Aleksander Mendyk
Department of Pharmaceutical Technology and Biopharmaceutics Faculty of Pharmacy Jagiellonian University Medical College, Medyczna 9 Street, Kraków 30-688, Poland
e-mail: mfmendyk@cyf-kr.edu.pl
Barbara Wiśniowska
Unit of Pharmacoepidemiology and Pharmacoeconomics Faculty of Pharmacy Jagiellonian University Medical College, Medyczna 9 Street, Kraków 30-688, Poland
Miłosz Polak
Unit of Pharmacoepidemiology and Pharmacoeconomics Faculty of Pharmacy Jagiellonian University Medical College, Medyczna 9 Street, Kraków 30-688, Poland
Jakub Szlęk
Department of Pharmaceutical Technology and Biopharmaceutics Faculty of Pharmacy Jagiellonian University Medical College, Medyczna 9 Street, Kraków 30-688, Poland
Anna Glinka
Unit of Pharmacoepidemiology and Pharmacoeconomics Faculty of Pharmacy Jagiellonian University Medical College, Medyczna 9 Street, Kraków 30-688, Poland
Sebastian Polak
Unit of Pharmacoepidemiology and Pharmacoeconomics Faculty of Pharmacy Jagiellonian University Medical College, Medyczna 9 Street, Kraków 30-688, Poland

Artificial neural networks for carditoxicity prediction of drugs - practical considerations

Introduction Early toxicity prediction for potential drugs is considered as a necessary safety measure regarding recent withdrawals of many substances from the pharmaceutical market. The latter was substantially based on the identified cardiotoxicity related to the inhibition of the potassium channels encoded by hERG (the human ether-a-go-go related gene). Thus, the drugs affinity to hERG channels is considered now as one of the major screening factors for potentially dangerous substances. There are theories describing relationships between hERG channels blocking activity and chemical structure but they often lack of physiological/pathological factors and drug concentration influence. Thus, it is feasible to use empirical modeling to fill this gap. The aim of this work was to create predictive model for chemical substances affinity to hERG channels by means of artificial neural networks (ANNs).

Materials and methods Database used for the modeling purposes was recently published and is freely available from the CompTox project website (www.toxportal.net). Input data were derived from the published in vitro experiments. Inputs represented in vitro experiment settings, chemical descriptors of drugs and drug concentration. Output was simply percent of hERG channel inhibition (range 0 to 1). Final set contained 1969 records describing 200 drugs. Initial number of inputs was 109. Enhanced 10-fold cross validation (10-cv) was applied, where whole

drugs information was excluded from test sets. For external validation a test set of 193 records (25 substances) for drugs both previously present (different in vitro settings) and absent in the native dataset was used. Drugs chemical structures were drawn in MarvinSketch or downloaded from PubChem Compound database. The molecules were structurally optimized with use of molconvert command-line program included in Marvin Beans package. Resulting *.sdf files were the subject to descriptor calculations by cxcalc program with selected 41 plugins. The default parameters were used in both cxcalc and molconvert programs. Multi-layer perceptrons (MLPs) and neuro-fuzzy ANNs (NFs) were trained with use of back-propagation (BP) algorithm with momentum, delta-bar-delta and jog-of-weights modifications. Various activation functions were tested: hyperbolic tangent, logarithmic, logistic and linear. MLPs architectures were varied from 1 to 6 hidden layers and up to 200 nodes in each layer. For NFs of Mamdani (multiple input single output) MISO type only one layer was applied. Adjacent layers were fully interconnected. Sensitivity analysis was performed in order to reduce initial number of inputs to the crucial variables set by means of iterative algorithm with gradual inputs reduction and models predictive performance assessment. The latter was generalization error estimated by means of 10-cv with root mean squared error (RMSE) measure. Ensemble ANNs systems were applied and combined by simple average of their outputs in order to improve predictability of the model.

Results The input reduction procedure resulted in 39 parameters describing in vitro setting (8), drug physico-chemical properties (30), and concentration (1). The best ANNs architectures found were as follows: (1) ANN with 3 hidden layers with 15, 7 and 5 nodes in each one respectively and logistic activation function; 2) ANN with 2 hidden layers with 20 and 10 nodes. The resulting 10-cv RMSE was 0.22 with respect to the validation data set RMSE = 0.2. This result, although not satisfactory seems to be final with the available data representation. Future research will be devoted to the improvement of the model by enhancing input data by new factors/variables, if available.