A. Khakimova [1]
R. Nugmanov [1]
T. Madzhidov [1]
A. Varnek [2]

PREDICTION OF METABOLIC TRANSFORMATIONS OF ORGANIC COMPOUNDS BY CYP1A2 ISOENZYME USING CONDENSED GRAPH OF REACTIONS

[1]A.M. Butlerov Insitute of Chemistry, Kazan Federal University, Russia
[2]University of Strasbourg, Strasbourg, France
aigul03.14@gmail.com

Analysis of metabolites is one of important stages of pharmacokinetics investigations of a drug molecule. Bad ADME profile, including undesired metabolism, could lead to rejection of drug design project on late stages, causing substancial financial and time costs. Thus, taking into account enourmous costs of drug discovery process, computational prediction of metabolic transformations of a given molecule is becoming more and more important.

The goal of the present study was to predict the metabolic transformations of organic compounds by the cytochrome P450 1A2 isoenzyme. During this research, it was suggested to consider metabolic transformation as a chemical reaction. The latter can be represented as a Condensed Graph of Reaction [1] (CGR), for which fragment descriptors can be calculated. Classification model that distinguish "real" and "unreal" metabolyc transformations could be built using resulting descriptor vector.

Three datasets were collected: (1) data that were used to create the XenoSite model [2], hereafter called XS, (2) the data from the Metabolite Database [3], MDB, and (3) the joint set of XS and MDB data, hereafter called ALL. Thus, three models were obtained using these datasets. For external validation, a set from the MDB database was assembled. With the use of the ChemAxon Standardizer program, the tautomeric forms of the compound were converted to a canonical form, functional groups were standardized, aromatization was carried out, and explicitly defined hydrogen atoms were removed.

Aromatic hydroxylation reactions were generated for the database molecules using ChemAxon Reactor program. Then they were divided into "real" and "unreal" transformations. The reactions were encoded in CGR and the reaction descriptors were calculated using the ISIDA Fragmenter program [1]. SVM [4] with Tanimoto kernel was used as machine learning tool. A five-fold cross-validation (CV) procedure was carried out. To verify the quality of the model, the IAP [5] metric was used. The results are presented in the Table below.

	CV	TEST
MDB	0.82	0.88
XS	0.68	0.68
ALL	0.85	0.82
CON		0.91

Results in Table show that quite robust models were obtained, that was supported by external test set as well. Consensus model where probability of oxidation is calculated as geometric mean over MDB, XS, and ALL model-based predictions, shows the best IAP value. This approach has superior predictive power in comparison with other open-access tools, like XenoSite server (IAP on the test set is 0.83) and Way2Drug (IAP on the test set 0.84).

1. Varnek A. et al. *J Comput Aided Mol Des*, *2005*, **19:** 693.
2. Zaretzki J. et al. *J. Chem. Inf. Model*, 2013, **53:** 3373.
3. Accelrys, Inc., Accelrys Metabolite, San Diego, 2009 (http://accelrys.com).
4. Chang C.-C. et al. *LIBSVM : a library for support vector machines*, 2001.
5. Filimonov D. et al. *in Varnek A. et al. Chemoinformatics Approaches to Virtual Screening*, 2008, RSC Publishing: 182-216.

The research was supported by Russian Science Foundation, grant 14-43-00024