Computers in Biology and Medicine 45 (2014) 20–25

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Comprehension of drug toxicity: Software and databases

Andrey A. Toropov [a,*], Alla P. Toropova [a], Ivan Raska Jr [b], Danuta Leszczynska [c], Jerzy Leszczynski [d]

[a] IRCCS, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via La Masa 19, Milano 20156, Italy
[b] 3rd Department of Medicine, Department of Endocrinology and Metabolism, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, U Nemocnice 1, 12808 Prague 2, Czech Republic
[c] Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA
[d] Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, PO Box 17910, Jackson, MS 39217, USA

ARTICLE INFO

Article history:
Received 30 August 2013
Accepted 18 November 2013

Keywords:
In silico toxicology
In silico methods
QSAR
Computational toxicology
Drug toxicity

ABSTRACT

Quantitative structure–property/activity relationships (QSPRs/QSARs) are a tool (*in silico*) to rapidly predict various endpoints in general, and drug toxicity in particular. However, this dynamic evolution of experimental data (expansion of existing experimental data on drugs toxicity) leads to the problem of critical estimation of the data. The carcinogenicity, mutagenicity, liver effects and cardiac toxicity should be evaluated as the most important aspects of the drug toxicity. The toxicity is a multidimensional phenomenon. It is apparent that the main reasons for the increase in applications of *in silico* prediction of toxicity include the following: (i) the need to reduce animal testing; (ii) computational models provide reliable toxicity prediction; (iii) development of legislation that is related to use of new substances; (iv) filling data gaps; (v) reduction of cost and time; (vi) designing of new compounds; (vii) advancement of understanding of biology and chemistry. This mini-review provides analysis of existing databases and software which are necessary for use of robust computational assessments and robust prediction of potential drug toxicities by means of *in silico* methods.

1. Introduction

There are various tools that one could use for prediction of properties (activities) of chemical compounds. Among them the quantitative structure–property/activity relationships (QSPRs/QSARs) methods hold important place. The history of evolution of the QSPRs/QSARs techniques contains three basic periods. The first period involved design of molecular descriptors which are correlated with important physicochemical parameters and/or with various biological endpoints. The statistical quality of QSPR/QSAR for all available compounds which were used to build up the model was considered as the main result. The second period, the statistical quality of model for external "invisible" compounds which were not involved in building up model, become the main criterion of quality of QSPR/QSAR. On the other hand, the third period, the definition of chemical space where QSPR/QSAR can be used with satisfactory accuracy, *i.e.* so-called the domain of applicability has become the measure of quality for QSPR/QSAR models. These criteria are briefly described by Organization for Economic Co-operation and Development (OECD) as Setubal principles: QSARs for regulatory application should: (1) be associated with a defined endpoint of regulatory importance, (2) take the form of an unambiguous algorithm, (3) have a defined domain of applicability, (4) be associated with appropriate measures of goodness of fit, robustness, and predictivity, and (5) have a mechanistic basis [1].

The drug discovery establishment has been probably one of the original industries to appreciate the QSPR/QSAR technology and still remains its the most important user. In fact, the drug discovery protocol needs to define two groups of endpoints related to new molecular entities (NMEs) [1] which relate to both the therapeutic and toxic effects [2].

Physico-chemical indicators have been increasingly used during the early stages of drug discovery to provide a more comprehensive understanding of the key properties that affect the biological functions (*i.e.* ADME—absorption, distribution, metabolism, and excretion). The most commonly measured physicochemical properties are permeability and solubility (due to their importance in the gastrointestinal absorption of orally administered drugs), and also lipophilicity, integrity, and stability (since these properties generally define the pharmaceutical potential of

* Corresponding author. Tel.: +39 2 39014595; fax: +39 2 39014735.
E-mail addresses: andrey.toropov@marionegri.it, aatoropov@yahoo.com (A.A. Toropov).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.11.013

a compound) [3–8]. Often, the ADME concept is expanded by toxicity [5,6]. Apparently, this potential hazardous quality of NME should be taken into account with no less care than possible therapeutic effects.

There are several endpoints that relate to potential hazardous effects(such as the liver effects of drug candidates [9]; cardiac toxicity [10], and blood–brain barrier of compounds that has influence upon both drug efficiency and drug toxicity [11–13]) which should be estimated during the early stages of drug discovery. Finally carcinogenicity [14] and mutagenicity [15] of a drug also should be estimated during the initial stages of drug discovery.

Due to huge cost and time necessary for research and development related to drug design, many *in silico* methods have been developed to provide accurate prediction of pharmacokinetic properties, such as Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET), in very early stage of drug discovery [3–8,14,15].

Recently [16,17], the physiologically based pharmacokinetic/pharmacodynamic (PBPK/PD) approaches (for quantitatively description of the metabolism) have been suggested. In fact, PBPK models represent synthesis of mathematical calculations and experimental measurements [17]. Nevertheless, this approach becomes an necessary tool for the drug discovery [18] because allows to take into account chemical structures of potential metabolites [19].

Various approaches could be used to analyze possible toxicity of drugs. Apparently, classes of substances which are potential NMEs define the methodological details of the drug toxicity analysis. It appears that organic compounds are the most known source of the NMEs. However, this group of NME contains many sub-classes [2,5,7]. Peptides [20,21] are also source of NMEs, and this case often involves application of "unclassic" QSAR approaches. Relatively new class of potential NMEs consist of various nano materials such as fullerene derivatives [22–25] nanoparticles [26,27], and others [28]. Finally, one also needs to consider an important challenge of the drug discovery—so-called drug–drug interactions [29–31].

The average success rate for NMEs in all therapeutic areas combined, starting from first-in-human studies to registration during 1991–2000 period was approximately 11%. In 2003, the US Food and Drug Administration (FDA) approved 21 NMEs, however, during next years, this number has decreased (only 15 NMEs approved in 2010) [32]. Although lack of efficacy is a major contributor to a disappointment, toxicity can also be a cause of failure in drug development [1].

Thus, the systematization of the information that is necessary for drug discovery is possible only with involving of multimillion databases and reliable software [33]. In this minireview, we discuss possible ways of the systematization of the *in silico* methods that can be used for fast, preliminary estimation of toxicity of compounds with the possible therapeutically effects.

2. Sources of drug toxicity data

Apparently, as concluded by the results of a recent study, collaboration between industry, computational software developers and regulatory researchers led to the development of a toxicity database and classification rules for NMEs [34]. Table 1 contains the basic principles of classification for substances available via database [34].

There are various kinds of toxic endpoints. An endpoint can be related to different organisms, *e.g.* rats, mice, fishes, birds, and human. From point of view of praxis most important toxicity is one that is related to human, but these data are very limited. Therefore, there are attempts to select organism with some similarity to organism of human (rat, dog, monkey, etc.). An endpoint can be related to various routes of adsorption (inhalation, oral, skin, etc.). Consequently, a toxic endpoint can be measured by different units (mg/kg, mg/m^3, ppm, etc.). Thus, the search for data on different kinds of toxicity for various substances is complex enough task.

Table 1
Classification scheme for substances involved in the drug discovery [34].

Class	Comments
1	Known to be both mutagenic and carcinogenic
2	Known to be mutagenic but unknown carcinogenic potential
3	Structurally alerting compound, unrelated to the active pharmaceutical ingredient (API) and of unknown mutagenic potential
4	Structurally alerting compound related to the API
5	No structural alerts or sufficient evidence for absence of mutagenicity

However, there are a plethora of private and public data resources available for developing toxicity models. Recent reviews summarized available public toxicity databases [35,36] (Table 2).

On the other hand, the large number and variety of sources of drug toxicity (their number is increasing day by day) lead to two questions: (i) where one can get information on toxicity of the potential therapeutic agent; and (ii) how one can estimate accuracy and reliability of the data. The problem of adequate and fast estimation of drug toxicity lead to creation of international organizations such as Food and Drug Administration (FDA), Center of Drug Evaluation and Research (CDER), and the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) legislation of the EU [32,35,36]. In spite of a number of organizations involved, harmonizing the reporting of chemical toxicity data to facilitate comparison between available data sources and databases remains a critical need [37].

It should be noted that these international organizations encourage the use of *in silico* methods in order to answer two questions mentioned above, as well as for the harmonizing of available data on the toxicity in general, and drug toxicity in particular [37].

One can obtain practical skills of estimation of the data related to toxicity after visiting web sites listed in Table 2.

3. *In silico* toxicology tools

In silico toxicology is generally, but not exclusively, a predictive science. The approaches used for helping to define safety and discovery efforts in therapeutics represent a large number of chemical–biological informatics-based programs. Toxicology-oriented computational approaches as a rule are based on building toxicity databases. This gives possibility to carry out the QSAR analysis (modeling) [35].

Fig. 1 shows the general scheme of the QSAR analysis and illustrates the roles of the databases and QSAR models.

The main reasons for the *in silico* prediction of toxicity in general are [36]: (i) Pressure to reduce animal testing; (ii) Computational models provide suitable toxicity prediction; (iii) Legislation (Governmental policies in both the European Union (EU) and North America) has encouraged and, in some cases, mandated the use of computational techniques to predict toxicity. For example, the US EPA has utilized QSAR to assist in the pre-manufactory notification of new chemicals, especially where no toxicity data are physically in hand. This requirement for models has inspired considerable advancement in the prediction of acute toxicity for environmental endpoints); (iv) Filling data gaps; (v) Cost and time reduction; (vi) Identification of new toxicological problems; (vii) Designing of new compounds; (viii) Higher throughput and *in silico* approaches have higher reproducibility if the same model is used. Again, this *in silico* approaches have low compound synthesis requirement; (ix) *In silico* models have the