IN SILICO PREDICTIONS OF COMPLEX BIOLOGICAL EFFECTS

Milen Todorov
Assen Zlatarov University, Department of Inorganic Chemistry, 8010 Burgas, Bulgaria,
mtodorov@btu.bg

ABSTRACT

Determining the toxicity of chemicals is necessary to identify their harmful effects on humans. One of the complex biological effects - carcinogenicity is constantly used for development of new methodologies for identification of carcinogenic chemical compounds. The testing strategy for identification chemical carcinogens largely relies on the 2-year rodent bioassay, which is time-consuming and labor-intensive. Thus, there is an increasing effort to develop alternative approaches for identification of chemicals with carcinogenic potential. In silico approaches based on quantitative structure-activity relationships (QSAR) are rapid and inexpensive and thus have been investigated for such purposes. Since the correlation between carcinogenicity and variety genotoxic and non-genotoxic endpoints (mutagenicity, receptor mediated effects) was found it is expected that models based on such data could be used successfully for identification of chemical carcinogens acting via both mechanisms. In the current study the implemented profiling schemes in non-commercial software tool associated to both mechanisms was used to predict the carcinogenic effect for large number of chemicals.

Keywords*: complex biological effects, computational toxicology, QSAR*

INTRODUCTION

Carcinogenicity is among the toxicological endpoints that pose the highest public concern. The standard bioassays in rodents used to assess the carcinogenic potency of chemicals are time-consuming, costly and require the sacrifice of a large number of animals. In order to avoid the huge number of the experimental tests both statistically based and knowledge-based methods are used for prediction of the carcinogenic potential for diverse chemical compounds. The first group - statistically-based methods (MultiCASE, Leadscope, TOPKAT, LAZAR and CAESAR systems) [1-5] rely on techniques such as multivariate analysis, rule-induction, artificial intelligence, cluster analysis, pattern recognition, etc.). They deal with limited or no prior chemical or biological classification according to mechanism of carcinogenicity [6]. The knowledge-based (or rule-based) methods (HazardExpert, OncoLogic, Toxtree, and DEREK systems) [7-10] include toxicological knowledge, expert judgment and fuzzy logic taking into consideration toxicokinetics, toxicodynamics and metabolism related to processes with cellular macromolecules or receptors. Each of the above mentioned approaches has the potentials and limitations described in the literature [11, 12**]**. In addition it should be emphasized that the larger part of the programs are commercial products which could be used under expensive license agreements. On the other hand as a result of joint efforts of computational toxicologist from different organizations a unified public free *in silico* tool has been created [14]. It allows predictions of diverse chemical compounds for variety endpoints including carcinogenicity.

The aim of this study is to investigate the possibility to predict carcinogenicity by making use of public tool with encoded structural rules for specific genotoxic and non-genotoxic biological endpoints.

Mechanisms of carcinogenic effect

Carcinogenic chemicals have conventionally been divided into two broad categories based of the presumed mode of action: genotoxic or non-genotoxic. Genotoxic carcinogens cause damage by interacting directly with DNA – many known mutagens are in this category. In contrast, non-genotoxic carcinogens cause “epigenetic” changes, i.e. effects that do not involve alterations in DNA but that may influence the carcinogenic process. The mechanistic understanding of the carcinogenic process differs considerably between the two modes of action. The distinction is not absolute – chemicals can be carcinogenic by both models of action.

A unifying scientific theory for the mode of action of epigenetic carcinogens is still missing, because they act through a wide variety of different and specific mechanisms. For this reason, QSARs for epigenetic carcinogenicity are still in an early stage of development. A number of structural alerts (SAs) and characteristics of several types of non-genotoxic carcinogens have been summarized by Woo & Lai [15].

Mutagenicity as a surrogate endpoint

As discussed in previous section, one of the mechanisms associated with carcinogenicity is thought to be the formation of a covalent bond between the electrophilic chemicals and the DNA bases. Schultz et al. [16] described a conceptual framework for predicting the toxicity of reactive chemicals where plausible molecular initiating events were based on covalent reactions with nucleophiles in proteins or DNA and would ultimately lead to a variety of different adverse outcomes including mutagenicity. Based on this observation it is expected that the interactions associated this effect can be considered as an outcome of a presence of same toxicophores (structural alerts) responsible for DNA damages. As a logical sequence it is also expected that QSAR models based on mutagenicity experimental data (Ames test) can be used successfully in prediction of carcinogenic effect of xenobiotics.

Nongenotoxic mechanisms for carcinogenicity

Epigenetic (or nongenotoxic) carcinogens refer to natural or man-made chemical compounds that may evoke carcinogenic processes without giving rise to heritable DNA sequence mutations. It is likely that, at odds with genotoxic agents, multiple pathways need to be altered for cancer induction by epigenetic/nongenotoxic carcinogens. Accordingly, molecular targets of epigenetic/nongenotoxic carcinogenic activities may include a variety of cellular and extracellular constituents of various organs, with the only exclusion of DNA. Carcinogenic responses, such as enhanced cell proliferation, unregulated cell growth, and aberrant cell cycle kinetics, may be determined by altered sensitivity to a wide variety of cytokines, hormones, growth factors, and other cell growth mediators that may operate via membrane receptors, signal transduction pathways, and intracellular communication processes.

MATERIALS AND METHODS

Carcinogenicity database

The ISSCAN database (http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7), developed by the Istituto Superiore di Sanità (Rome, Italy), contains information on more than 1150 chemical compounds tested with the long-term carcinogenicity bioassay on rodents (rat, mouse). Historically, this database was developed to support the development of (Q)SAR models for chemical carcinogenicity. ISSCAN is downloadable in pdf, xls and sdf formats, and is searchable by chemical name and CAS number.

OECD QSAR Toolbox

The *Toolbox* [14] is a expert system that incorporates the OECD guidance related to categorization, read-across, and QSAR models. It also incorporates a large number of data sets containing physical and chemical property data, molecular descriptors, mammalian and non-mammalian toxicity test data, *in vitro* and high throughput data, and categorical and endpoint/mechanistic descriptors derived by variety organizations for thousands of chemicals. A GUI allows the user to enter or retrieve data on individual chemicals on a point-and-click basis; define category criteria; and conduct read-across, trend analyses or run QSAR models to fill data gaps for untested chemicals.

Another advantage of the system is the opportunity to investigate a chemical with account to its metabolic fate. It is well known that the chemical in its parent form may not exert toxic effect

however after metabolism a reactive metabolite can be produced which may damage biological macromolecules. This became extremely important in assessment of the carcinogenic potential of various type of chemicals.

Carcinogenicity (genotox and nongenotox) alerts by ISS

The profiler is an expanded and updated version of the correspondent module of the software Toxtree [9]. It works as a decision tree for estimating carcinogenicity, based on a list of 55 structural alerts (SAs). Most of the new SAs are relative to nongenotoxic carcinogenicity, whereas the SAs in the initial list mainly coded genotoxic carcinogenicity. The SAs for carcinogenicity are molecular functional groups or substructures known to be linked to the carcinogenic activity of chemicals. As one or more SAs embedded in a molecular structure are recognised, the system flags the potential carcinogenicity of the chemical.

Profiling schemes for DNA damages. OASIS DNA v. 1.3 and ISS v.2.3

The profiler OASIS DNA v. 1.3 is based on Ames mutagenicity model part of OASIS TIMES system [17]. The profiler contains exact definitions of 78 structural alerts responsible for interaction of chemicals with DNA. The scope of this profiler is to investigate the presence of alerts within the target molecules responsible for interaction with DNA, especially related to Ames mutagenicity.

The second ISS v.2.3profiler contains a list of 30 structural alerts (SAs). The SAs for mutagenicity are molecular functional groups or substructures known to be linked to the mutagenic activity of chemicals. As one or more SAs embedded in a molecular structure are recognized, the system flags the potential mutagenicity of the chemical.

RESULTS AND DISCUSSION

Initially the ISSCAN database was analyzed in order to select one of the experimental protocols for data consistency. Based on largest number of experimental results carcinogenicity was investigated by using data obtained for rats by dosed feed. As a result a total number of 890 chemicals (381 carcinogenic and 509 non carcinogenic) have been investigated. The results were obtained by application of the following profiles: Carcinogenicity alerts by ISS; OASIS DNA v. 1.3 and ISS v. 2.3. The prediction results are summarized in Table 1.

Table 1. Prediction results for carcinogenicity obtained by - Carcinogenicity alerts by ISS; OASIS DNA v. 1.3 and ISS v. 2.3

Observed carcinogenicity/ number of chemicals		Predictions for carcinogenicity			
		Positive Carcinogenicity ISS	*Positive* by OASIS 1.3	*Positive* by ISS 2.3	*Negative*
Positives	389	303	7	-	76
Negatives	509	233	22	-	268

As it can be seen sensitivity (correct predicted carcinogens) exceeds 80%. Moreover it was found that Carcinogenicity ISS profile demonstrate alone satisfactory predictive performance. Seven chemicals only are predicted to be carcinogenic as a result of positive prediction for mutagenic effect. On the other hand 76 chemicals are wrongly predicted to be non-carcinogens. Since there is also negative result for mutagenic effect for these 76 chemicals one possibility that may be useful is application of other endpoint profiles related to carcinogenicity. Prolonged estrogen exposure, both from endogenous and exogenous sources, induces cell proliferation in cells that express estrogen receptors, such as those found in the endometrium, breast, and liver. The response of these organs, which normal growth is under hormonal control, to the estrogen-induced proliferative effects, is a progression from normal growth to hyperplasia to neoplasia. Considering the role of estrogenic effect in relation

with carcinogenicity a profile for this endpoint in the Toolbox is applied for prediction of the 76 false negatives. In 21 chemicals specific structural characteristic associated with positive estrogenic effect were found. This finding shows the need of expansion of the carcinogenic profile mainly in the section of non-genotoxic (epigenetic) categories. Development of new predictive rules associated with receptor mediated affects is expected to improve the predictions and such rules should be investigated and further implemented in the library of profile schemes.

The specificity which is a measure for correct predictions for non-carcinogenic effect was found to be 53% (Table 1; 268/509 correct predicted chemicals). This result shows that there is urgent need for improvement of the profile. In this respect the most problematic categories have been identified and the contribution for false positive prediction among them is shown in Table 2.

Table 2. False positive predictions for carcinogenicity obtained by - Carcinogenicity alerts by ISS; OASIS DNA v. 1.3 and ISS v. 2.3

#	Number of chemicals	Category (functional group)	Mechanism
1	42	Primary aromatic amines	Genotoxic
2	22	Nitroaromatic	Genotoxic
3	13	Halogenated benzenes	Non genotoxic
4	13	Thiocarbonyls	Non genotoxic
5	10	Hydrazines	Genotoxic
6	10	Aromatic diazo	Genotoxic
7	10	Alpha, beta unsaturated carbonyls	Genotoxic

As it can be seen the largest number of incorrect predicted chemicals is found in category Primary aromatic amines. Theoretically, it is known that chemicals which contain this structural feature are reactive toward biological macromolecules. On the other hand from practical point of view there is a need for precise definition of the alert in order to be adapted to identify those chemicals with positive effect. More specifically, its definition should represent structural variation of the neighbor substituents. Such improvement could be achieved by contrasting correct predicted chemicals which possess same structural alert. In this way it is possible to be defined additional supporting masks which ultimately will decrease the number of false positive predictions.

Another thing that should be pointed out is the role of metabolic transformations. Aromatic amines have to be metabolized to reactive electrophiles in order to exert their carcinogenic potential. For aromatic amines and amides, this typically involves an initial N-oxidation to N-hydroxyarylamines and N-hydroxyarylamides, which is mediated by cytochrome P-450. Upon further activation by enzymatic esterification, nitrenium ions are formed. These highly reactive intermediates bind covalently to biomolecules, generating aminoaryl derivatives [18]. Due to availability of false predictions for chemicals in this category, obviously this mechanism not always leads to positive carcinogenic effect. An explanation and further prediction of the prediction results could be achieved by investigation of the metabolism of all chemicals of interest. For example one of the wrongly predicted chemical as carcinogen is 1,4-Benzenediamine (CAS: 105-30-6). Additional investigation of the metabolism for this chemical has been performed. It should be pointed out that there is a metabolism database incorporated in the Toolbox which is suitable for comparison of observed and simulated metabolic maps. It was found that there are two observed metabolites produced as a result of acylation transformation [19]. More likely the role of metabolism is responsible for deactivation the effect of known reactive metabolites. Therefore, additional analysis is needed in order to account metabolism as a mandatory element in the predictive scheme for carcinogenic effect.

The second group with large number of false prediction is Nitroaromatic compounds. Again,

the role of metabolism as activating or deactivating step should be investigated for improvement of the predictions related to this chemical category.

Concerning non-genotoxic mechanism – two chemical categories were found to contribute for false positive predictions. The first one Halogenated benzenes represents chemical structures like dioxins and polychlorinated biphenyls. After the binding of dioxins, the activated Ah receptor forms a heterodimer with another transcription factor. This ternary complex binds to regulatory sequences on DNA and specifically activates transcription of a battery of dioxin-inducible genes. Primary target genes identified so far are a number of genes encoding drug metabolizing enzymes. Such complex sequence of events is difficult to be associated with precisely defined structural rules due to diverse structures of the compounds in this category. In addition current definitions in the profile were found to be very general (halogen atom connected to aromatic ring with two restriction masks). A way for improvement is to redefinition of these structural rules to be more precise by analyzing new chemical structures.

The same limitation which affects correct predictions was found also for the second category - Thiocarbonyls. Generally, the effect of these chemicals has been attributed to their ‘‘anti thyroid’’ activity which leads to a disruption of the pituitary-thyroid hormonal regulatory system. Further analysis of new compounds from this category is suggested for improvement of the prediction performance of the carcinogenicity profile.

CONCLUSIONS

Carcinogenicity is one of the effects which should be evaluated as it is required by many directives for safety assessment of chemicals. A huge variety of chemicals exist in the environment, and their potential to exert carcinogenic effect is unknown. In the current study the predictive performance of profile used for identification of carcinogenic compounds has been used to predict the effect of 890 experimentally tested chemicals.

The overall prediction for carcinogens shows high performance – 80% correct predictions. Regarding non-carcinogens it was found that the role of metabolic deactivation should be taken into account. In addition the need of more precise definition of structural characteristics (functional groups) for several chemical classes is discussed.

REFERENCES

1. Chakravarti, S.K., R.D. Saiakhov, and G. Klopman. Optimizing Predictive Performance of CASE Ultra Expert System Models Using the Applicability Domains of Individual Toxicity Alerts. Journal of Chemical In-formation and Modeling, (52), (2012), p. 2609-2618.

2. Valerio, L. G., K. P. Cross. Characterization and validation of an in silico toxicology model to predict the mutagenic potential of drug impurities. Toxicology and Applied Pharmacology, (2012), p. 209-221.

3. Benigni R. Computational prediction of drug toxicity: the case of mutagenicity and carcinogenicity. Drug Discovery Today: Technologies, (1), 4, (2004), p. 457-463.

4. LAZAR toxicity predictions https://lazar.in-silico.ch/predict

5. CAESAR http://www.caesar-project.eu/

6. E Benfenati, R Benigni, DM DeMarini, C Helma, D Kirkland, TM Martin, P Mazzatorta, G Ouedraogo-Arras, AM Richard, B Schilter, WGE Schoonen, RD Snyder, C Yang

Predictive models for carcinogenicity: frameworks, state-of-the-art, and perspectives

J. Environ. Sci. Health C, 27 (2009), pp. 57–90

7. HazardExpert http://www.compudrug.com

8. OncoLogic https://www.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals

9. ToxTree http://toxtree.sourceforge.net/

10. Judson P., S. Stalford and J. Vessey. Assessing confidence in predictions made by knowledge-based systems Toxicology Research, (2), 1, (2013), p. 70-79

11. R Serafimova, MF Gatnik, A Worth Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity, EUR 24427 EN:58

http://ecb.jrc.ec.europa.eu/DOCUMENTS/QSAR/EUR_24427_EN.pdf (2010)

12. Petkova S., and Ramon S. Nanotoxicology and regulatory issues of emerging technologies. Organisation for the Prohibition of Chemical Weapons International Cooperation and Assistance / ICB, 2013, AP2013/RP/06

14. OECD QSAR Toolbox https://www.qsartoolbox.org/

15. Woo YT & Lai DY (2003). Mechanisms of action of chemical carcinogens, and their role in Structure-Activity Relationships (SAR) analysis and risk assessment. In: Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens (R Benigni, ed). Boca Raton: CRC Press: 41–80.

16. Schultz, T. W., R. E.Carlson, M. T.Cronin, , J. L.Hermens, , R.Johnson, , P. J.O'Brien, D. W. Roberts, A. Siraki, K. B. Wallace, and G. D. Veith, A conceptual framework for predicting the toxicity of reactive chemicals: modeling soft electrophilicity. SAR QSAR Environ. Res. 2006, 17(4), 413–428.

17. Mekenyan O., S. Dimitrov, T. Pavlov, Dimitrova G., Todorov M., Petkov P., Kotov S. Simulation of chemical metabolism for fate and hazard assessment. V. Mammalian hazard assessment. SAR and QSAR in Environmental Research (23) (5-6), (2012), p. 553-606.

18. Benigni, R., Giuliani, A., Franke, R., and Gruska, A. (2000). Quantitative structure-activity relationships of mutagenic and carcinogenic aromatic amines. Chem.Revs. 100, 3697-3714.

19. Dressler, W., T. Appelqvist, Food Chem. Toxicol., 44(3), (2006). (in vivo), (p. 371 - 379)