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

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 isofoms (3A4, 2D6, 1A2, 2C9, and 2C19). These models provide the group with the highest likelihood of inhibition will include a certain CYP450 isoform with IC50 < 50 µM. General "inhibition" models estimate whether the compound will exhibit any inhibition at all (IC50 < 50 µM), while "Efficient inhibition" models predict the probability that the compound will inhibit a selected enzyme with clinically significant IC50 < 10 µM.

DATA SET

Datasets of up to 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.

MODEL DEVELOPMENT

CYP450 inhibition models have been derived using a novel GALAS (Global Adjusltely Localized According to Similarity) 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 function)
• 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 achieved by introducing 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 complexes with high reliability predictions (RI>0.5), accuracy exceeds 95%.

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 an IC50 < 10 µM threshold were used in this evaluation.

REFERENCES