Using docking-derived protein-ligand atom pair descriptors to increase performance of QSAR models for human CYP450 inhibition

Sergii Novotarskyi[†‡], Iurii Sushko[†‡], Robert Koerner[†‡] and Igor Tetko[†‡]

[†]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health (GmbH), Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany
[‡]eADMET GmbH, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany

Cytochromes P450 (CYP450) are a superfamily of enzymes, involved in metabolism of a large number of xenobiotic compounds.[1] CYP450 are involved in metabolism of a large amount of drugs, currently present on the market.[2] Individual CYP enzymes in families 1, 2 and 3 metabolize xenobiotics, including the majority of small molecule drugs currently in use. The distinctive feature of CYP450 enzymes is broad and overlapping substrate specificity.[3] Therefore, prediction of CYP450 inhibition activity of small molecules poses an important task due to high risk of drug-drug interactions.

It is estimated that over 75% of currently marketed drugs are metabolized by CYP450. Of these reactions over 90% are facilitated by CYP1A2, CYP2C9, CYP2D6 and CYP3A4.[4] This makes these enzymes particularly interesting targets for in-silico inhibition prediction.

In this work the quality of novel docking-derived protein-ligand atom pair descriptors is benchmarked in QSAR modeling of HTS data for human CYP450 inhibition. The calculation of descriptors involves a flexible docking of the molecule to the rigid binding cite of the cytochrome (in this study the AutoDock Vina tool was used). The obtained top-ranked conformation is then processed to obtain the descriptors.

The training sets for the benchmarked models were obtained from PubChem BioAssay database. The identifiers of PubChem BioAssays with training data for this study are AID410 for CYP1A2, AID883 for CYP2C9, AID884 for CYP3A4, and AID891 for CYP2D6. These assays include the molecules from Molecular Libraries Small Molecule Repository (MLSMR). The test sets are obtained from the AID1851 assay by excluding all molecules present in the training set. They include molecules from biofocused libraries (including FDA-approved drugs), combinatorial chemistry libraries and libraries of purified natural products or related structures.[5] This makes the test sets structurally diverse and different to the training sets.

The models presented in the study achieved 82 - 87% of correctly classified compounds on the validated training set and 65-75% of correctly classified instances on the test sets. The use of docking-derived atop pair descriptors allowed a statistically significant increase in model performance. The dramatic decrease in model performance on the test sett compared to the validated training sets can be explained b structural dissimilarity of the training and test sets. The use of applicability domain approaches to select only confident predictions allowed to achieve the accuracy of 90% of correctly classified instances on the subset of 20% most confident predictions.

The datasets and the benchmarked models are available on the Online Chemical Modeling Environment (http://ochem.eu)[6]

(1) Rendic, S. Summary of information on human CYP enzymes: human P450 metabolism data *Drug Metab. Rev.* **2002**, *34*, 83-448.
(2) Masimirembwa, C. M. Thompson, R.; Andersson, T. B. In vitro high throughput screening of compounds for favorable metabolic properties in drug discovery *Comb. Chem. High T. Scr.* **2001**, *4*, 245-263.
(3) Pelkonen, O. Turpeinen, M. Hakkola, J. Honkakoski, P. Hukkanen, J.; Raunio, H. Inhibition and induction of human cytochrome P450 enzymes: current status *Arch. Toxicol.* **2008**, *82*, 667-715.
(4) Guengerich, F. P. Cytochrome p450 and chemical toxicology *Chem. Res. Toxicol* **2008**, *21*, 70-83.
(5) 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*, 1050-1055.
(6) Sushko et al Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information *J. Comput. Aided Mol. Des* **2011**, *25*, 533-554.