Optimization of compounds ADMET properties *via* machine-learning-based tools

Sabina Podlewska[1,2]*, Rafał Kafel[1]

[1]Maj Institute of Pharmacology Polish Academy of Sciences, 12 Smętna Street, 31-343 Kraków, Poland

[2] Department of Technology and Biotechnology of Drugs, Jagiellonian University, Medical College, Medyczna Street 9, 30-688 Cracow, Poland

*e-mail: smusz@if-pan.krakow.pl

Background

The complex process of development of new potential drugs is based not only on the provision of the activity towards the desired set of receptors, but the compounds should also possess favourable physicochemical and pharmacokinetic properties, metabolic stability and a lack of toxicity – the parameters that disqualify compounds from further consideration despite a preferential activity profile. Currently, a number of approaches for prediction of physicochemical and ADMET properties are available. They are mostly ligand-based tools and two classes of models are constructed – classification ones (mutagenesis/non-mutagenesis, stable/unstable, soluble/insoluble, etc.) or regression tools are applied and QSAR-type models are formed, in which quantitative impact of particular structural moieties on considered
parameters is examined.

Aim of the study

The aim of the project is composed of two main parts: construction of a ML-based tool for evaluation of physicochemical and ADMET properties of compounds and development of methodology for optimization of chemical structures (in terms of physicochemical and ADMET properties) on the basis of the constructed predictive models (Figure 1). Here, we focus on the results obtained during the optimization of metabolic stability.

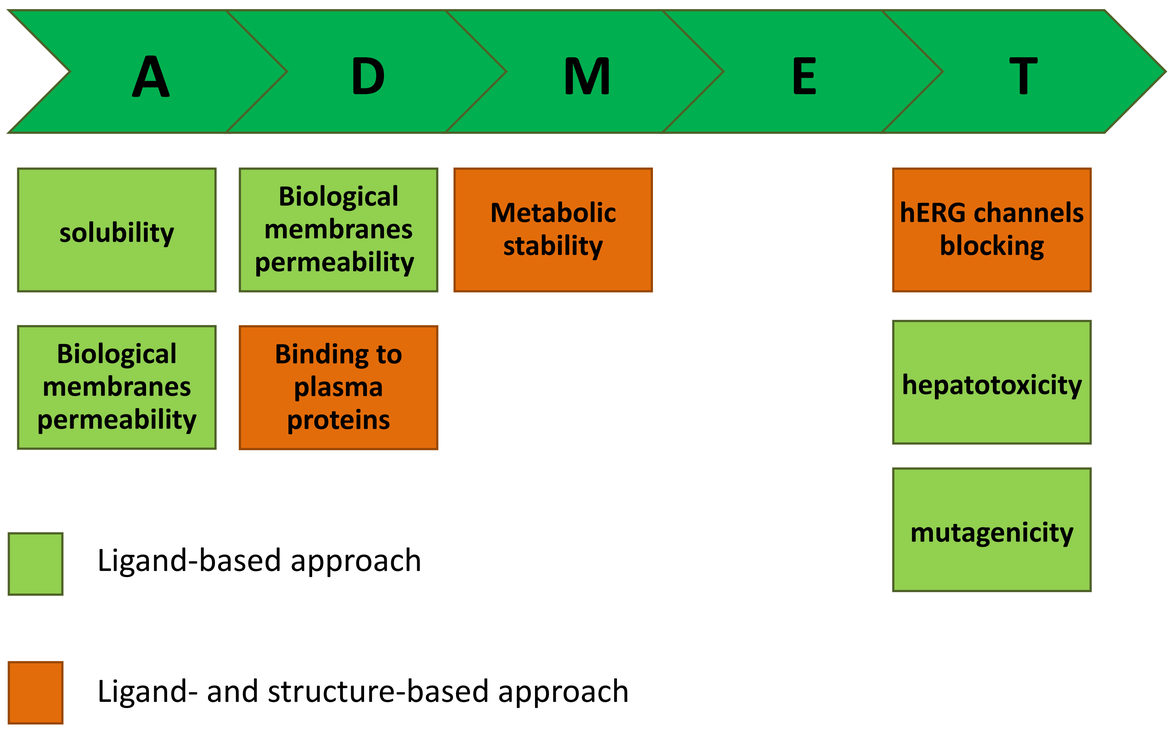


Figure 1. Properties evaluated via the constructed tool

Existing tools

Existing tools for ADMET properties predictions (both commercial and freely available) are gathered in Table 1. Due to the high number of tools of such a type, only the most popular examples are provided. They are mostly ligand-based tools, and two classes of models are constructed – classification ones (mutagenesic/non-mutagenesic, stable/unstable, soluble/insoluble, etc.) or regression tools. The application of the tools of the latter type is connected with the formation of the QSAR-type models, in which the quantitative impact of particular structural moieties on considered parameters is examined. Many comprehensive software packages for ADMET properties evaluation are available, such as ADMET Predictor, CASE ULTRA, DEREK, META-PC, METEOR, ONCOLOGIC, PASS, TOPKAT, and VIRTUALTOXLAB. Moreover, the initial characteristics of physicochemical and pharmacokinetic properties are offered in most packages of software for molecular modeling, such as QikProp in the Schrödinger Suite, Molecular Descriptors in the MOE, or ADMET and Predictive Toxicology from the BIOVIA Discovery Studio. A number of individual ADMET properties can also be evaluated via various online servers, such as ALOGPS, Molinspiration, PreADMET, MetaPrint2D, MetaPred or Pred-hERG.

Methods

The data for the construction of the tool for metabolic stability predictions were collected from the ChEMBL database. All records with the T1/2 parameter reported were downloaded, and separate sets referring to human, rat and mouse experiments were prepared (standard deviation of data obtained for human, mouse, and rat data for justification of preparation of separate models is presented in Figure 2). The compounds were represented with the use of the 1- and 2-dimensional PaDEL-Descriptors (1d2d descriptors) and Extended Fingerprint (ExtFP) from the same software package.

The constructed tool predicts the numerical value of metabolic stability with the predictive model based on the application of the two types of machine learning algorithms. The first one, SMOreg which is a modification of the very popular and efficient algorithm Support Vector Machine (SVM) into Sequential Minimal Optimization (SMO) and adjusted for performing regression tasks and two classification algorithms – SMO and Random Forest. However, in order to enable easier interpretability of the outcome of regression experiments, compounds are also divided into three classes according to metabolic stability values – low, medium, and high – and the results are colored accordingly. For each of the analyzed structures, the ten most similar compounds from the training set (Tanimoto metric, topological fingerprint from RDKit package) are found and provided in separate files for manual inspection (the particular chemical structure is provided only once and the median half-lifetime value is given). Structures for analysis online can be submitted as a sdf file or drawn using the MarvinJS plugin.

Table 1. A summary of some of the available tools for ADMET properties predictions.

Package name	Link	Availability	Description
ADMET Predictor	http://www.simulations-plus.com/	commercial software	Comprehensive characteristic of physicochemical and ADMET properties of compounds, including cancerogenity, mutagenicity, overall toxicity and possibility of interactions with 5 selected CYP isoforms
CASE ULTRA	http://www.multicase.com/case-ultra-models	commercial software	A set of statistical and expert tools for evaluation of compounds toxicity
DEREK	http://www.lhasalimited.org	commercial software	Expert system for predicting toxicity of compounds, including cancerogenity, mutagenicity, genotoxicity, teratogenicity, influence on fertility, irritating influence on skin or allergic effect
META-PC	http://www.multicase.com/meta-pc	commercial software	Expert system for predicting products of compounds metabolism
METEOR	http://www.lhasalimited.org	commercial software	Expert system for predicting metabolic transformations
ONCOLOGIC	http://www2.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals	free software	Predicting of cancerogenicity of compounds
PASS	http://www.pharmaexpert.ru	commercial software (simplified version is freely available online)	Qualitative evaluation of above 3500 properties, including mechanisms of action, side and toxic effects, interaction with various enzymes and transport proteins, influence on genes expression
TOPKAT	http://accelrys.com/products/collaborative-science/biovia-discovery-studio/qsar-admet-and-predictive-toxicology.html	commercial software	Mutagenicity, cancerogenity, irritating action on skin, eyes, etc.
VIRTUALTOXLAB	http://www.biograf.ch/	commercial software (free license for non-commercial use)	Evaluation of compounds toxicity and possible side effects, uses docking and multi-dimensional QSAR analysis

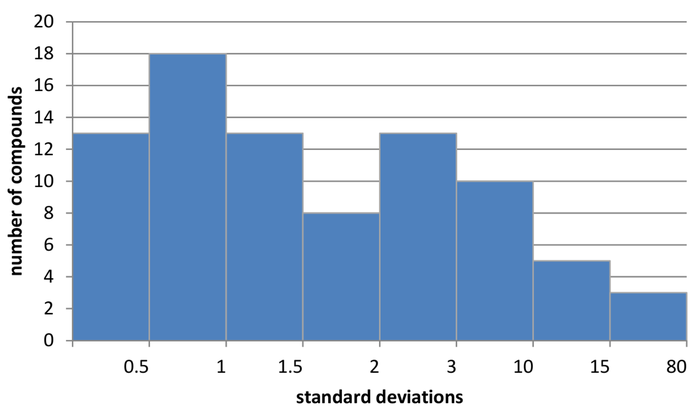
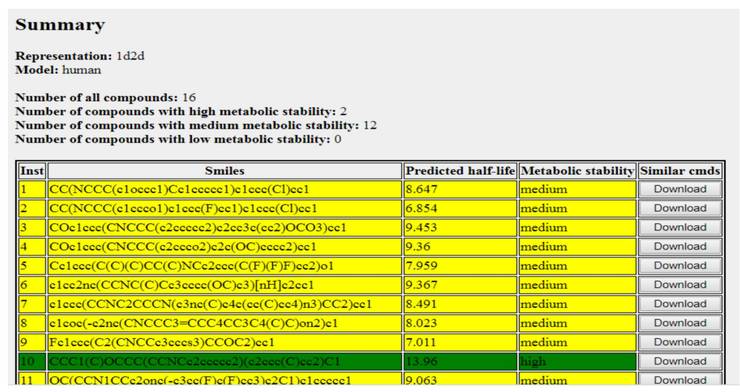


Figure 2. Standard deviation values of half-lifetimes between human, rat and mouse data

Example output

An example output obtained by the constructed tool is presented in Figure 3. The summary describes the compound representation used, the predicted value of half-lifetime and the metabolic stability class to which a compound was assigned. Additionally, in order to perform a more detailed analysis, the ten most similar compounds from the training set (in terms of Tanimoto metric-based similarity) can be downloaded for each of the analyzed structures.



Summary

Representation: 1d2d
Model: human

Number of all compounds: 16
Number of compounds with high metabolic stability: 2
Number of compounds with medium metabolic stability: 12
Number of compounds with low metabolic stability: 0

Inst	Smiles	Predicted half-life	Metabolic stability	Similar cmds
1	CC(NCCC(c1ccec1)Cc1ccccc1)c1ccc(Cl)cc1	8.647	medium	Download
2	CC(NCCC(c1ccco1)c1ccc(F)cc1)c1ccc(Cl)cc1	6.854	medium	Download
3	COc1ccc(CNCCC(c2ccccc2)c2cc3c(cc2)OCO3)cc1	9.453	medium	Download
4	COc1ccc(CNCCC(c2ccco2)c2c(OC)cccc2)cc1	9.36	medium	Download
5	Cc1ccc(C(C)(C)CC(C)NCc2ccc(C(F)(F)F)cc2)o1	7.959	medium	Download
6	c1cc2nc(CCNC(C)Cc3cccc(OC)c3)[nH]c2cc1	9.367	medium	Download
7	c1ccc(CCNC2CCCN(c3nc(C)c4c(cc(C)cc4)n3)CC2)cc1	8.491	medium	Download
8	c1coc(-c2nc(CNCCC3=CCC4CC3C4(C)C)on2)c1	8.023	medium	Download
9	Fc1ccc(C2(CNCCc3cccs3)CCOC2)cc1	7.011	medium	Download
10	CCC1(C)OCCC(CCNCc2ccccc2)(c2ccc(C)cc2)C1	13.96	high	Download
11	OC(CCN1CCc2onc(-c3cc(F)c(F)cc3)c2C1)c1ccccc1	9.063	medium	Download

Figure 3. Example output.

Acknowledgments

The study was supported by the grant OPUS 2018/31/B/NZ2/00165 financed by the National Science Centre, Poland (www.ncn.gov.pl).

6th International Electronic Conference on Medicinal Chemistry
1-30 November 2020
sponsored: MDPI pharmaceuticals