Empirical Modeling of the Sodium Channel Inhibition Caused by Drugs

Aleksander Mendyk[1], Barbara Wiśniowska[1], Kamil Fijorek[2], Anna Glinka[1], Miłosz Polak[1], Jakub Szlęk[1], Sebastian Polak[1]

[1]Jagiellonian University Medical College, Kraków, Poland
[2]Cracow University of Economics, Kraków, Poland

Abstract

The aim of this work was to create extended QSAR model of the relationship between sodium channel blocking activity of the particular compound and its chemical structure together with the in vitro assay conditions. Artificial neural networks (ANNs) were chosen as modeling tools. Chemoinformatics software was used for calculation of the molecular descriptors describing the structure of the interest. Drug concentration causing 50% of the channel inhibition (IC_{50}) was used as the modeling endpoint. The data was based on the literature search and consisted of 38 drugs and 108 records. Initial number of inputs was 110 and during the sensitivity analysis was reduced to 20. ANNs models were optimized in the extended 10-fold cross-validation scheme yielding RMSE = 0.68, NRMSE = 20.7% and R^2 = 0.35. Best models were ANNs ensembles combining three ANNs with their outputs averaged as a collective output of the system.

1. Introduction

According to the ICH S7B guidelines in vitro currents inhibition assessment is a compulsory element of the drug candidates non-clinical evaluation [1]. It has to be noted although that the screening includes hERG channel only as this channel blockade and subsequent I_{Kr} current inhibition are recognized as the potential QT prolongation surrogates. As it was described elsewhere in majority of cases clinically observed QT prolongation and Torsade de Pointes events concerns drugs which inhibit the I_{Kr} current in vitro. There are although compounds which are potent in vitro hERG channel inhibitors and don't express the proarrhythmic consequences in the clinical settings. Such phenomenon is mainly explained by the multiple channels inhibition and subsequent QT prolongation compensation [2]. To be able to assess the drug triggered arrhythmia risk and predict in vivo human situation wide in vitro inhibition results are needed. In the situation when either no in vitro studies were conducted or the results are reduced to the hERG channel only, reliable QSAR models are necessary for other ionic currents.

The study was aimed to create predictive model taking into account chemical structure of the particular compound and its ability to block the sodium channel in the cell membrane.

2. Materials and methods

The data was based on the literature search. Several databases were searched: Medline, Scopus and Google Scholar. The key phrases were: 'I_{Na}', or 'sodium current', or 'sodium channel', or 'Na^+ channel' and 'half-maximal inhibitory concentration' (IC_{50}) either in the article title, keywords or abstract. There were no limits for search results except publication language (English). After careful papers examination and manual data extraction the final dataset contained 108 records describing 38 drugs. Initial number of inputs was 110. Input vector contained in vitro experimental settings combined with molecular descriptors of particular drug compound. Including in vitro research setting as the input data allows the prediction abilities enhancement and makes the derived model more flexible. Extended QSAR methodology was previously successfully applied for other channels [3, 4].

The output was a single variable encoding IC_{50} value as its negative logarithm (pIC_{50}). Molecular descriptors were computed by chemoinformatics software Marvin (ChemAxon, UK) [5]. Drugs chemical structures were structurally optimized with use of Marvin "molconvert" tool. Resulting *.sdf files were the subject to descriptor calculations by "cxcalc" tool with selected 41 plugins.

Artificial neural networks (ANNs) were chosen as modeling tools. Two major types of ANNs were applied: multi-layer perceptrons (MLPs) and neuro-fuzzy systems (NFs) of Mamdani type. ANNs were trained with use of back-propagation 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, their hidden layer size was optimized between 5 to 100 nodes. Sensitivity analysis was performed in order to reduce initial number of inputs to the crucial variables set. The procedure was carried out with modified Żurada method [6, 7]. The generalization error was estimated by means of enhanced 10-fold cross validation (10-cv), where whole drugs information was excluded from the test sets in order to simulate the most difficult task for a model to perform: to predict unknown structure behavior. Generalization error was expressed as root mean squared error (RMSE, Eq. 1), normalized root mean squared error (NRMSE, Eq. 2) and coefficient of determination (R^2) of predicted vs. observed values.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(PRED_i - OBS_i)^2}{n}} \qquad (1)$$

where:

OBS – observed value
PRED – predicted value
n – total number of records

$$NRMSE = \frac{RMSE}{(OBS_{max} - OBS_{min})} \qquad (2)$$

where:

RMSE – root mean squared error
OBS_{max} – maximal value of the observed results
OBS_{min} – minimal value of the observed results

In order to improve model predictive performance ensemble ANNs systems were applied, where ANNs were combined by simple average of their outputs.

3. Results

Sensitivity analysis allowed to reduce inputs number from initial value of 110 to 20 – their labels and meanings were presented in the Table 1.
Variables 1-5 encode in vitro experimental settings, whereas 6 – 20 are molecular descriptors of the chemical compound of interest. It is noticeable that the molecular descriptors might be classified to the following groups:

- topology (6-10, 14, 16, 18)
- geometry (17, 19)
- surface area (11)
- physicochemical properties (12, 13, 15, 20)

The best predictive system found in this study was an ensemble containing three MLP ANNs:

- single hidden layer with 20 nodes and logistic activation function
- four hidden layers with 15, 7, 5 and 3 nodes respectively and hyperbolic tangent activation function
- four hidden layers with 120, 80, 40 and 20 nodes respectively and logarithmic activation function

A simple average of the outputs of the above described ANNs was the final response of the system. The overall 10-cv generalization error of the ensemble was RMSE = 0.68, NRMSE = 20.7% and R^2 = 0.35 (Fig. 1).

Table 1. Sensitivity analysis results and description of the selected crucial variables.

No.	*[Label]* and description
1	*[Hz]* depolarization pulse frequency
2	*[model]* cell line type (XO / CHO / HEK / GP)
3	*[t1_pulse]* duration of depolarization pulse
4	*[holding_pot]* holding potential
5	*[depol_puls]* depolarization pulse voltage
6	*[Heteroaromatic_ring_count]* no. of heteroaromatic rings
7	*[Stereoisomer_count]* no. of double-bond stereoisomers
8	*[resonantcount]* no. of resonant structures
9	*[Balaban_index]* topological information
10	*[Smallest_ring_size]* no. of atoms included in the smallest ring
11	*[ASA-]* solvent accessible surface area of all atoms with negative partial charge
12	*[bpKa1]* first basic pKa
13	*[apKa2]* second acidic pKa
14	*[Asymmetric_atom_count]* no. of asymmetric atoms
15	*[bpKa2]* second basic pKa
16	*[Chiral_center_count]* no. of tetrahedral stereogenic center atoms
17	*[Minimal_projection_radius]* minimal radius of the compound projected on the planar surface
18	*[Ring_atom_count]* no. of atoms in the ring
19	*[Maximal_projection_radius]* maximal radius of the compound projected on the planar surface
20	*[Mass]* molecular mass

XO – Xenopus oocyte; CHO – Chinese hamster ovaries; HEK – human embryonic kidney cell line; GP – guinea pig cardiomyocytes

4. Discussion

To our knowledge the proposed model is the first available model for the drug triggered in vitro cardiac sodium channels inhibition prediction. Apart of that the main advantage of the proposed solution lies in the applied methodology which includes in vitro data utilization. Such approach was previously used to develop similar models for other non-hERG currents namely I_{Ks} and I_{Ca} [3, 4].

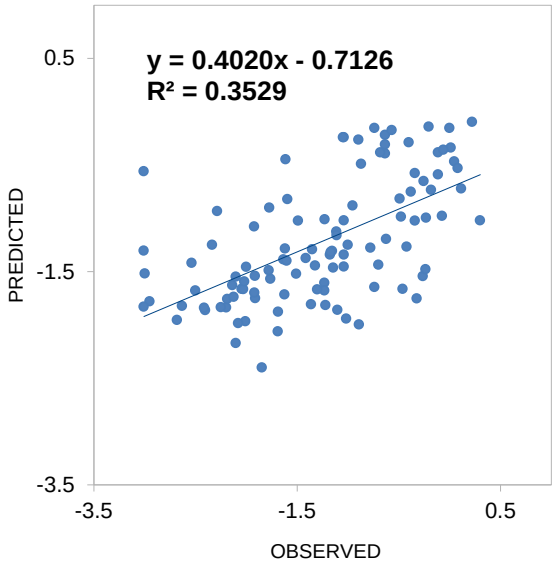

Fig. 1. Predicted vs. observed pIC_{50} for the best ANNs ensemble.

This work is an example of the data-mining procedures combined with predictive modeling. The latter was the major goal to be achieved by the ANNs modeling. The model was created as a future part of the ToxComp system in order to provide assessment of the multi-channel block on the cardiomyocyte and its consequence to the heart rhythm. For the data-mining ANNs provide information about the analyzed problem by their autonomous selection of the crucial variables. Here, it might be noticed, that there is a lot of variables associated with the compound's geometry, specifically rings size and stereochemistry-related information. Thus, it might be hypothesized that particular compound's geometry and stereoisomeric properties are crucial to its action on the sodium channel. Protonation related properties seem to be also important. The above relationships were discovered empirically by ANNs and being in accordance with the common knowledge on the subject (QSAR) might be the basis to the more classical quantitative models development. It has to be stressed here that these relationships are hidden as a part of the 'black-box' ANNs model and so far cannot be presented in other form. However, even the quoted above character of the crucial variables is a valuable information about potential cardiotoxic effects caused by drugs.

The results are mostly determined by the data used at the model development stage. All of the information describing drugs triggered I_{Na} current inhibition was derived from the available literature and as the consequence it is inhomogeneous. Moreover, lack of standardization of the in vitro assays conditions results in the significant degeneration of any non-linear functions applied for such a small dataset.

All data used for the model development are freely available for download and further processing [8].

5. Conclusions

Regarding harsh testing conditions and the empirical nature of the models, their predictability was found to be acceptable, thus the above models are to be included into the ToxComp in silico carditoxicity prediction software (www.tox-portal.net). Future work will be devoted to the task of switching from the 'black-box' ANNs models to the more classical mathematical formulas.

Acknowledgements

Project financed by Polish National Center for Research and Development LIDER project number LIDER/02/187/L-1/09.

References

[1] ICH S7B (http://www.ich.org).
[2] Redfern W.S., Carlsson L., Davis A.S., Lynch W.G., MacKenzie I., Palethorpe S., Siegl P.K.S., Strang I., Sullivan A.T., Wallis R., Camm A.J., and Hammond T.G.. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development, Cardiovasc Res 2003; 58: 32-45.
[3] Polak S, Wisniowska B, Glinka A, Fijorek K, Mendyk A. Slow delayed rectifying potassium current (IKs) – analysis of the in vitro inhibition data and predictive model development. Journal of Applied Toxicology, Article first published online: 14 Feb 2012 [Epub ehead of print].
[4] Wisniowska B, Mendyk A, Fijorek K, Glinka A, Polak S. Predictive model for the L-type channel inhibition - multichannel block in QT prolongation risk assessment. Journal of Applied Toxicology,Article first published online: 3 Jul 2012 [Epub ehead of print].
[5] ChemAxon, http://www.chemaxon.com
[6] Mendyk, A., Jachowicz, R. Unified methodology of neural analysis in decision support systems built for pharmaceutical technology. Expert Syst Appl 2007;32: 1124–1131.
[7] Żurada JM, Malinowski A, Usui S. Perturbation Method for Deleting Redundant Inputs of Perceptron Networks. Neurocomputing 1997;14: 177-193.
[8] www.tox-database.net

Address for correspondence.

Aleksander Mendyk
Department of Pharmaceutical Technology and Biopharmaceutics, Jagiellonian University - Medical College, Medyczna 9 St., 30-688 Kraków, Poland
mfmendyk@cyf-kr.edu.pl