

Optimization of pharmacokinetic compound profile of serotonin receptor ligands via machine learning

Sabina Podlewska^{1,2*}, Rafał Kafel¹

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

²Department 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

At the beginning of the process of drug design and development, the main emphasis is put on the provision of compound activity towards desired panel of targets. However, at the same time, or in the subsequent stages, the compound needs to be adequately profiled in terms of its physicochemistry and ADMET parameters.

Aim of the study

The aim of this project is to construct a machine-learning-based tool for the evaluation and optimization of compound physicochemical and pharmacokinetic properties.

Results

The parameters, which undergo evaluation and optimization via the developed tool is presented in Figure 1.

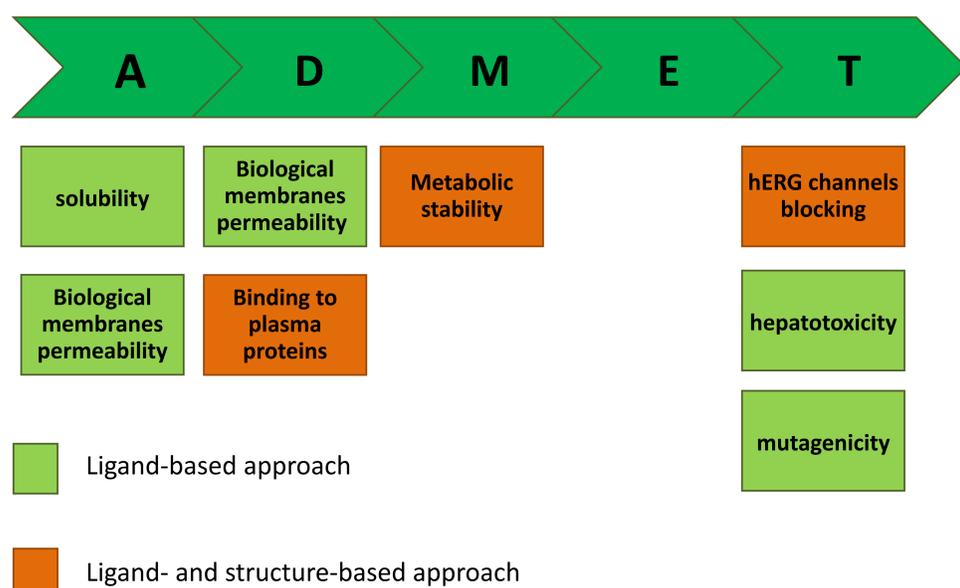


Figure 1. Properties evaluated via the constructed tool

Results

The optimization process was started from the focus on metabolic stability. The data for the construction of the tool for metabolic stability predictions were collected from the ChEMBL database. All records with the T_{1/2} and clearance parameters 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). Respective workflow is presented in Figure 2.

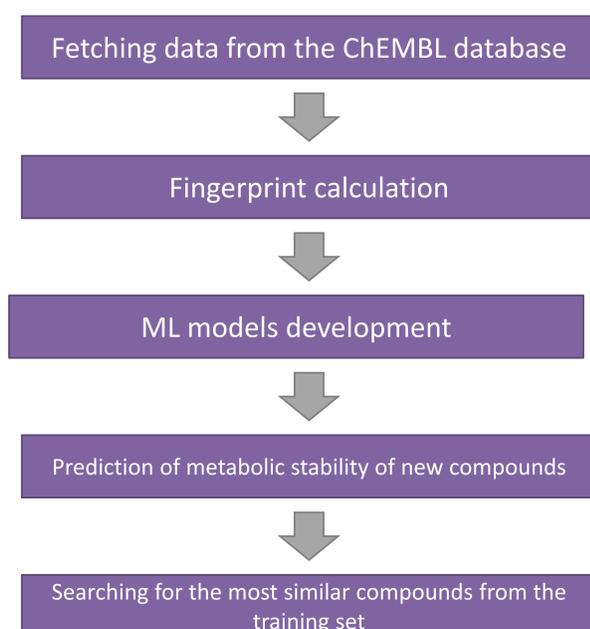


Figure 2. Scheme of the developed protocol for optimization of compound metabolic stability

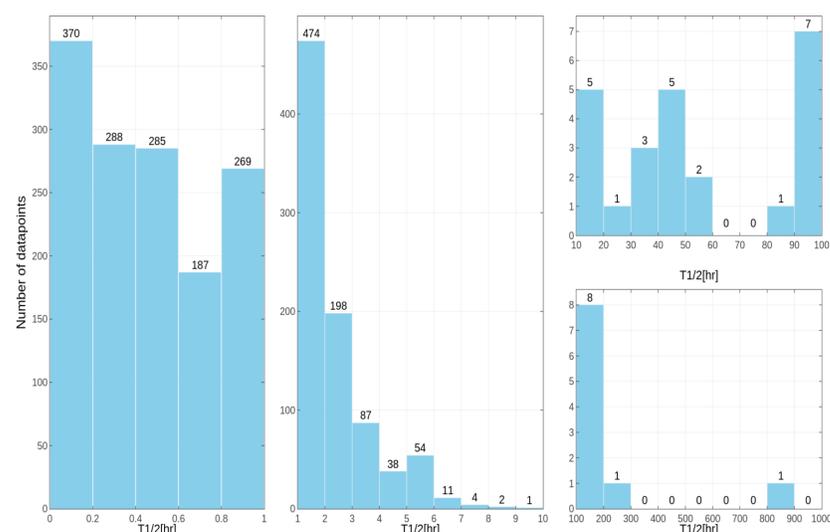


Figure 3. Distribution of compound half-lifetimes in the constructed datasets referring to experiments performed on human samples. For better visualization, the dataset was divided into several parts.

The scheme of the prepared ML models is presented in Figure 4. In the classification studies, we divided the data into three stability classes using the following thresholds for half-lifetime values:

- ≤ 0.6 – low
- $(0.6 - 2.32)$ – medium
- > 2.32 – high

In general, the ML-based predictions of metabolic stability were accurate, and values of different evaluating parameters obtained in 10-fold cross-validation studies are presented in Figure 5.

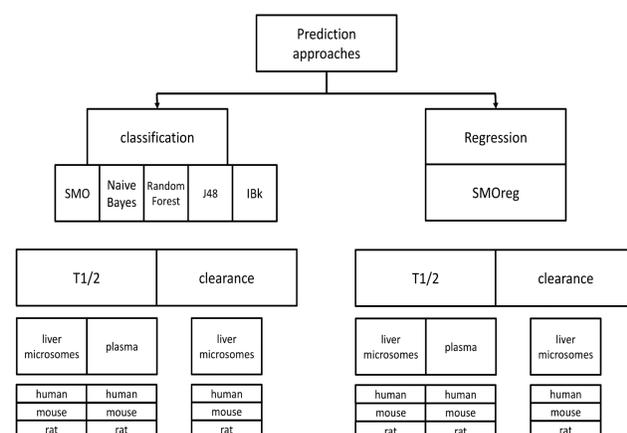


Figure 4. Prediction approaches covered when performing evaluation of metabolic stability

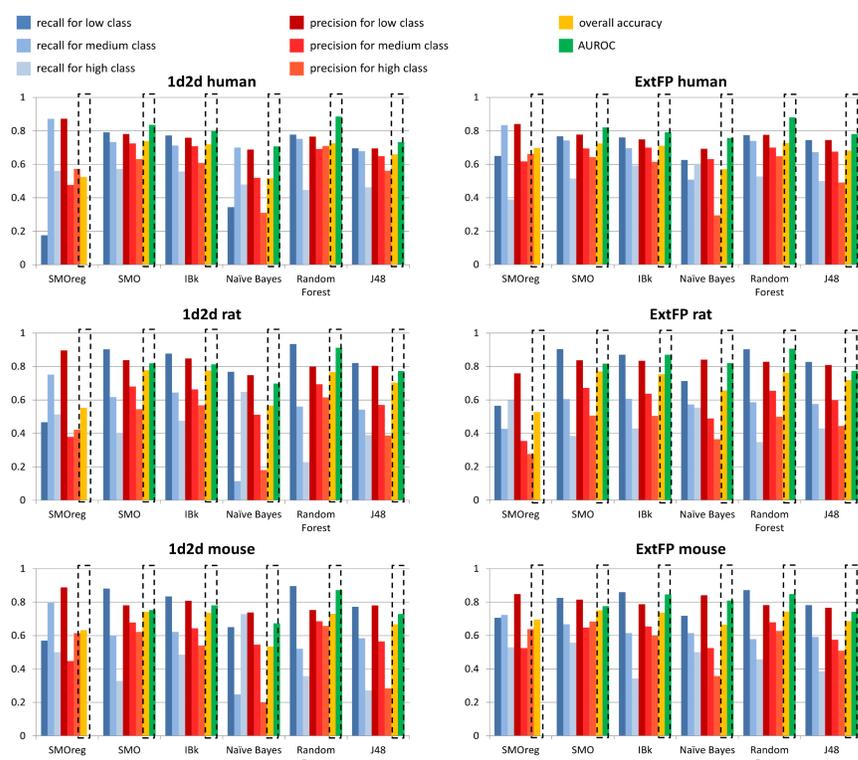


Figure 5. Evaluating parameters values obtained in the cross-validation studies.

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).