JRC Scientific and Technical Reports

Review of Software Tools for Toxicity Prediction

Mojca Fuart Gatnik and Andrew Worth

EUR 24489 EN - 2010

JRC EUROPEAN COMMISSION

ihcp Institute for Health and Consumer Protection

The mission of the IHCP is to provide scientific support to the development and implementation of EU policies related to health and consumer protection.

The IHCP carries out research to improve the understanding of potential health risks posed by chemical, physical and biological agents from various sources to which consumers are exposed.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information
Address: Via E. Fermi 2749, 21027 Ispra (VA), Italy
E-mail: andrew.worth@ec.europa.eu
Tel.: +39 0332 789566
Fax: +39 0332 786717

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 59865

EUR 24489 EN
ISBN 978-92-79-16395-1
ISSN 1018-5593
doi:10.2788/60101

Luxembourg: Publications Office of the European Union

Printed in Italy

ABSTRACT

When assessing the properties of chemicals, the easiest and most consistent way of applying (Quantitative) Structure-Activity Relationship ([Q]SAR) models is to use ready-made software that implements the models via a user-friendly interface. A wide range of software tools are available for predicting physicochemical properties, toxicological endpoints and other biological effects, as well as fate in the environment and biological organisms. Typically, a given software package predicts multiple properties and endpoints, and some are extensible, allowing the user to develop new models or include new knowledge. In addition to (Q)SAR models and rulebases that are incorporated in software tools, there is a growing scientific literature which reports thousands of (Q)SARs.

In this report, we give an overview of the software packages that are commonly used in the assessment of chemical toxicity. These software packages are potentially useful in the hazard and risk assessment of chemicals, including for regulatory purposes. However, the applicability of any given software tool needs to be carefully evaluated and documented.

LIST OF ABBREVIATIONS

ADME	Absorption, Distribution, Metabolism, Elimination
ADMET	Absorption, Distribution, Metabolism, Elimination and Toxicity
CASE	Computer Automated Structure Evaluation
EC	European Commission
ECHA	European Chemicals Agency
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
EPI	Estimation Programs Interface
EU	European Union
IC50	Half Maximal Inhibitory Concentration
JRC	Joint Research Centre
LD50	Median Lethal Dose
LMC	Laboratory of Mathematical Chemistry
LOAEL	Lowest Observed Adverse Effect Level
MCASE	Multi Computer Automated Structure Evaluation
MC4PC	Multi CASE for Personal Computer
MRDD	Maximum Recommended Daily Dose
MTD	Maximum Tolerated Dose
OECD	Organisation for Economic Cooperation and Development
OPS	Optimal Predictive Space
PASS	Prediction of Activity Spectra for Substances
(Q)SAR	(Quantitative) Structure-Activity Relationship
QMRF	QSAR Model Reporting Format
QPRF	QSAR Prediction Reporting Format
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
SDF	Structure Data file
SRC	Syracuse Research Corporation
TEST	Toxicity Estimation Software Tool
TIMES	Tissue MEtabolism Simulator
TTC	Threshold of Toxicological Concern
US EPA	United States Environmental Protection Agency

CONTENTS

1. Introduction

The following sections briefly describe software tools for toxicity prediction that are either in the public domain or commercially available. Some of the freely available software tools have been developed under the terms of open-source licenses, which means that other experts can further develop and disseminate the software (Jeliazkova et al, 2010). Websites for the various tools are given in Tables 1 and 2, and their ability to predict properties and endpoints relevant to toxicity assessment is highlighted in Table 3. Some of these tools were evaluated by an ECETOC Task Force (ECETOC, 2003).

This review focuses on software tools for toxicity prediction, and complements a recent review on software tools for predicting biokinetic (ADME) properties (Mostrag-Szlichtyng & Worth, 2010).

2. Freely available software

A summary of freely available software is given in Table 1. The following paragraphs describe these tools in general terms.

2.1. Caesar models

A series of statistically-based models, developed within EU-funded CAESAR project (http://www.caesar-project.eu), have been implemented into open-source software and made available for online use via the web. Predictions are made for five endpoints: mutagenicity (Ames), carcinogenicity, developmental toxicity, skin sensitisation, and the bioconcentration factor.

2.2. EPI Suite

EPI (Estimation Programs Interface) Suite estimates a range of physicochemical properties, environmental fate parameters and ecotoxicity. It has been developed by the US EPA in collaboration with Syracuse Research Corporation (SRC), and is used widely by governmental and industry organisations to support the assessment of new and existing industrial chemicals. EPI Suite is freely downloadable from the US EPA website: http://www.epa.gov/oppt/exposure/pubs/episuite.htm

2.3. Lazar

Lazar is an open-source software programme that makes predictions of toxicological endpoints (currently, mutagenicity, human liver toxicity, rodent and hamster carcinogenicity, MRDD) by analysing structural fragments in a training set (Helma, 2006; Maunz & Helma, 2008). It is based on the use of statistical algorithms for classification (k-nearest neighbours and kernel models) and regression (multi-linear regression and kernel models). In contrast to traditional k-NN techniques, Lazar treats chemical similarities not in absolute values, but as toxicity dependent values, thereby capturing only those fragments that are relevant for the toxic endpoint under investigation. Lazar performs automatic applicability domain estimation and provides a confidence index for each prediction, and is usable without expert knowledge. Lazar runs under Linux and a web-based prototype is also freely accessible: http://lazar.in-silico.de/

2.4. OECD QSAR Application Toolbox

The OECD QSAR Application Toolbox is a standalone software application for gaps in (eco)toxicity data needed for assessing the hazards of chemicals. Data gaps are filled by following a flexible workflow in which chemical categories are built and missing data are estimated by read-across or by applying local QSARs (trends within the category). The Toolbox also includes a range of profilers to quickly evaluate chemicals for common mechanisms or modes of action. In order to support read-across and trend analysis, the Toolbox contains numerous databases with results from experimental studies. The first version of the Toolbox, released in March 2008, was a proof-of-concept version. The first update (version 1.1) was released in December 2008. The second phase of the project to develop a more comprehensive Toolbox which fully implements the capabilities of the first version was launched in November 2008 with a four-year timeline. The OECD Toolbox is freely available from the OECD website: http://www.oecd.org/env/existingchemicals/qsar

2.5. OncoLogic

This is an expert system that assesses the potential of chemicals to cause cancer. OncoLogic® was developed by the US EPA in collaboration with LogiChem, Inc. It predicts the potential carcinogenicity of chemicals by applying the rules of SAR analysis and incorporating what is known about the mechanisms of action and human epidemiological studies. The software reveals its line of reasoning, like human experts, to support predictions made. It also includes a database of toxicological information relevant to carcinogenicity assessment. The Cancer Expert System is comprised of four subsystems that evaluate fibres, metals, polymers, and organic chemicals of diverse chemical structures. Chemicals are entered one-by-one and the user needs a limited knowledge of chemistry in order to select the appropriate subsystem. OncoLogic is freely downloadable from the US EPA website: http://www.epa.gov/oppt/sf/pubs/oncologic.htm

2.6. Toxtree

Toxtree is a flexible and user-friendly open-source application that places chemicals into categories and predicts various kinds of toxic effect by applying decision tree approaches. It is freely available from the JRC website: http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE

Toxtree has been developed by the JRC in collaboration with various consultants, in particular Ideaconsult Ltd (Sofia, Bulgaria). A key feature of Toxtree is the transparent reporting of the reasoning underlying each prediction. Toxtree v 1.60 (July 2009) includes classification schemes for systemic toxicity (Cramer scheme and extended Cramer scheme), as well as mutagenicity and carcinogenicity (Benigni-Bossa rulebase and the ToxMic rulebase on the in vivo micronucleus assay). The Cramer scheme is probably the most widely used approach for structuring chemicals in order to make an estimation of the Threshold of Toxicological Concern (TTC).

The current version of Toxtree (v2.1.0, June 2010) also applies the TTC scheme of Kroes et al. (2004), alerts for skin sensitisation alerts (Enoch et al, 2008), and SMARTCyp, a two-dimensional method for the prediction of cytochrome P450-mediated metabolism (Rydberg et al, 2010). SMARTCyp predicts which sites in a molecule are labile for metabolism by Cytochromes P450.

2.7. PASS

This tool, developed by the Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Moscow is a computerised system for the Prediction of Activity Spectra for Substances. It predicts several specific toxicities among them mutagenicity, carcinogenicity, teratogenicity and embryotoxicity, and also mechanisms of action and pharmacological effects. The system predicts the probability (Pa) of a biological activity for a new compound, by estimating the similarity/dissimilarity of the new substance to substances with well known biological activities present in the training set (70 000 compounds). The tool also gives a cross reference between biological activities on the basis of the knowledgebase of mechanism-effect relationships. An online version of PASS is available at: http://195.178.207.233/PASS/index.html

2.8. T.E.S.T.

The Toxicity Estimation Software Tool is an open-source application developed by the US EPA. It estimates the toxicity of a compound by applying several QSAR methodologies thus allowing the user to have greater confidence in predicted toxicities. Among other toxicities it predicts rat oral LD50, Ames mutagenicity, developmental toxicity, as well as acute toxicity to fish (fathead minnow), Daphnia magna and Tetrahymena pyriformis. The tool is freely downloadable from the EPA website (http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST). The models are well documented and the training set is made available as structure files (SDF file).

3. Commercially available software

A summary of commercially available software is given in Table 2. The following paragraphs describe these tools in general terms.

3.1. ACD/Tox Suite

The ACD/Tox Suite (formerly called ToxBoxes), provided by ACD/Labs and Pharma Algorithms, provides predictions of various toxicity endpoints including hERG inhibition, genotoxicity, CYP3A4 inhibition, ER binding affinity, irritation, rodent LD50, aquatic toxicity, and organ-specific health effects (http://www.acdlabs.com/products/admet/tox/). The predictions are associated with confidence intervals and probabilities, thereby providing a numerical expression of prediction reliability. The software incorporates the ability to identify and visualize specific structural toxicophores, giving insight as to which parts of the molecule are responsible for the toxic effect. It also identifies analogues from its training set, which can also increase confidence in the prediction. Predictions are based on data from over 100,000 compounds. The algorithms and datasets not disclosed. A web version of the software is freely accessible at http://www.pharma-algorithms.com/webboxes/

3.2. ADMET Predictor

This is software developed by Simulations Plus (http://www.simulations-plus.com/) for the predictive modelling of ADMET (Absorption, Distribution, Metabolism, Elimination, and Toxicity) properties. It includes a number of in-built models for ADMET, and allows new predictive models to be built from the user's data.

3.3. BioEpisteme

This is primarily a research tool developed by the Prous Institute for Biomedical Research (http://www.prousresearch.com/). It is organised into two main modules: a model building module and a data prediction module. The model building module provides a range of 2D and 3D descriptors; the data prediction module predicts adverse effects. It appears to have been developed mainly for applications in the pharmaceutical industry.

3.4. Derek

Derek for Windows (DfW) is a SAR-based system is developed by Lhasa Ltd, a non-profit company and educational charity (https://www.lhasalimited.org/). DfW contains over 50 alerts covering a wide range of toxicological endpoints in humans, other mammals and bacteria. An alert consists of a toxicophore (a substructure known or thought to be responsible for the toxicity) and is associated with literature references, comments and examples. A key feature of DfW is the transparent reporting of the reasoning underlying each prediction.

All the rules in DfW are based either on hypotheses relating to mechanisms of action of a chemical class or on observed empirical relationships. Information used in the development of rules includes published data and suggestions from toxicological experts in industry, regulatory bodies and academia. The toxicity predictions are the result of two processes. The program first checks whether any alerts in the knowledge base match toxicophores in the query structure. The reasoning engine then assesses the likelihood of a structure being toxic. There are nine levels of confidence: certain, probable, plausible, equivocal, doubted, improbably, impossible, open, contradicted. Derek can be integrated with Lhasa's Meteor

software, which makes predictions of fate, thereby providing predictions of toxicity for both parent compounds and their metabolites.

3.5. HazardExpert

This is a module of the Pallas software developed by CompuDrug (http://compudrug.com/). It predicts the toxicity of organic compounds based on toxic fragments, and it also calculates bioavailability parameters (from logP and pKa). It is a rule-based system with an open knowledge base, allowing the user to expand or modify the data on which the toxicity estimation relies. It covers the following endpoints relevant to dietary toxicity assessment: carcinogenicity, mutagenicity, teratogenicity, membrane irritation, immunotoxicity and neurotoxicity. A further application of the program is prediction the toxicity of the parent compound and its metabolites by linking with MetabolExpert, another module of the Pallas software.

3.6. MDL QSAR

This is primarily a research tool, originally developed and marketed by MDL, and now by Symyx (http://www.symyx.com/). It enables the user to build and apply new QSARs, supporting model development by providing over 400 built-in 2D and 3D molecular descriptor calculators. It includes a variety of predictive modules, including rodent carcinogenicity (FDA model).

3.7. Molcode Toolbox

This is a commercial tool developed and marketed by Molcode Ltd (http://molcode.com/). It has a range of modules for predicting toxicological endpoints and ADME properties. The models are well documented and the underlying experimental data is made available with references and structure files (MDL molfile).

3.8. MultiCASE

This software, developed by MultiCASE Inc. (http://multicase.com/), implements the so-called CASE (Computer Automated Structure Evaluation) approach, and is referred to in different ways (MCASE or MC4PC), depending on the software version and computer platform and its successor. The program automatically generates predictive models from datasets provided by the user. It is based on a fragment-based technology sometimes referred to as the CASE approach (Klopman & Rosenkranz 1994). The program performs a hierarchical statistical analysis of a database to discover substructures that appear mostly in active molecules thus being with high probability responsible for the observed activity. Initially, it identifies the statistically most significant substructure within the training set. This fragment, labelled the top biophore, is considered responsible for the activity of the largest possible number of active molecules. The active molecules containing this biophore are then removed from the database, and the remaining ones are submitted to a new analysis for identification of the next biophore. The procedure is repeated until either the activity of all the molecules in the training set has been accounted for or no additional statistically significant substructure can be found. Then for each set of molecules containing a specific biophore, the program identifies additional parameters called modulators, which can be used to derive QSAR within the reduced set of congeneric molecules. The modulators consist of certain substructures or physicochemical parameters that significantly enhance or diminish the activity attributable to the biophore. QSARs are then derived by incorporating the biophores

and the modulators into the model. The program includes modules to predict physicochemical properties and a range of toxicological endpoints, including carcinogenicity, mutagenicity, teratogenicity, irritation, developmental toxicity, and acute toxicity. For the endpoints, the software uses it own toxicity scale, from 0 to 100 CASE units, to cover the range from inactive, marginally active and active. In many cases, it is difficult to relate these CASE units to traditional measures of toxicity.

3.9. OASIS TIMES

The Tissue MEtabolism Simulator (TIMES), developed by LMC (Bourgas University, Bulgaria; http://oasis-lmc.org/) integrates on the same platform a metabolic simulator and QSAR models for predicting toxicity of selected metabolites. The metabolic simulator generates plausible metabolic maps from a comprehensive library of biotransformations and abiotic reactions. It allows prioritization of chemicals according to toxicity of their metabolites. OASIS TIMES can be used to predict a range of endpoints, including acute toxicity for different species, receptor-binding affinities (oestrogen, androgen and aryl hydrocarbon receptors), mutagenicity and chromosomal aberration, while also accounting for the metabolic activation of chemicals.

3.10. TOPKAT

This QSAR-based system, developed by Accelrys Inc. (http://accelrys.com/), makes predictions of a range of toxicological endpoints, including mutagenicity, developmental toxicity, rodent carcinogenicity, rat chronic LOAEL, rat Maximum Tolerated Dose (MTD) and rat oral LD_{50}. The QSARs are developed by regression analysis for continuous endpoints and by discriminant analysis for categorical endpoints. TOPKAT models are derived by using a range of two-dimensional molecular, electronic and spatial descriptors. TOPKAT estimates the confidence in the prediction by applying the patented Optimal Predictive Space (OPS) validation method. The OPS is TOPKAT's formulation of the model applicability domain - a unique multivariate descriptor space in which a given model is considered to applicable. Any prediction generated for a query structure outside of the OPS space is considered unreliable.

3.11. ToxAlert

This tool, also a module of the Pallas suite, flags compounds for hazards associated with specific pharmacophores (structural alerts). The prediction is based on an improved version of the knowledge base implemented in HazardExpert, and in addition to the overall toxicity profile, it provides probability percentages for different toxicity endpoints. Like HazardExpert, it has an open knowledge base, allowing additions and modifications to the underlying data.

3.12. q-Tox

A tool developed by Quantum Pharmaceuticals (http://q-pharm.com/) utilises a novel approach for the prediction of toxicity. It is based on the premise that biological activity results from the capacity of small molecules to modulate the activity of the proteome. Publically available IC50 values for several proteins were used to build interpretation models. The tool predicts several toxicity endpoints, mouse, rat, dog rabbit LD50 and also side effects. The draw back of the tool is that the estimated calculation time is 5 to 10 hours per molecule.

3.13. CSGenoTox

Is a tool which predicts Ames mutagenicity, developed by ChemSilico (http://chemsilico.com/). Topological molecular descriptors were selected with neural network analysis to optimize the relationship between experimental and calculated mutagenic index. Mutagenicity is expressed as 1 for a mutagen and 0 for a non-mutagen.

4. Conclusions

The aim of this review is to summarise the availability of software models for toxicity prediction. The models are based on a wide variety of approaches, including models that are mechanistically-based or at least mechanistically plausible, and models that have no apparent mechanistic basis.

The availability and quality of the models varies depending on the endpoint: in general, models for acute toxicity are more reliable than "complex" endpoints which comprise a large number of partially understood mechanisms, such as chronic toxicity, systemic toxicities, and reproductive toxicity. For mutagenicity and carcinogenicity, there is a relative abundance of reliable models, mainly based on the fact that these toxic effects are driven by chemical reactivity (electrophilic binding to DNA and/or proteins).

The applicability of individual software models to specific groups of chemicals is outside the scope of this review. Often, the validation characteristics of software models are insufficiently documented to provide sufficient confidence that the models can be used reliably for the chemical(s) of interest. In many cases, details of the predictive algorithm are not transparent. One the one hand, it can be argued that this is not essential since the algorithm is implemented directly by the software; on the other hand, it can be argued that the absence of the algorithm undermines confidence in the predictions, since the basis of prediction is not transparent. Furthermore, the applicability domains of the models are not always transparent, although some software tools provide assessments of prediction reliability that are based on applicability domain considerations.

It is concluded that further research is needed to assess the applicability of currently available software models to chemical groups / inventories of interest, and these assessments should be documented according to internationally agreed principles and reporting formats, such as the QSAR Model Reporting Format (QMRF) and the QSAR Prediction Reporting Format (QPRF). An increasingly range of QMRFs for both software models and literature-based models are freely accessible from the JRC QSAR Model database (http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=QRF). For further information on the documentation of models and their predictions, the reader is referred to other publications (ECHA, 2008; Worth, 2010) as well as guidance notes published by the European Chemicals Agency (ECHA 2010a,b).

5. References

ECETOC (2003). (Q)SARs: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications. Technical Report No. 89, Brussels, Belgium.

ECHA (2008). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R6. European Chemicals Agency, Helsinki, Finland. Available at: http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_en.htm?time=1252064523#r6

ECHA (2010a). Practical guide 5. How to report (Q)SARs. European Chemicals Agency, Helsinki, Finland. Available at: http://echa.europa.eu/doc/publications/practical_guides/pg_report_qsars.pdf

ECHA (2010b). Practical guide 6. How to report categories and read-across. European Chemicals Agency, Helsinki, Finland. Available at: http://echa.europa.eu/doc/publications/practical_guides/pg_report_readacross_categ.pdf

Enoch SJ, Madden JC & Cronin MTD (2008). Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. SAR and QSAR in Environmental Research **19**, 555-578.

Helma C (2006). Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Molecular Diversity **10**, 147-158.

Jeliazkova N, Jaworska J & Worth AP (2010). Open source tools for read across and category formation, in *In Silico Toxicology. Principles and Applications.* MTD Cronin & J Madden (Eds). Royal Society of Chemistry, Cambridge, UK, in press.

Klopman G & Rosenkranz H (1994). Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTI-CASE. Mutation Research **305**, 33-46.

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG & Würtzen G (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food & Chemical Toxicology **42**, 65-83.

Maunz A & Helma C (2008). Prediction of chemical toxicity with local support vector regression and activity-specific kernels. SAR and QSAR in Environmental Research **19**, 413-431.

Mostrag-Szlichtyng A & Worth A (2010). Review of QSAR Models and Software Tools for predicting Biokinetic Properties. JRC Technical Report, EUR 24377 EN. Available from: http://ecb.jrc.ec.europa.eu/qsar/publications/

Rydberg P, Gloriam DE, Zaretzki J, Breneman C & Olsen L (2010). SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. ACS Medicinal Chemistry Letters **1**, 96-100.

Worth AP (2010). The role of QSAR methodology in the regulatory assessment of chemicals, Chapter 13 in *Recent Advances in QSAR Studies: Methods and Applications.* T Puzyn, J Leszczynski & MTD Cronin (Eds). Springer, Heidelberg, Germany, pp. 367-382.

6. Tables

Table 1. Commonly used freely available software tools

Software and developer	Availability	Methodology	Comment
EPI Suite; US EPA http://www.epa.gov/oppt/exposure/pubs/episuite.htm	Freely available	Statistical	Downloadable tool suitable for non-specialised users.
OncoLogic®; US EPA http://www.epa.gov/oppt/newchems/tools/oncologic.htm	Freely available	Knowledge-based	Downloadable tool suitable for users with a limited knowledge of chemistry. Transparent predictions.
Toxtree; EC - JRC http://ecb.jrc.ec.europa.eu/qsar/qsar-tools	Freely available	Hybrid - Statistical and knowledge-based	Downloadable and open source tool suitable for non-specialised users.
Toxmatch; EC - JRC http://ecb.jrc.ec.europa.eu/qsar/qsar-tools	Freely available	Statistical	Downloadable and open source *research* tool for chemical similarity analysis. Supports chemical grouping and read-across. Specialised expertise required.
OECD QSAR Toolbox http://www.oecd.org	Freely available	Hybrid - Statistical and knowledge-based	Downloadable *research* tool for profiling mechanisms, chemical grouping and read-across. Specialised expertise required.
Lazar; In Silico Toxicology (Freiburg university) http://lazar.in-silico.de	Freely available	Statistical	Web-accessible and open source tool under development in EU OpenTox project. Suitable for non-specialised users.
Caesar project models http://www.caesar-project.eu/software/index.htm	Freely available	Statistical	Web-accessible and open source tool developed in EU Caesar project. Suitable for non-specialised users.
PASS http://195.178.207.233/PASS/index.html	Freely available	Statistical	Web-accessible, generates predictions on line upon registration.
T.E.S.T. http://www.epa.gov/nrmrl/std/cppb/qsar/#TEST	Freely available	Statistical	Downloadable and open source tool for toxicity estimation developed by US EPA. Suitable for non-specialised users.

Table 2. Commonly used commercial software tools

Software and developer	Availability	Methodology	Comment
ADMET Predictor; Simulations Plus http://www.simulations-plus.com	Commercial	Statistical	
TOPKAT; Accelrys Inc http://www.accelrys.com	Commercial	Statistical	Algorithms are not transparent.
Pallas software (HazardExpert, ToxAlert; MetabolExpert); CompuDrug Ltd http://www.compudrug.com	Commercial	Knowledge-based	
Derek; Lhasa Ltd http://www.lhasalimited.org	Commercial	Knowledge-based	Knowledge base is transparent.
MultiCASE; MultiCASE Inc http://www.multicase.com	Commercial	Statistical	
MDL QSAR http://www.symyx.com/	Commercial	Statistical	Research tool.
BioEpisteme http://www.prousresearch.com/	Commercial	Statistical	Research tool.
ACD ToxSuite (ToxBoxes); ACDLabs and Pharma Algorithms Product description: http://www.acdlabs.com/products/admet/tox/ Free web application: http://www.pharma-algorithms.com/webboxes/	Commercial (free web application)	Statistical (neural networks)	Easy to use. Algorithms are not transparent.
OASIS TIMES; LMC, Bourgas University, Bulgaria http://www.oasis-lmc.org	Commercial	Hybrid - Statistical and knowledge-based	
Molcode Toolbox; Molcode Ltd, Estonia http://molcode.com/	Commercial	Statistical	Easy to use. Algorithms and underlying data are transparent.
q-Tox	Commercial	Statistical	
CSGenoTox	Commercial	Statistical	

Table 3. Software capable of predicting toxicological endpoints

SOFTWARE (AND DEVELOPER)	AVAILABILITY	ENDPOINT										
		Acute oral toxicity	Repeat dose (chronic) oral toxicity	Genotoxicity (including mutagenicity)	Carcinogenicity	Reproductive (including developmental) toxicity	Endocrine activity / disruptiopn	Hepatotoxicity	Nephrotoxicity (+ urinary tract toxicity)	Neurotoxicity	Cytotoxicity	Immunotoxicity (3)
ACD/Tox Suite (ToxBoxes)	Commercial	•		•			•					
ADMET Predictor (Simulations Plus Inc.)	Commercial		• (1)	•	•		•	•				
BioEpisteme	Commercial				•			•	•			
Caesar project models (Mario Negri Institute)	Freely available			•	•	•						
Derek (Lhasa Ltd)	Commercial			•	•	•	•	•	•	•		•
HazardExpert (CompuDrug)	Commercial			•	•					•		•
Lazar (In Silico Toxicology; Freiburg university)	Freely available		• (1)	•	•			•				
Leadscope (Leadscope)	Commercial			•	•	•		•	•	•		
MCASE/MC4PC (MultiCASE)	Commercial	•	•		•	•	•	•	•		•	
MDL QSAR (MDL)	Commercial	•	• (1)	•	•			•	•			
OASIS-TIMES (Laboratory of Mathematical Chemistry, Bourgas University)	Commercial			•			•					
OncoLogic (US EPA)	Freely available				•							
Pallas Suite including ToxAlert, Cytotoxicity (CompuDrug)	Commercial			•	•					•	•	
TerraQSAR (TerraBase)	Commercial	•					•					
TOPKAT (Accelrys)	Commercial	•	•	•	•	•						
Toxtree (JRC)	Freely available		• (2)	•	•							
Molcode Toolbox (Molcode Ltd)	Commercial		•	•	•		•				•	

SOFTWARE (AND DEVELOPER)	AVAILABILITY	ENDPOINT										
		Acute oral toxicity	Repeat dose (chronic) oral toxicity	Genotoxicity (including mutagenicity)	Carcinogenicity	Reproductive (including developmental) toxicity	Endocrine activity / disruptiopn	Hepatotoxicity	Nephrotoxicity (+ urinary tract toxicity)	Neurotoxicity	Cytotoxicity	Immunotoxicity (3)
PASS (Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Moscow)	Freely available			•	•	•		•	•	•	•	
q-Tox (Quantum Pharmaceuticals)	Commercial	•										
T.E.S.T. (US EPA)	Freely available	•				•						
CSGenoTox (ChemSilico)	Commercial			• (4)								

(1) maximum tolderated dose in humans; (2) Cramer classification tree; (3) immunotoxicity other than skin sensitisaton; (4) prediction of the mutagenic index for Ames test mutagenicity

Table 4. Software capable of predicting toxicological endpoints

SOFTWARE (AND DEVELOPER)	AVAILABILITY	ENDPOINT						
		Skin sensitisation	Skin irritation	Skin corrosion	Eye irritation	Respiratory sensitisation	Eye corrosion	Phototoxicity
ACD/Tox Suite (ToxBoxes)	Commercial		•		•			
ADMET Predictor (Simulations Plus Inc.)	Commercial							
BioEpisteme	Commercial							
Caesar project models (Mario Negri Institute)	Freely available	•						
Derek (Lhasa Ltd)	Commercial	•	•		•	•		•
HazardExpert (CompuDrug)	Commercial		•		•			
Lazar (In Silico Toxicology; Freiburg university)	Freely available							
Leadscope (Leadscope)	Commercial							
MCASE/MC4PC (MultiCASE)	Commercial		•		•			
MDL QSAR (MDL)	Commercial		•		•			
OASIS-TIMES (Laboratory of Mathematical Chemistry, Bourgas University)	Commercial	•						
OncoLogic (US EPA)	Freely available							
Pallas Suite including ToxAlert, Cytotoxicity (CompuDrug)	Commercial							
TerraQSAR (TerraBase)	Commercial		•					
TOPKAT (Accelrys)	Commercial	•	•		•	•		
Toxtree (JRC)	Freely available		•	•	•		•	
Molcode Toolbox (Molcode Ltd)	Commercial				•		•	
PASS	Freely available		•					

European Commission

EUR 24489 EN – Joint Research Centre – Institute for Health and Consumer Protection

Title: Review of Software Tools for Toxicity Prediction

Author(s): Fuart Gatnik M and Worth A

Luxembourg: Publications Office of the European Union

2010 – 13 pp. – 21 x 29,7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-16395-1
doi:10.2788/60101

Abstract

When assessing the properties of chemicals, the easiest and most consistent way of applying (Quantitative) Structure-Activity Relationship ([Q]SAR) models is to use ready-made software that implements the models via a user-friendly interface. A wide range of software tools are available for predicting physicochemical properties, toxicological endpoints and other biological effects, as well as fate in the environment and biological organisms. Typically, a given software package predicts multiple properties and endpoints, and some are extensible, allowing the user to develop new models or include new knowledge. In addition to (Q)SAR models and rulebases that are incorporated in software tools, there is a growing scientific literature which reports thousands of (Q)SARs.

In this report, we give an overview of the software packages that are commonly used in the assessment of chemical toxicity. These software packages are potentially useful in the hazard and risk assessment of chemicals, including for regulatory purposes. However, the applicability of any given software tool needs to be carefully evaluated and documented.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-24889-EN-C

Publications Office

ISBN 978-92-79-16395-1