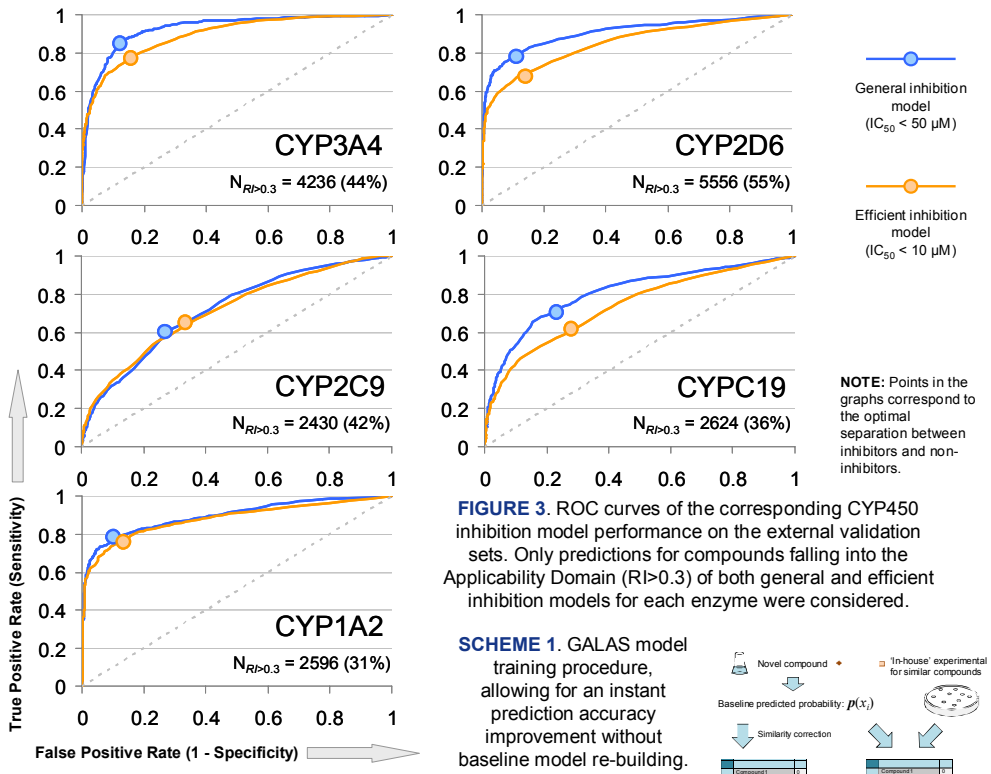
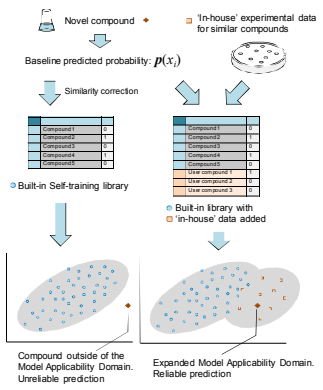


FIGURE 3. ROC curves of the corresponding CYP450 inhibition model performance on the external validation sets. Only predictions for compounds falling into the Applicability Domain (RI>0.3) of both general and efficient inhibition models for each enzyme were considered.

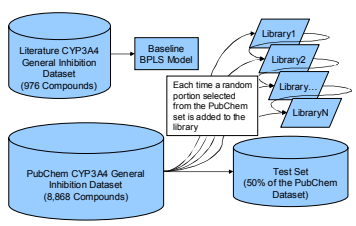
SCHEME 1. GALAS model training procedure, allowing for an instant prediction accuracy improvement without baseline model re-building.



Analysis of the provided ROC curves allows identification of the probability threshold ensuring the best classification of compounds. This threshold is not necessarily 0.5 and varies depending on the model and nature of experimental data (in our case it varied between 0.1 and 0.5). While the general inhibition model proved more useful in identifying compounds with IC_{50} <10 μM, interestingly enough, its performance for the non-inhibitors remained very good as well. However, this result is most likely caused by the specifics of the qHTS data used in the validation.

TRAINABILITY OF THE CYP450 INHIBITION MODELS

Every model has an applicability domain, beyond which the quality of predictions becomes highly questionable. To address this issue, all of the models derived using GALAS modeling methodology provide the possibility to expand their Applicability Domain by incorporating new experimental data into Self-training Libraries.



SCHEME 2. The scenario for the demonstration of the training functionality of the GALAS models.

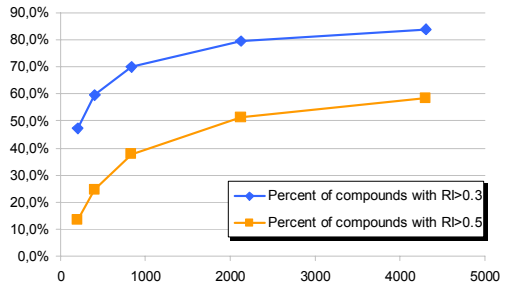


FIGURE 4. Percent of compounds within the Model Applicability Domain (RI>0.3) and obtaining high quality predictions (RI>0.5) when the size of the PubChem based Self-training Library increases.

CYTOCHROME P450 INHIBITION PREDICTION SOFTWARE

All of the described predictors are commercially available as part of ACD/ADME Suite and ACD/Tox Suite software (www.acdlabs.com/pc_admet). It provides both a standard model interface and Trainable module interface that can be used to expand the Applicability Domain of the provided models with the help of 'in-house' CYP450 inhibition data.

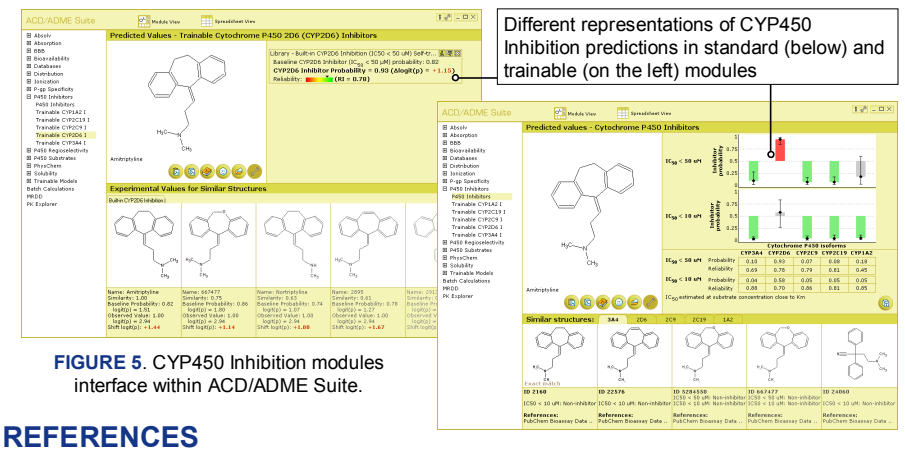


FIGURE 5. CYP450 Inhibition modules interface within ACD/ADME Suite.

REFERENCES

[1] http://pubchem.ncbi.nlm.nih.gov/
[2] Sazonovas, A. et al. *SAR QSAR Environ. Res.* **2010**, 21, 127-148.
[3] Didziapetris, R. et al. *J. Comput. Aid. Mol. Des.* **2010**, in press. DOI: 10.1007/s10822-010-9381-1
[4] Veith, H. et al. *Nat. Biotechnol.* **2009**, 27,1050-1055.