Marie Curie Initial Training Network

Environmental Chemoinformatics (ECO)

Project report 1 / 2012

24 October 2012

Modeling and interpretation of toxicity of chemical compounds using toxicophores

Early stage researcher:
Elena Salmina

Project supervisor:
Dr. Igor V. Tetko

Research Institution:
Helmholtz Zentrum München

INTRODUCTION

Although Quantitative Structure Activity Relationship (QSAR) modeling is a powerful technique for prediction of biological activity and toxicity of chemicals, the major problem is the interpretability of such models. The models based on many powerful methods, such as neural networks or support vector machines, often use hundreds of descriptors and are particularly difficult to interpret. These models frequently raise a concern of the end-users, *i.e.* regulators, because of the inability to get a clear mechanistic interpretation of the observed structure-toxicity relationships. The lack of interpretability makes difficulty of these models to comply with the rules on the validation of models [1], can limit the use of such models for the legislative purposes.

Usage of molecular fragments that are associated with particular type of toxicity and known as “toxicophores” or “structure alerts” provides a mechanistic explanation of toxic nature of compounds and could be utilized to complement QSAR models and to help interpreting their predictions.

GOALS AND OBJECTIVES

According to the above-mentioned background the research incudes two main goals:

1. A creation of toxicophore database (first stage of the research);
2. Usage of the created toxicophore database for development, interpretation and improvement QSAR-models describing toxic properties of chemicals (second stage of the research).

and the next objectives:

1. Collection of data about toxicity of chemical compounds, namely compilation of molecular fragments associated with different toxicological endpoints (structural alerts) along with data about associated biotransformations and pathways, etc.

2. Creation of “toxicophores” catalog that will serve as a guide for development and interpretation of models

3. An integration of the collected data to create QSAR models to predict toxicity of chemical compounds. Sensitivity analysis of descriptors and calculation of their relationships with chemical structure of toxicophoric fragments.

4. Validation and explanation of developed toxicity models using toxicophores.

FIRST STAGE RESEARCH RESULTS (9th January – 9th April)

During the first stage of the research I focused on goal 1 and objectives 1-2. I studied the literature in the field of structural alert [e.g., articles 2-48]. The "evolution" of toxicophores was investigated starting from the seminal work by James and Elizabeth Miller, who highlighted the electrophilic nature of chemical carcinogenesis [2, 3]. Structural alerts associated with different toxicological endpoints, e.g. genotoxic and non-genotoxic carcinogenicity, mutagenicity, skin sensitization, acute aquatic toxicity, idiosyncratic toxicity (reactive metabolite formation) were identified and grouped according to the endpoint (Table 1).

Table 1

Extracted structural alerts

Toxicological endpoint	Number of described structural alerts	Publications	Source of alerts
Genotoxic carcinogenicity / mutagenicity	117	2-19	5, 11,13,17
Non-genotoxic carcinogenicity	5	17	17
Skin sensitization	161	20-33	20, 22, 25, 31
Acute aquatic toxicity	99	34-39	35, 36
Idiosyncratic toxicity	35	40-42	41, 42

The information on structural alerts was used to create a new online module to screen the toxicophores. This module, named ToxAlerts, was developed as an extension to the Online Chemical Modeling Environment system (OCHEM) (http://ochem.eu) [49]. The ToxAlerts was implemented in collaboration with the chemoinformatics research group of eADMET GmbH. Encoding of molecular fragments responsible for toxicity was implemented in SMiles ARbitrary Target Specification (SMARTS) language (http://www.daylight.com).

ToxAlerts module description

ToxAlerts module consists of two main components: a database to store structural alerts for various toxicological endpoints in an organized manner and an on-line facility to screen chemical compounds against these alerts.

Database structure

The central entity in the ToxAlerts database is a structural alert, which is uniquely identified by:

1. A structural pattern represented by a SMARTS string;
2. A publication where the alert was mentioned;
3. A toxicological endpoint associated with this alert (*e.g.*, carcinogenicity or skin sensitization).

Additionally to this obligatory and fundamental information required for each alert, the database allows storing supplementary information, such as:

1. A chemical name of the alert (*e.g.*, “Acid halides”, “Sulphonyl azides”, *etc.*);
2. A visual depiction of the alert (can be uploaded in PNG format manually);
3. Position of the alert in the publication (page, table / figure);
4. Arbitrary supplementary information (*e.g.,* mechanism of action associated with the alert, information whether metabolic bioactivation is required to observe the toxic effect, *etc.*).

A schema of the database is presented in Figure 1.

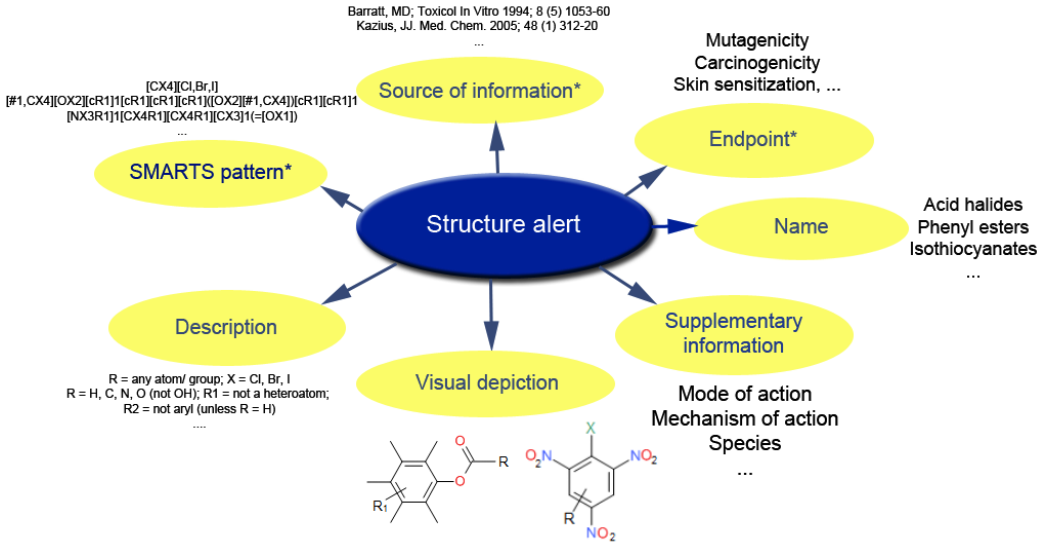

Figure 1. A schema of the ToxAlerts database. The main features describing alerts are shown. In addition to this information, each alert has several other properties, i.e. introducer and modifier of the alert, creation and modification time, modification history, as well as visibility and user rights control inherited from the OCHEM.

Screening facilities

Structural alerts can be filtered using particular criteria (*e.g.*, endpoint, publication or alert name), marked as selected and grouped into an alert set. Such alert set can be saved in a browser of alerts sets by specifying its name. The saved alert sets can be used for screening purposes.

ToxAlerts is tightly integrated into OCHEM: the platforms share the same database of users, publications and chemical compounds. Moreover, ToxAlerts promotes the same design principles: the data and alerts are contributed by users and can be freely accessed by other users. Both ToxAlerts and OCHEM strictly require all data to be complemented with the original source of information – a reference to a scientific publication where the alert was described.

Integration of ToxAlerts with OCHEM provides many benefits for the usage of this database. Firstly, the compound sets (referred to as “baskets” and “tags”) prepared in the OCHEM database can be readily used for screening against toxicological alerts in ToxAlert. Secondly, the compounds filtered by alerts can be exported from the OCHEM in a wide variety of formats and/or can be used for any further research, e.g. confirmation of the potentially toxic compounds with QSAR models. Thirdly, presence or absence of particular preselected alerts can be used as molecular descriptors for the development of QSAR models.

ToxAlerts database was extended by uploading several sets of functional groups and molecular moieties (so-called „filters“) used by five companies to identify unstable, toxic and reactive compounds. Description of these filters is given in Table 2.

Table 2

Filters are stored in ToxAlerts system

Company	Number of filters
Enamine (http://enamine.emolecules.com)	108
ChemDiv (http://eu.chemdiv.com)	100
Maybridge (http://www.maybridge.com)	22
LifeChemicals (http://www.lifechemicals.com)	72
Pfizer (http://www.pfizer.com)	38

The database also includes 120 alerts for electrophilic agent recognition [43]. The formation of a covalent adduct has been recently described as the molecular initiating event for toxicity endpoints [50, 51] and therefore these alerts are interesting for interpretation of toxicity of chemical compounds. In summary, during my work I collected and introduced as SMARTS data for 877 alerts collected from 17 sources.

MEETINGS AND CONFERENCES

During the first stage of my project I participated and presented results of my work at:

1. ITN-ECO Madrid Winter School 2012 "In vitro and in silico toxicology: Biological activity and computational rationalization" [27th February-2nd March] (INIA headquarters, Ctra. De la Coruña Km 7, E-28040 Madrid); poster.

2. The Munich Life Science Symposium for Young Scientists <INTERACT-2012> [29th March-30th March] (poster abstract is represented in appendix A);

3. 15th International Workshop on Quantitative Structure-Activity Relationships (QSAR 2012) in Environmental and Health Sciences [June 18-22, Tallinn, Estonia] (abstract which has been accepted for oral presentation).

FURTHER WORK

The goal of the first stage of my project, a comprehensive literature review and development of the toxicophore database and collection of toxicophores, was fully achieved. During the second stage, which is planned for autumn of 2012, I will work towards application of the developed alerts database for development, interpretation and improvement QSAR-models to predict toxicity of chemicals.

REFERENCES:

1. OECD, The report from the Expert Group on (quantitative) structure-activity relationship ([Q]SARs) on the principles for the validation of (Q)SARs. OECD Series on Testing and Assessment. *ENV/JM/MONO.* **2004**/24*,* 49.
2. Miller J.A., Miller EC.. Ultimate chemical carcinogens as reactive mutagenic electrophiles. In: Hiatt HH, Watson JD, Winsten JA, editors. *Origins of Human Cancer.* Cold Spring Harbor Laboratory; Huntington, NY, USA: **1977**. pp. 605-628.
3. Miller E.C., Miller J.A. Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. *Cancer*. **1981**, 47:2327-45.
4. Ashby J. Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. Environmental mutagenesis*.* **1985**, 7:919-921.
5. Ashby J., Tennant R.W. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. Mutat Res. **1988**, 204(1):17-115.
6. Tennant RW, Ashby J. Classification according to chemical structure, mutagenicity to Salmonella and level of carcinogenicity of a further 39 chemicals tested for carcinogenicity by the U.S. National Toxicology Program. Mutat Res. **1991**, 257(3):209–227.
7. Ashby J., Tennant R.W. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutat Res*. **1991**, 257(3):229-306.
8. Ashby J., Paton D. The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. *Mutat Res*. **1993**, 286(1):3-74.
9. Munro I.C., Ford R.A., Kennepohl E., Sprenger J.G. Correlation of structural class with no-observed-effect levels: A proposal for establishing a threshold of concern. *Food and Chemical Toxicology*. **1996**, 34(9):829-867.
10. Sawatari K, Nakanishi Y, Matsushima T. Relationships between chemical structures and mutagenicity: a preliminary survey for a database of mutagenicity test results of new work place chemicals*. Ind Health*. **2001**, 39(4):341-5.
11. Bailey A.B., Chanderbhan N., Collazo-Braier N., Cheeseman M.A., Twaroski M.L. The use of structure-activity relationship analysis in the food contact notification program. *Regulat Pharmacol Toxicol.* **2005**, 42:225-35.
12. Benigni R. Structure-activity relationship studies of chemical mutagens and carcinogens: mechanistic investigations and prediction approaches*. Chem Revs.* **2005**, 105:1767-800.
13. Kazius J, McGuire R, Bursi R. Derivation and validation of toxicophores for mutagenicity prediction*. J Med Chem*. **2005**, 48:312-20.

14. Kazius J, Nijssen S, Kok J, Bäck T, Ijzerman AP. Substructure mining using elaborate chemical representation. *J Chem Inf Model*. **2006**,46(2):597-605.
15. Benigni R., Bossa C. Structural alerts of mutagens and carcinogens. *Current Computer-Aided Drug Design*. **2006**, 2(2), 169-176.
16. Benigni R, Bossa C, Netzeva T, Rodomonte A, Tsakovska I. Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. *Environ Mol Mutagen*. **2007**, 48(9):754-71.
17. Benigni R., Bossa C. Structure alerts for carcinogenicity, and the *Salmonella* assay system: A novel insight through the chemical relational databases technology. *Mutat Res*. **2008**, 659(3):248-261.
18. Enoch S.J., Cronin M.T., Ellison C.M. The use of a chemistry-based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity. *Altern Lab Anim.* **2011**, 39(2):131-45.
19. Putz M.V, Ionaşcu C., Putz A.-M., Ostafe V. Alert-QSAR. Implications for electrophilic theory of chemical carcinogenesis. *Int. J. Mol. Sci.* **2011**, *12*(8), 5098-5134.
20. Payne M.P., Walsh P.T. Structure-activity relationships for skin sensitization potential: development of structural alerts for use in knowledge-based toxicity prediction systems. *J Chem Inf Comput Sci*. **1994**, 34(1):154-61.
21. Barratt M.D., Basketter D.A., Chamberlain M., Payne M.P., Admans G.D., Langowski J.J. Development of an expert system rulebase for identifying contact allergens. *Toxicol In Vitro*. **1994**, 8(4):837-9.
22. Barratt M.D., Basketter D.A., Chamberlain M., Admans G.D., Langowski J.J. An expert system rulebase for identifying contact allergens. *Toxicol In Vitro*. **1994**, 8(5):1053-60.
23. Barratt M.D., Langowski J.J. Validation and subsequent development of the DEREK skin sensitization rulebase by analysis of the BgVV list of contact allergens. *J Chem Inf Comput Sci*. **1999**, 39(2):294-8.
24. Zinke S., Gerner I., Schlede E. Evaluation of a rule base for identifying contact allergens by using a regulatory database: comparison of data on chemicals notified in the European Union with "structural alerts" used in the DEREK expert system. *Altern Lab Anim*. **2002**, 30(3):285-98.
25. Gerner I., Barratt M.D., Zinke S., Schlegel K., Schlede E. Development and prevalidation of a list of structure-activity relationship rules to be used in expert systems for prediction of the skin-sensitising properties of chemicals. *Altern Lab Anim.* **2004**, 32(5):487-509.
26. Gerner I., Schlegel K., Walker J.D., Hulzebos E. Use of physicochemical property limits to develop rules for identifying chemical substances with no skin irritation or corrosion potential. *QSAR & Combinatorial Science.* **2004**, 23(9): 726-733.
27. Hulzebos E., Walker J.D., Gerner I., Schlegel K. Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential. QSAR & Combinatorial Science. **2004**, 24(3): 332-342.
28. Gerner I., Schlegel K., Walker J.D., Hulzebos E .Use of Physicochemical Property Limits to Develop Rules for Identifying Chemical Substances with no Skin Irritation or Corrosion Potential. *QSAR & Combinatorial Science.* **2004**, 23(9): 726–733.
29. Walker J.D., Gerner I., Hulzebos E., Schlegel K. The Skin Irritation Corrosion Rules Estimation Tool (SICRET). QSAR & Combinatorial Science. **2005**, 24(3): 378-384.
30. Sabcho D. Dimitrov S.D., Low L.K., Patlewicz G.Y., Kern P.S., Dimitrova G.D., Comber M.H.I., Phillips R.D., Niemela J., Bailey P.T., Mekenyan O.G. Skin Sensitization: Modeling based on skin metabolism simulation and formation of protein conjugates**.** *Int J Toxic***. 2005,** 24: 189-204.
31. Enoch S.J., Madden J.C., Cronin M.T. Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ Res*. **2008**, 19(5-6):555-78.
32. Schultz T.W., Rogers K., Aptula A.O. Read-across to rank skin sensitization potential: subcategories for the Michael acceptor domain. *Contact Dermatitis*. **2009**, 60(1):21-31.

33. Aptula A.O., Enoch S.J., Roberts D.W. Chemical mechanisms for skin sensitization by aromatic compounds with hydroxy and amino groups. *Chem Res Toxicol*. **2009**, 22(9):1541-7.
34. Franks N.P., Lieb W.R. Mechanisms of general anesthesia. *Environ Health Perspect*. **1990**, 87: 199–205
35. Hermens J.L. Electrophiles and acute toxicity to fish. *Environ Health Perspect*. **1990**, 87: 219–225.
36. Verhaar H.J.M, Leeuwen C.J., Hermens J.L.M. Classifying environmental pollutants. Part 1: Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere*. **1992**, 25(4):471-491.
37. Christine L. Russom C.L., Bradbury S.P., Broderius S.J., Dean E. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ Tox Chem*. **1997**, 16(5):948-967.
38. Verhaar H.J., Solbé J., Speksnijder J., Leeuwen C.J., Hermens J.L. Classifying environmental pollutants: Part 3: External validation of the classification system. *Chemosphere.* **2000**, 40(8): 875-883.
39. Comber M.H.I., Walker J.D., Watts C., Hermens J. Quantitative structure-activity relationships for predicting potential ecological hazard of organic chemicals for use in regulatory risk assessments. *Environ* Toxic *Chem.* **2003**, 22(3): 1822-1828.
40. Kalgutkar A.S., Gardner I., Obach R.S., Shaffer C.L., Callegari E., Henne K.R., Mutlib A.E., Dalvie D.K., Lee J.S., Nakai Y., O'Donnell J.P., Boer J.,Harriman S.P. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr Drug Metab*. **2005**, 6(3):161-225.
41. Kalgutkar A.S., Soglia J.R. Minimising the potential for metabolic activation in drug discovery. *Expert Opin Drug Metab Toxicol.* **2005**,1(1):91-142.
42. Stepan A.F., Walker D.P., Bauman J., Price D.A., Baillie T.A., Kalgutkar A.S., Aleo M.D. Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. *Chem Res Toxicol*. 2011, 24(9):1345-410.
43. Enoch S.J., Ellison C.M., Schultz T.W., Cronin M.T. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit Rev Toxicol.* **2011**, 41(9):783-802.
44. Sanderson D.M., Earnshaw C.G. Computer prediction of possible toxic action from chemical structure; the DEREK system. Hum Exp Toxicol. **1991**, 10(4):261-73.
45. Ridings J.E., Barratt M.D., Cary R., Earnshaw C.G., Eggington C.E., Ellis M.K., Judson P.N., Langowski J.J., Marchant C.A., Payne M.P., Watson W.P., Yih T.D. Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. *Toxicology*. **1996**, 106(1-3):267-79.
46. Aptula A.O., Roberts D.W. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem Res Toxicol*. **2006**, 19(8):1097-105.
47. Roberts D.W., Patlewicz G., Dimitrov S.D., Low L.K., Aptula A.O., Kern P.S., Dimitrova G.D., Comber M.I., Phillips R.D., Niemelä J., Madsen C.,Wedebye E.B., Bailey P.T., Mekenyan O.G. TIMES-SS--a mechanistic evaluation of an external validation study using reaction chemistry principles. *Chem Res Toxicol*. **2007**, 20(9):1321-30.
48. Enoch S.J., Roberts D.W., Cronin M.T.D. Electrophilic reaction chemistry of low molecular weight respiratory sensitizers. *Chem Res Toxicol*. **2009**, 22, 1447–1453.
49. Sushko I., Pandey A., Novotarskyi S., Körner R., Rupp M., Teetz W., Brandmaier S., Abdelaziz A., Prokopenko V., Tanchuk V., Todeschini R., Varnek A., Marcou G., Ertl P., Potemkin V., Grishina M., Gasteiger J., Baskin I., Palyulin V., Radchenko E., Welsh W., Kholodovych V., Chekmarev D., Cherkasov A., Aires-de-Sousa J., Zhang Q-Y., Bender A., Nigsch F., Patiny L., Williams A., Tkachenko V., Tetko I. Online chemical modeling environment (OCHEM): web

platform for data storage, model development and publishing of chemical information. *J Cheminform*. **2011**, 19 (3, Suppl 1):20.

50. Schultz T.W., Sparfkin C.L., Aptula A.O. Reactivity-based toxicity modelling of five-membered heterocyclic compounds: application to Tetrahymena pyriformis. *SAR QSAR Environ Res.* **2010**, 21(7-8):681-91.

51. Schwöbel J.A., Koleva Y.K., Enoch S.J., Bajot F., Hewitt M., Madden J.C., Roberts D.W., Schultz T.W., Cronin M.T. Measurement and estimation of electrophilic reactivity for predictive toxicology. *Chem Rev*. **2011**, 111(4):2562-96.

APPENDIX A

Abstract for the Munich Life Science Symposium for Young Scientists <INTERACT-2012>, Germany on 29 – 30 March 2012

ONLINE DATABASE OF TOXICOLOGICAL STRUCTURE ALERTS

Elena Salmina, Iurii Sushko, Vladimir A. Potemkin, Igor V. Tetko

Identification of chemicals that are able to exert adverse effects on human, other living organism or environment is an important feature of the modern toxicology. This problem is especially important considering the new legislations such as REACH in European Union.

Assessment of toxicity with direct *in-vivo* animal tests is very costly and time consuming. An alternative, which could reduce costs and avoid unnecessary animal tests, is using predictive computational models. One of the simplest yet powerful techniques for detecting potentially toxic chemicals is using substructure patterns also known as structure alerts or toxicophores.

We have developed a platform for collecting and storing toxicological alerts from literature and for screening chemical datasets against these alerts. In our system, an alert is uniquely identified by a SMARTS pattern, a toxicological endpoint and a publication where the alert was described. Additionally, the system allows storing complementary information such as name, comments, mechanism of action, etc.

Most importantly, the platform can be easily employed for fast screening of large chemical datasets against the toxicological alerts, providing a detailed profile of the chemicals grouped by alerts and endpoints. Such a facility can be used for decision making regarding whether a compound should be tested experimentally, validated with available QSAR models or eliminated from consideration altogether.

The system is open and tightly integrated with the Online Chemical Modeling Environment (OCHEM) [1]. Any user on the Web can introduce new alerts, browse and edit alerts introduced by other users or screen his/her datasets against all or some alerts. The datasets filtered by alerts can be used at OCHEM for other typical tasks: export in a wide variety of formats, creation of QSAR models, additional filtering by other criteria, etc.

The platform can be accessed on the Web at http://ochem.eu/alerts and it is open for any user following a simple registration procedure.

[1] – Sushko et al. J Comput Aided Mol Des. 2011 Jun;25(6):533-54.