Optimization of ADMET properties – ligand- and structure-based approach

Rafał Kafel, Sabina Podlewska*

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

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

INTRODUCTION

During the drug design process, usually main focus is put on the provision of the adequate activity of a compound towards considered set of targets. However, it is insufficient, as despite possessing desired affinity profile, a compound might not be further considered in the drug discovery pipeline, due to unfavorable ADMET properties.

In the study, we develop a platform for comprehensive evaluation of a compound in terms of ADMET features. We consider solubility, metabolic stability, biological membranes permeability, hERG channels blocking, and mutagenicity. For all evaluated features, ligand-based models were developed (with the use of machine learning algorithms and key-based fingerprints for compound representation). In addition, metabolic stability, and hERG channels blocking can also be evaluated in the structure-based mode. It involves docking to the respective proteins (8 CYP subtypes in the case of metabolic stability), representation of obtained ligand-protein complexes via the Structural Interaction Fingerprints and their automatic evaluation with the use of machine learning methods.

The parameters, which underwent evaluation are presented in Figure 1.

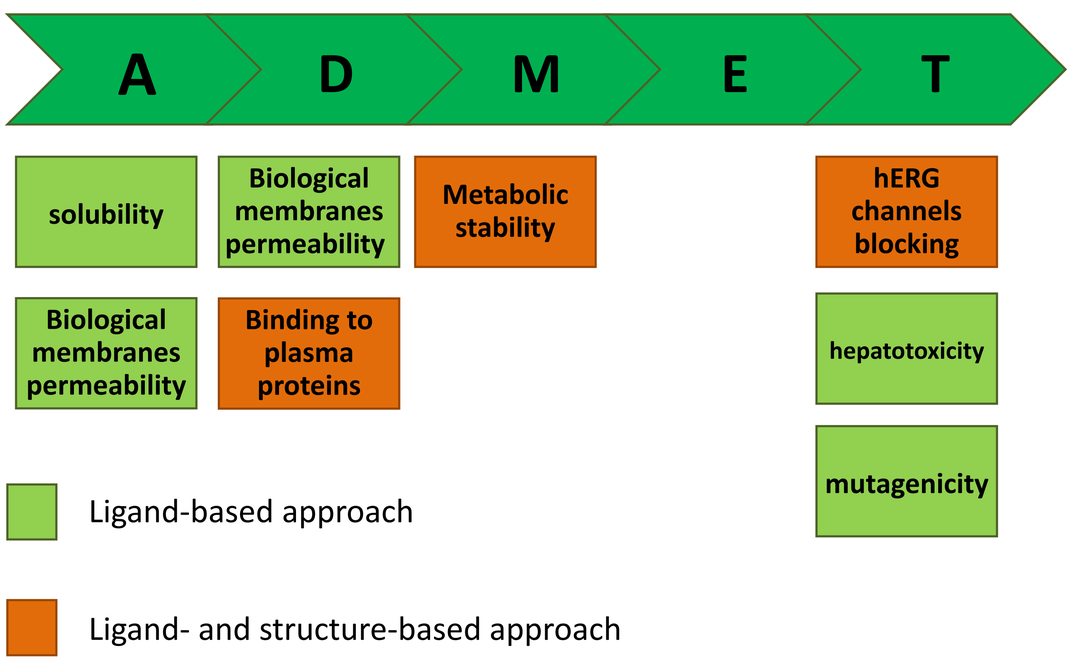

Figure 1. Properties evaluated via the constructed tool

SCHEME OF THE PROTOCOL

Scheme ot the protocol for the evaluation of the selected properties in the ligand-based approach is presented in Figure 2.

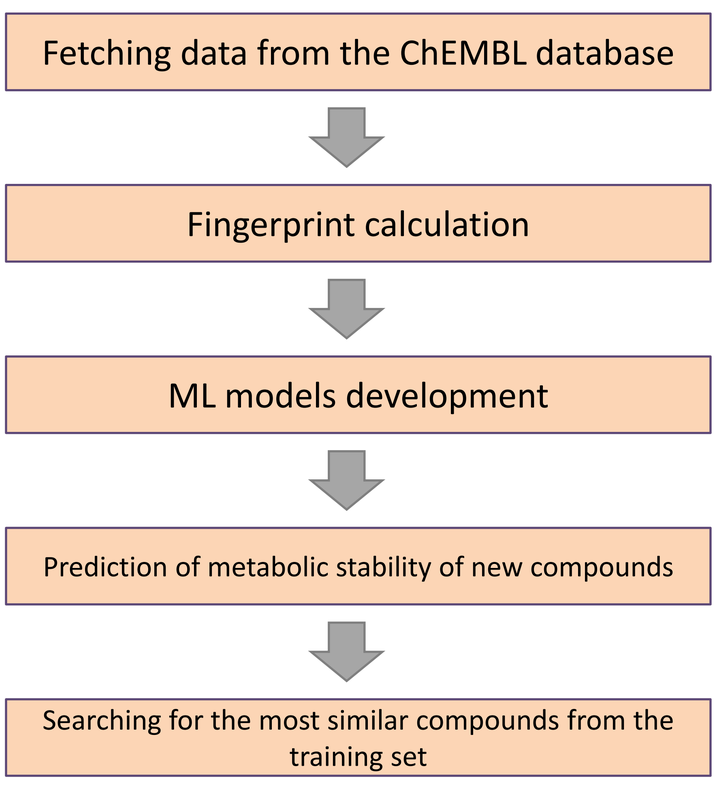

Figure 2. Scheme of the developed protocol for optimization of compound ADMET properties.

RESULTS

The results are presented on the example of the metabolic stability predictions. The data for the construction of the tool for metabolic stability predictions were collected from the ChEMBL database. All records with the T1/2 and clearance parameters reported were downloaded, and separate sets referring to human, rat and mouse experiments were prepared. 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 data distribution is presented in Figure 3. 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 (Figure 4).

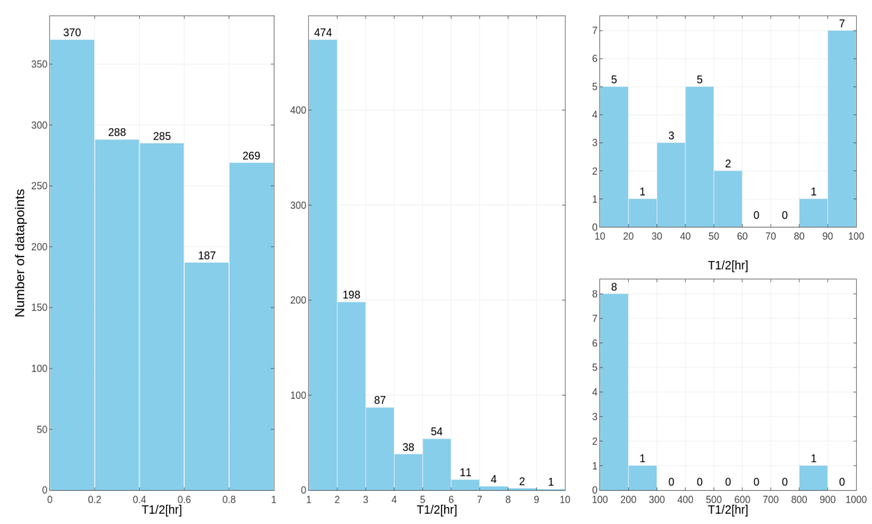

Figure 3. Distribution of compound half-lifetimes in the constructed datasets referring to experiments performed on human samples. For better visualization, the dataset was divided into several parts.

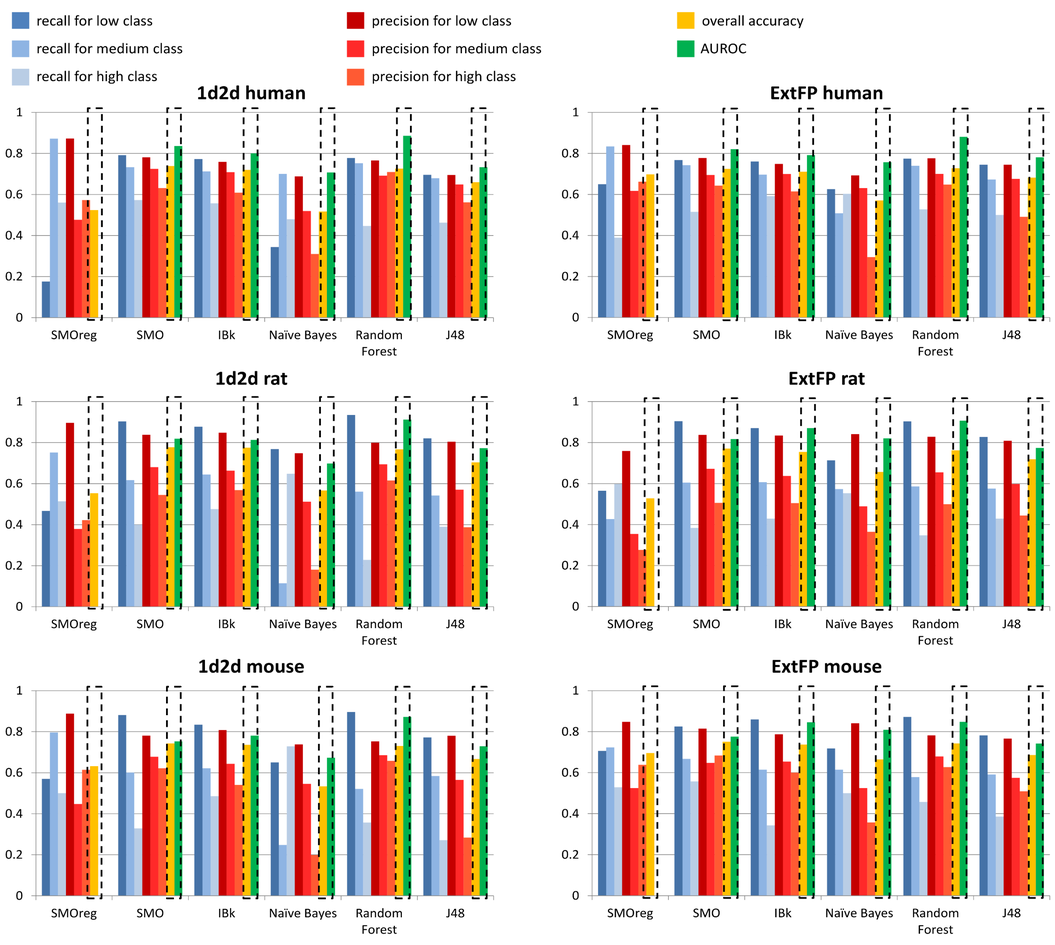

Figure 4. Evaluating parameters values obtained in the cross-validation studies.

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)