J Comput Aided Mol Des (2010) 24:891–906
DOI 10.1007/s10822-010-9381-1

Trainable structure–activity relationship model for virtual screening of CYP3A4 inhibition

Remigijus Didziapetris · Justas Dapkunas · Andrius Sazonovas · Pranas Japertas

Received: 16 March 2010 / Accepted: 14 August 2010 / Published online: 1 September 2010

Abstract A new structure–activity relationship model predicting the probability for a compound to inhibit human cytochrome P450 3A4 has been developed using data for >800 compounds from various literature sources and tested on PubChem screening data. Novel GALAS (Global, Adjusted Locally According to Similarity) modeling methodology has been used, which is a combination of baseline global QSAR model and local similarity based corrections. GALAS modeling method allows forecasting the reliability of prediction thus defining the model applicability domain. For compounds within this domain the statistical results of the final model approach the data consistency between experimental data from literature and PubChem datasets with the overall accuracy of 89%. However, the original model is applicable only for less than a half of PubChem database. Since the similarity correction procedure of GALAS modeling method allows straightforward model training, the possibility to expand the applicability domain has been investigated. Experimental data from PubChem dataset served as an example of in-house high-throughput screening data. The model successfully adapted itself to both data classified using the same and different IC_{50} threshold compared with the training set. In addition, adjustment of the CYP3A4 inhibition model to compounds with a novel chemical scaffold has been demonstrated. The reported GALAS model is proposed as a useful tool for virtual screening of compounds for possible drug-drug interactions even prior to the actual synthesis.

Keywords Drug-drug interactions · CYP3A4 inhibition · QSAR · GALAS model · Model applicability domain · Trainable model

R. Didziapetris · J. Dapkunas · A. Sazonovas · P. Japertas
ACD/Labs Inc., A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania

J. Dapkunas (✉)
Department of Biochemistry and Biophysics, Vilnius University, M.K.Ciurlionio g. 21/27, LT-03101 Vilnius, Lithuania
e-mail: justas.dapkunas@gmail.com

Introduction

Metabolism related drug-drug interactions, predominantly being caused by the inhibition of drug metabolizing enzymes, are among the main problems in modern drug discovery. Cytochrome P450 3A4 (CYP3A4) is the most relevant of such enzymes in human organism, responsible for more than 50% of drug metabolism [1]. It is a broad specificity oxygenase which is able to metabolize compounds belonging to many diverse drug classes [2]. Inhibition of CYP3A4 can lead to undesired accumulation of its substrates in the organism potentially resulting in toxic side effects. Up to date a number of drugs (mibefradil, terfenadine, astemizole) has been withdrawn from the market because of drug-drug interactions [1]. As a result, testing novel compounds for CYP3A4 inhibition has become a common practice in pharmaceutical industry [3].

Many in vitro methods have been developed and are used today to screen large libraries of synthesized compounds, such as inhibition of standard CYP3A4 probes in human liver microsomes or high-throughput screening (HTS) assays based on metabolism of fluorescent or luminogenic substrates [4–7]. The data acquired from such studies have been used later to develop in silico structure–activity relationship models of CYP3A4 inhibition (Table 1) [8–17], which can serve as virtual screening tools

Table 1 Summary of classification structure–activity relationship models of CYP3A4 inhibition based on large data sets of diverse compounds

Model and data	Modeling details[a]	Results[b]
Zuegge et al. 2002 [8][c]		
Inhibitors: $IC_{50} < 1$ μM (421, 36.5%)	*Methods:* PLS, ANN	*Accuracy:* 90%
Non-inhibitors: $IC_{50} > 50$ μM (160, 14.5%)	*Descriptors:* fragmental, topological, physicochemical	*Sensitivity:* 93%
		Specificity: 86%
Kriegl et al. 2005 [10, 11][c]		
Strong inhibitors: $IC_{50} < 2$ μM (243, 18%);	*Methods:* PLS, SVM	*Accuracy:* 70% (three class model)
Medium inhibitors: 2 μM $< IC_{50} <$ 20 μM (561, 42%)	*Descriptors:* physicochemical, topological, quantum chemical, 3D structural	*Sensitivity:* strong: 68%, medium: 63%
Non-inhibitors: $IC_{50} > 20$ μM (541, 40%)		
Arimoto et al. 2005 [12][d]		
Inhibitors: $IC_{50} < 3$ μM (1578, 35%)	*Methods:* RT, kNN, BC, LR, SVM	*Accuracy:* 83%
Non-inhibitors: $IC_{50} > 3$ μM (2892, 65%)	*Descriptors:* topological, fragmental	*Sensitivity:* 82%
		Specificity: 81%
Jensen et al. 2007 [14][e]		
Inhibitors: $IC_{50} < 20$ μM (361, 26%)	*Method:* kNN	*Accuracy:* 88%
Non-inhibitors: $IC_{50} > 20$ μM (1021, 74%)	*Descriptors:* fragmental	*Sensitivity:* 65%
		Specificity: 94%
		Not classified: 14%
Gleeson et al. 2007 [15][d]		
Inhibitors: $IC_{50} < 6.3$ μM (145, 19.8%)	*Methods:* PLS, RT	*Accuracy:* 89%
Non-inhibitors: $IC_{50} > 15.8$ μM (420, 57.2%)	*Descriptors:* fragmental, physicochemical	*Sensitivity:* 67%
		Specificity: 96%
Choi et al. 2009 [16]		
Inhibitors: not defined (394, 42.5%)	*Method:* RT	*Accuracy:* 73%
Non-inhibitors: not defined (533, 57.5%)	*Descriptors:* topological, physicochemical, 3D structural	*Sensitivity:* 83%
		Specificity: 54%

[a] Abbreviations used: *PLS* Partial Least Squares/Projection to Latent Structures, *ANN* Artificial Neural Networks, *SVM* Support Vector Machines, *RT* Regression Trees, *kNN* k-Nearest Neighbors, *BC* naive Bayesian Classifier, *LR* Logistic Regression

[b] Test set classification results for the best model is reported in case of several models

[c] Inhibition of erythromycin metabolism by recombinant CYP3A4

[d] Inhibition of 7-benzyloxy-4-trifluormethylcoumarin (BFC) metabolism by recombinant CYP3A4

[e] Inhibition of erythromycin metabolism in human liver microsomes

in evaluating the possibility for new compounds to inhibit CYP3A4. Such approach is very attractive because in silico models may be applied in early stages of drug discovery at a very small cost. Predictions are very fast and can be obtained for virtual compounds prior to their synthesis.

Despite the relative successes in CYP3A4 inhibition modeling, a need still exists for new computational approaches for identification of CYP3A4 inhibitors. All computational models have their limitations. For example, widely accepted 3D-QSAR models analyze spatial ligand-enzyme interactions assuming that binding mode for all compounds is the same. The predictive power of such CYP3A4 inhibition models is limited [18], because in reality a great variety of ligand binding modes to CYP3A4 exists as well as a large conformational degree of freedom in the active site is possible [19–22]. In this work we have applied 2D structural descriptors which proved to be suitable for CYP3A4 specificity modeling.

One of the main limitations of all predictive structure–activity relationship models is that they are valid only in the part of chemical space closely related to the training set, called model applicability domain. According to OECD principles for (Q)SAR validation, any model that is proposed for regulatory use should be based on a defined experimental endpoint, tested on data that were not used for its development, and after all associated with a defined domain of applicability [23]. The importance of model applicability domain has been also shown for CYP3A4 inhibition model recently [17].

Most of the previously published models are based on proprietary datasets, which automatically raises several issues regarding their potential application:

- The effective assessment of the applicability domain of such models is not possible as the actual training set structures are not available in public;
- The modeling techniques used in model development do not allow estimation of prediction reliability;
- In-house datasets usually consist of specific compounds that a particular institution is working with, rendering such models practically useless for anyone dealing with different compound classes.

Given these facts, we have developed a structure–activity relationship model, predicting the probability for a compound to inhibit CYP3A4, based on publically available diverse data. A novel GALAS (Global, Adjusted Locally According to Similarity) modeling methodology was used, which allowed forecasting the reliability of each prediction. The successful applications of this method in predicting continuous properties, such as Log*P* and acute toxicity in terms of LD_{50}, have been recently described in detail [24, 25]. In case of CYP3A4 inhibition models it was adapted for binary data.

GALAS model is a combination of two approaches: a global model for the prediction of the property of interest, and a similarity based local correction model. This methodology not only allows the estimation of reliability of predictions, but also makes it possible to expand the applicability domain of a resulting model in a very straightforward manner, i.e. without time consuming full statistical reparameterization of the model. These features will be demonstrated in an example involving the training of the CYP3A4 inhibition model developed using data from literature sources with the data from PubChem screening project. Finally, it will be shown how the model adapts to a series of compounds belonging to a novel chemical class.

Data

Literature dataset

Two datasets have been used in the development and validation of CYP3A4 inhibition models. CYP3A4 inhibition data in the first set ("Literature dataset") were collected from various literature sources (scientific publications, drug prescribing information). Inhibition of metabolism of probe CYP3A4 substrates (midazolam, testosterone, erythromycin, 7-benzyloxy-4-trifluormethylcoumarin (BFC), nifedipine, and others) has been considered. Classification of compounds was performed only after critical analysis of original literature in order to identify any cases of contradictions in experimental CYP3A4 inhibition data, such as substrate dependency [19, 26] or inconsistency between data obtained using human liver microsomal and recombinant enzyme [27]. Only IC_{50} values determined at substrate concentration close to K_m value were used for classification. Compounds having $IC_{50} < 40$ μM were classified as CYP3A4 inhibitors, having $IC_{50} > 60$ μM were classified as non-inhibitors, compounds with intermediate IC_{50} values (40–60 μM) or discrepant results in different assays were marked as inconclusive. In cases when detailed analysis of CYP3A4 inhibition was available and inhibition constant K_i has been reported, compounds having $K_i < 20$ μM were classified as CYP3A4 inhibitors.

PubChem dataset

The second set ("PubChem dataset") was downloaded from National Center for Biotechnology Information (NCBI) PubChem database (assay ID 884) on 10 June 2008 [28]. PubChem database contained 14,127 entries concerning CYP3A4 inhibition, determined using luminogenic CYP3A4 inhibition screening method [5]. PubChem data were pre-processed prior to further analysis: entries containing inorganic compounds, non-covalent complexes, and mixtures were excluded; salts were converted to corresponding acids or bases; water molecules were removed from hydrates. All compounds that were marked as CYP3A4 activators in PubChem database were excluded from further analysis. The resulting entries were classified as CYP3A4 inhibitors, non-inhibitors or inconclusive. To identify inhibitors in general the same rules were used as in PubChem: entry was classified as inhibitor if observed assay score was ≥40 and non-inhibitor if the score was 0, with inconclusive range lying between scores of 0 and 40. PubChem activity score is assigned from fitted IC_{50} value, with respect to completeness of dose–response curve and efficacy of inhibition (maximum inhibition response). For compounds having PubChem activity score > 40 the IC_{50} is less than 40 μM, therefore the classifications of PubChem and literature datasets in this case are consistent with each other. No specific effort was made to review any supporting experimental information and PubChem scores were used as provided.

Following the classification of the entire PubChem database, the attention was switched to the compounds provided with several experimental results, i.e. represented by multiple entries in the database (3,797 entries for 1,546 compounds with stereoisomers treated as duplicates). Only 55 compounds (4%) had contradicting classification as active (activity score > 40) in one experiment while inactive (activity score 0) in another. Such compounds were marked as inconclusive. Additional 398 compounds (26%) had inconclusive result in one of the experiments and therefore were also marked as inconclusive. For the

Table 2 Distribution of compounds in data sets with regard to CYP3A4 inhibition

Data set	No. of compounds	Inhibitors	Non-inhibitors	Inconclusive compounds
Literature dataset[a]	907	335 (36.9%)	497 (54.8%)	75 (8.3%)
PubChem dataset (general inhibition)[b]	11,060	3,032 (27.4%)	5,496 (49.7%)	2,532 (22.8%)
PubChem dataset (effective inhibition)[c]	11,060	1,238 (11.2%)	6,401 (57.9%)	3,421 (30.9%)

[a] Inhibitors: $IC_{50} < 40$ μM; non-inhibitors: $IC_{50} > 60$ μM; inconclusive compounds: IC_{50} 40–60 μM or contradicting results in different experiments

[b] Inhibitors: PubChem Activity score > 40; non-inhibitors: PubChem Activity score 0; inconclusive compounds: other (including contradictory results in repeating experiments)

[c] Inhibitors: $IC_{50} < 5$ μM and efficacy > 70%; non-inhibitors: $IC_{50} > 30$ μM; inconclusive compounds: other

remaining 1,093 compounds (71%) that had consistent classification one entry per compound was left in database.

Another classification scheme was applied to distinguish effective inhibitors in PubChem dataset: compounds with $IC_{50} < 5$ μM and maximal efficacy > 70% were classified as effective inhibitors, compounds with activity score of 0 or $IC_{50} > 30$ μM were treated as inactive, remaining compounds were classified as inconclusive.

Summary of datasets

Number of compounds and their distribution with regard to CYP3A4 inhibition classes for both datasets are summarized in Table 2. Inconclusive data from all datasets were not used in modeling. Compounds that were present in both literature and PubChem databases were excluded from the latter one, which was used as a validation set.

Consistency of experimental data for 297 compounds present in both Literature and correspondingly classified PubChem datasets has been checked before modeling. A total of 119 compounds have inconclusive classification in one of the sets. Detailed analysis of experimental values for remaining compounds is given in Table 3. For 156 out of 178 compounds (88%) consistent classification of inhibition was observed in both datasets.

Fragmental descriptors

Fragmental descriptors were chosen for the modeling of CYP3A4 inhibition. Molecules were fragmented using a set of 379 predefined fragments. The major part of this set is intended for the description of general chemical compound constitution and was comprised of conventional fragmental descriptors, such as atoms, functional groups, molecular shape fragments, and others. These descriptors have been already used and proved effective in previous projects, involving modeling various biological activities and chemical properties [24, 25]. Additionally several typical fragments describing CYP3A4 specificity were added (e.g., nitrogen containing heterocycles, methylenedioxybenzene, fragments representing possible cytochrome P450 metabolism sites, etc.).

Table 3 Consistency of experimental data present in both literature and PubChem datasets

	Literature	
	Inhibitor	Non-inhibitor
PubChem		
Inhibitor	36	13
Non-inhibitor	9	120
Agreement: 87.6%		

Methods

GALAS modeling methodology has been developed utilizing extensive experience from numerous QSAR modeling projects. Recently the successful application of this methodology in predicting various properties of continuous nature (e.g., Log*P*, LD_{50}) has been reported [24, 25]. Two main parts constituting the basis of the method are a global QSAR model providing baseline predictions for the property of interest and local corrections calculated according to the experimental data for the most similar compounds from the training set. The global model is intended to capture the general trends in CYP3A4 specificity and identify structural fragments tending to affect inhibition. The purpose of the second step is identifying any systematic errors made by the global model in the local chemical space of the test compound and compensating for it.

Global model

Baseline QSAR is a linear model built using logistic PLS. This method is a variation of ordinary PLS, possessing all its useful features in combination with the ability to analyze binary data. The predicted value in this case is the logit transformation of probability for a compound to be a CYP3A4 inhibitor (p) which is calculated as the sum of all the fragmental contributions:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \sum_i a_i f_i + c \quad (1)$$

here p is the probability for a compound to exhibit the CYP3A4 inhibition; f_i is the occurrence sum of a particular fragment in a molecule; a_i—statistical coefficient of the fragment, determined using logistic PLS; c—intercept.

This model was compared to those developed using linear Support Vector Machine (SVM) and nonlinear Random Forest methods. The statistical results were similar for all three models (see "Results and discussion" section). In fact, the implementation of GALAS modeling method described in this article is also possible using other global models as baseline. Logistic PLS was chosen for further works due to its easier interpretability and our experience in development of PLS-based QSAR models.

Dynamic similarity

Another important role of the baseline model is determining the "dynamic" similarity key used to compare compounds in prediction space, i.e. in terms of a particular analyzed property. Igor Tetko and his colleagues were among the first to realize the advantages of such approach over the use of typical methods that define the similarity of two molecules in descriptor space and a priori [29, 30]. The same predefined similarity key used in the models for different properties disregards the possibility that what is similar in case of property A can be significantly less relevant for property B.

Global model predicting CYP3A4 inhibition has been developed combining logistic PLS with the bootstrapping technique. The latter method implies random compound sampling from the initial training set [31], i.e. generation of new training sub-sets and derivation of independent model for each sub-set. Performing this procedure 100 times provides each compound with a vector of 100 predictions, each based on a slightly different part of the initial training set. This ensemble of models contains information about the influence and stability of every independent descriptor. Broad variability of the particular descriptor coefficient value across the models is an indication that the global QSAR cannot correctly estimate its influence. As a result, molecules containing the same highly variable descriptor are the best candidates to correct the potentially unreliable baseline prediction. The more variation is observed in the values of a particular descriptor coefficient, the more significant contribution the corresponding atom or functional group should have in the similarity assessment.

On the other hand, the most stable descriptors are those most widely encountered (–CH_2–, –CH_3, aromatic carbon and similar in case of fragmental descriptors). Their minimal contribution in the similarity assessment (i.e., the compounds that differ only by one –CH_2– or –CH_3 group will be treated as nearly identical by the model) conforms to the general chemical logic that compounds in homologous series are the most similar compounds among themselves.

The quantitative measure of the individual similarity between any two compounds (Similarity Index, SI_i) in the GALAS model is the square of correlation coefficient (r^2) between the prediction vectors. When comparing these vectors, the same trends in variability of 100 logistic PLS predictions indicate that the two compounds possess a very similar pattern of structural features found by the model to be of great influence in case of the analyzed property. This leads to a conclusion that two such compounds are indeed similar. Every difference in the set of significant fragments will inevitably reduce the correlation between the prediction vectors decreasing the SI_i.

Local model

Determining the similarity between any two compounds is a key process in the second layer of the GALAS modeling methodology. Here the predictions of the baseline model are compared to experimental values of the 5 training set compounds most similar to the query molecule. The final probability estimation is a combination of global prediction and local correction:

$$\ln\left(\frac{p}{1-p}\right) = \sum_i a_i f_i + c + \Delta \quad (2)$$

here Δ is the correction calculated according to the experimental data for the most similar compounds.

The Δ value itself is calculated as a weighted average from the differences between global QSAR predictions and experimental data for the five most similar compounds in the training set:

$$\Delta = \frac{\sum_{i=1}^{5}(a^{i-1} \cdot SI_i \cdot \Delta_i)}{\sum_{i=1}^{5} a^{i-1}} \quad (3)$$

here Δ is a correction that should be applied for the given prediction from the global model; a is a weighting constant (simple average is calculated if this constant is 1); SI_i is an individual Similarity Index between given compound and the i-th most similar compound in the training set; Δ_i is the difference between logit value of the experimental result and the value predicted by global model for the i-th most similar compound prior to its transformation into the baseline probability.

Since function logit approaches negative and positive infinity when its argument approaches 0 and 1, respectively

(the numbers used as experimental result indications in our case), 5% cut-offs are used at both ends of the probability value range to avoid unreasonable values while calculating Δ_i in Eq. 3, i.e., logit(>0.95) = logit(0.95) = 2.94 and logit(<0.05) = logit(0.05) = −2.94. For example, Δ_i for experimental positive result and predicted baseline probability 0.7 is calculated as the difference between logit(0.95) and logit(0.7), and in case of baseline probability > 0.95 Δ_i is zero.

The application of the Δ correction prior to the transformation using the logistic function (as shown in Eq. 2) and subsequent specifics of the Δ_i calculation (as described in Eq. 3) are prompted by necessity to confine the final corrected probability (p in Eq. 2) into the interval between 0 and 1. Straightforward application of the correction to the baseline probability would leave the possibility of the final probability exceeding 1 or falling below 0, in certain cases. Conversely, the above described workflow assures the different influence of the correction on the final probability depending on the initial baseline prediction. A proposed large positive correction will have a great impact if the baseline probability is low, but will not influence the prediction if the baseline probability is already close to 1.

Estimation of prediction reliability

GALAS modeling methodology also allows estimating the quality of prediction. This feature is even more significant than the accuracy improvement, compared to the global statistical methods. Knowing the prediction reliability is very important given the fact that any QSAR model is characterized by its applicability domain, outside of which, the performance of the model is usually poor [17, 23, 24]. No prediction can be considered reliable if there are no similar compounds in the training set. In the situations when such compounds do exist, but experimental data for them are inconsistent with regard to the global model, predictions based on such data cannot be confident as well. These two assumptions provide the basis for the calculation of Reliability Index (RI) which has been made directly dependent on these two factors.

The presence or absence of similar compounds in the training library is indicated by the compound Similarity Index (SI) to the entire data set. This index is calculated by weighted averaging of all the individual Similarity Indices (SI_i) for the test molecule and each of the five most similar compounds from the training library. Data-Model Consistency Index (DMCI) is used to quantitatively evaluate the consistency of experimental data of similar compounds with the global baseline model. DMCI value compares the individual differences between experimental and predicted baseline property values (Δ_i) for the same most similar compounds from the training library with the overall local correction for the compound of interest calculated by the Eq. 3. The more individual differences are scattered around the calculated average (Δ), the more inconsistent are the data for the similar compounds with regards to the global baseline model and vice versa.

The necessity to include a measure for consistency of experimental data has been described in our recent publication covering the topic of acute toxicity modeling [24]. In case of CYP3A4 inhibition, compound classes also exist where some representatives are potent inhibitors while others do not inhibit this enzyme at all despite being very similar. Such situation was observed for 1,4-dihydropyridine calcium channel antagonists [32].

The final prediction Reliability Index characterizing the applicability domain of the model is calculated in the following manner:

$$RI = SI \cdot DMCI. \quad (4)$$

It is a value set to vary between 0 and 1 with larger RI values indicating more reliable predictions. A more in-depth consideration of some of the mathematical background of the GALAS modeling methodology is available in our recent publication [24].

Training of the GALAS model

Another important feature of GALAS modeling methodology is the possibility for an easy and straightforward expansion of the applicability domain of resulting models. Training of the GALAS models is performed by simply adding new compounds with experimental data to the similarity correction (local) part of the modeling. If the baseline model is unable to predict accurately for a certain class of compounds, the local similarity correction can compensate this inaccuracy by calculating the appropriate Δ value. Introduction of the new data results in the adaptation of the local model to a new part of the chemical space, represented by the newly imported compounds, while the same global model is used. The latter fact allows performing model training "on-the-fly" without time consuming full statistical reparameterization of the model.

Development and validation of models

The model predicting CYP3A4 inhibition is based on the Literature dataset and predicts probability for a compound to be a CYP3A4 inhibitor with $IC_{50} < 40$ μM. Literature dataset was randomly split into the training and test set (190 compounds in the test set, ca. 20%). Initially the model was tested on this internal test set, followed by a more rigorous validation using the PubChem dataset as an external test set, which was published after the model has been developed. As it was already mentioned, compounds that were present in both databases were excluded from the external test set.

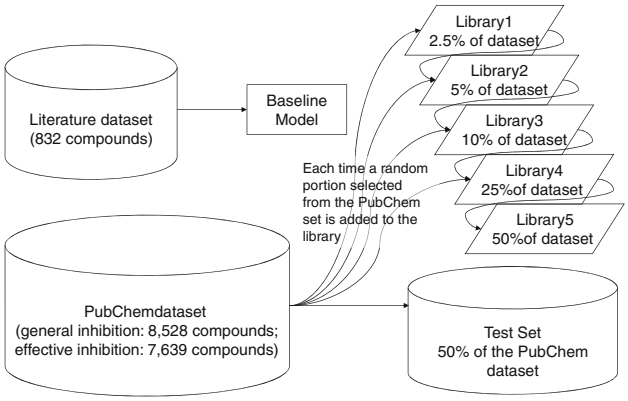


Fig. 1 The workflow of the GALAS model trainability testing

Compounds with predicted probability >0.5 were considered as inhibitors, whereas probability of <0.5 indicated non inhibitors. Additionally, the performance of global and local models was adjusted using receiver operating characteristic (ROC) curves [33].

Following performance test the possibility to train GALAS model has been investigated. In order to verify whether the model can be trained using data from a different assay two different trainability experiments have been conducted on PubChem data classified as CYP3A4 inhibitors using different thresholds. In the first scenario the model was trained using PubChem dataset, classified according to PubChem activity score, i.e. data from similar experimental assay. The second trainability example used PubChem dataset with only effective CYP3A4 inhibitors classified as positive compounds, mimicking a different experimental assay with a different potency threshold to identify inhibitors. In both cases the PubChem dataset was divided into several parts: one half of it was reserved as the validation set, while the remaining parts were added one by one to the training library (see Fig. 1 for details).

The third trainability example involved compounds having a completely new structural scaffold. CYP3A4 inhibition data for 10 insulin-like growth factor-1 receptor (IGF-1R) inhibitors have been recently published [34]. All these compounds were classified as CYP3A4 inhibitors having $IC_{50} < 40$ μM. Five compounds were randomly selected and added one by one to the similarity correction library which in case of original model consisted of the literature dataset. After the addition of each training molecule the probability to inhibit CYP3A4 and corresponding RI values were calculated for five remaining compounds.

Software

Molecule fragmentation and all subsequent statistical analysis were performed using Algorithm Builder 1.8 software [35, 36], except SVM and Random Forest models which were developed using R 2.6.2 [37]. ROC curves were generated by a web-based calculator [38]. The trainability of GALAS model was tested on ACD/ADME Suite 4.95 software application [39].

Results and discussion

Classification scheme, distinguishing compounds between inhibitors and non-inhibitors according to their experimental data is the first aspect of any attempt to develop statistical classification models of enzyme inhibition. Inhibition potency thresholds selected for classification in earlier structure–activity relationship studies of CYP3A4 inhibition cannot be directly compared between each other, because experimental estimation of CYP3A4 inhibition depends on the methods used [7, 19, 20]. The percent of active compounds identified in different screening programs also confirms that; and the distribution of compounds according to CYP3A4 inhibition potency is not the same in available databases. PubChem screening program identified 27% of compounds with IC_{50} values less that 40 μM (see Table 2). The study performed on *Novo Nordisk* in-house database, utilizing the assay of erythromycin metabolism inhibition in human liver microsomes, identified a similar fraction of 26% of compounds as active, yet a lower threshold was used ($IC_{50} = 20$ μM, instead of $IC_{50} = 40$ μM) [14]. Even greater proportions of active compounds were reported in other studies using IC_{50} thresholds as low as <10 μM for compound classification (see Table 1) [8, 10, 12, 15].

The aforementioned facts suggest that universal and objective thresholds for classification of compounds into CYP3A4 inhibitors and non-inhibitors cannot be established [3]. In this work two classification thresholds for diversifying the compounds according to their experimental CYP3A4 inhibition data were chosen. The first one is used to identify inhibitors in general ($IC_{50} < 40$ μM), while the second one distinguishes only effective inhibitors ($IC_{50} < 5$ μM) from the rest. The structure-activity relationship model described in this article predicts general CYP3A4 inhibitory properties ($IC_{50} < 40$ μM) and was built using data from literature sources (Literature dataset). This unusually high threshold to classify inhibitors is chosen for consistency of classification of literature and PubChem data. It allows identification of the most general properties related to CYP3A4 inhibition.

Further it will be demonstrated how this model, built using GALAS modeling methodology, is able to adapt itself even to the binary data obtained using IC_{50} thresholds other than in the construction of the training set, i.e. when only effective inhibitors with $IC_{50} < 5$ μM were classified as active compounds. Consequently, proposed classification

model based on any selected threshold can later serve as baseline model in training with CYP3A4 inhibition data from any available inhibition assay: either based on fluorescent substrate metabolism by recombinant CYP3A4, or inhibition studies with human liver microsomes, or other.

Global model

The first step of the presented GALAS model is a logistic PLS with predefined set of fragments as independent variables (see "Methods" section)—a baseline model of CYP3A4 inhibition. This is a linear and additive model. It produces relevant predictions for both internal and external test sets with accuracy of about 80%. The overall results are comparable to other standard machine learning methods like SVM and Random Forest (Table 4). The fact that an additive model accurately describes probability for a compound to inhibit CYP3A4 is in agreement with the very broad specificity of this enzyme [2]. The active site cavity of CYP3A4 is considerably larger than that of any other cytochrome P450 isoform. It also has the potential to expand considerably on ligand binding [21, 40] and is even able to accept several molecules simultaneously [21, 41]. The dependence of CYP3A4 inhibition potency on molecular weight has also been shown to be markedly different from that of the other cytochrome P450 isoforms with no drop in the mean potency for large compounds with molecular weight of > 750 Da [42]. Therefore, additivity of the model still persists for compounds with high molecular weight.

The most attractive feature of in silico screening tools is the ability to use them in the assessment of the properties for virtual compound libraries. Large databases of structures can be screened even prior to the actual synthesis of the substances. A part of this study has been devoted to the analysis of the baseline model predictions for some virtual compounds in order to understand how changes in chemical structure affect the probability of drug-like compound to inhibit CYP3A4. Table 5 presents predicted probabilities to inhibit CYP3A4 for virtual ipriflavone analogues. Ipriflavone itself is not a CYP3A4 inhibitor, but its metabolites are weak inhibitors [43, 44]. Similar flavones that inhibit this enzyme are present in PubChem database. Baseline model predicts ipriflavone as non inhibitor of CYP3A4—the probability for this compound to have $IC_{50} < 40$ μM is 0.3 ($\text{logit}(p) = -0.85$).

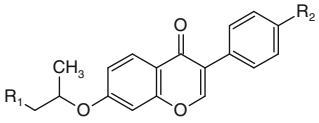
Two studies have been published recently that are dedicated to the analysis of the influence of physicochemical properties and most popular substituents of organic compounds on ADME and toxicity parameters, including CYP3A4 inhibition [42, 45]. The predictions of the baseline model obtained for virtual compounds in Table 5 are within a good agreement with results of these studies. The predicted probabilities were converted to corresponding logit(p) values in order to maintain linear scale before comparison of the published changes in average pIC_{50} [45] with the predicted values of our models. Then the difference between the logit(p) of a virtual compound and the logit(p) of ipriflavone was calculated ($\Delta\text{logit}(p)$). This value shows the influence of a particular substituent in our baseline model. The overall correlation of published changes in pIC_{50} and $\Delta\text{logit}(p)$ for considered virtual ipriflavone analogues is high ($r^2 = 0.70$, Fig. 2). Some notable substituents are analysed further in the text.

Introduction of an acidic group makes the inhibition of CYP3A4 enzyme almost impossible (predicted probability for virtual analogue $p = 0.05$, $\Delta\text{logit}(p) = -2.1$). The negative impact of any acidic group on IC_{50} for CYP3A4 inhibition value has been also shown using proprietary data ($\Delta pIC_{50} = -0.55$) [42, 45]. Data on HMG-CoA reductase inhibitors also show that, while in acidic form, these compounds show no activity towards CYP3A4, whereas

Table 4 Classification results of CYP3A4 inhibition provided by the global models for internal and external test sets

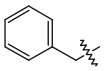
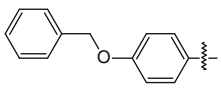
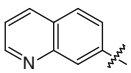
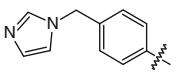
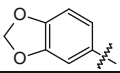
	Internal test set (190 compounds)				External test set (8,528 compounds)			
Logistic PLS		Pred. True	Pred. False			Pred. True	Pred. False	
	Obs. True	52	17	Sensitivity: 75.4%	Obs. True	1,843	1,189	Sensitivity: 60.8%
	Obs. False	17	104	Specificity: 86.0%	Obs. False	855	4,641	Specificity: 84.4%
				Accuracy: 82.1%				Accuracy: 76.0%
SVM		Pred. True	Pred. False			Pred. True	Pred. False	
	Obs. True	55	14	Sensitivity: 79.7%	Obs. True	2,071	961	Sensitivity: 68.3%
	Obs. False	24	97	Specificity: 80.1%	Obs. False	1,173	4,323	Specificity: 78.7%
				Accuracy: 80,0%				Accuracy: 74.7%
Random forest		Pred. True	Pred. False			Pred. True	Pred. False	
	Obs. True	53	16	Sensitivity: 76.8%	Obs. True	1,750	1,282	Sensitivity: 57.7%
	Obs. False	17	104	Specificity: 86.0%	Obs. False	684	4,812	Specificity: 87.6%
				Accuracy: 82.6%				Accuracy: 76.9%

Table 5 Effects of structure modifications of virtual ipriflavone analogues on predicted baseline probability for a compound to inhibit CYP3A4



R_1	R_2	ΔpIC_{50}[a]	p	logit(p)	Δlogit(p)
Ipriflavone:					
-H	-H	0	0.3	-0.85	0.00
Ionogenic substituents:					
-COOH	-H	-0.55	0.05	-2.94	-2.10
-H	-COOH	-0.55	0.04	-3.18	-2.33
$-NH_2$	-H	0.07	0.23	-1.21	-0.36
-H	$-NH_2$	0.07	0.26	-1.05	-0.20
$-N(CH_3)_2$	-H	0.09	0.15	-1.73	-0.89
-H	$-N(CH_3)_2$	0.36	0.49	-0.04	0.81
	-H	0.00	0.19	-1.45	-0.60
Hydrophylic substituents:					
-OH	-H	-0.23	0.25	-1.10	-0.25
-H	-OH	-0.23	0.29	-0.90	-0.05
$-CH_2OH$	-H	-0.05	0.33	-0.71	0.14
$-CONH_2$	-H	-0.01	0.21	-1.32	-0.48
$-SO_2CH_3$	-H	0.12	0.28	-0.94	-0.10
-H	$-SO_2CH_3$	0.12	0.25	-1.10	-0.25
-H	$-NHSO_2CH_3$	0.28	0.23	-1.21	-0.36
Ethers, esters and amides:					
$-COOCH_3$	-H	0.26	0.48	-0.08	0.77
-H	$-COOCH_3$	0.26	0.34	-0.66	0.18
-H	$-OCH_3$	0.11	0.37	-0.53	0.32
-H	$-OCF_3$	0.44	0.53	0.12	0.97
-H	$-OCH_2CH_3$	0.22	0.42	-0.32	0.52
$-SCH_3$	-H	0.24	0.35	-0.62	0.23
-H	$-SCH_3$	0.24	0.33	-0.71	0.14
$-CH_2OCH_3$	-H	0.19	0.45	-0.20	0.65
$-OCOCH_3$	-H	0.16	0.48	-0.08	0.77
$-NHCOCH_3$	-H	0.23	0.39	-0.45	0.40
Aliphatic substituents:					
$-CH_3$	-H	0.11	0.36	-0.58	0.27
$-CH_2CH_3$	-H	0.32	0.39	-0.45	0.40
$-CH_2CH_2CH_3$	-H	0.28	0.43	-0.28	0.57
$-CH(CH_3)_2$	-H	0.29	0.39	-0.45	0.40
$-CH_2CH_2CH_2CH_3$	-H	0.20	0.48	-0.08	0.77
$-CH(CH_3)(CH_2CH_3)$	-H	0.50	0.41	-0.36	0.48
$-C(CH_3)_3$	-H	0.36	0.57	0.28	1.13
-H		0.51	0.53	0.12	0.97
-H		0.42	0.63	0.53	1.38

Table 5 continued

Aromatic substituents:					
	-H	0.46	0.77	1.21	2.06
	-H	0.65	0.88	1.99	2.84
	-H	0.54	0.57	0.28	1.13
	-H	0.73	0.54	0.16	1.01
	-H	0.86	0.68	0.75	1.60
	-H	-[b]	0.97	3.48	4.32
	-H	-[b]	0.86	1.82	2.66
	-H	-[b]	0.91	2.31	3.16
	-H	-[b]	0.9	2.20	3.04
Halogen containing substituents:					
-H	-F	0.07	0.27	-0.99	-0.15
-H	-Cl	0.21	0.37	-0.53	0.32
-H	-Br	0.27	0.35	-0.62	0.23
-H	$-CF_3$	0.24	0.48	-0.08	0.77
$-CH_2F$	-H	0.52	0.39	-0.45	0.40
-CN	-H	0.08	0.36	-0.58	0.27
-H	-CN	0.08	0.23	-1.21	-0.36

[a] ΔpIC_{50} values are taken from a previous publication [45]

[b] The ΔpIC_{50} for these substituents was not reported in article [45]

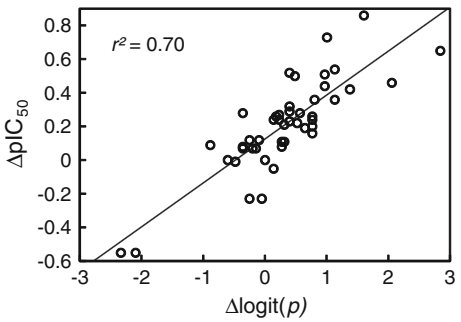


Fig. 2 The correlation between published changes in average pIC_{50} (ΔpIC_{50}) [45] and changes in values predicted by the baseline model ($\Delta logit(p)$)

their lactone form inhibits this enzyme [46]. Similar effect is observed in a novel series of arginine vasopressin V2 receptor agonists. The compound with carboxygroup does not inhibit CYP3A4 while others do [47]. Introduction of a methyl ester shows positive impact on CYP3A4 inhibition both analyzing experimental data and predictions of our model (p in the range from 0.34 to 0.48 for esters, corresponding $\Delta logit(p)$ in the range between 0.18 and 0.77). Therefore it may be concluded that the negative charge of an acidic group is the most significant property reducing the interaction between inhibitor and CYP3A4.

Introduction of a strong basic group (e.g., aliphatic amine) slightly decreases the probability for a compound to inhibit CYP3A4, but the effect is nowhere comparable to that of an acidic group (p varies between 0.15 and 0.23, corresponding $\Delta logit(p)$ in the range of -0.89 to -0.36). Similarly, average inhibition potency for basic compounds has been found to decrease slightly compared with neutral compounds in large datasets of CYP3A4 inhibitors analyzed in other studies [42].

Negative impact of various hydrophylic groups (hydroxyl or amide, p ranges between 0.21 and 0.33, $\Delta logit(p)$ varies within -0.48 and -0.05 with an exception for hydroxymethyl substituent having $\Delta logit(p) = 0.14$) is possibly related to the decrease of lipophilicity of the compound. On the contrary, increasing the size of the molecule with hydrophobic aliphatic or aromatic residue

results in the rise of the probability for a compound to inhibit CYP3A4 (p in the range 0.36 to 0.97, $\Delta\text{logit}(p)$ varies between 0.27 and 4.32). The majority of previously reported CYP3A4 specificity models highlighted the importance of multiple hydrophobic groups for effective interaction with this enzyme [15]. This characteristic dependency of CYP3A4 inhibition potency on molecular weight and lipophilicity was as well observed in a recent analysis of a large amount of CYP3A4 inhibition data [42]. The significance of hydrophobic interactions for effective CYP3A4 inhibition might be expected from the known crystal structure of this enzyme, in which a phenylalanine cluster plays an important role in defining the CYP3A4 active site [40]. This particular fact explains why aromatic substituents inflict a larger influence on predicted probabilities compared to an aliphatic chain (probabilities 0.36 to 0.57 for aliphatic chains and 0.57 to 0.88 for aromatic rings). The introduction of nitrogen containing aromatic rings increases the probability to inhibit enzyme even more (predicted probability in the range from 0.68 to 0.97). Possession of the nitrogen-containing heterocyclic moieties (such as imidazole, quinoline, pyrimidine) gives the compound a possibility to form complexes with the heme iron inside cytochrome P450 enzymes [48].

Methoxy- and ethoxygroups ($R–OCH_3$, $R–OC_2H_5$) connected to the aromatic ring increase probability to inhibit CYP3A4 (p varies from 0.37 to 0.42, $\Delta\text{logit}(p)$ ranges from 0.32 to 0.52). Methoxygroup connected to aromatic ring is a possible site of CYP3A4 mediated metabolism [49]. Substituents of this type increase the probability for a compound to become a mechanism-based CYP3A4 inhibitor. One of the biggest impacts on predicted CYP3A4 inhibition probability was observed following the addition of a methylenedioxybenzene substituent ($p = 0.9$, $\Delta\text{logit}(p) = 3.04$). This group, well known for its relevance in mechanism-based inhibition, is frequent among inhibitors [48, 49].

Although predictive models described in this article are not able to distinguish mechanism-based inhibitors from competitive ones, this shortcoming is not likely to be important in practical applications. Many of the known clinically relevant inhibitors like azole type drugs interact with CYP3A4 as mixed type inhibitors—both competitive and mechanism-based [50]. High-throughput screening experiments used in determination of cytochrome P450 inhibitors cannot identify mechanism-based inhibition as well.

Local model

The global model described above successfully reflects the general trends of CYP3A4 inhibition. However, the sensitivity of the model is lower than specificity (75% in the internal test set and 61% in the external test set, see Table 4). This can be attributed to the features of some CYP3A4 inhibitors which could not be described using linear model. Under these circumstances a method is needed that is able to account for possible nonlinear effects. Local correction of the baseline model predictions according to experimental data for the most similar compounds (a second layer of the GALAS model) has been developed to deal with this kind of problems. Moreover, this routine allows estimation of the prediction reliability in the form of calculated Reliability Index (RI), as well as training of the model using new experimental data.

Detailed results of the validation of the GALAS model for CYP3A4 inhibition are presented in Table 6. Predictions having RI < 0.3 fall outside the applicability domain of the model [24] and are not considered here. This affects only a small fraction of the internal test set. The majority (73%) of its compounds obtain predictions of acceptable reliability. The statistical parameters of the GALAS model for such compounds are superior compared to those of the baseline model (overall accuracy—89%, sensitivity—83%,

Table 6 The final results of the GALAS model of CYP3A4 inhibition for the internal and external test set compounds falling within the model applicability domain (RI > 0.3) and obtaining high reliability predictions (RI > 0.5)

	Internal test set (190 compounds)				External test set (8,528 compounds)			
RI > 0.3		Pred. True	Pred. False			Pred. True	Pred. False	
	Obs. True	43	9	Sensitivity: 82.7%	Obs. True	477	214	Sensitivity: 69.0%
	Obs. False	6	80	Specificity: 93.0%	Obs. False	189	2638	Specificity: 93.3%
				Accuracy: 89.1%				Accuracy: 88.5%
	% of compounds within RI range: 72.6%				% of compounds within RI range: 41.3%			
RI > 0.5		Pred. True	Pred. False			Pred. True	Pred. False	
	Obs. True	30	3	Sensitivity: 90.9%	Obs. True	110	59	Sensitivity: 65.1%
	Obs. False	1	41	Specificity: 97.6%	Obs. False	19	1012	Specificity: 98.2%
				Accuracy: 94.7%				Accuracy: 93.5%
	% of compounds within RI range: 39.5%				% of compounds within RI range: 14.1%			

specificity—93%). This is actually greater than classification consistency between the datasets compiled from literature sources and PubChem database (see Table 3). When considering only predictions with high reliability (RI > 0.5), the accuracy increases to 95%, with both sensitivity and specificity being higher than 90%. These values approach the accuracy of experimental measurements that were observed in the analysis of compounds having multiple experimental values reported in the PubChem database. Such results serve as the first confirmation that the employed Reliability Index calculation methodology effectively assesses the quality of prediction and defines the applicability domain of the model.

The benefits of correcting baseline predictions according to experimental data for similar compounds can be seen while analyzing receiver operating characteristic (ROC) curves for the internal test set (Fig. 3). These graphs show the dependence between false positive rate (1 – specificity) and true positive rate (sensitivity) as the discrimination threshold of the model is varied [33]. The point (0; 1) corresponds to perfect classification. The closer to this point gets the curve, the better is the model. As it can be seen from the ROC curves for compounds in the applicability domain of the model (Fig. 3a), local modeling shifts the baseline predictions towards better classification. The influence of similarity correction is bigger for compounds obtaining high reliability predictions (Fig. 3b). This can be explained by the fact that significant Δ values are calculated only in the cases when highly similar compounds with consistent experimental data are present in the training library.

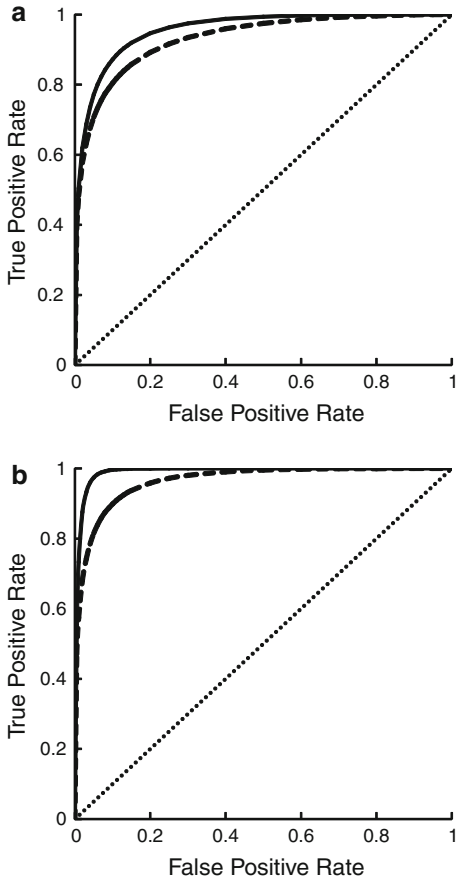


Fig. 3 ROC curves comparing the performance of global and similarity corrected models on the internal test set: **a** compounds within the applicability domain of the model (RI > 0.3); **b** compounds obtaining high reliability predictions (RI > 0.5). *Dotted line* indicates random classifier, *dashed line*—baseline model, *solid line*—similarity corrected model

Comparison of the GALAS model results for internal test set compounds within different RI ranges reveal the improvement of all the statistical parameters with the increase of RI values (see Fig. 4). Similar trends are observed when testing the model on the external set—PubChem dataset. For compounds that belong to the applicability domain of the model (RI > 0.3) the statistical results are also better than those of the baseline model (Table 6 vs. Table 4). The accuracy of predictions is 89% with acceptable sensitivity (69%) and very good specificity (93%). Overall prediction accuracy is even better (93%) for compounds having high Reliability Index (RI > 0.5). However, only less than half (41%) of the PubChem dataset actually belongs to the applicability domain of the model, and only a small fraction of compounds (14%) obtain high reliability predictions. In situations like this, the possibility to expand the applicability domain of the model would be of the utmost importance.

Training of the GALAS model

The similarity correction procedure described in the methodological part plays the major role in the ability to train the model with external data. In addition to the original training library, initially containing the same training set compounds used in the development of a baseline model, new compounds can be added allowing the model to cover a novel, previously unknown, part of the chemical space.

The above described procedure does not introduce large changes if the baseline predictions are already correct. Yet if a query molecule representing a part of the chemical space just added to the model is mispredicted by the global model which remains unchanged, the similar compounds just added to the similarity correction library effectively correct this baseline prediction based on the newly available experimental data.

Training with data from a similar assay

PubChem dataset was used to demonstrate how training of the model is able to improve the predictions for new compounds. A randomly selected part of this database was

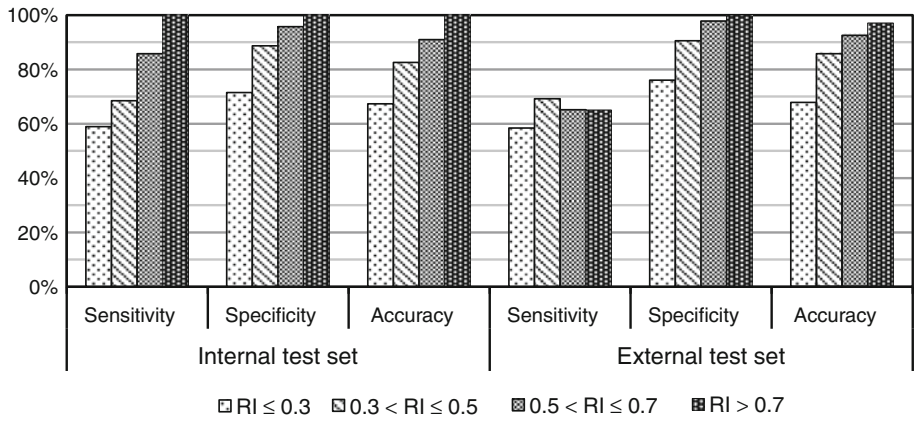


Fig. 4 The statistical parameters of GALAS models for internal and external test set compounds classified according to the Reliability Index values

used as a validation set for trainability testing, while remaining compounds were added in several portions to form a new training library of the model, as described in the "Methods" section. The results for models containing different fractions of the PubChem dataset in a training library are presented in Fig. 5.

The baseline model predicts CYP3A4 inhibition specificity for the trainability validation part of the PubChem dataset with overall accuracy of 76%, sensitivity of 59%, and specificity of 85% (see Fig. 5a). These parameters are comparable to the baseline model performance on the whole PubChem dataset. Addition of just a small part of the remaining dataset (ca. 200 compounds) as the training library markedly improves the statistical performance (accuracy 79%, sensitivity 65%, specificity 87%), whereas further expansion of this library with PubChem compounds keeps increasing the values of statistical parameters of the model (see Fig. 5a), reaching a maximum accuracy of 86%, sensitivity of 77%, and specificity of 90% when $> 4{,}000$ compounds are added to the library.

Notably, in all cases, the specificity of the model was significantly better than sensitivity. This means that a compound dissimilar to the training set is always more likely to be classified as a non-inhibitor. The small number of correctly identified inhibitors can be named as a major drawback of the baseline model. This number noticeably increases with the addition of new compounds to the training library. Just 200 compounds added result in 35% of inhibitors being identified with $RI > 0.3$, and this percent increases to 46% when 400 compounds are added (see Fig. 5b). The fraction of inhibitors identified with high reliability ($RI > 0.5$) increases approximately twofold comparing the results for 200 and 400 compounds added to the library (13 and 25%, respectively). Even more inhibitors can be identified if larger training libraries are used, because of expanding of the model applicability domain to cover more compound structural classes. Following the addition of all available (>4,000) PubChem compounds to the training library 50% of CYP3A4 inhibitors are found with high reliability of prediction.

In order to directly illustrate the actual expansion of the GALAS model applicability domain as a result of new compounds added to the training library; let us consider the following results. Only 48% of the trainability validation set belongs to the applicability domain ($RI > 0.3$) of the model trained with just 200 compounds. The number of acceptable reliability predictions increases to 60% after adding another 200 compounds (400 in total) (see Fig. 5c). Further enlargement of the training library eventually results in applicability domain of the model covering 89% of the test set at $> 4{,}000$ compounds added (ca. twofold increase in total). The relative growth in the number of high reliability predictions ($RI > 0.5$) is even bigger, starting from 14% at 200 compounds added, and reaching 59% at $> 4{,}000$ compounds added (ca. four times more). These results definitely show the adaptation of the model, based on the literature data, to the part of the chemical space occupied by the PubChem dataset compounds.

Training with data from an assay with a different potency threshold

Following the successful training of CYP3A4 inhibition model with data from a similar experimental assay classified using the same IC_{50} threshold, a natural question arises: can the trainable model described above be adapted for the discrimination between CYP3A4 inhibitors and non-inhibitors identified using a different scale? In other words, do the binary data (inhibitor/non-inhibitor) used in model training necessarily have to be obtained using exact same criteria as for the training set, or data produced by any screening programs can be used to train the model, no matter what experimental protocol or inhibition potency threshold was used to classify compounds in terms of CYP3A4 inhibition? To test this possibility, the initial model was trained with the PubChem dataset classified using significantly different inhibition potency threshold ($IC_{50} = 5$ μM versus $IC_{50} = 40$ μM used to classify the

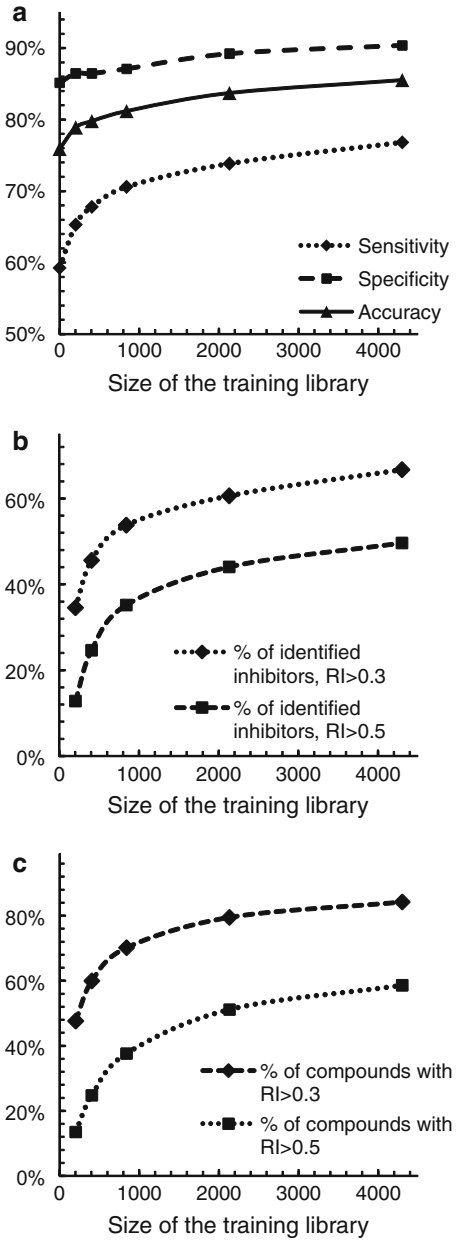


Fig. 5 Training of the GALAS model using data from a similar assay: **a** statistical parameters of the model; **b** percentage of identified inhibitors; **c** percentage of compounds predicted with RI > 0.3 and RI > 0.5

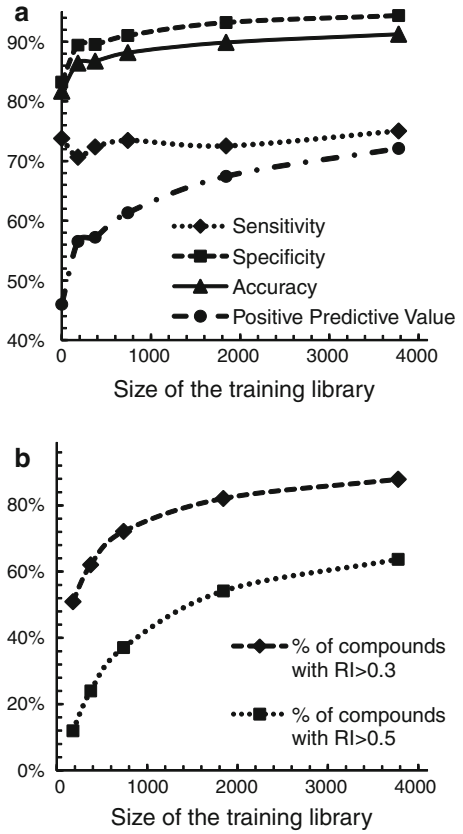


Fig. 6 Training of the GALAS model using data from an assay with different potency threshold: **a** statistical parameters of the model; **b** percentage of compounds predicted with RI > 0.3 and RI > 0.5

training set data). The results for the trained models obtained during this experiment are presented in Fig. 6.

The first look at the results reveals that initially many false positive predictions are observed. Using the baseline model for the classification of the PubChem validation set, the positive predictive value is only 46%, i.e. only about a half of compounds predicted as active are indeed experimentally determined as effective CYP3A4 inhibitors. This is an inevitable consequence of the differences in classification thresholds used. Percent of false positive predictions decreases and positive predictive value of the model increases following the addition of the new compounds from the PubChem set to the training library (see Fig. 6a). Improvements in other statistical parameters are observed as well (Fig. 6a). The fraction of reliable and high reliability predictions depends on the number of new compounds added to the library analogously as observed in the previous training example (Fig. 6b). These results demonstrate the ability of the GALAS modeling methodology, employed in this work, to handle data from different types of experimental CYP3A4 inhibition studies.

Training with compounds from a new structural class

The above examples show the adaptation of the model to experimental data obtained using a different method or even a different classification threshold. Obviously, having such a large library with experimental data for new compounds (exceeding the original training set several times) is a reasonable argument for the revision of the model. Still a simple statistical reparameterization of the existing model not necessarily produces better results than the described training procedure. In contrast, full remodeling is a long

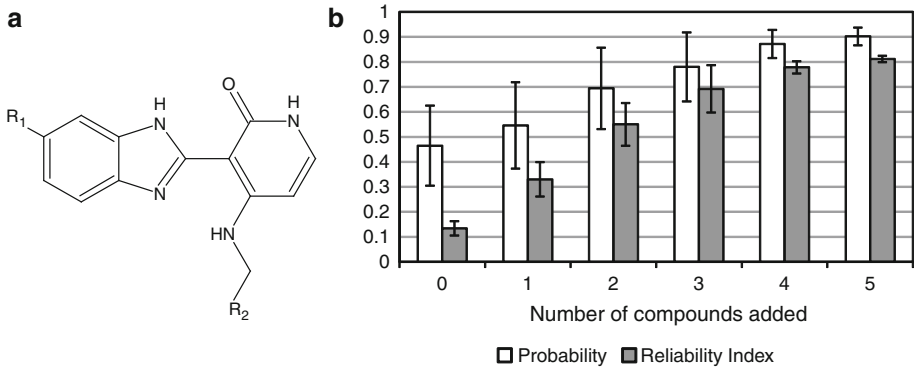


Fig. 7 Training of the GALAS model with compounds belonging to a novel chemical class: **a** general scaffold of compounds (R_1—aliphatic heterocycle, R_2—aromatic substituent); **b** increase in average predicted probability and RI values (*error bars* indicate standard deviation)

process, involving a more in-depth analysis of data, calculation of new descriptors, etc. When new experimental data are obtained for only a few compounds belonging to a novel chemical class, it is usually desirable to avoid these time consuming procedures.

We have decided to test the trainability feature with recently published inhibitors of insulin-like growth factor-1 receptor (IGF-1R) as a model of such a novel drug class [34]. The common scaffold of these compounds is not present among the molecules in the Literature and PubChem datasets and is shown in Fig. 7a. According to the classification rules used in this article, all 10 published IGF-1R inhibitors are CYP3A4 inhibitors having $IC_{50} < 40$ μM. Five randomly selected compounds were added one by one to the similarity correction library of the model initially containing literature dataset. The changes of average predicted probability and average Reliability Index values for the remaining five compounds are shown in Fig. 7b.

The original model (0 compounds added) predicts three compounds as non-inhibitors with probabilities ranging from 0.24 to 0.39, and two inconclusively with probabilities around 0.5. Average probability is 0.46 for all five compounds. The RIs are low for all predictions with average of 0.13 (values range from 0.09 to 0.16), indicating that novel IGF-1R inhibitors fall outside the applicability domain of original CYP3A4 inhibition model. As expected, after adding similar compounds to the training library both predicted probabilities and RIs increase, indicating successful model training. In this case adding three compounds is enough to adapt the model to the novel class of IGF-1R inhibitors as calculated probabilities to inhibit CYP3A4 and corresponding RI values become higher than 0.5 for all test compounds.

Recently, the importance of the applicability domain evaluation in QSAR modeling was illustrated in the study involving models for the prediction of plasma protein binding and CYP3A4 inhibition generated using *GlaxoSmithKline* in-house data [17]. Using internal test set the average error of prediction of CYP3A4 inhibition potency (expressed as IC_{50}) was close to experimental variability. Switching to the external validation set significantly increased the error which kept steadily growing following each subsequent expansion of this set with newly available data measured after the model was built. The bigger was the time gap between model development and new CYP3A4 inhibition measurements, the larger prediction errors were observed [17]. Such time dependence is a direct consequence of structural diversity changes in the in-house databases of experimentally tested compounds. Training the model using above described methodology would be useful in keeping the models up to date and adjusting them to the parts of chemical space of researcher's interest. While full revisions of QSAR models are inevitable from time to time having a substantial amount of new experimental data, the proposed training procedure of the GALAS model can be used on a daily basis.

Conclusions

A structure–activity relationship model of CYP3A4 inhibition has been developed using the novel GALAS modeling method. It exhibits good agreement between experimental and predicted values and can be useful in the identification of compounds possessing the risk to cause drug-drug interactions due to the inhibition of CYP3A4. The trainability feature of this model enables adjusting predictions to the needs of any particular project in pharmaceutical industry.

In silico tools of this type allow rapid screening of virtual libraries prior to their actual synthesis. In addition to calculating the probability for a compound to be a CYP3A4 inhibitor the described methodology also provides the reliability of each prediction. Among other applications this feature enables compound prioritization before experimental testing. Compounds predicted with low reliability, even after model training procedure involving proprietary databases, naturally become the first candidates for experimental studies on interaction with CYP3A4. Other

compounds with estimated low probability of inhibition and high prediction reliability can be treated as safe regarding CYP3A4 inhibition.

The reported GALAS model is a valuable tool in drug discovery projects as it allows consideration of possible drug-drug interactions caused by CYP3A4 inhibition in the earliest stages of drug discovery. The obtained results show good promise for the application of the same methodology in the development of analogous predictive models for inhibition of other human enzymes belonging to cytochrome P450 family (CYP1A2, CYP2C9, CYP2C19, CYP2D6) using PubChem screening and public literature data.

References

1. Wienkers LC, Heath TG (2005) Nat Rev Drug Discov 4:825–833
2. Rendic S, Di Carlo FJ (1997) Drug Metab Rev 29:413–580
3. Kerns E, Di L (2008) Drug-like properties: concepts, structure design and methods: from ADME to toxicity optimization. Academic Press, San Diego
4. Miller VP, Stresser DM, Blanchard AP, Turner S, Crespi CL (2000) Ann NY Acad Sci 919:26–32
5. Cali JJ, Ma D, Sobol M, Simpson DJ, Frackman S, Good TD, Daily WJ, Liu D (2006) Expert Opin Drug Metab Toxicol 2:629–645
6. Zlokarnik G, Grootenhuis PDJ, Watson JB (2005) Drug Discov Today 10:1443–1450
7. Walsky RL, Obach RS (2004) Drug Metab Dispos 32:647–660
8. Zuegge J, Fechner U, Roche O, Parrott N, Engkvist O, Schneider G (2002) Quant Struct Act Relat 21:249–256
9. Ekins S, Berbaum J, Harrison RK (2003) Drug Metab Dispos 31:1077–1080
10. Kriegl JM, Arnhold T, Beck B, Fox T (2005) J Comput Aided Mol Des 19:189–201
11. Kriegl JM, Eriksson L, Arnhold T, Beck B, Johansson E, Fox T (2005) Eur J Pharm Sci 24:451–463
12. Arimoto R, Prasad M, Gifford EM (2005) J Biomol Screen 10:197–205
13. Mao B, Gozalbes R, Barbosa F, Migeon J, Merrick S, Kamm K, Wong E, Costales C, Shi W, Wu C, Froloff N (2006) J Chem Inf Model 46:2125–2134
14. Jensen BF, Vind C, Padkjaer SB, Brockhoff PB, Refsgaard HHF (2007) J Med Chem 50:501–511
15. Gleeson MP, Davis AM, Chohan KK, Paine SW, Boyer S, Gavaghan CL, Arnby CH, Kankkonen C, Albertson N (2007) J Comput Aided Mol Des 21:559–573
16. Choi I, Kim SY, Kim H, Kang NS, Bae MA, Yoo S, Jung J, No KT (2009) Eur J Med Chem 44:2354–2360
17. Weaver S, Gleeson MP (2008) J Mol Graph Model 26:1315–1326
18. Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA (1999) J Pharmacol Exp Ther 290:429–438
19. Wang RW, Newton DJ, Liu N, Atkins WM, Lu AY (2000) Drug Metab Dispos 28:360–366
20. Lu P, Lin Y, Rodrigues AD, Rushmore TH, Baillie TA, Shou M (2001) Drug Metab Dispos 29:1473–1479
21. Ekroos M, Sjögren T (2006) Proc Natl Acad Sci USA 103:13682–13687
22. Galetin A, Clarke SE, Houston JB (2003) Drug Metab Dispos 31:1108–1116
23. Worth AP, Hartung T, Van Leeuwen CJ (2004) SAR QSAR Environ Res 15:345–358
24. Sazonovas A, Japertas P, Didziapetris R (2010) SAR QSAR Environ Res 21:127–148
25. Japertas P, Sazonovas A, Didziapetris R, Petrauskas A (2008) The 235th ACS National Meeting, New Orleans, LA
26. Stresser DM, Blanchard AP, Turner SD, Erve JC, Dandeneau AA, Miller VP, Crespi CL (2000) Drug Metab Dispos 28:1440–1448
27. Nomeir AA, Ruegg C, Shoemaker M, Favreau LV, Palamanda JR, Silber P, Lin CC (2001) Drug Metab Dispos 29:748–753
28. The PubChem Project. http://pubchem.ncbi.nlm.nih.gov/. Accessed 10 Jun 2008
29. Tetko IV (2002) J Chem Inf Comput Sci 42:717–728
30. Tetko IV (2002) Neur Proc Lett 16:187–199
31. Efron B (1979) Ann Statist 7:1–26
32. Katoh M, Nakajima M, Shimada N, Yamazaki H, Yokoi T (2000) Eur J Clin Pharmacol 55:843–852
33. Fawcett T (2006) Patt Recog Lett 27:861–874
34. Velaparthi U, Wittman M, Liu P, Carboni JM, Lee FY, Attar R, Balimane P, Clarke W, Sinz MW, Hurlburt W, Patel K, Discenza L, Kim S, Gottardis M, Greer A, Li A, Saulnier M, Yang Z, Zimmermann K, Trainor G, Vyas D (2008) J Med Chem 51:5897–5900
35. Algorithm Builder. ACD/Labs Inc., Toronto, ON, Canada. http://www.acdlabs.com
36. Japertas P, Didziapetris R, Petrauskas A (2002) Quant Struct Act Relat 21:23–37
37. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.r-project.org/
38. Eng J (2006) ROC analysis: web-based calculator for ROC curves. http://www.jrocfit.org. Accessed 19 Jul 2010
39. ADME Suite. ACD/Labs Inc., Toronto, ON, Canada. http://www.acdlabs.com
40. Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD, Johnson EF (2004) J Biol Chem 279:38091–38094
41. Korzekwa KR, Krishnamachary N, Shou M, Ogai A, Parise RA, Rettie AE, Gonzalez FJ, Tracy TS (1998) Biochemistry 37:4137–4147
42. Gleeson MP (2008) J Med Chem 51:817–834
43. Monostory K, Vereczkey L, Lévai F, Szatmári I (1998) Br J Pharmacol 123:605–610
44. Moon Y, Kim SY, Ji HY, Kim YK, Chae HJ, Chae SW, Lee HS (2007) Xenobiotica 37:246–259
45. Gleeson P, Bravi G, Modi S, Lowe D (2009) Bioorg Med Chem 17:5906–5919
46. Ishigami M, Honda T, Takasaki W, Ikeda T, Komai T, Ito K, Sugiyama Y (2001) Drug Metab Dispos 29:282–288
47. Tsukamoto I, Koshio H, Kuramochi T, Saitoh C, Yanai-Inamura H, Kitada-Nozawa C, Yamamoto E, Yatsu T, Shimada Y, Sakamoto S, Tsukamoto S (2009) Bioorg Med Chem 17:3130–3141
48. Ortiz de Montellano PR (1995) In: Cytochrome P450: structure, mechanism and biochemistry. Plenum Press, New York
49. Testa B (1995) The metabolism of drugs and other xenobiotics: biochemistry of redox reactions. Academic Press, San Diego
50. Zhou S, Yung Chan S, Cher Goh B, Chan E, Duan W, Huang M, McLeod HL (2005) Clin Pharmacokinet 44:279–304