

# The open-access dataset for *insilico* cardiotoxicity prediction system

Sebastian Polak<sup>1\*</sup>, Barbara Wisniowska<sup>1</sup>, Kamil Fijorek<sup>2</sup>, Anna Glinka<sup>1</sup>, Miłosz Polak<sup>1</sup>, Aleksander Mendyk<sup>3</sup>

<sup>1</sup>Unit of Pharmacoepidemiology and Pharmacoconomics, Jagiellonian University, Medical College, Cracow, Poland, Medyczna 9 Str., 30-688 Kraków, Poland; <sup>2</sup>Department of Statistics, Faculty of Management Cracow University of Economics Rakowicka 27 Str., 31-510 Kraków, Poland; <sup>3</sup>Department of Pharmaceutical Technology and Biopharmaceutics, Faculty of Pharmacy, Jagiellonian University, Medical College, Cracow, Poland, Medyczna 9 Str., 30-688 Kraków, Poland; Sebastian Polak - Email: spolak@cm-uj.krakow.pl; Phone: (+) 4812 6205 517; Fax: (+) 4812 6205 820; \*Corresponding author

Received May 05, 2011; Accepted May 09, 2011; Published June 06, 2011

## Abstract:

Drug cardiotoxicity is one of the main reasons of fatal drug related problem events and the subsequent withdrawals. Therefore, its early assessment is a crucial element of the drug development process. For the drug driven hERG inhibition assessment, which is assumed to be the main reason for toxicity, in vitro techniques are used. Gold standards are based on the Patch Clamp method with the use of various cell models but due to its low throughput, *insilico* models have become more appreciated. To develop a reliable empirical QSAR model, wide dataset containing a variety of cases has to be available. In this article, a freely available for download, set of data is described. It is based on literature peer-reviewed reports and contains hERG inhibition information expressed as IC50 value for 263 molecules described in 642 records. All studies were done with the use of three cell models (XO, CHO, HEK) and other elements describe the electrophysiological settings of the in vitro study. The above mentioned set was used for the successful development of the predictive models.

## Background:

QT prolongation with possible fatal TdP arrhythmia is a common cause of drug attrition and relabeling or withdrawals from the market [1]. Thus cardiac liability testing of the compounds during the development process has gained increased regulatory and public attention. The major cause of QT prolongation is connected with a direct block of the hERG channel responsible for the delayed rectifier potassium current ( $I_{Kr}$ ) in cardiomyocytes by non-antiarrhythmic drugs. A wide variety of preclinical in vitro studies are employed to evaluate the drug-hERG interactions potential. The accepted and recommended standard assay is the measurement of the patch clamp current obtained by the expression of hERG in heterologous mammalian cells. However the Patch Clamp technique is time-consuming, labor-intensive, provides relatively low throughput and generates high costs thus is not appropriate at the early stage of the drug development process. Therefore *insilico* screening tests become appreciable and are widely used [2]. Numerous molecules or concentrations can be processed simultaneously. Although the availability and quality of the datasets used at the models development stage are crucial factors. The high sensitivity of the measured channel's inhibition on the experimental settings (i.e. cell models, temperature) is a well-known phenomenon which cannot be bypassed and has to be considered in the data [3, 4]. It can be done either at the data preprocessing level with the use of the inter-settings extrapolation factors or during the modeling stage by introducing additional parameters.

## Methodology of development:

As the first step, a preliminary list of drugs known or suspected to have cardiotoxic potential has been created, using Fenichel's database [5], the International Registry for Drug Induced Arrhythmias - Arizona resources [6] and

Roche [7]. Drugs reported to block the hERG channel or  $I_{Kr}$  were included, whereas compounds influencing the cardiac action potential in any other way were excluded from the dataset. In order to collect experimental IC50 values for hERG of listed drugs Scopus, Medline and Google Scholar searches were performed for each drug. There was no time-limit for the search query. Key phrases were: the name of the compound of interest and 'IC50', or 'hERG', or 'Human Ether-A-Go-Go', or 'potassium channel', or 'potassium current' or ' $I_{Kr}$ ', either in the article title, keywords or abstract. If there were no results for a combination of a compound name with any keyword, the compound's class name was used in the query. The recovered reports were scrutinized for additional references not covered by the electronic search. In addition to half-maximal inhibitory concentration (IC50 value), papers were revised for additional information describing experimental settings. The complete set of information was gained only for experiments performed with one of the three most popular expression systems: *Xenopus* oocytes, Human Embryonic Kidney cells, Chinese Hamster Ovary cells, and if they employed electrophysiology assessments. All information was tracked down to the original source, validated over the citing article and described by both of them to allow the users to recognize them unequivocally [8].

## Utility:

The obtained dataset is an up to date collection of hERG IC50 values with experimental details culled from available scientific literature sources. It contains 263 molecular entries, totaling 642 records concerning different in vitro measurements. The dataset provides information about the factors with a possible influence on IC50 value such as the model applied in experiments, measurement techniques, transfection types, experimental conditions (temperature and  $K^+$  concentration in external solution) and voltage protocol

(protocol type: step, ramp, step-ramp; holding potential, depolarization potential, measurement potential; depolarization pulse time). This allows the generation of a consistent, homogenous input dataset in order to obtain high quality and efficient mathematical models. The molecules included in the dataset cover a wide range of IC<sub>50</sub> values (10 orders of magnitude) and a broad chemical space encompassing drug-specific physico-chemical properties.

#### Caveats:

The main limitations of the dataset result from the assumptions made – there is a limited number of cell models included, heterogeneity of the measurement methods used is high (lack of a standard procedure).

#### Availability and future developments:

The described dataset is freely available for download from the project website for commercial and non-commercial use [9]. Its development is part of the larger project aiming to deliver software for hERG channel inhibition prediction. The models used are based on the described dataset completed with the drugs physico-chemical parameters. The algorithms utilized cover artificial neural networks and RandomForests [10]. Future development plans include a regular dataset upgrade based on the newly published peer-reviewed articles.

Also being considered is the expansion of the dataset by adding new cell models (i.e. guinea pig ventricular myocytes). As other ionic currents are nowadays recognized as important (i.e. I<sub>Ks</sub>, I<sub>Na</sub>) it is planned to expand the project from a single dataset to a database by adding new sets of information.

#### References:

- [1] Al-Khatib S *et al.* *JAMA*. 2003 **289**: 2120 [PMID: 12709470]
- [2] Van de Waterbeemd H & Gifford E. *Nat Rev Drug Discov*. 2003 **2**: 192 [PMID: 12612645]
- [3] Wiśniowska B & Polak S. *Toxicol Mech Methods*. 2009 **19**: 278 [PMID: 19778217]
- [4] Witchel HJ *et al.* *J Pharmacol Toxicol Methods*. 2002 **48**: 65 [PMID: 14565563]
- [5] <http://www.fenichel.net/pages/Professional/subpages/QT/Tables/pbydrug.htm>
- [6] <http://www.qtdrugs.org>
- [7] Roche O *et al.* *ChemBioChem*. 2002 **3**: 455 [PMID: 12007180]
- [8] Polak S *et al.* *J Appl Toxicol*. 2009 **29**: 183 [PMID: 18988205]
- [9] [www.tox-portal.net](http://www.tox-portal.net)
- [10] [www.tox-comp.net](http://www.tox-comp.net)

Edited by P Kanguane

Citation: Polak *et al.* *Bioinformation* 6(6): 244-245 (2011)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.