Toxicology Letters 250–251 (2016) 42–46

Contents lists available at ScienceDirect

Toxicology Letters

journal homepage: www.elsevier.com/locate/toxlet

Monte Carlo method for predicting of cardiac toxicity: hERG blocker compounds

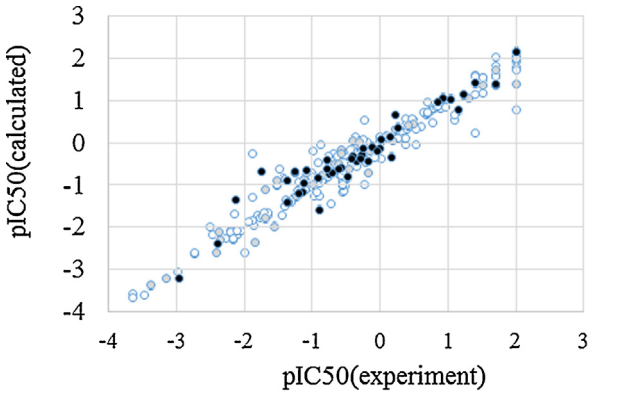
Marco Gobbi[a], Marten Beeg[a], Mariya A. Toropova[a,b], Andrey A. Toropov[c,*], Mario Salmona[d]

[a] Department of Molecular Biochemistry and Pharmacology, Laboratory of Pharmacodynamics and Pharmacokinetics, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milan, Italy
[b] Dipartimento di Scienze Farmaceutiche, Università degli Studi di Milano, Via L. Mangiagalli, 25, 20133 Milan, Italy
[c] Department of Environmental Health Science, Laboratory of Environmental Chemistry and Toxicology, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milan, Italy
[d] Department of Molecular Biochemistry and Pharmacology, Laboratory of Biochemistry and Protein Chemistry, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milan, Italy

HIGHLIGHTS

- Predictive models for cardiac toxicity are built up.
- The models are calculated with the Monte Carlo method.
- The CORAL software is utilized to build up the models.
- The statistical quality of the models is quite good.
- The models are built up in accordance with the OECD principles.

GRAPHICAL ABSTRACT



Training set (◦); Calibration set (◦); Validation set (•)

ARTICLE INFO

Article history:
Received 21 December 2015
Received in revised form 5 April 2016
Accepted 7 April 2016
Available online 8 April 2016

Keywords:
Monte Carlo method
Cardiac toxicity
QSAR
OECD principles
CORAL software

ABSTRACT

The estimation of the cardiotoxicity of compounds is an important task for the drug discovery as well as for the risk assessment in ecological aspect. The experimental estimation of the above endpoint is complex and expensive. Hence, the theoretical computational methods are very attractive alternative of the direct experiment. A model for cardiac toxicity of 400 hERG blocker compounds (pIC_{50}) is built up using the Monte Carlo method. Three different splits into the visible training set (in fact, the training set plus the calibration set) and invisible validation sets examined. The predictive potential is very good for all examined splits. The statistical characteristics for the external validation set are (i) the coefficient of determination r^2 = (0.90-0.93); and (ii) root-mean squared error s = (0.30–0.40).

* Corresponding author.
E-mail addresses: andrey.toropov@marionegri.it, aatoropov@yahoo.com (A.A. Toropov).

http://dx.doi.org/10.1016/j.toxlet.2016.04.010
0378-4274/

1. Introduction

Identification of potential human Ether-a-go-go Related-Gene (hERG) potassium channel blockers is a very important task of the drug discovery and drug safety process in pharmaceutical industries as well as for academic drug discovery centers. Recent works suggest starting to address such issues at the hit selection stage (Park et al., 2013; Anwar-Mohamed et al., 2014; Villoutreix and Taboureau, 2015).

In order to prioritize molecules during the early drug discovery phase and to reduce the risk of the preliminary checking up of pharmaceutical agents, the computational approaches have been developed to predict the potential of hERG blockage of new drug candidates (Villoutreix and Taboureau, 2015).

Thus, the estimation of the cardiac toxicity of organic hERG blockers is an important theoretical and practical task of medicinal chemistry (Obiol-Pardo et al., 2011). The quantitative structure – activity relationships (QSARs) is a possible tool to predict the cardiac toxicity (Frid and Matthews, 2010). The aim of this work is to build up QSAR models for the cardiac toxicity using the CORAL software (http://www.insilico.eu/coral). Simplified molecular input-line entry system (SMILES) (Weininger, 1988; Weininger, 1990; Weininger et al., 1989) is used to represent the molecular structure of examined organic hERG blockers.

2. Method

2.1. Data

The numerical data on cardiac toxicity (IC_{50}, half-maximal response dose (Obiol-Pardo et al., 2011), the endpoint is pIC_{50}, i.e. the negative decimal logarithm of the *IC50*) for 400 compounds taken in the literature (Obiol-Pardo et al., 2011). These four hundred compounds are distributed into three sets: the training set (≈73%), the calibration set (≈13.5%), and the external validation set (≈13.5%). The distributions are prepared according to the following principles: (i) these distributions are random; (ii) these distributions are different; and (iii) the range of the endpoint for the training, calibration, and validation sets are similar.

2.2. Optimal descriptors

Optimal descriptors, which are involved to build up the QSAR model for the pIC_{50}, are the following:

$$DCW(T*,N*) = \sum\{CW(s_k) + CW(ss_k) + CW(sss_k)\} + CW(BOND) + CW(NOSP) + CW(HALO) + CW(PAIR) \quad (1)$$

In Eq. (1), the s_k, ss_k, and sss_k are combinations of one, two, and three "SMILES atoms". The "SMILES atom" is a fragment of the SMILES notation, which contains one symbol or two symbols, which cannot be examined separately (e.g. 'Cl', 'Br', etc.). The *CW* (s_k), *CW*(ss_k), and *CW*(sss_k) are correlation weights of the above-mentioned "SMILES atoms". The correlation weights are coefficients, which used to calculate the descriptor. The numerical data for the correlation weights are obtained by the Monte Carlo method optimization procedure, which gives maximum of correlation coefficient between endpoint and the optimal descriptor. The BOND, NOSP, HALO, and PAIR are global attributes of SMILES which reflects the presence of various kinds of chemical bonds (BOND); the presence of nitrogen, oxygen, sulphur, and phosphorus (NOSP); the presence of halogens, i.e. fluorine, chlorine, bromine, and iodine (HALO); and presence of various combinations of SMILES atoms(PAIR). The *CW(BOND), CW(NOSP), CW(HALO),* and *CW(PAIR)* are correlation weights of the global attributes of SMILES. The detailed description of the above listed local (s_k, ss_k, and sss_k) and global (BOND, NOSP, HALO, and PAIR) attributes of SMILES is available in the literature (Toropova et al., 2015) as well as at web site of the CORAL software (CORAL, 2015).

The T is the threshold, i.e. a coefficient used to classify SMILES attributes into two classes (i) rare or noise; and (ii) active. The rare attributes are blocked (their correlation weights are fixed zero). The coefficient can be 1, 2, . . . , M. The T* is threshold which gives preferable statistical quality of the model for the calibration set. The N is the number of epochs of the Monte Carlo optimization. The N* is the number which gives preferable statistical quality for the calibration set. The T* and N* are calculated according to scheme suggested in works (Toropova et al., 2015). Having the numerical data for the correlation weights, one can calculate the *DCW(T*,N*)* for the training set and define regression parameters C_0 and C_1 for the following model.

$$pIC_{50} = C_0 + C_1 \times DCW(T*,N*) \quad (2)$$

The predictive potential of the model calculated with Eq. (2) should be checked up with external validation set (Toropova et al., 2015).

3. Results and discussion

3.1. QSAR models

The Monte Carlo optimization with T* and N* which are selected according to scheme suggested in work (Toropova et al., 2015) gives the following models:

$$pIC_{50} = -5.5431(\pm 0.0035) + 0.1418(\pm 0.0001) \times \mathrm{DCW}(1.22) \quad (3)$$

$$pIC_{50} = -5.4618(\pm 0.0036) + 0.1331(\pm 0.0001) \times \mathrm{DCW}(1.17) \quad (4)$$

$$pIC_{50} = -5.5541(\pm 0.0033) + 0.1527(\pm 0.0001) \times \mathrm{DCW}(1.28) \quad (5)$$

Table 1 contains the statistical characteristics of the models for pIC_{50} calculated with Eqs. (3)–(5). Fig. 1 contains the graphical representation of these models.

Table 2 contains comparison of the predictive potential of models suggested here with the models described in the literature. In fact, the only one model (Tan et al., 2012) is characterized by larger determination coefficient, in comparison with models calculated with Eqs. (3)–(5). However, the above-mentioned model (Tan et al., 2012) related to 133 compounds, whereas the models calculated with Eqs. (3)–(5) are related to 400 compounds.

Table 1
The statistical characteristics of QSAR models for organic hERG blockers, calculated with Eqs. (3)–(5).

Distribution	n_{train}	r^2_{train}	q^2_{train}	s_{train}	n_{calib}	r^2_{calib}	s_{calib}	n_{valid}	r^2_{valid}	s_{valid}	DD
1, Eq. (3)	294	0.9446	0.9440	0.283	52	0.7583	0.518	54	0.8618	0.396	342.8
2, Eq. (4)	290	0.9353	0.9345	0.301	54	0.8843	0.410	56	0.9359	0.298	327.7
3, Eq. (5)	299	0.9469	0.9463	0.265	51	0.9464	0.295	50	0.9035	0.343	337.6