QSAR-Models, Validation, and IIIC-Paradox for Drug Toxicity

Alla P. Toropova, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy
https://orcid.org/0000-0002-4194-9963
Andrey A. Toropov, Istituto di Ricerche farmacologiche Mario Negri IRCCS, Milan, Italy
https://orcid.org/0000-0001-6864-6340
Emilio Benfenati, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

ABSTRACT

Three kinds of drug toxicities are examined in this modeling analysis. These are: (i) toxicity of psychotropic drugs; (ii) cardiac toxicity; and (iii) drug carcinogenicity. Predictive models for the toxicity data are built up by the Monte Carlo technique. The simplified molecular input-line entry system (SMILES) is used for the representation of the molecular structure. Quantitative structure – activity relationships (QSAR) developed here are mathematical functions of corresponding SMILES. The index of ideality of correlation was tested as a tool to improve predictive potential of these models.

KEYWORDS

Drug Toxicity, Index of Ideality of Correlation, Monte Carlo Method, QSAR

INTRODUCTION

The majority of drugs are toxic (Robciuc et al., 2017; Erdem Kuruca et al., 2019). Consequently, the assessment of drug toxicity is an important task at the interface of chemistry, biology, medicine. Quantitative structure – property / activity relationships (QSPRs/QSARs) are useful tools to solve the task (Zhao & Li, 2018; Zhang et al., 2018; Kar & Roy, 2015; Toropov et al., 2014; Gissi et al., 2014; Raevsky et al., 2012). It is to be noted that modelling of both therapeutic efficacy and toxicity can be performed using the QSPR/QSAR technique. There are different categories of drug toxicity, e.g., cardiac toxicity (Frid & Matthews, 2010), liver-related adverse effects of drugs (Wong, 2006; Rodgers et al., 2010; Toropov et al., 2012), toxicity of psychotropic drugs (Esteki & Khayamian, 2008; Gissi et al., 2014; Gómez-Lumbrañas et al., 2018).

One can find newer toxicity of existing drugs from further medicinal experiments as well as exchange of old drug forms by new drug forms. In addition, simultaneous application of different therapeutic agents can lead to detection of unexpected effects. Sometimes, the detection can be made during long term observations, e.g. years, or even tens of years (Ziuganov et al., 2005; Mizuno et al.,

DOI: 10.4018/IJQSPR.2020010102

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
In order to obtain a more stable and reliable system for registration of drugs, so-called ADMET conception (absorption, distribution, metabolism, excretion, and toxicity) has been developed and suggested (Gola et al., 2006; Dearden, 2007; Cheng et al., 2013; Vora et al., 2019; Zaki et al., 2019).

In the framework of the ADMET conception as well as in the frame of generalized pharmaceutical applications, several approaches are formulated to solve the tasks related to sphere of pharmaceutical sciences. These are: (i) linear regression analysis (Roy & Mitra, 2012); (ii) artificial neural networks (Fjodorova et al., 2010; Verma & Matthews, 2015); (iii) random forest (Schöning et al., 2018); (iv) super vector machine (Yu et al., 2017); (v) Monte Carlo technique (Veselinović et al., 2016; Bhargava et al., 2017).

Each of the above approaches is characterized by a list of advantages as well as by a list of disadvantages. The common practice in the QSPR/QSAR studies is to search for improvement of models. Naturally, some criteria to compare QSPR/QSAR models are necessary. Some widely used criteria contradict common sense and results of manifold computational experiments (Hartung & Hoffmann, 2009) but nonetheless they are often applied for estimation of the QSPR/QSAR models. For instance, well-known $q^2$ being calculated with the training set (Golbraikh & Tropsha, 2002) factually does not correlate with $R^2$ (determination coefficient) for external validation set (Kubinyi, 2004). This situation is known as “Kubinyi paradox” (Hartung & Hoffmann, 2009). The Kubinyi paradox indicated the necessity of external validation for any QSPR/QSAR models. In addition, the Kubinyi paradox questioned reliability of internal validation (i.e. validation via $q^2$).

Every molecular feature will have some influence on one or other endpoints. However, different features can have special (weak or strong) impact for an endpoint under consideration. Presence of halogens as well as oxygen, nitrogen, sulphur, and other chemical elements has apparent fundamental impact practically on any endpoint. Similarly, the branching in a molecular skeleton, presence / absence of the double or triple covalent bonds, and aromaticity are important factors for physicochemical and biochemical behaviour of substances at the molecular level. These circumstances play an important role in the development of the computational chemistry to define strategy and tactics of building up a predictive model.

An analogy is a specific kind of similarity that involves the interpretation of a given system via some other system that is more clear. The analogy has played and continues to play a key role in the development of all the sciences (Rouvray, 1994). Genetic algorithm (GA) in multiple linear regression analysis (MLRA) is well-known analogy between algorithms aimed to build up a predictive model using some pool of molecular descriptors and biologic evolution processes (Turabekova et al., 2014). Artificial neural networks are computing systems that are analogy of the biological neural networks that constitute animal brains (Fjodorova et al., 2010; Verma & Matthews, 2015). The balance of correlations is another QSPR/QSAR approach developed with analogy with a system that is far from the computational chemistry (Toropova et al., 2011). The optimization of correlation weights can be interpreted as the acceptance of some “generalized” decision, that has impact to all substances, not only for the so-called active training and for passive training sets. This is similar to acceptance of generalized decision by bicameral legislature in parliament. Bicameral legislature is a way to avoid odious solution preferable from point of view of separated group of deputies. Similarly, two groups of substances, which have different influences on the final decision, being used separately as the active training set and the passive training set, can give possibility to avoid odious decision preferable for visible substances distributed into the active training set. Experiments described in the literature confirm that the “bicameral” building up a model (decision) can be more “suitable” decision in comparison with a decision that is obtained without the “bicameral” process of decision-making. One more analogy examined here is so-called index of ideality of correlation developed by analogy with conception of ideal gas, ideal solvent, and ideal symmetry (Toropov & Toropova, 2017).

The aim of this study is to check the ability of the Index of ideality of correlation (IIC) as a tool to improve QSPR/QSAR models for drug toxicity built up by the balance of correlations.
Strategies of Virtual Screening in Medicinal Chemistry