JRC Scientific and Technical Reports

Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity

Rositsa Serafimova, Mojca Fuart Gatnik and Andrew Worth

EUR 24427 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 59068

EUR 24427 EN
ISBN 978-92-79-16103-2
ISSN 1018-5593
doi:10.2788/26123

Luxembourg: Publications Office of the European Union

ABSTRACT

The review of QSARs for genotoxicity and carcinogenicity was performed in a broad sense, considering both models available in software tools and models that are published in the literature. The review considered the potential applicability of diverse models to pesticides as well as to other types of regulated chemicals and pharmaceuticals. The availability of models and information on their applicability is summarised in tables, and a range of illustrative or informative examples are described in more detail in the text. In many cases, promising models were identified but they are still at the research stage. For routine application in a regulatory setting, further efforts will be needed to explore the applicability of such models for specific purposes, and to implement them in a practically useful form (i.e. user-friendly software). It is also noted that a range of software tools are research tools suitable for model development, and these require more specialised expertise than other tools that are aimed primarily at end-users such as risk assessors. It is concluded that the most useful models are those which are implemented in software tools and associated with transparent documentation on the model development and validation process. However, it is emphasised that the assessment of model predictions requires a reasonable amount of QSAR knowledge, even if it is not necessary to be a QSAR practitioner.

LIST OF ABBREVIATIONS

AhR	Aryl Hydrocarbon Receptor
CHL	Chinese Hamster Lung cells
CHO	Chinese Hamster Ovary cells
DfW	Derek for Windows (Lhasa Ltd)
DNA	Deoxyribonucleic acid
ECHA	European Chemicals Agency
EU	European Union
GHS	Globally Harmonized System of Classification and Labelling of Chemicals
JRC	Joint Research Centre
NCCT	National Center for Computational Toxicology
NTP	National Toxicology Program
OECD	Organisation for Economic Cooperation and Development
QMRF	QSAR Model Reporting Format
QSAR	Quantitative Structure-Activity Relationship
SAR	Structure-Activity Relationship
SAs	Structural Alerts
SCE	Sister Chromatid Exchange
TTC	Threshold of Toxicological Concern
UDS	Unscheduled DNA synthesis

CONTENTS

1. Introduction

To date, hundreds of (Q)SAR models have been published in the literature for predicting genotoxicity and carcinogenicity. The most commonly modelled endpoint for genotoxicity has been Ames test mutagenicity, whereas carcinogenicity models have focused mostly on the rodent bioassay. In this report describes the background biology, the various methodologies used, and summarises some of the key conclusions from the extensive literature concerning the predictivity and applicability of existing models.

2. Background biology

Mutagenicity refers to the induction of permanent transmissible changes in the amount or structure of the genetic material in cells or organisms. These changes may involve a single gene (point mutations), a block of genes or entire chromosomes (structural or numerical chromosome aberrations). Genotoxicity is a broader term and refers to processes that alter the structure, information content or segregation of DNA and which are not necessarily associated with mutagenicity. Such processes include unscheduled DNA synthesis (UDS), sister chromatid exchange (SCE), DNA strandbreaks, DNA adduct formation, and mitotic recombination. In many cases, genotoxicity may lead to cancer. Thus, genotoxicity testing is performed to assess the potential of substances to induce genotoxic effects which may cause heritable damage or lead to cancer in humans. A summary of different types of genotoxicity test is given in Table 1.

Chemicals are defined as carcinogenic if they induce tumours, increase tumour incidence and/or malignancy or shorten the time to tumour occurrence (ECHA, 2008). Traditionally, carcinogens have been identified from epidemiological studies or from animal experiments. 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 summarised and (Woo & Lai, 2003). Recognised mechanisms of non-genotoxic carcinogenicity include peroxisome proliferation, aryl hydrocarbon receptor (AhR) binding, inhibition of gap junctional intercellular communication, oxidative stress, alteration of DNA methylation, endocrine disruption and renerative cell proliferation (Woo & Lai, 2003).

In contrast, in the case of genotoxic carcinogens, the electrophilic theory was introduced more than 25 years ago by James and Elizabeth Miller (Miller & Miller, 1981) who also led the way for the use of (Q)SAR in the prediction of genotoxicity and carcinogenicity. In general, genotoxic carcinogens have the unifying feature that they are either electrophiles or can be activated to electrophilic reactive intermediates (pro-electrophiles). The electrophilic theory of genotoxic carcinogenicity has led to two main (Q)SAR approaches for modelling genotoxic chemicals: a) to identify the electrophilic functional groups or substructures, i.e. to develop SAR models based on structural alerts (SAs); and b) to find molecular descriptors which can be quantitatively related to the activity of the chemicals, i.e. to develop QSARs. Most studies have provided qualitative models (SARs), which provide a “coarse-grain” approach for the identification of genotoxic potential. In addition, although more challenging,

numerous studies have attempted to develop quantitative models (QSARs), which provide a more precise means of assessing genotoxicity and carcinogenicity, mainly for congeneric sets of chemicals.

3. Regulatory classification of mutagens and carcinogens

A summary of the former EU and new GHS criteria for classifying substances on the basis of mutagenicity and carcinogenicity is given in Table 2. These schemes classify substances into different levels of concern based on the strength and weight of available evidence. Expertise in ergaultory toxicology is required to apply the classification criteria in a consistent manner.

4. Databases

A number of web-based databases provide access to experimental data for mutagenicity carcinogenicity, and are thus useful for (Q)SAR development and assessment. Until recently, public toxicity databases were constructed primarily as "look-up-tables" of existing data, and most often did not contain chemical structures. However, modern technologies are now providing powerful tools to create new types of searchable databases, providing an effective means of linking toxicity with chemical structure. Some databases only allow information to be retrieved chemical-by-chemical but others provide the possibility to download an entire database. Several reviews have surveyed the status of public toxicity databases (Richard & Williams 2003; Benigni *et al.*, 2008a). In this section, a short explanation is given of the main databases, and a summary is presented in Table 3.

CPDB: The Carcinogenic Potency Database (CPDB) (http://potency.berkeley.edu/cpdb.html) provides a unique resource of the results of 6540 chronic, long-term animal cancer tests on 1547 chemicals. The CPDB provides easy access to the bioassay literature, with qualitative and quantitative analyses of both positive and negative experiments that have been published over the past 50 years in the general literature through 2001 and by the National Cancer Institute/National Toxicology Program through 2004. The CPDB is downloadable in pdf, xls and txt formats, and is searchable by chemical name, CAS number, or author.

Danish QSAR database: The Danish EPA has developed a (Q)SAR database as a free source of predicted toxicities (not experimental data) for over 166,000 chemicals. For information on genotoxicity, the database contains predictions for various types of Ames test as well as a range of in vitro endpoints: chromosomal aberrations (CHO and CHL cells), gene mutation assays (mouse lymphoma/tk, CHO/hprt) and Unscheduled DNA Synthesis (UDS) in rat hepatocytes. A range of in vivo models are also included (Drosophila SLRL, mouse micronucleus, rodent dominant lethal assay, mouse Sister Chromatid Exchange (SCE) in bone marrow and mouse Comet assay). All these models were derived using the MULTICASE software. For information on carcinogenicity, the database includes (in addition to the genotoxicity models), eight MULTICASE FDA cancer models, rodent carcinogenic potency, hepatospecificity, oestrogenicity and aryl 33 hydrocarbon (AH) receptor binding. The Danish QSAR database can be freely accessed over the internet from the JRC website (http://ecbqsar.jrc.it/). The database includes a flexible system for chemical structure and parameter searching. This database should be used with caution, since the data are not experimental data but predictions, many of which will not resultfrom use of the more recent models.

DSSTOX: Both the CPDB and the online NTP database have been "chemically-indexed" in the DSSTox (Distributed Structure-searchableToxicity) database (http://www.epa.gov/ncct/dsstox), developed by US EPA's National Center for Computational Toxicology (NCCT). DSSTOX emphasises quality procedures for accurate and consistent chemical structure annotation of toxicological experiments. Chemical structures and summary mutagenicity and carcinogenicity data have been published for the entire CPDB inventory (www.epa.gov/ncct/dsstox/sdf_cpdbas.html), along with the URL address locating the specific chemical data webpage on the CPDB website provided for each indexed chemical substance. Chemical structures and indicators of data availability

have also been provided for the entire chemical inventory of the online NTP database, for each of the four main NTP study areas (developmental, immunological, genetox, and chronic cancer bioassays).

ECHA CHEM: Information on susbtances evaluated under REACH are provided by ECHA CHEM, which is hosted by the European Chemicals Agency (ECHA) (http://echa.europa.eu/chem_data_en.asp).

ESIS: The European chemical Substances Information System (ESIS) is a freely accessible data via the JRC ex-ECB website (http://ecb.jrc.ec.europa.eu/esis/) providing information on chemicals related to: EINECS (European Inventory of Existing Commercial chemical Substances); ELINCS (European List of Notified Chemical Substances); NLP (No-Longer Polymers); the Biocidal Products Directive (BPD) active substances listed in Annex I or IA of Directive 98/8/EC or listed in the so-called list of "non-inclusions"; PBT (Persistent, Bioaccumulative, and Toxic) or vPvB (very Persistent and very Bioaccumulative) assessments of Existing Substances; Classification and Labelling (C&L), the Export and Import of Dangerous Chemicals listed in Annex I of Regulation (EC) No 689/2008; High Production Volume Chemicals (HPVCs) and Low Production Volume Chemicals (LPVCs), including EU Producers/Importers lists; IUCLID Chemical Data Sheets; EU Priority Lists and EU Risk Assessments produced under the Existing Substances Regulation (ESR).

EXCHEM: This database (http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp) was developed by the Chemicals Investigation Promoting Council, Japan and was supervised by Office of Chemicals Safety Evaluation and Licensing Bureau Pharmaceutical and Food Safety Bureau Ministry of Health, Labour and Welfare, Japan. EXCHEM contains data for Ames mutagenicity, chromosomal aberrations and mouse micronucleus assays for more than 250 HPV chemicals. Most of the information is in Japanese but there is also information in English. The database is searchable by CAS number and name.

GAP: The Genetic Activity Profile Database was initially developed by US EPA and IARC, and now by ILS (http://www.ils-inc.com). Data on approx 300 chemicals were compiled from volumes 1-50 of the IARC Monographs and on 115 compounds identified as Superfund Priority Substances. The data (qualitative and quantitative) are displayed as graphic profiles and data tables for up to 200 short-term assays that range from bacterial tests to human studies in vivo. The latest version was produced in 2000 (GAP2000). A CD rom is available on request from ILS.

IARC: The International Agency for Research on cancer (IARC) website provides access to the *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* (http://monographs.iarc.fr/index.php). The IARC Monographs have reviewed more than 900 chemcials and have identified more than 400 known, probable and possible carcinogens. The monographs are searchable by key word, CAS number, synonym or chemical name.

ISSCAN: This 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.

NTP: The US National Toxicology Program (NTP) (http://ntp.niehs.nih.gov) provides access to publicly available data from more than 500 two-year, two species, toxicology and carcinogenesis studies collected by the NTP and its predecessor, the National Cancer Institute's Carcinogenesis Testing Program. The NTP database also contains results relating to approximately 300 toxicity studies from shorter duration tests and from more than 2000 genetic toxicity studies, including both in vitro and in vivo tests. In addition, test data from the immunotoxicity, developmental toxicity and reproductive toxicity studies are continually being added to this database. The data can be accessed as technical reports; the user can browse them directly, make searches (by chemical name or CAS number, for example), or download the reports in pdf form.

ToxRefDB: This database (http://www.epa.gov/ncct/toxrefdb/) was developed by the NCCT, in partnership with EPA's Office of Pesticide Programs (OPP), to store data from in vivo animal toxicity studies. The original aim was to populate ToxRefDB with pesticide registration toxicity data that has been historically stored as hard-copy and scanned documents by OPP. ToxRefDB currently includes chronic, cancer, sub-chronic, developmental, and reproductive studies on 330 chemicals, many of which are pesticide active ingredients. ToxRefDB is downloadable in xls format but without structural information.

TOXNET: The TOXNET database of the US National Library of Medicine (NLM) (http://toxnet.nlm.nih.gov) is a cluster of different databases, collecting information on toxicology, hazardous chemicals, environmental health, and toxic releases. From the website, it is possible to search within and across the databases by several identifiers, such as chemical name, CAS number, molecular formula, classification code, locator code, and structure or substructure. Among the TOXNET databases, the Chemical Carcinogenesis Research Information System (**CCRIS**) and the **GENE-TOX** databases deal specifically with mutagenicity and carcinogenicity data. CCRIS contains over 9000 chemical records with animal carcinogenicity, mutagenicity, tumour promotion, and tumor inhibition test results provided by the National Cancer Institute (NCI). Test results have been reviewed by experts in carcinogenesis and mutagenesis. GENE-TOX was developed by the US EPA and contains genetic toxicology (mutagenicity) test data, resulting from expert peer review of the open scientific literature, on over 3000 chemicals.

5. Structure-activity relationships for non-congeneric chemicals

One of the simplest and best known approaches to predict genotoxicity and carcinogenicity for structurally diverse chemicals is based on the use of SAs, sometimes accompanied by modulating factors. This section traces the development of the main SA-based approaches.

The first list SAs for mutagenicity was proposed by Ashby (Ashby 1985), who subsequently extended the lists with additional SAs as well as some detoxifying functionalities (Ashby and Tennant 1988). The resulting 19 SAs are referred to collectively as the Ashby poly-carcinogen model, often represented by a fictitious chemical structure containing all of the alerts (Figure 1).

Bailey *et al*. (Bailey *et al*. 2005) generated a list of 33 SAs for regulatory use in the US FDA. This list was based on the Ashby alerts and on a list compiled by Munro *et al*. (Munro *et al*., 1996).

Kazius *et al*. (Kazius *et al*. 2005) produced another list of SAs by using a combination of data mining and expert knowledge. This list contains 29 SAs accompanied with detoxifying fragments and is reported to classify its training set (2401 mutagens and 1936 non-mutagens) with an accuracy of 82%.

The Laboratory of Mathematical Chemistry (LMC, Bourgas, Bulgaria) has developed a list of 17 SAs. These are implemented in the OASIS TIMES software (see below).

More recently, Benigni and Bossa (2008) combined the above sources and some information from the OncoLogic software to generate a list of 33 SAs. Five of the Benigni-Bossa alerts refer to non-genotoxic mechanisms of action and several of them have accompanying modulating factors. The reported accuracy of prediction is 78% for mutagenicity and 70% for carcinogenicity, based on an analysis of the ISSCAN database. The Benigni-Bossa SAs is implemented in the Toxtree software and in the OECD Toolbox (see below).

The relationships (overlaps) between the different lists of SAs are illustrated in Figure 2. For the purpose of this Venn diagram, comparison between alerts was based simply on the main functional group in the SA and not on exact matches between whole alerts. In the different lists, SAs are described with different levels of detail. At present, the most comprehensive list of SAs is the Benigni-Bossa list (the number of the SAs containing in the Bailey list is the same but many of these alerts contain the same functional group and different substitutes).

6. Types of expert systems

An expert system has been defined as any formalised system that is often, but not necessarily, computer based, and that can be used to make predictions on the basis of prior information (Deaden *et al*., 1997). Expert systems (and their implementation in software tools) are based on three main modelling approaches referred to rule-based, statistically-based, or hybrid methods.

Rule-based systems contain "if-then-else" rules that combine toxicological knowledge, expert judgment and fuzzy logic. Commonly used software tools based on this approach include OncoLogic (Woo *et al*. 1995), Derek (Sanderson & Earnshaw 1991; Ridings *et al*., 1996) and HazardExpert (Smithing & Darvas 1992). Derek and HazardExpert can be used in conjunction with their sister programs Meteor and Metabolexpert to predict the genotoxicity and carcinogenicity potential of metabolites as well as parent compounds. In addition to these commercial tools, models included in the freely available Toxtree software and the OECD Toolbox are rule-based.

Statistically-based systems use a variety of statistical, rule-induction, artificial intelligence, and pattern recognition techniques to build models from non-congeneric databases. Statistically based systems are included in the commercial tools MultiCASE and TOPKAT, and the publicly available Lazar and CAESAR models. In addition, many models published in the literature and not implemented in software are statistically based.

Hybrid models are based on a combination of knowledge-based rules and statistically-derived models. These are based on the general idea that, within the structural space of a single SA (considered to represent a single interaction mechanism), statistically derived models can quantitatively predict the variation in the reactivity of the alert conditioned by the rest of the molecular structure. Examples of the hydrid approach include models implemented in the OASIS TIMES (Mekenyan *et al*,. 2004; Mekenyan *et al*., 2007; Serafimova *et al*., 2007) as well as some literature-based models not implemented in software.

A good example of hydrid model in the literature is the Purdy model for carcinogenic potential (1996). Purdy's model is a decision tree with 11 individual rules or QSARs for different chemical classes. The QSARs are relatively simple in form, ranging from an allowed range or cut-off of a computed property value (for example, logP, molecular volume, E(LUMO), partial atomic charges and superdelocalisibilities), to a three dimensional constraint specifying a fixed distance between lone pairs of electrons. The author reported a high accuracy of classification (92% and 88% for the training and test sets, respectively) The main advantages of the Purdy model is that is based on simple rules, and the reasoning underlying the prediction can be seen. The main disadvantages are that the rules is restricted to a few chemical classes and do not have clear mechanistically interpretations. In addition, the model is not yet automated so is not practically available for routine use. However, this limitation could easily be addressed by a software programmer.

The advantages and disadvantages of the three main approaches are summarised in Table 4.

7. Software tools

Genotoxicity and carcinogenicity prediction is featured in a wide range of commercial and freely available software tools, the most commonly used of which are described below. A summary is given in Table 5.

CAESAR: A statistical model for mutagenicity was developed and released as an open source software tool in the frame of the EU CAESAR project (http://www.caesar-project.eu/). Gini and colleagues (Ferrari *et al*., 2009) used the Support Vector Machine (SVM) classification method to develop a model based on the 4225 compounds from the Kazius-Bursi mutagenicity database. The authors reported correct classification rates of 92.3% and 83.2% for the training and test sets, respectively. The results were considered to be in the same order of magnitude as experimental error. No information was provided about the applicability domain. In order to minimise the number of false

negatives, the authors investigated the combined use of their model with some SAs from the Benigni-Bossa rulebase (using Toxtree). As expected, the results showed that the number of false negatives could be reduced but this was at the expense of increasing the number of the false positives. This resulted in a slight change in accuracy of 1.1% less. The authors concluded that by using the so-called "cascade model", a classification accuracy close to the reliability of the Ames test data could be achieved. In fact, they achieved this accuracy by using just the SVM classification algorithm.

In the CAESAR project, two complementary approaches (regression and classification) were applied to develop models for carcinogenicity. The original dataset, extracted from the CPDB, consisted of 805 chemicals with rat TD50 values. This dataset was spit into training (n=644) and test (n=161) sets. The regression model was developed by applying a Monte Carlo method to TD50 data. The classification model was developed by applying the Counter-Propagation Artificial Neural Network (CP-ANN) method and a set of MDL descriptors. The authors reported an accuracy of classification of 91-96% for the training set and 68-74% for the test set.

Derek: This is a commercial system developed and marketed by Lhasa Ltd. The development of knowledge-based rules in Derek is overseen by collaborative group which consists of representatives from commercial, educational and non-profit organisations. The current version of Derek (v. 12; released in December 2009) contains 89 alerts for mutagenicity, 77 for chromosome damage, and 61 for carcinogenicity. The chromosome damage alerts are based primarily on data from the in vitro chromosome aberration test, however additional assays (in vivo chromosome aberration test, in vitro and in vivo micronucleus test and L5178Y TK+/- assay) have been considered when writing alerts, and some alerts are entirely based on alternative assays. The chromosome damage alerts cover both direct DNA damage and other genotoxic mechanisms.

The hazard assessment in Derek is usually justified with relevant literature references, which give the user more confidence in the predictions. The main advantages of the system are the transparency in the predictions, the fact that the rule development is peer-reviewed by a user group, and new rules can be added easily. It should be noted that DfW does not provide negative predictions (the absence of a predicted hazard simply means that no relevant alerts were identified; it does not necessarily mean the absence of hazard). A QMRF for the DfW mutagenicity model is available in the JRC QSAR Model Database.

In a recent study, Crettaz and Benigni (2005) assessed the ability of DfW to qualitatively predict the rodent carcinogenicity and the genotoxic potential of 60 pesticides registered in Switzerland. The percentage of false negatives was 31% for carcinogenicity. The associated sensitivity of 69% indicates that most of the pesticides with positive rodent bioassay results were detected by DfW. On the other hand, the low specificity of 47% is equivalent to a false positive rate of 53%. Such chemicals would be predicted as carcinogenic while rodent bioassays would not confirm this potential.

In a recent EFSA-sponsored study on the applicability of TTC concept to pesticides and their metabolites carried out by the Chemicals Regulation Directorate (UK), DfW was used to predict the genotoxicity and carcinogenicity of 100 randomly selected pesticide active substances (CRD, 2009). It was concluded that DfW is not reliable predictor for these two endpoints. However, authors note that the dataset of 100 compounds is inevitably biased by excluding compounds with high genotoxic potential. When the analysis focused on compounds of greatest potential concern, those with positive study results for both tumours and genotoxicity, the predictivity based on an alert for either genotoxicity or carcinogenicity was good (10/12 correct), although the ratioanle for the prediction was often incorrect. It was concluded that additional work should be performed on the reliability of genotoxicity predictions from DfW and other (Q)SAR programs.

HazardExpert: The HazardExpert models (Smithing and Darvas, 1992) are proprietary, the software now being marketed by CompuDrug Ltd. The program works by searching the query structure for known toxicophores that are derived from the literature in the field of QSAR or from the US EPA and

Interagency Testing Committee (ITC) monographs. Predictions are made in four levels of toxicity, taking into account the effects of bioavailability and bioaccumulation.

As an evaluation of its ability to predict human and animal carcinogenicity, 192 agents evaluated in the IARC Monographs (volumes 1-42) were processed through Hazardexpert (Dearden *et al*., 1997). The difference between the classification in the IARC list and that assessed by Hazardexpert was used for the analysis. As a result, some important fragments were found to be missing from the toxic fragments database, including vinyl chlorides, organophosphates, organometallic compounds, and isocyanates.

In a separate evaluation study based on 80 NTP chemicals (56 rodent carcinogens; 24 non-carcinogens), HazardExpert was found to have an overall concordance of 51%, and to be good at identifying non-carcinogens (specificity of 81%), but poor at identifying carcinogens (sensitivity of 36%).

Lazar: The predictive performance of Lazar was assessed by Helma (2006), who used a training set of 1447 chemicals from the CPDB and 4337 chemicals from the Kazius/Bursi database for external validation. Leave-one-out and external validation experiments indicated that Salmonella mutagenicity can be predicted with 85% accuracy for compounds within the applicability domain of the CPDB. The LOO accuracy of Lazar predictions for rodent carcinogenicity was reported as 86%, and the accuracies for other carcinogenicity endpoints varied between 78 and 95% for structures within the applicability domain. A QMRF for Lazar mutagenicity is under preparation.

MDL QSAR: This is a commercial software tool originally developed by MDL and now marketed by Symyx Ltd (see above). The software has been used by Contrera *et al*. (2005a) to develop discriminant models for bacterial mutagenicity using a dataset of over 3000 chemicals and with sensitivity, specificity and concordance of 81%, 76% and 81%, respectively. These models are not readily transferable and thus of limited practical used.

Valerio *et al*. (2007) evaluated the utility of a discriminant analysis modelling approach (MDL-QSAR) to estimate the carcinogenic potential of small, organic, naturally occurring chemicals found in the human diet. They used as a training set of over 1200 chemicals, comprised primarily of pharmaceuticals, industrial chemicals and some natural products. A sample set of 123 naturally occurring chemicals found in the human diet with known low and high risk potential as rodent carcinogens, and a control group of 19 synthetic dietary chemicals with known high carcinogenic potential were use as a test set. The predictive performance based on this test set was an overall concordance of 80%, a sensitivity of 97%, and a marginal specificity of 53%. These results support the usefulness of the MDL-QSAR software in identifying the rodent carcinogenic potential of naturally occurring organic chemicals. As also noted by the authors, further assessment of the software will be needed for a wider range of dietary chemicals.

MultiCASE: The MultiCASE models are proprietary. The software has been widely used by the Danish EPA to build models for a range of genotoxicity and carcinogenicity endpoints. Genotoxicity models include Ames mutagenicity (two models), direct mutagenicity, base-pair mutagenicity, frame-shift mutagenicity, chromosomal aberrations (two models), mouse micronucleus assay, mouse sister chromosomal exchange. Carcinogenicity models include rat, mouse, female, male carcinogenicity, TD50 rat, mouse carcinogenicity. The Danish EPA reported concordances between 56-100% for the different models (http://www.mst.dk/English/Chemicals/Substances_and_materials/QSAR/). More information on these models, and pre-generated predictions for over 166,000 chemicals and can be found at the EPA website (http://www.mst.dk/English/Chemicals/) as well the as JRC website (http://ecbqsar.jrc.it/). The JRC version of the Danish database includes a flexible system for chemical structure and property searching.

In a study by Matthews and Contrera (1998), MCASE was used with numerous in-house modifications of the system, including: a) enhancement of the size of the control database modules; b) optimization of MCASE SAR assay evaluation criteria; c) incorporation of a carcinogenic potency scale for control compound activity and MCASE biophores; d) construction of individual rodent gender and species-

specific modules; and e) use of assay acceptance criteria for query and control database compounds. The optimised system was reported to demonstrate excellent sensitivity for carcinogens (97%), and specificity for non-carcinogens (98%), in a test set of 126 chemicals. While these seem like very promising results, they are not verifiable: the MCASE model is not readily transferable, and the data used are confidential and therefore are not available for use in the development other modelling methodologies or to assist in the assessment of the improved MCASE system. Similar studies have been carried out more recently by Matthews and co-workers, with more extensive datasets (Matthews *et al*., 2006a, 2006b).

OASIS/TIMES: The hybrid approach has been used by Mekenyan and colleagues to develop models for Ames mutagenicity and chromosomal aberration. These models are implemented in the OASIS TIMES software. Each SA is accompanied by modulating factors, to account for the influence of the rest of the molecule, as well as with defined and documented mechanism of interaction with DNA (for the mutagenicity model) and/or nuclear proteins and enzymes (for the chromosomal aberration model). Expert knowledge was used to define the SAs and the mechanistic basis for prediction (interaction with biological macromolecules) is well documented. A pattern recognition approach (COREPA) was used to derive modulating factors for each SA.

In contrast to other models for genotoxicity, the OASIS models include a liver metabolic simulator based on documented metabolic pathways. The training sets used for the models were split into chemicals that are mutagenic without metabolic activation, mutagenic after metabolic activation, and non mutagenic with and without metabolic activation. This is an important advantage of the OASIS/TIMES software, because the role of metabolism is rarely accounted for. To demonstrate the importance of metabolism, the authors showed that when predictions are obtained without using the metabolic simulator for chemicals known to be active after metabolic activation, the sensitivity was dramatically decreased to 22%. The main disadvantage of the OASIS/TIMES software is that it is a little bit slower than other software.

OECD Toolbox: The current version of this software (http://toolbox.oasis-lmc.org) implements two so-called "profilers" connected with genotoxicity and carcinogenicity. The first one is the Benigni-Bossa rulebase (Benigni *et al*., 2008b) and the second is the OASIS DNA binding profiler developed by LMC Bourgas (Serafimova *et al*., 2007). The OECD Toolbox also includes a few databases with experimental data that can be used to support grouping and read-across: a) the ISSCAN database – 1149 chemicals containing data for carcinogenicity and Ames mutagenicity; b) the EXCHEM database – 256 chemicals containing data for Ames mutagenicity, chromosomal aberrations and mouse micronucleus assay; c) the OASIS Genotox database – 2684 chemicals with data for Ames mutagenicity and chromosomal aberrations as well as data for metabolism. The Toolbox also includes the Danish EPA database containing predicted data of different genotoxicity and carcinogenicity endpoints for more than 166,000 chemicals.

Oncologic: Oncologic is a knowledge-based system developed by LogicChem Inc (Woo & Lai, 2005). It can be freely download from the US EPA website (http://www.epa.gov/oppt/newchems/tools/oncologic.htm). It uses a series of hierarchically ordered rules to describe and predict the carcinogenic potential of chemicals. These rules have been developed in collaboration with the structure-activity team at the US EPA's Office of Pollutions Prevention and Toxics. The current version (December 2009) includes over 40,000 rules based on knowledge and generalisations derived from the examination of more than 10,000 chemicals belonging to approximately 50 chemical classes. The main advantages of the system are that it includes a large amount of human knowledge, the predictions are restricted to those classes for which adequate knowledge is available, and reports usually include supporting information to justify the prediction. The main disadvantages are that there is no possibility for batch calculations, and the system requires some chemistry expertise, with the user needed to take decisions step-by-step during the prediction.

TOPKAT: The TOPKAT models are proprietary. According to Enslein (1994), the accuracy of mutagenicity and carcinogenicity predictions are extremely high: 98% (against a mutagenicity dataset

of 1083 chemicals) and 99.6% (against a carcinogenicity dataset of 705 chemicals), respectively. However, some subsequent evaluation studies (Prival 2001) indicate that for external sets of chemicals, the accuracy of TOPKAT prediction is considerably lower (40-75% against datasets of 30-40 chemicals) and these results were not significantly better when the analyses were restricted to predictions made inside the OPS.

Toxtree: Toxtree currently includes two modules for mutagenicity and carcinogenicity prediction – the Benigni-Bossa rulebase (which expands on the Ashby supermutagen model; see above) and the ToxMic rulebase for the in vivo micronucleus assay (Benigni *et al*., 2010). The developers have reported an accuracy of prediction around 70% for carcinogenicity, 78% for mutagenicity and 59% for the *in vivo* micronucleus assay (Benigni *et al*., 2009).

MolCode Toolbox: This commercial tool developed and marketed by Molcode Ltd includes modules for Ames mutagenicity and female rat carcinogenicity.

8. Literature reviews and comparative evaluation studies

The literature relating to the *in silico* prediction of genotoxicity and carcinogenicity is huge, with more than 100 papers dedicated to (Q)SARs. A list of reviews, expert opinions and evaluation studies is given in Table 6, whereas details of primary research studies published in the past 10 years for genotoxicity and carcinogenicity models are given in Tables 7 and 8, respectively. Given the extent of the literature in this field, this section focuses on key findings from evaluation studies that have compared the performances of different models, including software models. A number of such "multi-model evaluation studies" have been published (Table 6), which have been well summarised by Benigni *et al*. (2007). A representative selection of evaluation studies is described below.

Zeiger *et al*. (1996) used 100 NTP chemicals to compare ability of two computer systems (TOPKAT and CASE), one physicochemical screening test and one human expert system to predict Salmonella whereas the physicochemical system produced a lower (61%) concordance. Similar results for DfW and TOPKAT were reported by Cariello *et al*. (2002) - the accuracy of prediction of Ames mutagenicity by DfW was 65% (against a dataset of 400 GlaxoSmithKline chemicals). The overall concordance for TOPKAT was 73% but it should also be noted that TOPKAT was capable to predict 300 out of the 400 chemicals.

Two other evaluation exercises were devised by the NTP. In the first exercise (Benigni 1997), regarding the prediction of rodent carcinogenicity for 44 chemicals, different approaches were compared: computer-based systems (CASE, TOPKAT, DfW, COMPACT), human experts (Benigni, Tennant and Ashby, Weisburger and Lijinsky) and experimental data. For the structure-based approaches the overall accuracy was in the range 50-65%, whereas the Tennant and Ashby approach attained an accuracy of 75%. In the second exercise, based on 30 chemicals, the list of methods was extended. In this second exercise, the highest overall accuracy achieved was 60-65% (Benigni and Zito, 2004).

An informative survey was performed by Benigni and Bossa (2008). They summarised the outcomes of a series of external prediction exercises performed by various investigators with three non-local models in the commercial domain: MultiCase, TOPKAT, and DfW. The results included those obtained in the prospective prediction exercises by the NTP as well as several studies performed by companies using in-house datasets. The common characteristic of these studies is that the chemicals to be predicted were different from those used in the training sets by the model developers, and were performed independently. It was found that the predictions for external chemicals vary considerably both in terms of overall accuracy and in terms of relative proportions of true and false positives. As an example, Figure 3. shows a Receiver Operating Curve (ROC) graph which summarises the predictive performance of the DfW software system in a series of external prediction exercises for carcinogenicity (a) and mutagenicity (b). In the ROC Curve, a perfect performance would be in the top-left corner (100% true positives, 0% false positives). The line represents performances that could

be obtained by chance, so for a model to be better than chance, it needs to be in the top-left quadrant of the figure. The observations for TOPKAT and MultiCase were similar to those for DfW. These findings contrast with the usually good performances reported by the model developers, as assessed on large non-congeneric databases.

Mayer et.al. (2008) compared the abilities of several computer-based models (OncoLogic, MultiCASE, Ashby-Tennant structural alerts) to predict carcinogenicity with several genotoxic tests (Ames, mouse lymphoma assay and chromosomal aberration). Using data for 650 chemicals from the CPDB database, the authors found that the (Q)SAR methods produced a higher concordance frequency (71% to 88% versus 62% to 75% for genetic tests) and lower percentage of false negatives (8.6% to 27% versus 20% to 39% for genetic tests).

Similar findings were reported by Snyder (2009) who compared the carcinogenicity test results of 545 marketed drugs with genotoxicity assay results. The data were taken primarily from the Physicians Desk Reference (PDR; 1999-2008). The analysis included an evaluation of the predictivity of Derek and MCASE/MC4PC. The authors reported a low predictability of carcinogenicity based on the genotoxic assays. The two software programs performed reasonably well, and better than the in vitro genotoxic assays, in terms of high specificity (low percentage of false positives) and overall concordance. The weakness of the software was the low sensitivity of both programs, but it was still higher than that performed from in vitro assays.

Building on the study using MDL QSAR study by Valerio and colleagues (2007), Mazzatorta *et al*. (2009) examined the performance of a wider series of *in silico* tools for predicting the carcinogenicity of natural chemicals. They extracted 50 chemicals from the Valerio data set, the majority of which were pyrrolidine alkaloids and phenolic-type compounds (20 high-risk and 30 low-risk chemicals in terms of carcinogenicity) and they applied two statistical models (MC4C and Lazar) and three knowledge-based expert systems (Toxtree, Derek and OncoLogic). Based on the results, the authors categorised the models into three performance groups. The first group - high sensitivity (>90%) and low to medium specificity (<68%) - includes OncoLogic. The second group - medium sensitivity and specificity (between 58 and 80%) - includes MC4PC and Lazar. The third group - low sensitivity (<41%) and high specificity (>74%) includes Derek and Toxtree. These results indicate that the carcinogenicity potential of naturally occurring chemicals can be reliably predicted by using a battery of software tools that combine high sensitivity (thereby minimising false negatives) and high specificity (thereby minimising false positives).

The battery approach was also investigated by Matthews *et al*. (2008), who explored the combined use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM and Derek in predicting carcinogenic potential. They found that the use positive predictions from any two programs showed better overall performance than use of the single programs alone, with a sensitivity of about 85% and specificity of 58%. When focussing on defined modes of action, the authors reported that consensus positive predictions of carcinogenicity by two QSAR programs could detect 99% of the carcinogens (including both genotoxic and nongenotoxic carcinogens) in the study.

The results of (Q)SAR evaluation studies such as those described here can also be placed into context by considering the results of a study by Kirkland *et al*. (2005) who evaluated the abilities of some of the most commonly used *in vitro* genotoxicity tests (Ames, mouse lymphoma assay [MLA], in vitro micronucleus [MN] and chromosomal aberrations [CA] as well as battery of three of these tests) to discriminate rodent carcinogens from non-carcinogens. The authors based their comparison on a large dataset of over 700 chemicals compiled from the CPDB, NTP and IARC databases as well as other publications. It was found that combinations of two and three test systems had greater sensitivity than individual tests resulting in sensitivities of around 90% or more, depending on the test combination The sensitivity of individual methods was between 59% (for Ames for over 500 chemicals) and 79% (for MN for over 80 chemicals). The specificity of the Ames test was reasonable (73.9%), but all mammalian cell tests had a low specificity (below 45%), and this was reduced in combinations of two and three test systems. When a battery of three tests was investigated, 75–95% of the non-carcinogens

were incorrectly predicted (i.e. were false positives) results in at least one test in the battery. This highlights deficiencies in the current ability to extrapolate from in vitro mutagenicity results to in vivo carcinogenicity.

In a recent study by Hansen et. al. (2009), and a large Ames mutagenicity data set comprising about 6500 non-confidential compounds was compiled and made publicly available (http://ml.cs.tu-berlin.de/toxbenchmark/). They used the dataset to compare the predictive performances of three commercial tools (Derek, MultiCASE, and an off-the-shelf Bayesian machine learner in Pipeline Pilot) with four non-commercial machine learning implementations (Support Vector Machines, Random Forests, k-Nearest Neighbours, and Gaussian Processes). PipelinePilot, trained with the developed data set, showed the best predictive perfromance of the three commercial tools followed by MultiCASE. The expert system Derek gave the lowest sensitivity and specificity of all considered models. However, closer examination of the results reveals that the difference between the best commercial model (Pipeline Pilot) and the best machine learning approach (SVM) is a sensitivity of just a few percent, so it is difficult to draw firm conclusions. In general, machine learning algorithms are expected to perform better in cases such as this where they derive their knowledge exclusively from the training data, as opposed to models such as MultiCASE and Derek, which have rules derived from other datasets ot based on expert knowledge. This study is useful not only in terms of the dataset which is made publicly available, but also because it demonstrates the power of machine learning approaches. Such approaches are particularly useful in model discovery, after which optimal models could be used as the basis for developing models with a mechanistic basis.

9. Conclusions

When considering computational models for genotoxicity and carcinogenicity prediction, it should be remembered that these endpoints are based on multiple mechanisms of action, and are experimentally assessed by multiple tests, the results of which require expert interpretation. Thus, the in silico models are often modelling the “higher-level” interpretation of one or more experimental results rather than the “lower-level” experimental data themselves. This is different to models for some other endpoints (e.g. acute toxicity) where the models can be based directly on experimental data (e.g. LD50 values).

At present, (Q)SAR methods are more reliable for predicting genotoxic potential than carcinogenic potential. Carcinogenicity prediction represents a considerable challenge due to the multitude of possible mechanisms of toxic action. The prediction of non-genotoxic carcinogenicity and carcinogenicity in humans is especially problematic. Models for predicting carcinogenic potency are lacking.

The accuracy of Ames mutagenicity prediction is typically 70-75%, whereas for carcinogenicity it is generally between 50-75%, depending on the (Q)SAR and dataset used. This is reasonable taking into account the complexity of the carcinogenicity endpoint, and the fact that models do not explicitly include ADME properties, which could be critical steps in the carcinogenic process. An important direction for future research would be to incorporate ADME considerations in the overall prediction. It will also be important to build more models for non-genotoxic mechanisms of action.

When evaluating (Q)SARs and software models on the basis of published papers, it is easy to obtain mixed messages. Thus, it is important to critically evaluate the design of the study. The accuracy of model prediction reported by the model developers is usually quite high for both training and test sets. However, this can be deceiving and is generally a consequence of the way in which the training and test sets were formed by splitting available datasets. In contrast, the accuracy of the prediction for external and independently chosen test sets is not so high.

When using computational models for regulatory purposes, it is concluded that predictions of genotoxicity and carcinogenicity should not be based on the use of any single model alone, but on a Weight of Evidence approach including information is possible from all available sources (QSARs, read across, in vitro test methods). Studies such as those performed by Valerio *et al*. (2007), Coterrill

et al. (2008) and Mazzatorta *et al*. (2009) support the usefulness of computational tools, especially when used in batteries that combine high sensitivity models (to minimise false negatives) with high specificity models (thereby minimising false positives). Building on such studies, there is a need for further research aimed at developing and assessing model batteries and integrated testing strategies for genotoxicity and carcinogenicity.

As with all endpoints, predictions should always be interpreted by an expert with knowledge of the endpoint and an appreciation of the strengths and limitations of the specific model applied. An essential piece of information is the applicability domain of the model, and the reliability of prediction for the chemical of interest. Unfortunately, this information is often not available or easily obtained.

10. References

Andrews LE, Banks TM, Bonin AM, Clay SF, Gillson AE & Glover SA (2004). Mutagenic N-Acyloxy-N-alkoxyamides: Probes for Drug-DNA Interactions. Australian Journal of Chemistry 57(4), 377-381.

Ashby J (1985). Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. Environmental Mutagenesis 7, 919–921.

Ashby J & Paton D (1993). The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. *Mutation Research* **286**, 3-74.

Ashby J & Tennant RW (1988). 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. Mutation Research 204(1), 17-115.

Bacha PA, Gruver HS, Den Hartog BK, Tamura SY & Nutt RF (2002). Rule Extraction from a Mutagenicity Data Set Using Adaptively Grown Phylogenetic-like Trees. Journal of Chemical Information and Computer Sciences 42(5), 1104-1111.

Bahram H (2005). Optimal QSAR analysis of the carcinogenic activity of drugs by correlation ranking and genetic algorithm-based PCR. Journal of Chemometrics 18(11), 475-485.

Baeten A, Tafazoli M, Kirsch-Volders M & Geerlings P (1999). Use of the HSAB principle in quantitative structure-activity relationships in toxicological research: application to the genotoxicity of chlorinated hydrocarbons. International Journal of Quantum Chemistry 74(3), 351-355.

Bailey A, Chanderbhan N, Collazo-Braier N, Cheeseman M & Twaroski M (2005). The use of structure-activity relationship analysis in the food contact notification program. Regulatory Toxicology and Pharmacology 42, 225–235

Bang SJ & Cho SJ (2004). Comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) study of mutagen X. Bulletin of the Korean Chemical Society 25(10), 1525-1530.

Basak SC, Balasubramanian K, Gute BD, Mills D, Gorczynska A & Roszak S (2003). Prediction of Cellular Toxicity of Halocarbons from Computed Chemodescriptors: A Hierarchical QSAR Approach. Journal of Chemical Information and Computer Sciences 43(4), 1103-1109.

Benigni R (1997). The first US National Toxicology Program exercise on the prediction of rodent carcinogenicity: definitive results. Mutation Research 387, 35–45.

Benigni R (2004). Prediction of human health endpoints: mutagenicity and carcinogenicity. In Predicting Toxicity and Fate (Cronin M & Livingstone D, ed.), CRC Press, 173-192.

Benigni R (2004). Chemical structure of mutagens and carcinogens and the relationship with biological activity. Journal of Experimental & Clinical Cancer Research 23(1), 5-8.

Benigni R (2005). Structure• activity relationship studies of chemical mutagens and carcinogens: Mechanistic investigations and prediction approaches. Chemical Reviews 105 (5), 1767–1800.

Benigni R & Bossa C (2006). Structure-activity models of chemical carcinogens: state of the art, and new directions. Annali dell'Istituto Superiore di Sanita 42(2), 118-126

Benigni R & Bossa C (2008). Predictivity of QSAR. Journal of Chemical Information and Modeling 48, 971-980.

Benigni R & Bossa C (2008). Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutation Research 659, 248–261.

Benigni R & Bossa C (2008). Predictivity and reliability of QSAR models: The case of mutagens and carcinogens. Toxicology Mechanisms and Methods 18(2-3), 137-147

Benigni R, Bossa C, Richard AM & Yang C (2008a). A novel approach: chemical relational databases, and the role of the iss can database on assessing chemical carcinogenicity. Ann Ist Super Sanità 44(1), 48-56.

Benigni R, Bossa C, Jeliazkova N, Netzeva T & Worth A (2008b). The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. EUR 23241 EN. http://ecb.jrc.ec.europa.eu/qsar/publications/

Benigni R, Bossa C, Netzeva T, Rodomonte A & Tsakovska I (2007). Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. Environmental and Molecular Mutagenesis 48(9), 754-771.

Benigni R, Bossa C, Tcheremenskaia O & Worth A (2009). Development of structural alerts for the in vivo micronucleus assay in rodents. EUR 23844 EN. http://ecb.jrc.ec.europa.eu/qsar/publications/

Benigni R, Bossa C & Worth A (2010). Structural analysis and predictive value of the rodent in vivo micronucleus assay results Mutagenesis, in press. Available online: doi:10.1093/mutage/geq010

Benigni R, Cotta-Ramusino M, Andreoli C & Giuliani A (1992). Electrophilicity as measured by Ke: molecular determinants, relationship with other physical-chemical and quantum mechanical parameters, and ability to predict rodent carcinogenicity. Carcinogenesis 13(4), 547-53.

Benigni R, Passerini L & Rodomonte A (2003). Structure-activity relationships for the mutagenicity and carcinogenicity of simple and •-• unsaturated aldehydes. Environmental and Molecular Mutagenesis 42(3), 136-143.

Benigni R, Conti L, Crebelli R, Rodomonte A & Vari MR (2005). Simple and •,•-unsaturated aldehydes: Correct prediction of genotoxic activity through structure-activity relationship models. Environmental and Molecular Mutagenesis 46(4), 268-280.

Benigni R & Giuliani A (1992). QSAR studies in genetic toxicology: congeneric and noncongeneric chemicals. Archives of Toxicology, Supplement 15, 228-237.

Benigni R & Giuliani A (1994). Quantitative structure-activity relationship (QSAR) studies in genetic toxicology: mathematical models and the "biological activity" term of the relationship, Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 306(2), 181-6.

Benigni R, Netzeva TI, Benfenati E, Bossa C, Franke R, Helma C, Hulzebos E, Marchant C, Richard A, Woo Y-T & Yang C (2007). The Expanding Role of Predictive Toxicology: An Update on the (Q)SAR Models for Mutagens and Carcinogens. Journal of Environmental Science and Health 25, 53–97.

Benigni R & Zito R (2003). Designing safer drugs: (Q)SAR-based identification of mutagens and carcinogens. Current Topics in Medicinal Chemistry 3(11), 1289-1300.

Benigni R & Zito R (2004). The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. Mutation Research 566, 49-63.

Bentzien J, Hickey E, Kemper R, Brewer M, Dyekjaer J, East S, Whittaker M (2010). An in silico method for predicting ames activities of primary aromatic amines by calculating the stabilities of nitrenium ions**.** Journal of Chemical Information and Modeling 50(2), 274-97.

Bercu JP, Morton SM, Deahl JT, Gombar VK, Callis CM, van Lier RB. (2010). In silico approaches to predicting cancer potency for risk assessment of genotoxic impurities in drug substances. Regulatory Toxicology and Pharmacology. (in press)

Bhat KL, Hayik S, Sztandera L & Bock CW (2005). Mutagenicity of aromatic and heteroaromatic amines and related compounds: A QSAR investigation. QSAR & Combinatorial Science 24(7), 831-843.

Blake BW, Enslein K, Gombar VK & Borgstedt HH (1990). Salmonella mutagenicity and rodent carcinogenicity: quantitative structure-activity relationships. Mutation research 241(3), 261-71.

Boerth DW (2007). Molecular modeling for screening of pesticide-nucleotide binding potentials. Abstracts of Papers, 233rd ACS National Meeting, Chicago, IL, United States, March 25-29, 2007 (2007), AGRO-104.

Borosky GL (2007). Ultimate Carcinogenic Metabolites from Aromatic and Heterocyclic Aromatic Amines: A Computational Study in Relation to Their Mutagenic Potency. Chemical Research in Toxicology 20(2), 171-180.

Brown, S.J., Raja, A.A. & Lewis, D.F.V. (1994). A comparison between COMPACT and Hazardexpert evaluations for 80 chemicals tested by the NTP/NCI rodent bioassay. ATLA 22, 482-500.

Bruce E, Autenrieth R, Burghardt R, Donnelly K & McDonald T (2008). Using quantitative structure-activity relationships (QSAR) to predict toxic endpoints for polycyclic aromatic hydrocarbons (PAH). Journal of Toxicology and Environmental Health, A 71(16),1073-84.

Bull RJ & Reckhow DA (2007). Use of chemical models and structure-activity relationships to identify novel disinfection by-products of potential toxicological concern. Preprints of Extended Abstracts presented at the ACS National Meeting, American Chemical Society, Division of Environmental Chemistry 47(1), 695-700.

Buttingsrud B, Ryeng E, King RD & Alsberg BK (2006). Representation of molecular structure using quantum topology with inductive logic programming in structure-activity relationships. Journal of Computer-Aided Molecular Design 20(6), 361-373.

Cariello NF, Wilson JD, Britt BH, Wedd DJ, Burlinson B & Gombar V (2002). Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. Mutagenesis 17(4), 321-329.

Carlsson L, Helgee EA & Boyer S (2009). Interpretation of Nonlinear QSAR Models Applied to Ames Mutagenicity Data. Journal of Chemical Information and Modeling 49(11), 2551-2558.

Casalegno M, Benfenati E & Sello G (2006). Application of fragment-based model to the prediction of the genotoxicity of aromatic amines. Internet Electronic Journal of Molecular Design 5(8), 431-446.

Cash GG, Anderson B, Mayo K, Bogaczyk S & Tunkel J (2005). Predicting genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 585(1-2), 170-183.

Cash GG (2001). Prediction of the genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 491(1-2), 31-37.

Castro EA, Toropova AP, Toropov AA & Maksudov SK (2007). SMILES based optimal descriptors: QSAR modeling mutagenicity. Physical Chemistry: An Indian Journal 2(3), 171-175.

Chen S, Kao C, Huang M, Shih M, Chen Y, Huang S & Liu T (2003). Assessment of Genotoxicity of Benzidine and Its Structural Analogues to Human Lymphocytes Using Comet Assay. Toxicological Sciences 72(2), 283-288.

Chen JJ, Tsai CA, Young JF & Kodell RL (2005). Classification ensembles for unbalanced class sizes in predictive toxicology. SAR and QSAR in Environmental Research 16(6), 517-529.

Cho SJ (2003). Computational study of mutagen X. Bulletin of the Korean Chemical Society 24(6), 731-732.

Cho SJ (2005). Hologram quantitative structure activity relationship (HQSAR) study of mutagen X. Bulletin of the Korean Chemical Society 26(1), 85-90.

Chroust K, Pavlova M, Prokop Z, Mendel J, Bozkova K, Kubat Z, Zajickova V & Damborsky J (2007). Quantitative structure-activity relationships for toxicity and genotoxicity of halogenated aliphatic compounds: wing spot test of Drosophila melanogaster. Chemosphere 67(1), 152-159.

Combes R. (2000). The use of structure-activity relationships and markers of cell toxicity to detect non-genotoxic carcinogens. Toxicology in Vitro 14(4), 387-399.

Combes R & Rodford R (2004). The use of expert systems for toxicity prediction - illustrated with reference to the DEREK program. In Predicting Toxicity and Fate (MTD Cronin and DJ Livingstone, Eds), CRC Press, 193–204.

CRD (2009). Applicability of thresholds of toxicological concern in the dietary risk assessment of metabolites, degradation and reaction products of pesticides, Final report to EFSA in relation to EFSA Grant Agreement EFSA/PPR/2008/01. Chemicals Regulation Directorate, UK.

Contrera JF, Matthews EJ, Kruhlak, NL & Benz RD (2005a). In silico screening of chemicals for bacterial mutagenicity using electrotopological E-state indices and MDL QSAR software. Regulatory Toxicology and Pharmacology 43(3), 313-323.

Contrera J, Maclaughlin P, Hall L & Kier L (2005b). QSAR modeling of carcinogenic risk using discriminant analysis and topological molecular descriptors. Current Drug Discovery Technologies 2(2), 55-67.

Contrera JF, Kruhlak NL, Matthews EJ & Benz RD (2007). Comparison of MC4PC and MDL-QSAR rodent carcinogenicity predictions and the enhancement of predictive performance by combining QSAR models. Regulatory Toxicology and Pharmacology 49(3), 172-182.

Contrera JF, Matthews EJ, Kruhlak NL & Benz RD (2008). In Silico Screening of Chemicals for Genetic Toxicity Using MDL-QSAR, Nonparametric Discriminant Analysis, E-State, Connectivity, and Molecular Property Descriptors. Toxicology Mechanisms and Methods 18(2-3), 207-216.

Cotterill JV, Chaudhry MQ, Matthews W & Watkins RW (2008). In silico assessment of toxicity of heat-generated food contaminants. Food and Chemical Toxicology 46, 1905–1918

Crettaz P & Benigni R (2005). Prediction of the rodent carcinogenicity of 60 pesticides by the DEREKfW expert system. Journal of chemical information and modeling 45(6), 1864-1873.

Cronin M (2006). The role of hydrophobicity in toxicity prediction. Current Computer-Aided Drug Design, 2(4), 405-413.

Cronin MT, Jaworska J, Walker J, Comber M, Watts C & Worth A (2003). Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. Environmental Health Perspectives 111(10), 1391-1401.

Cronin MT, Dearden J, Walker J & Worth A (2003). Quantitative structure-activity relationships for human health effects: Commonalities with other endpoints. Environmental Toxicology and Chemistry 22(8), 1829-1843.

Cunningham AR, Moss ST, Iype SA, Qian G, Qamar S & Cunningham SL (2008). Structure-Activity Relationship Analysis of Rat Mammary Carcinogens. Chemical Research in Toxicology 21(10), 1970-1982.

Custer LL & Sweder KS (2008). The role of genetic toxicology in drug discovery and optimization. Current Drug Metabolism 9, 978-985.

Dearden JC (2003). In silico prediction of drug toxicity. Journal of Computer-Aided Molecular Design 17(2-4), 119-127.

Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PN, Payne MP, Richard AM, Tichy M, Worth AP & Yourick JJ (1997). The Development and Validation of Expert Systems for Predicting Toxicity. Alternatives to Laboratory Animals 25, 223-252.

Deeb O, Hemmateenejad B, Jaber A, Garduno-Juarez R & Miri R (2007). Effect of the electronic and physicochemical parameters on the carcinogenesis activity of some sulfa drugs using QSAR analysis based on genetic-MLR and genetic-PLS. Chemosphere. 67(11):2122-30. Epub 2007 Feb 20.

Dhar K (2006). Qsar for carcinogenicity in methyl benz[a]anthracenes and benz[c]acridines using •-electronic data in a self-consistent field MO theory. Polycyclic Aromatic Compounds 26(5), 385-397.

Didziapetris R, Lanevskij K & Japertas P (2008). Trainable QSAR model of Ames genotoxicity. Abstracts of Papers, 236th ACS National Meeting, Philadelphia, PA, United States, August 17-21, 2008 (2008), TOXI-088.

Dobo KL, Greene N, Cyr MO, Caron S & Ku WW (2006). The application of structure-based assessment to support safety and chemistry diligence to manage genotoxic impurities in active pharmaceutical ingredients during drug development. Regulatory Toxicology and Pharmacology 44(3), 282-293.

Dorn SB, Degen GH, Bolt HM, van dLJ, van AFAA, van dDDJ, den DDJ & Lommerse JPM (2008). Some molecular descriptors for non-specific chromosomal genotoxicity based on hydrophobic interactions. Archives of Toxicology 82(5), 333-8.

Dorn SB, Degen GH, Mueller T, Bonacker D, Joosten HFP, van der Louw J, van Acker FAA & Bolt HM (2007). Proposed criteria for specific and non-specific chromosomal genotoxicity based on hydrophobic interactions. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 628(2), 67-75.

Du H, Wang J, Watzl J, Zhang X & Hu Z (2008). Classification structure-activity relationship (CSAR) studies for prediction of genotoxicity of thiophene derivatives. Toxicology letters 177(1), 10-9.

Duarte Silva I, Gaspar J, Gomes da Costa G, Rodrigues AS, Laires A & Rueff J (2000). Chemical features of flavonols affecting their genotoxicity. Potential implications in their use as therapeutical agents. Chemico-Biological Interactions 124(1), 29-51.

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

Eder E, Hoffman C, Deininger C & Scheckenbach S (1994). Risk assessment for mutagenic and carcinogenic activities of •,•-unsaturated carbonyl compounds by a screening strategy based on structure-activity relationships. Toxicology in Vitro 8(4), 707-710.

Enslein K, Gombar VK & Blake BW (1994). International Commission for Protection Against Environmental Mutagens and Carcinogens. Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program Mutation Research 205, 47-61.

Estrada E & Molina E (2006). Automatic extraction of structural alerts for predicting chromosome aberrations of organic compounds. Journal of Molecular Graphics & Modelling 25(3), 275-288.

Fang Y, Feng Y & Li M (2007). Optimal QSAR analysis of the carcinogenic activity of aromatic and heteroaromatic amines. QSAR & Combinatorial Science 27(5), 543-554.

Ferrari T, Gini G & Benfenati E (2009). Support Vector Machines in the Prediction of Mutagenicity of Chemical Compounds. CAESAR Workshop on QSAR Models for REACH, Mario Negri Institute, Milan, Italy, March 10-11, 2009.

Fjodorova N, Vra•ko M, Tušar M, Jezierska A, Novi• M, Kühne R & Schüürmann G (2009). Quantitative and qualitative models for carcinogenicity prediction for non-congeneric chemicals using CP ANN method for regulatory uses. Molecular Diversity 2009 Aug 15. [Epub ahead of print]

Fjodorova N, Vracko M, Jezierska A, Novic M. (2010) Counter propagation artificial neural network categorical models for prediction of carcinogenicity for non-congeneric chemicals. SAR QSAR Environ Res. 21(1), 57-75.

Fratev F & Benfenati E (2008). A combination of 3D-QSAR, docking, local-binding energy (LBE) and GRID study of the species differences in the carcinogenicity of benzene derivatives chemicals. Journal of Molecular Graphics & Modelling 27(2), 147-160.

Fratev F, Mihayova E, Tadjer A & Benfenati E (2007). Study of species-specific carcinogenicity of benzene derivatives. 1. Combination of CoMFA and GRID analysis. Oxidation Communications 30(4), 891-911.

Fu P, Von Tungeln L, Chiu L-H, Own Z (1999). Halogenated-polycyclic aromatic hydrocarbons: A class of genotoxic environmental pollutants. Journal of Environmental Science and Health, Part C: Environmental Carcinogenesis & Ecotoxicology Review, C17(2), 71-109.

Gallegos Saliner A, Robert D, Gironés X & Carbó-Dorca R (2001). Structure-toxicity relationships of polycyclic aromatic hydrocarbons using molecular quantum similarity. Journal of Computer-Aided Molecular Design 15(1), 67-80.

Garg A, Bhat KL & Bock CW (2002). Mutagenicity of aminoazobenzene dyes and related structures: a QSAR/QPAR investigation. Dyes and Pigments 55(1), 35-52.

Glowienke S, Frieauff W, Allmendinger T, Martus H, Suter W & Mueller L (2005). Structure-activity considerations and in vitro approaches to assess the genotoxicity of 19 methane-, benzene- and toluenesulfonic acid esters. Mutation research 581(1-2), 23-34.

Gonzalez MP, Moldes MdCT, Fall Y, Dias LC & Helguera AM (2005). A topological sub-structural approach to the mutagenic activity in dental monomers. 3. Heterogeneous set of compounds. Polymer 46(8), 2783-2790.

Gonzalez MP, Dias LC & Helguera AM (2004). A topological sub-structural approach to the mutagenic activity in dental monomers. 2. Cycloaliphatic epoxides. Polymer 45(15), 5353-5359.

Gonzalez MP, Helguera AM, Ruiz RM & Garcia Fardales JR (2004). A topological sub-structural approach of the mutagenic activity in dental monomers. 1. Aromatic epoxides. Polymer 45(8), 2773-2779.

Gonzalez-Mancebo S, Gaspar J, Calle E, Pereira S, Mariano A, Rueff J & Casado J (2004). Stereochemical effects in the metabolic activation of nitrosopiperidines: correlations with genotoxicity. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 558(1-2), 45-51.

Gramatica P, Papa E, Marrocchi A, Minuti L & Taticchi A (2007). Quantitative structure–activity relationship modeling of polycyclic aromatic hydrocarbon mutagenicity by classification methods based on holistic theoretical molecular descriptors. Ecotoxicology and Environmental Safety 66 (3), 353-361.

Gramatica P, Pilutti P & Papa E (2007). Approaches for externally validated QSAR modelling of Nitrated Polycyclic Aromatic Hydrocarbon mutagenicity. SAR and QSAR in Environmental Research 18(1-2), 169-178.

Gramatica P, Consonni V & Pavan M (2003). Prediction of aromatic amines mutagenicity from theoretical molecular descriptors. SAR and QSAR in Environmental Research 14(4), 237-250.

Granville CA, Ross MK, Tornero-Velez R, Hanley NM, Grindstaff RD, Gold A, Richard AM, Funasaka K, Tennant AH, Kligerman AD & et a (2005). Genotoxicity and metabolism of the source-water contaminant 1,1-dichloropropene: activation by GSTT1-1 and structure-activity considerations. Mutation research 572(1-2), 98-112.

Greene N (2002). Computer systems for the prediction of toxicity: an update. Advanced Drug Delivery Reviews 54, 417–431.

Greene N (2006). Computational models to predict toxicity. In Comprehensive Medicinal Chemistry II (Eds Taylor JB & Triggle DJ) 5, 909-932

Hall LH & Hall LM (2005). QSAR Modeling Based on Structure-Information for Properties of Interest in Human Health. SAR and QSAR in Environmental Research 16(1-2), 13-41.

Halova J & Aust M (2003). SAR of mutagenes by fingerprint descriptors. Annals of West University of Timisoara, Series of Chemistry 12, 163-166.

Hansch C (1991). Structure-activity relationships of chemical mutagens and carcinogens Science of the Total Environment 109-110, 17-29.

Hansen K, Mika S, Schroeter T, Sutter A, Laak At, Steger-Hartmann T, Heinrich N & Mu¨ller K-R (2009). Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. Journal of Chemical Information and Modeling 49, 2077–2081.

He L, Jurs PC, Custer LL, Durham SK & Pearl GM (2003). Predicting the Genotoxicity of Polycyclic Aromatic Compounds from Molecular Structure with Different Classifiers. Chemical Research in Toxicology 16(12), 1567-1580.

Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T, Heinrich N & Mueller K-R (2009). Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. Journal of Chemical Information and Modeling 49(9), 2077-2081.

Hayashi M, Kamata E, Hirose A, Takahashi M, Morita T & Ema M (2005). In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 588(2), 129-135.

Hayashi M, Kamata E, Hirose A, Takahashi M, Morita T & Ema M (2006). In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. Mutation Research 588, 129-135.

Helguera A, Perez M, Combes R & Gonzalez M (2005). The prediction of carcinogenicity from molecular structure. Current Computer-Aided Drug Design 1(3), 237-255.

Helguera AM, Gonzalez MP, Cordeiro MNDS & Perez MAC (2007). Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds. Toxicology and Applied Pharmacology 221(2), 189-202.

Helguera AM, Gonzalez MP & Briones JR (2004). TOPS-MODE approach to predict mutagenicity in dental monomers. Polymer 45(6), 2045-2050.

Helguera AM, Cordeiro MNDS, Perez MAC, Combes RD & Gonzalez MP (2008). Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds. Toxicology and Applied Pharmacology 231(2), 197-207.

Helguera AM, Cordeiro MNDS, Perez MAC, Combes RD & Gonzalez MP (2008). QSAR modeling of the rodent carcinogenicity of nitro compounds. Bioorganic & Medicinal Chemistry 16(6), 3395-3407.

Helguera AM, Cabrera Perez MA, Gonzalez MP, Ruiz RM & Gonzalez Diaz H (2005). A topological substructural approach applied to the computational prediction of rodent carcinogenicity. Bioorganic & Medicinal Chemistry 13(7), 2477-2488.

Helguera Morales A, Cabrera Perez MA, Combes RD & Perez Gonzalez M (2006). Quantitative structure activity relationship for the computational prediction of nitrocompounds carcinogenicity. Toxicology 220(1), 51-62.

Helma C, Cramer T, Kramer S & De Raedt L (2004). Data Mining and Machine Learning Techniques for the Identification of Mutagenicity Inducing Substructures and Structure Activity Relationships of Noncongeneric Compounds. Journal of Chemical Information and Computer Sciences 44(4), 1402-1411.

Helma C (2006). Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Molecular Diversity 10(2), 147-158.

Hemmateenejad B, Safarpour MA, Miri R & Nesari N (2005). Toward an optimal procedure for PC-ANN model building: prediction of the carcinogenic activity of a large set of drugs. Journal of Chemical Information and Modeling **45**(1), 190-199.

Henry B, Grantv S, Klopman G & Rosenkranz H (1998). Induction of forward mutations at the thymidine kinase locus of mouse lymphoma cells: evidence for electrophilic and non-electrophilic mechanisms,.Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 397(2), 313-335.

Holder AJ & Ye L (2009). Quantum mechanical quantitative structure-activity relationships to avoid mutagenicity. Dental Materials 25(1), 20-25.

Helma C (2006). Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Molecular Diversity 10(2), 147-158.

Hu J, Wang W, Zhu Z, Chang H, Pan F & Lin B (2007). Quantitative structure-activity relationship model for prediction of genotoxic potential for quinolone antibacterials. Environmental science & technology 41(13), 4806-12.

Hulzebos E, Sijm D, Traas T, Posthumus R & Maslankiewicz L. (2005). Validity and validation of expert (Q)SAR systems. SAR and QSAR in Environmental Research 16(4), 385-401.

Ivanciuc O (2002). Support vector machine classification of the carcinogenic activity of polycyclic aromatic hydrocarbons. Internet Electronic Journal of Molecular Design [online computer file] 1(4), 203-218.

Janzowski C, Glaab V, Samimi E, Schlatter J, Pool-Zobel BL & Eisenbrand GEbEG (2000). Food relevant •,•-unsaturated carbonyl compounds: In vitro toxicity, genotoxic (mutagenic) effectiveness and reactivity towards glutathione. Carcinogenic and Anticarcinogenic Factors in Food, Symposium [Carcinogenic/Anticarcinogenic Factors in Food: Novel Concepts?], 3rd, Kaiserslautern, Germany, Oct. 4-7, 1998 469-473.

Jezierska A, Vracko M & Basak SC (2004). Counter-propagation artificial neural network as a tool for the independent variable selection: Structure-mutagenicity study on aromatic amines. Molecular Diversity 8(4), 371-377.

Jacobs A (2005). Prediction of 2-Year Carcinogenicity Study Results for Pharmaceutical Products: How Are We Doing?. Toxicological Sciences 88(1), 18-23.

Joshi S, Chourey VR, Lakhwani M & Thakur P (2006). QSAR study on the mutagenials of phenyltriazenes. Bioinformatics Trends 1(3), 47-50.

Kabankin AS & Kurlyandskii BA (2001). Discriminant analysis of the relationship between topological molecular structure and carcinogenicity of aromatic amines. Pharmaceutical Chemistry Journal (Translation of Khimiko-Farmatsevticheskii Zhurnal) 35(5), 257-259.

Kauffman GW & Jurs PC (2002). Predicting the genotoxicity of secondary and aromatic amines from molecular strucure. Abstracts of Papers, 223rd ACS National Meeting, Orlando, FL, United States, April 7-11, 2002 (2002), COMP-129.

Karelson M, Sild S & Maran U (2000). Non-linear QSAR treatment of genotoxicity. Molecular Simulation 24(4-6), 229-242.

Kazius J, McGuire R & Bursi R (2005). Derivation and validation of toxicophores for mutagenicity prediction. Journal of Medicinal Chemistry 48(1), 312-320.

Kim D, Hong S & Lee D (2006). The quantitative structure-mutagenicity relationship of polycyclic aromatic hydrocarbon metabolites. International Journal of Molecular Sciences 7(12), 556-570.

Kirkland D, Aardema M, Henderson L & Müller L (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 584(1-2), 1-256.

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

Klopman G, Zhang Z, Woodgate SD & Rosenkranz HS (1995). The structure-toxicity relationship challenge at hazardous waste sites. Chemosphere 31(1), 2511-9.

Klopman G, Zhu H, Fuller MA & Saiakhov RD (2004). Searching for an enhanced predictive tool for mutagenicity. SAR and QSAR in Environmental Research 15(4), 251-263.

Klopman G, Chakravarti SK, Harris N, Ivanov J & Saiakhov RD (2003). In-silico screening of high production volume chemicals for mutagenicity using the MCASE QSAR expert system. SAR and QSAR in environmental research 14(2), 165-80

Kubo T, Urano K & Utsumi H (2002). Mutagenicity characteristics of 255 environmental chemicals. Journal of Health Science 48(6), 545-554.

Kundu B, Richardson SD, Granville CA, Shaughnessy DT, Hanley NM, Swartz PD, Richard AM & DeMarini DM (2004). Comparative mutagenicity of halomethanes and halonitromethanes in Salmonella TA100: structure-activity analysis and mutation spectra. Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 554(1-2), 335-350.

Knize MG, Hatch FT, Tanga MJ, Lau EY & Colvin ME (2006). A QSAR for the mutagenic potencies of twelve 2-amino-trimethylimidazopyridine isomers: structural, quantum chemical, and hydropathic factors. Environmental and Molecular Mutagenesis 47(2), 132-146.

Koleva YK, Madden JC & Cronin MTD (2008). Formation of Categories from Structure-Activity Relationships To Allow Read-Across for Risk Assessment: Toxicity of •,•-Unsaturated Carbonyl Compounds. Chemical Research in Toxicology 21(12), 2300-2312.

Kruhlak N, Contrera J, Benz D & Matthews E (2007). Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. Advanced Drug Delivery Reviews (2007), 59(1), 43-55.

Kulkarni SA, Moir D & Zhu J (2007). Influence of structural and functional modifications of selected genotoxic carcinogens on metabolism and mutagenicity - a review. SAR and QSAR in Environmental Research 18(5-6), 459-514

Kulkarni SA & Zhu J (2008). Integrated approach to assess the domain of applicability of some commercial (Q)SAR models. SAR and QSAR in Environmental Research 19(1-2), 39-54.

Lagunin AA, Dearden JC, Filimonov DA & Poroikov VV (2005). Computer-aided rodent carcinogenicity prediction. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 586(2), 138-146.

Lai D, Woo Y-T, Argus M & Arcos J (1996). Carcinogenic potential of organic peroxides: Prediction based on structure-activity relationships (SAR) and mechanism-based short-term tests. Environmental Carcinogenesis & Ecotoxicology Reviews, C14(1), 63-80.

Langham JJ & Jain AN (2008). Accurate and Interpretable Computational Modeling of Chemical Mutagenicity. Journal of Chemical Information and Modeling 48(9), 1833-1839.

Lanevskij K, Japertas P, Didziapetris R & Petrauskas A (2006). Fragmental QSAR model for the prediction of AMES genotoxicity. Abstracts of Papers, 232nd ACS National Meeting, San Francisco, CA, United States, Sept. 10-14, 2006 (2006), CINF-053.

Lewis DFV, Ioannides C & Parke DV (2003). A quantitative structure-activity relationship (QSAR) study of mutagenicity in several series of organic chemicals likely to be activated by cytochrome P450 enzymes. Teratogenesis, Carcinogenesis, and Mutagenesis (Suppl.), 187-193.

Livingstone DJ, Greenwood R, Rees R & Smith MD (2002). Modelling mutagenicity using properties calculated by computational chemistry. SAR and QSAR in Environmental Research 13(1), 21-33.

Long A & Combes RD (1995). Using DEREK to predict the activity of some carcinogens/mutagens found in foods. Toxicology in Vitro 9(4), 563-9.

Lozano JJ, Pastor M, Cruciani G, Gaedt K, Centeno NB, Gago F & Sanz F (2000). 3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. Journal of Computer-Aided Molecular Design 14(4), 341-353.

Maatta JAE, Airenne TT, Nordlund HR, Janis J, Paldanius TA, Vainiotalo P, Johnson MS, Kulomaa MS & Hytonen VP (2008). Rational modification of ligand-binding preference of avidin by circular permutation and mutagenesis. ChemBioChem 9(7), 1124-1135.

Mahe P, Ueda N, Akutsu T, Perret J & Vert J (2005). Graph Kernels for Molecular Structure-Activity Relationship Analysis with Support Vector Machines. Journal of Chemical Information and Modeling 45(4), 939-951.

Maran U, Karelson M & Katritzky AR (1999). A comprehensive QSAR treatment of the genotoxicity of heteroaromatic and aromatic amines. Quantitative Structure-Activity Relationships 18(1), 3-10.

Marchant CA & Collaborative TD (1996). Prediction of Rodent Carcinogenicity Using the DEREK System for 30 Chemicals Currently Being Tested by the National Toxicology Program. Environmental Health Perspectives 104(5), 1065-1073.

Marino DJG, Peruzzo PJ, Castro EA & Toropov AA (2002). QSAR carcinogenic study of methylated polycyclic aromatic hydrocarbons based on topological descriptors derived from distance matrices and correlation weights of local graph invariants. Internet Electronic Journal of Molecular Design 1(3), 115-133.

Massarelli I, Imbriani M, Coi A, Saraceno M, Carli N & Bianucci AM (2009). Development of QSAR models for predicting hepatocarcinogenic toxicity of chemicals. European Journal of Medicinal Chemistry 44(9), 3658-3664.

Matthews EJ & Contrera JF (1998). A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodents using enhanced MCASE QSAR-ES software. Regulatory Toxicology and Pharmacology 28(3), 242-64.

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD & Contrera JF (2006a). An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regulatory Toxicology and Pharmacology 44, 83-96.

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD & Contrera JF (2006b). An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. Regulatory Toxicology and Pharmacology 44(2), 97-110.

Matthews EJ, Kruhlak NL, Benz RD, Contrera JF, Marchant CA & Yang C (2008). Combined Use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows Software to Achieve High-Performance, High-Confidence, Mode of Action-Based Predictions of Chemical Carcinogenesis in Rodents. Toxicology Mechanisms and Methods 18(2-3), 189-206.

Matsuoka A, Takeshita K, Furuta A, Ozaki M, Fukuhara K & Miyata N (2002). The 4'-hydroxy group is responsible for the in vitro cytogenetic activity of resveratrol. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 521(1-2), 29-35.

Mattioni BE, Kauffman GW, Jurs PC, Custer LL, Durham SK & Pearl GM (2003). Predicting the genotoxicity of secondary and aromatic amines using data subsetting to generate a model ensemble. Journal of chemical information and computer sciences 43(3), 949-63.

Mayer J, Cheeseman MA & Twaroski ML (2008). Structure–activity relationship analysis tools: Validationand applicability in predicting carcinogens. Regulatory Toxicology and Pharmacology 50, 50-58.

Mazzatorta P, Ringeissen S, Note R, Schilter B & Meunier JR (2009). In silico models to predict rodent carcinogenicity of naturally-occurring chemicals: comparative study and first insights into modes of action. Poster presentation at the Lhasa International Collaborative Group Meeting, November 2009.

Mazzatorta P, Tran L, Schilter B & Grigorov M (2007). Integration of Structure-Activity Relationship and Artificial Intelligence Systems To Improve in Silico Prediction of Ames Test Mutagenicity. Journal of Chemical Information and Modeling 47(1), 34-38.

McElroy NR, Thompson ED & Jurs PC (2003). Classification of Diverse Organic Compounds That Induce Chromosomal Aberrations in Chinese Hamster Cells. Journal of Chemical Information and Computer Sciences 43(6), 2111-2119.

McGregor DB, Pangrekar J, Rosenkranz HS & Klopman G (1994). A reexamination of the low prevalence of carcinogens in an early carcinogen screen. Regulatory Toxicology and Pharmacology 19(1), 97-105.

Mekenyan O, Dimitrov S, Schmieder P & Veith G (2003). In silico modelling of hazard endpoints: current problems and perspectives. SAR and QSAR in Environmental Research 14(5-6), 361-371.

Mekenyan O, Dimitrov S, Serafimova R, Thompson E, Kotov S, Dimitrova N & Walker JD (2004). Identification of the Structural Requirements for Mutagenicity by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals I: TA100 Model. Chemical Research in Toxicology 17, 753-766.

Mekenyan OG, Dimitrov SD, Pavlov TS & Veith GD (2005). POPs: A QSAR System for Developing Categories for Persistent, Bioaccumulative and Toxic Chemicals and their Metabolites POPS software. SAR and QSAR in Environmental Research 16(1-2), 103-133.

Mekenyan O, Todorov M, Serafimova R, Stoeva S, Aptula A, Finking R & Jacob E (2007). Identifying the Structural Requirements for Chromosomal Aberration by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals. Chemical Research in Toxicology 20, 1927–1941.

Mersch-Sundermann V, Rosenkranz H, & Klopman G (1994). The structural basis of the genotoxicity of nitroarenofurans and related compounds. Mutation Research 304(2), 271-84.

Miertus S, Frecer V & Majekova M (1989). QSAR and mechanistic studies on the genotoxic compounds including environmental effects. International Journal of Quantum Chemistry 35(1), 153-165.

Miller E & Miller J (1981). Searches for Ultimate Chemical Carcinogens and Their Reactions with Cellular Macromolecules. Cancer(47), 2327-2345.

Morales AH, Duchowicz PR, Perez MAC, Castro EA, Cordeiro MNDS & Gonzalez MP (2006). Application of the replacement method as a novel variable selection strategy in QSAR. 1. Carcinogenic potential. Chemometrics and Intelligent Laboratory Systems 81(2), 180-187.

Morales AH, Perez MAC & Gonzalez MP (2006). A radial distribution-function approach for predicting rodent carcinogenicity. Journal of Molecular Modeling 12(6), 769-780.

Morales Helguera A, Perez Gonzalez M, Dias Soeiro Cordeiro MN & Cabrera Perez MA (2008). Quantitative Structure-Carcinogenicity Relationship for Detecting Structural Alerts in Nitroso Compounds: Species, Rat; Sex, Female; Route of Administration, Gavage. Chemical Research in Toxicology 21(3), 633-642.

Mosier PD, Jurs PC, Custer LL, Durham SK & Pearl GM (2003). Predicting the Genotoxicity of Thiophene Derivatives from Molecular Structure. Chemical Research in Toxicology 16(6), 721-732.

Muster W, Albertini S & Gocke E (2003). Structure-activity relationship of oxadiazoles and allylic structures in the Ames test: an industry screening approach. Mutagenesis 18(4), 321-329.

Munro IC, Ford RA, Kennepohl E & Sprenger JG (1996). Thresholds of Toxicological Concern Based on Structure-Activity Relationships Drug Metabolism Reviews 28(1&2), 209-217

Nair PC & Sobhia ME (2008). Comparative QSTR studies for predicting mutagenicity of nitro compounds. Journal of Molecular Graphics & Modelling 26(6), 916-934.

Novak M & Rajagopal S (2002). Correlations of Nitrenium Ion Selectivities with Quantitative Mutagenicity and Carcinogenicity of the Corresponding Amines. Chemical Research in Toxicology 15(12), 1495-1503.

Onchoke KK (2009). Structure-mutagenicity relationships and energies of 1-, and 2-nitrotriphenylenes. THEOCHEM 899(1-3), 79-85.

Papa E, Pilutti P & Gramatica P (2008). Prediction of PAH mutagenicity in human cells by QSAR classification. SAR and QSAR in Environmental Research 19(1-2), 115-127.

Patlewicz G, Rodford R & Walker JD (2003). Quantitative structure-activity relationships for predicting mutagenicity and carcinogenicity. Environmental Toxicology and Chemistry 22(8), 1885-1893.

Parsons S & McBurney P (2005). The use of expert systems for toxicology risk prediction. In Predictive Toxicology (C Helma, ed), 135-175.

Passerini L (2003). QSARs for individual classes of chemical mutagens and carciongens. In Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens (R Benigni, ed), pp 81-123. CRC Press.

Perez-Garrido A, Gonzalez MP & Escudero AG (2008). Halogenated derivatives QSAR model using spectral moments to predict haloacetic acids (HAA) mutagenicity. Bioorganic & Medicinal Chemistry 16(10), 5720-5732.

Pérez-Garrido A, Helguera A, Rodríguez F, Cordeiro M. (2010). QSAR models to predict mutagenicity of acrylates, methacrylates and • ,• -unsaturated carbonyl compounds. Dental Materials 26 (5), 397-415

Pérez-Garrido A, Helguera A, López G, Cordeiro M, Escudero A. (2010). A topological substructural molecular design approach for predicting mutagenesis end-points of • , • -unsaturated carbonyl compounds. Toxicology 268 (1-2), 64-77

Platt KL, Grupe S & Fickler M (2008). The 3,4-oxide is responsible for the DNA binding of benzo[ghi]perylene, a polycyclic aromatic hydrocarbon without a "classic" bay-region. Chemico-Biological Interactions 176(2-3), 179-187.

Poletti A, Poletti L, Selvaggi R & Nottoli R (2002). Semiempirical and ab initio electronic structures and geometrical optimization of recently synthesized cyclopentaphenanthrene derivatives. Polycyclic Aromatic Compounds 22(2), 147-159.

Polloth C & Mangelsdorf I (1997). Commentary on the application of (Q)SAR to the toxicological evaluation of existing chemicals, Chemosphere 35(11), 2525-2542.

Popelier PLA, Smith PJ & Chaudry UA (2004). Quantitative structure-activity relationships of mutagenic activity from quantum topological descriptors: triazenes and halogenated hydroxyfuranones (mutagen-X) derivatives. Journal of Computer-Aided Molecular Design 18(11), 709-718.

Prival MJ (2001). Evaluation of the TOPKAT System for Predicting the Carcinogenicity of Chemicals. Environmental and Molecular Mutagenesis 37, 55-69.

Purdy R (1996). A Mechanism-mediated Model for Carcinogenicity: Model Content and Prediction of the Outcome of Rodent Carcinogenicity Bioassays Currently Being Conducted on 25 Organic Chemicals. Environmental Health Perspectives 104, 1085-1094.

Rallo R, Espinosa G & Giralt F (2005). Using an ensemble of neural based QSARs for the prediction of toxicological properties of chemical contaminants. Process Safety and Environmental Protection 83(B4), 387-392.

Richard A (1998). Structure-based methods for predicting mutagenicity and carcinogenicity: are we there yet? Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 400(1,2), 493-507

Richard A, Rabinowitz J & Waters M (1989). Strategies for the use of computational SAR methods in assessing genotoxicity. Mutation Research, Reviews in Genetic Toxicology 221(3), 181-96.

Richard AM & Williams CR (2003). Public Sources of Mutagenicity and Carcinogenicity Data: Use in Structure-Activity Relationship Models. QSARs of Mutagens and Carcinogens. R Benigni. Boca Raton, CRC Press: 151-177.

Ridings JE, Barratt MD, Cary R, Earnshaw CG, Eggington CE, Ellis MK, Judson PN, Langowski JJ, Marchant CA, Payne MP, Watson WP & Yih TD (1996). Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. Toxicology 106(1-3), 267-279.

Roberts JL, Dearden JC & Bilton RF (2002). QSAR study of genotoxicity and cytotoxicity of a series of naphthoquinones. Abstracts of Papers, 224th ACS National Meeting, Boston, MA, United States, August 18-22, 2002 (2002), COMP-173.

Rodriguez-Arnaiz R & Tellez GO (2002). Structure-activity relationships of several anisidine and dibenzanthracene isomers in the w/w+ somatic assay of Drosophila melanogaster. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 514(1-2), 193-200.

Rosenkranz HS (2004). SAR modeling of genotoxic phenomena: the consequence on predictive performance of deviation from a unity ratio of genotoxicants/non-genotoxicants. Mutation research 559(1-2), 67-71.

Rosenkranz HS (2004). Computers and humans in the SAR assessment of health hazards. Internet Electronic Journal of Molecular Design 3(8), 499-513.

Rosenkranz HS (2003). A data mining approach for the elucidation of the action of putative etiological agents: application to the non-genotoxic carcinogenicity of genistein. Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 526(1,2), 85-92.

Rosenkranz HS (2002). A paradigm for determining the relevance of short-term assays: application to oxidative mutagenesis. Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 508(1,2), 21-27.

Rosenkranz HS (2002). Relationship of sister chromatid exchanges, genotoxicity and carcinogenicity: An SAR-based analysis. Internet Electronic Journal of Molecular Design 1(5), 227-235.

Rosenkranz HS & Cunningham AR (2001). SAR modeling of genotoxic phenomena: the effect of supplementation with physiological chemicals. Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 476(1,2), 133-137.

Rosenkranz HS & Klopman G (1990). Identification of rodent carcinogens by an expert system. Progress in Clinical and Biological Research 340B, 23-48.

Rosenkranz HS & Klopman G (1990). The carcinogenic potential of cocaine. Cancer Letters 52(3), 243-6.

Rosenkranz HS & Klopman G (1990). Structural alerts to genotoxicity: the interaction of human and artificial intelligence. Mutagenesis 5(4), 333-61.

Rosenkranz H & Klopman G (1990). Evaluating the ability of CASE, an artificial intelligence structure-activity relational system, to predict structural alerts for genotoxicity. Mutagenesis 5(6), 525-527.

Rosenkranz HS & Klopman G (1990). Structural implications of the ICPEMC method for quantifying genotoxicity data. Mutation Research 305(1), 99-116.

Rosenkranz HS & Klopman G (1991). Predictions of the lack of genotoxicity and carcinogenicity in rodents of two gasoline additives: methyl- and ethyl-t-butyl ethers. In Vitro Toxicology 4(1), 49-54.

Rosenkranz HS & Klopman G (1992). A structural analysis of the genotoxic and carcinogenic potentials of cyclosporin A. Mutagenesis 7(2), 115-18.

Rosenkranz H & Klopman G (1995). Structure-activity relationships as alternatives in the study of carcinogenesis, Alternative Methods in .Toxicology and the Life Sciences, 11 (The World Congress on Alternatives and Animal Use in the Life Sciences: Education, Research, Testing, 1993), 379-390.

Rosenkranz HS, Liu M, Cunningham A & Klopman G (1996). Application of structural concepts to evaluate the potential carcinogenicity of natural products. SAR and QSAR in environmental research 5(2), 79-98.

Rosenkranz HS, Zhang YP & Klopman G (1995). Risk identification using structural concepts: the potential carcinogenicity of praziquantel. Regulatory Toxicology and Pharmacology 22(2), 152-161.

Rosenkranz HS, Zhang YP & Klopman G (1996). Studies on the potential for genotoxic carcinogenicity of fragrances and other chemicals. Food and Chemical Toxicology 36(8), 687-696.

Rothenbacher T & Schwack W (2009). Nontargeted multicomponent analytical screening of plastic food contact materials using fast interpretation of deliverables via expert structure-activity relationship software. Journal of AOAC International 92(3), 941-50.

Ruiz P, Faroon O, Moudgal CJ, Hansen H, De Rosa CT & Mumtaz M (2008). Prediction of the health effects of polychlorinated biphenyls (PCBs) and their metabolites using quantitative structure-activity relationship (QSAR). Toxicology Letters 181(1), 53-65.

Saiakhov RD & Klopman G (2008). MultiCASE Expert Systems and the REACH Initiative. Toxicology Mechanisms and Methods 18(2-3), 159-175.

Sanderson DM & Earnshaw CG (1991). Computer prediction of possible toxic action from chemical structure; the DEREK system. Human and Experimental Toxicology 10, 261-273.

Sangamwar AT, Deshpande UD, Pekamwar SS & Vadvalkar SM (2007). Improving decision making for drug candidates: a computational approach for benzthiazoles as antifungal. Indian Journal of Biotechnology 6(3), 389-396.

Schoenfeld HA & Witz G (2000). Structure-activity relationships in the induction of DNA-protein cross-links by hematotoxic ring-opened benzene metabolites and related compounds in HL60 cells. Toxicology Letters 116(1-2), 79-88.

Schultz T, Cronin M & Netzeva T (2003). The present status of QSAR in toxicology. Journal of Molecular Structure-THEOCHEM 622, 23-38.

Schultz T, Cronin M, Walker J & Aptula A (2003). Quantitative structure-activity relationships in (QSARs) in toxicology: A historical perspective. Journal of Molecular Structure-THEOCHEM 622, 1-22.

Serafimova R, Todorov M, Pavlov T, Kotov S, Jacob E, Aptula A & Mekenyan O (2007). Identification of the Structural Requirements for Mutagenicity, by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals. II. General Ames Mutagenicity Model. Chemical Research in Toxicology 20(4), 662-676.

Shen Q, Jiang J, Shen G & Yu R (2003). Variable selection by an evolution algorithm using modified Cp based on MLR and PLS modeling: QSAR studies of carcinogenicity of aromatic amines. Analytical and Bioanalytical Chemistry 375(2), 248-254.

Shoji R & Kawakami M (2006). Prediction of genotoxicity of various environmental pollutants by artificial neural network simulation. Molecular Diversity 10(2), 101-108.

Singh J, Singh S, Shaik B, Deeb O, Sohani N, Agrawal VK & Khadikar PV (2008). Mutagenicity of nitrated polycyclic aromatic hydrocarbons: a QSAR investigation. Chemical Biology & Drug Design 71(3), 230-243.

Smithing MP & Darvas F (1992). Hazardexpert:an expert system for predicting chemical toxicity. In: Food Safety Assessment (JW Finlay, SF Robinson and DJ Armstrong). Washington, American Chemical Society: 191-200.

Snyder RD & Hendry LB (2005). Toward a greater appreciation of noncovalent chemical/DNA interactions: Application of biological and computational approaches. Environmental and Molecular Mutagenesis 45(2/3), 100-105

Snyder RD & Smith MD (2005). Computational prediction of genotoxicity: room for improvement. Drug Discovery Today 10(16), 1119-1124.

Snyder RD, Ewing D & Hendry LB (2006). DNA intercalative potential of marketed drugs testing positive in in vitro cytogenetics assays. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 609(1), 47-59.

Snyder RD (2009). An Update on the Genotoxicity and Carcinogenicity of Marketed Pharmaceuticals with Reference to In Silico Predictivity. Environmental and Molecular Mutagenesis 50, 435-450.

Snyder RD, Pearl GS, Mandakas G, Choy WN, Goodsaid F & Rosenblum IY (2004). Assessment of the sensitivity of the computational programs DEREK, TOPKAT, and MCASE in the prediction of the genotoxicity of pharmaceutical molecules. Environmental and Molecular Mutagenesis 43(3), 143-158.

Sushko I, Novotarskyi S, Körner R, Pandey AK, Kovalishyn VV, Prokopenko VV, Tetko IV (2010). Applicability domain for in silico models to achieve accuracy of experimental measurements. Journal of Chemometrics 24 (3-4), 202-208

Suzuki T (2005). Prediction of carcinogenicity of noncongeneric chemicals using an artificial neural network. Toyo Daigaku Kiyo, Shizen Kagaku-hen 49, 1-11.

Sztandera L, Garg A, Hayik S, Bhat KL & Bock CW (2003). Mutagenicity of aminoazo dyes and their reductive-cleavage metabolites: a QSAR/QPAR investigation. Dyes and Pigments 59(2), 117-133.

Takamura-Enya T, Suzuki H & Hisamatsu Y (2006). Mutagenic activities and physicochemical properties of selected nitrobenzanthrones. Mutagenesis 21(6), 399-404.

Tanabe K, Lu•i• B, Ami• D, Kurita T, Kaihara M, Onodera N, Suzuki T. (2010) Prediction of carcinogenicity for diverse chemicals based on substructure grouping and SVM modelling. Molecular Diversity, 1-14

Tarasov VA, Mustafaev ON, Abilev SK & Mel'nik VA (2005). Use of Compound Structural Descriptors for Increasing the Efficiency of QSAR Study. Russian Journal of Genetics 41(7), 814-821.

Tekiner-Gulbas B, Temiz-Arpaci O, Oksuzoglu E, Eroglu H, Yildiz I, Diril N, Aki-Sener E & Yalcin I (2007). QSAR of genotoxic active benzazoles. SAR and QSAR in environmental research 18(3-4), 251-63.

Tenant RW & Ashby J (1991). 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. Mutation Research 257, 209-227.

Toivonen H, Srinivasan A, King RD, Kramer S & Helma C (2003). Statistical evaluation of the Predictive Toxicology Challenge 2000-2001. Bioinformatics 19(10), 1183-1193.

Tokiwa H, Sera N, Fukuhara K, Utsumi H, Sasaki S & Miyata N (2003). Structural activity relationship between Salmonella-mutagenicity and nitro-orientation of nitroazaphenanthrenes. Chemico-Biological Interactions 146(1), 19-25.

Toropov AA, Toropova AP, Benfenati E & Manganaro A (2009). QSAR modelling of carcinogenicity by balance of correlations. Molecular Diversity 13(3), 367-373.

Toropov AA & Toropova AP (2002). QSAR modeling of mutagenicity based on graphs of atomic orbitals. Internet Electronic Journal of Molecular Design 1(3), 108-114.

Toropov AA, Toropova AP & Benfenati E (2009). QSAR modelling for mutagenic potency of heteroaromatic amines by optimal SMILES-based descriptors. Chemical Biology & Drug Design 73(3), 301-312.

Toropov AA, Toropova AP & Benfenati E (2009). Simplified molecular input line entry system-based optimal descriptors: quantitative structure-activity relationship modeling mutagenicity of nitrated polycyclic aromatic hydrocarbons. Chemical Biology & Drug Design 73(5), 515-525.

Toropov AA & Benfenati E (2007). SMILES in QSPR/QSAR modeling: results and perspectives. Current Drug Discovery Technologies 4(2), 77-116.

Tozlovanu M, Faucet-Marquis V, Pfohl-Leszkowicz A & Manderville RA (2006). Genotoxicity of the hydroquinone metabolite of ochratoxin A: structure-activity relationships for covalent DNA adduction. Chemical research in toxicology 19(9), 1241-7.

Tuppurainen K (1999). Frontier orbital energies, hydrophobicity and steric factors as physical QSAR descriptors of molecular mutagenicity. A review with a case study: MX compounds. Chemosphere 38(13), 3015-30.

Valerio LG, Arvidson KB, Chanderbhan RF & Contrera JF (2007). Prediction of rodent carcinogenic potential of naturally occurring chemicals in the human diet using high-throughput QSAR predictive modeling. Toxicology and Applied Pharmacology 222(1), 1-16.

Valerio Jr L, Arvidson K, Busta E, Minnier B, Kruhlak N, Daniel Benz R. (2010). Testing computational toxicology models with phytochemicals. Molecular Nutrition and Food Research 54 (2), 186-194.

Valkova I, Vracko M & Basak SC (2004). Modeling of structure-mutagenicity relationships: counter propagation neural network approach using calculated structural descriptors. Analytica Chimica Acta 509(2), 179-186.

Veith G (2006). Roles for QSAR in risk assessment. ALTEX - Alternativen zu Tierexperimenten (2006), 23 Supplement, 369-72.

Venkatapathy R, Wang CY, Bruce RM & Moudgal C (2009). Development of quantitative structure-activity relationship (QSAR) models to predict the carcinogenic potency of chemicals. Toxicology and Applied Pharmacology 234(2), 209-221.

Vijayalakshmi KP & Suresh SC (2008). Theoretical studies on the carcinogenicity of polycyclic aromatic hydrocarbons. Journal of Computational Chemistry 29(11), 1808-17.

Vogel E & Ashby J (1994). Structure-activity relationships: experimental approaches. SCOPE 52, 231-54.

Votano JR, Parham M, Hall LH & Kier LB (2004). New predictors for several ADME/Tox properties: aqueous solubility, human oral absorption, and Ames genotoxicity using topological descriptors. Molecular diversity 8(4), 379-91.

Votano JR, Parham M, Hall LH, Kier LB, Oloff S, Tropsha A, Xie Q & Tong W (2004). Three new consensus QSAR models for the prediction of Ames genotoxicity. Mutagenesis 19(5), 365-77.

Venkatapathy R, Wang CY, Bruce RM & Moudgal C (2009). Development of quantitative structure-activity relationship (QSAR) models to predict the carcinogenic potency of chemicals. Toxicology and Applied Pharmacology 234(2), 209-221.

Vracko M (2006). QSAR approach in study of mutagenicity of aromatic and heteroaromatic amines. Topics in Heterocyclic Chemistry 4 (QSAR and Molecular Modeling Studies in Heterocyclic Drugs II), 85-106.

Vracko M, Szymoszek A & Barbieri P (2004). Structure-Mutagenicity Study of 12 Trimethylimidazopyridine Isomers Using Orbital Energies and "Spectrum-like Representation" As Descriptors. Journal of Chemical Information and Computer Sciences 44(2), 352-358.

Vracko M, Mills D & Basak SC (2004). Structure-mutagenicity modelling using counter propagation neural networks. Environmental Toxicology and Pharmacology 16(1-2), 25-36.

Wang K, Golbraikh A & Tropsha A (2009). Optimization of pattern recognition and classification by combinatorial QSAR modeling of the carcinogenic potency database. Abstracts of Papers, 237th ACS National Meeting, Salt Lake City, UT, United States, March 22-26, 2009, COMP-159.

Wang L, Wang Y, Chen J & Guo L (2009). A structure-based investigation on the binding interaction of hydroxylated polycyclic aromatic hydrocarbons with DNA. Toxicology 262(3), 250-257.

Wang X, Lin Z, Yin D, Liu S & Wang L (2005). 2D/3D-QSAR comparative study on mutagenicity of nitroaromatics. Science in China, Series B: Chemistry 48(3), 246-252.

Waters M, Richard A, Rabinowitz J, Stack H & Garrett N (1990). Structure-activity relationships. Computerized systems. EPA Report EPA/600/D-90/125; Order No. PB90-263476, 75

Waters M, Richard A, Rabinowitz J, Stack F, Garrett N, Lohman P & Rosenkranz H (1994). Structure-activity relationships: computerized systems. SCOPE 52, 201-29.

White AC, Mueller RA, Gallavan RH, Aaron S & Wilson AGE (2003). A multiple in silico program approach for the prediction of mutagenicity from chemical structure. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 539(1-2), 77-89.

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.

Woo Y-T & Lai DY (2005). OncoLogic: A mechanism-based expert system for predicting the carcinogenic potential of chemicals. In Predictive Toxicology (C Helma, ed), 385-413

Woo YT, Lai DY, Argus MF & Arcos JC (1995). Development of structure–activity relationship rules for predicting carcinogenic potential of chemicals. Toxicology Letters 79, 219–228.

Woo Y-T, Lai D, Argus M & Arcos J (1996). Carcinogenicity of organophosphorus pesticides/compounds: An analysis of their structure-activity relationships, Environmental Carcinogenesis & Ecotoxicology Reviews, C14(1), 1-42.

Woo Y, Lai D, McLain JL, Manibusan MK & Dellarco V (2002). Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products. Environmental Health Perspectives Supplements 110(1), 75-87.

Xiao Q, Dai X, Wang X & Wang L (2005). Studies on quantitative structure-genotoxicity of substituted benzenes on the human lymphocyte. Zhongguo Huanjing Kexue 25(1), 18-22. Language: Chinese

Xiao Q, Gao S, Wang X, Liu S & Wang L (2007). Quantitative structure-genotoxicity relationship of substituted phenols to human lymphocyte. Huanjing Huaxue 26(5), 582-587. In Chinese.

Yang C, Hasselgren CH, Boyer S, Arvidson K, Aveston S, Dierkes P, Benigni R, Benz RD, Contrera J & Kruhlak NL (2008). Understanding Genetic Toxicity Through Data Mining: The Process of Building Knowledge by Integrating Multiple Genetic Toxicity Databases. Toxicology Mechanisms and Methods 18(2-3), 277-295

Young R (2002). Genetic toxicology: Web resources. Toxicology 173(1-2), 103-121.

Yourtee D, Holder AJ, Smith R, Morrill JA, Kostoryz E, Brockmann W, Glaros A, Chappelow C & Eick D (2001). Quantum mechanical quantitative structure activity relationships to avoid mutagenicity in dental monomers. Journal of Biomaterials Science, Polymer Edition 12(1), 89-105.

Zani F, Pregnolato M & Zampollo F (1999). Antimicrobial and genotoxic activities of N-(2-hydroxyethyl)-1,2-benzisothiazol-3(2H)-thione carbamic esters. Farmaco 54(10), 643-647.

Zeiger E, Ashby J, Bakale G, Enslein K, Klopman G & Rosenkranz H (1996). Prediction of Salmonella mutagenicity. Mutagenesis 11(5), 471-484.

Zhang Q & Aires-de-Sousa J (2007). Random Forest Prediction of Mutagenicity from Empirical Physicochemical Descriptors. Journal of Chemical Information and Modeling 47(1), 1-8.

Zhang Z, Niu J & Zhi X (2008). A QSAR Model for Predicting Mutagenicity of Nitronaphthalenes and Methylnitronaphthalenes. Bulletin of Environmental Contamination and Toxicology 81(5), 498-502.

Zhang YP, Sussman N, Klopman G & Rosenkranz HS (1997). Development of methods to ascertain the predictivity and consistency of SAR models. Application to the U.S. national toxicology program rodent carcinogenicity bioassays. Quantitative Structure-Activity Relationships 16(4), 290-295.

Zhu X, Zhang YP, Klopman G & Rosenkranz HS (1999). Thalidomide and metabolites: indications of the absence of 'genotoxic' carcinogenic potentials. Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 425(1), 153-167.

Zhu H, Rusyn I, Richard A & Tropsha A (2008). Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. Environmental Health Perspectives 116(4), 506-513.

11. Figures

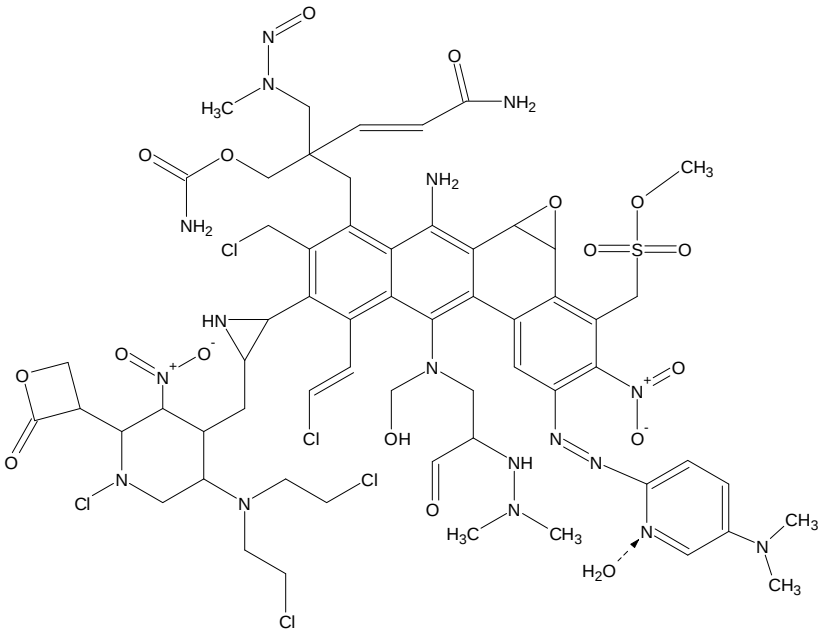

Figure 1. Ashby's poly-carcinogen model, modified after Ashby and Tennant (1988) and Tenant and Ashby (1991)

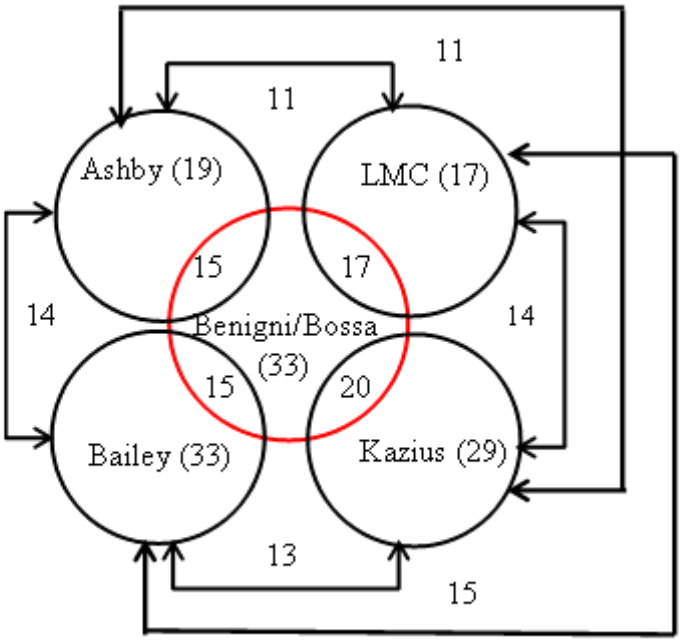

Figure 2. Comparison between the Ashby, Benigni-Bossa, LMC, Bailey and Kazius lists of SAs

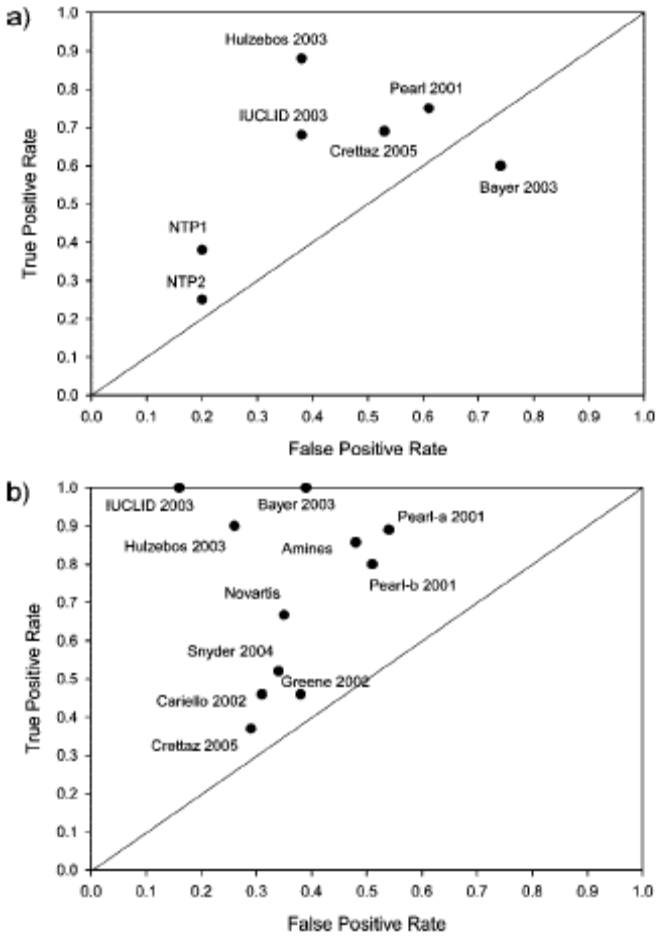

Figure 3. Receiver Operating Curve (ROC) graph for the performance of the Derek software system in a series of external prediction exercises for carcinogenicity (a) and mutagenicity (b). The codes make reference to the following datasets or original studies: (a): Hulzebos 2003, Bayer 2003, IUCLID 2003, Pearl 2001, NTP1, NTP2, and Crettaz 2005. (b) Hulzebos 2003, Bayer 2003, Amines, IUCLID 2003, Cariello 2002, Greene 2002, Snyder 2004, Pearl-a 2001 and Pearl-b 2001, and Crettaz 2005 (sfter Benigni and Bossa 2008)

12. Tables

Table 1. Genotoxicity test methods and endpoints

Test method	**Genotoxic endpoints**	**EU method / OECD guideline**
In vitro test methods		
Bacterial reverse mutation test - Ames	Mutagenicity: gene mutations	EU B.12/13 OECD 471
In vitro mammalian cell gene mutation test – *hprt* test	Mutagenicity: gene mutations	EU B.17 OECD 476
In vitro mammalian cell gene mutation test – Mouse lymphoma assay	Mutagenicity: gene mutations and structural chromosome aberrations	EU B.17 OECD 476
In vitro mammalian chromosome aberration test	Mutagenicity: structural and numerical chromosome aberrations	EU B.10 OECD 473
In vitro micronucleus test	Mutagenicity: structural and numerical chromosome aberrations	EU (none) OECD 487 (draft)
In vivo test methods, somatic cells		
In vivo mammalian bone marrow chromosome aberration test	Mutagenicity: structural and numerical chromosome aberrations	EU B.11 OECD 475
In vivo mammalian erythrocyte micronucleus test	Mutagenicity: structural and numerical chromosome aberrations	EU B.12 OECD 474
Unscheduled DNA synthesis (UDS) test in mammalian liver cells *in vivo*	Genotoxicity: DNA repair	EU B.39 OECD 486
Transgenic animal models	Mutagenicity: gene mutations	EU (none) OECD (none)
In vivo alkaline single-cell gel electrophoresis assay for DNA strand breaks (Comet assay)	Genotoxicity: DNA strand breaks	EU (none) OECD (none
Mammalian bone marrow Sister Chromatid Exchanges (SCE)	Genotoxicity: DNA strand breaks and DNA adduct formation	
In vivo test methods, germ cells		
Mammalian spermatogonial chromosome aberration test	Mutagenicity: structural and numerical chromosome aberrations	EU B.23 OECD 483
Rodent dominant lethal test	Mutagenicity: structural and numerical chromosome aberrations	EU B.22 OECD 478
Transgenic animal models	Mutagenicity: gene mutations	EU none OECD none
In vivo alkaline single-cell gel electrophoresis assay for DNA strand breaks (Comet assay)	Genotoxicity: DNA strand breaks	EU none OECD none
Unscheduled DNA synthesis (UDS) test in testicular cells *in vivo*	Genotoxicity: DNA repair	

Table 2. Regulatory classification criteria for mutagenicity and carcinogenicity

EU criteria	UN GHS criteria
MUTAGENICITY	
CATEGORY 1: substances known to be mutagenic to man. There is sufficient evidence from epidemiological studies to establish a causal association between human exposure to a substance and heritable genetic damage. **Risk phrase:** R46 May cause heritable genetic damage.	**CATEGORY 1:** known to induce heritable mutations or regarded as if they induce heritable mutations in the germ cells of humans.
	CATEGORY 1A: substances known to induce heritable mutations in the germ cells of humans. The classification in Category 1A is based on positive evidence from human epidemiological studies. **Health hazard code:** H340 May cause genetic defects
CATEGORY 2: substances which should be regarded as if they are mutagenic to man. There is sufficient evidence to provide a strong presumption that human exposure to the substance may result in the development of heritable genetic damage. **Risk phrase:** R46 May cause heritable genetic damage	**CATEGORY 1B:** substances regarded as if they induce heritable mutations in the germ cells of humans. **Health hazard code:** H340 May cause genetic defects
CATEGORY 3: substances which cause concern for man owing to possible mutagenic effects. There is evidence from appropriate mutagenicity studies, but this is insufficient to place the substance in category 2. **Risk phrase:** R68 Possible risk of irreversible effects.	**CATEGORY 2:** substances which cause concern for humans owing to the possibility that they may induce heritable mutations in the germ cells of humans. **Health hazard code:** H341 Suspcected of causing genetic defects
CARCINOGENICITY	
CATEGORY 1: substances known to be carcinogenic to man. There is sufficient evidence to establish a causal association between human exposure to a substance and the development of cancer. **Risk phrases:** R45 May cause cancer R49 May cause cancer by inhalation	**CATEGORY 1:** known or presumed human carcinogens. A substance is classified in Category 1 for carcinogenicity on the basis of epidemiological and/or animal data.
CATEGORY 2: substances which should be regarded as if they are carcinogenic to man. There is sufficient evidence to provide a strong presumption that human exposure to a substance may result in the development of cancer.	**CATEGORY 1A**: known to have carcinogenic potential for humans, classification is largely based on human evidence. **Health hazard code:** H350 May cause cancer
	CATEGORY 1B: presumed to have carcinogenic potential for humans, classification is largely based on animal evidence. **Health hazard code:** H350 May cause cancer

EU criteria	UN GHS criteria
CATEGORY 3: substances which cause concern for man owing to possible carcinogenic effects but in respect of which the available information is not adequate for making a satisfactory assessment. There is some evidence from appropriate animal studies, but this is insufficient to place the substance in category 2. **Risk phrase:** R40 Limited evidence of a carcinogenic effect	**CATEGORY 2:** suspected human carcinogens The placing of a substance in Category 2 is done on the basis of evidence obtained from human and/or animal studies, but which is not sufficiently convincing to place the substance in Category 1A or 1B. **Health hazard code:** H351 Suspected of causing cancer

Table 3. Public databases for genotoxicity and carcinogenicity

Database (name and link)	Information
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity http://ml.cs.tu-berlin.de/toxbenchmark/	Ames mutagenicity databaset for 6500 compounds, made freely available by Berlin University of Technology. Downloadable sdf files.
Carcinogenic Potency Database (CPDB) http://potency.berkeley.edu/cpdb.html	Contains of the results of 6540 chronic, long-term animal cancer tests on 1547 chemicals
Danish QSAR database EPA site: http://130.226.165.14/index.html JRC site: http://ecbqsar.jrc.ec.europa.eu/	Searchable database of *predictions* for approx 166,000 chemicals. The predictions are based on MulitCase models developed by the Danish EPA.
DSSTox (Distributed Structure-searchable Toxicity) database www.epa.gov/ncct/dsstox	The DSSTox website provides a public forum for publishing downloadable, structure- searchable, standardized chemical structure files associated with toxicity data
GAP – Genetic Activity Profile Database initially developed by US EPA and IARC, and now by ILS (http://www.ils-inc.com). CD rom available on request	Data on approx 300 chemicals from volumes 1-50 of the IARC Monographs and on 115 compounds identified as Superfund Priority Substances. Latest update in 2000.
European Chemical Substances Information System (ESIS). Freely accessible from the JRC ex-ECB website: http://ecb.jrc.ec.europa.eu/esis/	Information on chemicals related to: EINECS, the European List of Notified Chemical Substances (ELINCS); No-Longer Polymers (NLP) list; High Production Volume Chemicals (HPVCs); Low Production Volume Chemicals (LPVCs); Classification and Labelling (C&L); IUCLID chemical data sheets; EU priority lists and risk assessments performed under the Existing Substances Regulation (ESR); active substances listed on Annex 1 or 1A of the Biocidal Products Directive as well as substances that are "non-inclusions"; Existing Substance evaluated in relation to their PBT properties.
Existing Chemicals Examination (EXCHEM) database (Japan) http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp	Contains data for Ames mutagenicity, chromosomal aberrations and mouse micronucleus assays for more than 250 HPV chemicals
Istituto superiore di Sanità database (ISSCAN) http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7	Contains information on more than 1150 chemical compounds tested with the long-term carcinogenicity bioassay on rodents as well as mutagenicity data for more of them.
Monographs on the Evaluation of Carcinogenic Risks to Humans http://monographs.iarc.fr/index.php	A series of scientific reviews that studied more than 900 agents and have identified more than 400 known, probable and possible carcinogens.
National Toxicology Program (NTP) database http://ntp.niehs.nih.gov	Contains data from more than 500 two-year, two species, toxicology and carcinogenesis also contains results collected on approximately 300 toxicity studies from shorter duration tests and from more than 2000 genetic toxicity studies, some of which include both in vitro and in vivo tests
Toxicity Reference Database (ToxRefDB) http://www.epa.gov/ncct/toxrefdb/	Includes chronic, cancer, sub-chronic, developmental, and reproductive studies on 330 of chemicals, many of which are pesticide active ingredients
TOXNET database of the National Library of Medicine (NLM), including the Carcinogenesis Research Information System database (CCRIS) and the Genetic Toxicology Databank (GENE-TOX) http://toxnet.nlm.nih.gov/	CCRIS contains over 9000 chemical records with animal carcinogenicity, mutagenicity, tumour promotion, and tumor inhibition test results. GENE-TOX contains genetic toxicology (mutagenicity) test data, resulting from expert peer review of the open scientific literature, on over 3000 chemicals

Table 4. Comparison of three main approaches in expert systems

Approach	Advantages	Disadvantages
Rule-based	• mechanistically connected to the predicted endpoint • provide reasoning for the predictions • in many cases support the prediction with literature references or expert knowledge	• often restricted and/or ill-defined applicability domain • usually cannot explain differences of the activity within a chemical class • usually have lower accuracy of the prediction than statistical models
Statistical	• usually have high accuracy of the predictions • can be use for preliminary research when mechanism of action is unknown	• usually difficult to interpret the model predictions • often do not provide mechanistically reasoning of the predictions • often non-transparent to the end-user
Hybrid	• combines advantages of rule-based and statistical approaches, including mechanistic interpretability (for SA part), and overall accuracy	• likely to have restricted applicability domain

Table 5. Software for genotoxicity and carcinogenicity

Software	Availability	Comments (endpoints predicted, applicability and performance)
CAESAR http://www.caesar-project.eu/	Freely available	Mutagenicity, carcinogenicity
DfW (Lhasa Ltd.) http://www.lhasalimited.org	Commercial	Mutagenicity, chromosome damage, genotoxicity, carcinogenicity, peroxisome proliferation
GAP – Genetic Activity Profile Database developed by US EPA	Not readily available. Used in-house by US EPA	Data on 299 chemicals compiled by IARC and US EPA. Data are available on 299 compounds selected from volumes 1-50 of the IARC Monographs and on 115 compounds identified as Superfund Priority Substances.
HazardExpert http://www.compudrug.com	Commercial	Mutagenicity, oncogenicity
Lazar http://lazar.in-silico.de	Freely available	Ames mutagenicity, carcinogenicity
MDL-QSAR http://www.symyx.com/	Commercial	Carcinogenicity
MolCode Toolbox http://molcode.com/	Commercial	Ames mutagenicity, carcinogenicity
Multicase (MCASE/MC4PC) MultiCASE Inc http://www.multicase.com	Commercial	Research tool - applies a statistical approach that automatically identifies molecular substructures that have a high probability of being relevant to the observed biological activity. Requires a learning set comprised of a mix of active and inactive molecules of diverse composition.
OASIS – TIMES http://www.oasis-lmc.org	Commercial	Ames mutagenicity, chromosomal aberrations
OECD Toolbox http://toolbox.oasis-lmc.org	Freely available	Includes two so-called “profilers” associated with genotoxicity and carcinogenicity, as well as three databases with experimental data that can be used to support grouping and read-across
OncoLogic™ http://www.epa.gov/oppt/newchems/tools/oncologic.htm	Freely available	Carcinogenicity
PASS Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Moscow http://ibmc.p450.ru/PASS//	Commercial	Classification models giving probability of mutagenic effects. There are two models, one for Ames mutagenicity, and another covered multiplein vitro and in vivo mutagnicity endpoints in mammals.
TOPKAT (Accelrys) http://www.accelrys.com	Commercial	Ames mutagenicity, carcinogenicity
Toxtree http://ecb.jrc.ec.europa.eu/qsar/	Freely available	Includes modules for mutagenicity, carcinogenicity, and the *in vivo* micronucleus assay

Table 6. Reviews and single model evaluation studies on (Q)SARs for genotoxicity and carcinogenicity

Year	Reference	Comment
2010	Valerio Jr L, Arvidson K, Busta E, Minnier B, Kruhlak N, Daniel Benz R (2010). Testing computational toxicology models with phytochemicals. Molecular Nutrition and Food Research 54 (2), 186-194.	Comparative evaluation to screening phytochemicals
2010	Sushko I, Novotarskyi S, Körner R, Pandey A, Kovalishyn V, Prokopenko V, Tetko I (2010). Applicability domain for in silico models to achieve accuracy of experimental measurements. Journal of Chemometrics 24 (3-4), 202-208.	Applicability domain for AMES mutagenicity model
2010	Pérez-Garrido A, Helguera A, López G, Cordeiro M, Escudero A (2010). A topological substructural molecular design approach for predicting mutagenesis end-points of •, •-unsaturated carbonyl compounds. Toxicology 268 (1-2), 64-77.	Comparative evaluation of TOPS-MODE and Toxtree for •,•-unsaturated carbonyl compounds
2009	Snyder RD (2009). An update on the genotoxicity and carcinogenicity of marketed pharmaceuticals with reference to in silico predictivity. Environmental and Molecular Mutagenesis 50, 435-450.	Comparative evaluation of Derek, MC4PC, marketed pharmaceuticals
2009	Rothenbacher T & Schwack W (2009). Nontargeted multicomponent analytical screening of plastic food contact materials using fast interpretation of deliverables via expert structure-activity relationship software. Journal of AOAC International 92(3), 941-950	Industry use of QSAR, Application of Derek to screening of plastic food contact materials
2009	Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T, Heinrich N & Mueller K-R (2009). Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. Journal of Chemical Information and Modeling 49(9), 2077-2081.	Comparative evaluation
2009	Mazzatorta P, Ringeissen S, Note R, Schilter B & Meunier JR (2009). In silico models to predict rodent carcinogenicity of naturally-occurring chemicals: comparative study and first insights into modes of action. Poster presentation at the Lhasa International Collaborative Group Meeting, November 2009	Comparative evaluation
2008	Kulkarni SA & Zhu J (2008). Integrated approach to assess the domain of applicability of some commercial (Q)SAR models. SAR and QSAR in Environmental Research 19(1-2), 39-54.	Comparative evaluation
2008	Mayer JM *et al*. (2008). Structure–activity relationship analysis tools: Validation and applicability in predicting carcinogens. Regulatory Toxicology and Pharmacology 50: 50-58.	Comparative evaluation
2008	Custer LL & Sweder KS (2008). The role of genetic toxicology in drug discovery and optimization. Current Drug Metabolism 9, 978-985.	Drug discovery and optimisation
2008	Benigni R & Bossa C (2008). Predictivity of QSAR. Journal of Chemical Information and Modeling 48, 971-980.	Comparative evaluation of different models
2008	Benigni R & Bossa C (2008). Predictivity and reliability of QSAR models: The case of mutagens and carcinogens. Toxicology Mechanisms and Methods 18(2-3), 137-147	Comparative evaluation of non-commercial QSARs
2008	Benigni R, Bossa C, Richard A & Yang C (2008). A novel approach: chemical relational databases, and the role of the ISSCAN can database on assessing chemical carcinogenicity. Ann Ist Super Sanità 44(1): 48-56.	Databases for genotoxicity and mutagenicity
2008	Saiakhov RD & Klopman G (2008). MultiCASE Expert Systems and the REACH Initiative. Toxicology Mechanisms and Methods 18(2-3), 159-175.	
2008	Contrera JF, Matthews EJ, Kruhlak NL & Benz RD (2008). In Silico Screening of Chemicals for Genetic Toxicity Using MDL-QSAR, Nonparametric Discriminant Analysis, E-State, Connectivity, and Molecular Property Descriptors. Toxicology Mechanisms and Methods 18(2-3), 207-216.	MDL-QSAR model for mutagenicity, clastogenicity and DNA damage
2008	Matthews EJ, Kruhlak NL, Benz RD, Contrera JF, Marchant CA & Yang C (2008). Combined Use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows Software to Achieve High-Performance, High-	Comparative evaluation

Year	Reference	Comment
	Confidence, Mode of Action-Based Predictions of Chemical Carcinogenesis in Rodents. Toxicology Mechanisms and Methods 18(2-3), 189-206.	
2008	Yang C, Hasselgren CH, Boyer S, Arvidson K, Aveston S, Dierkes P, Benigni R, Benz RD, Contrera J & Kruhlak NL (2008). Understanding Genetic Toxicity Through Data Mining: The Process of Building Knowledge by Integrating Multiple Genetic Toxicity Databases. Toxicology Mechanisms and Methods 18(2-3), 277-295.	
2008	Cotterill JV, Chaudhry MQ, Matthews W & Watkins RW (2008). In silico assessment of toxicity of heat-generated food contaminants. Food and Chemical Toxicology 46, 1905–1918.	TOPKAT and Derek used for predicting carcinogenicity, mutagenicity and acute toxicity.
2007	Benigni R., Netzeva T, Benfenati E, Bossa C, Franke R, Helma C, Hulzebos E, Marchant C, Richard A, Woo Y-T & Yang C (2007). The expanding role of predictive toxicology: An update on the (Q)SAR models for mutagens and carcinogens. Journal of Environmental Science and Health 25, 53–97.	Non-commercial QSARs
2007	Kulkarni SA, Moir D & Zhu J (2007). Influence of structural and functional modifications of selected genotoxic carcinogens on metabolism and mutagenicity - a review. SAR and QSAR in Environmental Research 18(5-6), 459-514.	
2007	Kruhlak N, Contrera J, Benz D & Matthews E (2007). Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. Advanced Drug Delivery Reviews (2007), 59(1), 43-55.	FDA use of QSAR, application to pharmaceutical impurities
2007	Contrera JF, Kruhlak NL, Matthews EJ & Benz RD (2007). Comparison of MC4PC and MDL-QSAR rodent carcinogenicity predictions and the enhancement of predictive performance by combining QSAR models. Regulatory Toxicology and Pharmacology 49(3), 172-182.	Comparative evaluation
2007	Mazzatorta P, Tran L-A, Schilter B & Grigorov M (2007). Integration of Structure-Activity Relationship and Artificial Intelligence Systems To Improve in Silico Prediction of Ames Test Mutagenicity. Journal of Chemical Information and Modeling 47(1), 34-38.	Comparative evaluation
2006	Benigni R & Bossa C (2006). Structure-activity models of chemical carcinogens: state of the art, and new directions. Annali dell'Istituto Superiore di Sanita 42(2), 118-126	
2006	Greene N (2006). Computational models to predict toxicity. In Comprehensive Medicinal Chemistry II (Eds Taylor JB & Triggle DJ) 5, 909-932	
2006	Dobo KL, Greene N, Cyr MO, Caron S & Ku WW (2006). The application of structure-based assessment to support safety and chemistry diligence to manage genotoxic impurities in active pharmaceutical ingredients during drug development. Regulatory Toxicology and Pharmacology 44(3), 282-293.	Industry use. Identification of impurities in active pharmaceutical ingredients during drug development
2006	Veith G (2006). Roles for QSAR in risk assessment. ALTEX Alternativen zu Tierexperimenten (2006), 23 Suppl, 369-72.	
2006	Snyder RD, Ewing D & Hendry LB (2006). DNA intercalative potential of marketed drugs testing positive in in vitro cytogenetics assays. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 609(1), 47-59.	
2006	Cronin M (2006). The role of hydrophobicity in toxicity prediction. Current Computer-Aided Drug Design, 2(4), 405-413.	
2005	Woo Y-T & Lai DY (2005). OncoLogic: A mechanism-based expert system for predicting the carcinogenic potential of chemicals. In *Predictive Toxicology* (C Helma, ed), 385-413.	
2005	Parsons S & McBurney P (2005). The use of expert systems for toxicology risk prediction. In *Predictive Toxicology* (C Helma, ed), 135-175.	
2005	Helguera A, Perez M, Combes R & Gonzalez M (2005). The prediction of carcinogenicity from molecular structure. Current Computer-Aided Drug Design, 1(3), 237-255.	
2005	Jacobs A (2005). Prediction of 2-Year Carcinogenicity Study Results for Pharmaceutical Products: How Are We Doing? Toxicological Sciences 88(1), 18-	Pharmaceutical products

Year	Reference	Comment
	23.	
2005	Benigni R (2005). Structure• activity relationship studies of chemical mutagens and carcinogens: Mechanistic investigations and prediction approaches. Chemical Reviews 105 (5), 1767–1800.	
2005	Snyder RD & Hendry LB (2005). Toward a greater appreciation of noncovalent chemical/DNA interactions: Application of biological and computational approaches. Environmental and Molecular Mutagenesis 45(2/3), 100-105.	Non-covalent binding of chemicals to DNA and genotoxicity
2005	Snyder RD & Smith MD (2005). Computational prediction of genotoxicity: room for improvement. Drug Discovery Today 10(16), 1119-1124.	
2005	Hall LH & Hall LM (2005). QSAR Modeling Based on Structure-Information for Properties of Interest in Human Health. SAR and QSAR in Environmental Research 16(1-2), 13-41.	2963 compounds, including 290 therapeutic drugs, 400 in external validation
2005	Crettaz P & Benigni R (2005). Prediction of the Rodent Carcinogenicity of 60 Pesticides by the DEREKfW Expert System. Journal of Chemical Information and Modeling 45: 1864-1873.	Evaluation of Derek
2005	Hulzebos E, Sijm D, Traas T, Posthumus R & Maslankiewicz L. (2005). Validity and validation of expert (Q)SAR systems. SAR and QSAR in Environmental Research 16(4), 385-401.	Comparative evaluation
2005	Hayashi M, Kamata E, Hirose A, Takahashi M, Morita T & Ema M (2005). In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 588(2), 129-135.	Comparative evaluation
2004	Combes R & Rodford R (2004). The use of expert systems for toxicity prediction - illustrated with reference to the DEREK program. In Predicting Toxicity and Fate (MTD Cronin and DJ Livingstone, Eds), CRC Press, 193–204.	
2004	Benigni R (2004). Prediction of human health endpoints: mutagenicity and carcinogenicity. In Predicting Toxicity and Fate (Cronin M & Livingstone D, ed.), CRC Press, 173-192.	
2004	Snyder RD, Pearl GS, Mandakas G, Choy WN, Goodsaid F & Rosenblum IY (2004). Assessment of the sensitivity of the computational programs DEREK, TOPKAT, and MCASE in the prediction of the genotoxicity of pharmaceutical molecules. Environmental and Molecular Mutagenesis 43(3), 143-158.	Evaluation of Derek, TOPKAT, MCASE. Genotoxicity of 394 marketed pharmaceuticals.
2004	Benigni R & Zito R (2004). The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. Mutation Research, Reviews in Mutation Research, 566(1), 49-63.	Comparative evaluation
2004	Benigni R (2004). Chemical structure of mutagens and carcinogens and the relationship with biological activity. Journal of Experimental & Clinical Cancer Research 23(1), 5-8.	
2004	Rosenkranz HS (2004). SAR modeling of genotoxic phenomena: the consequence on predictive performance of deviation from a unity ratio of genotoxicants/non-genotoxicants. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 559(1-2), 67-71.	
2004	Helma C, Cramer T, Kramer S & De Raedt L (2004). Data Mining and Machine Learning Techniques for the Identification of Mutagenicity Inducing Substructures and Structure Activity Relationships of Noncongeneric Compounds. Journal of Chemical Information and Computer Sciences 44(4), 1402-1411.	Comparative evaluation
2004	Votano JR, Parham M, Hall LH, Kier LB, Oloff S, Tropsha A, Xie Q & Tong W (2004). Three new consensus QSAR models for the prediction of Ames genotoxicity. Mutagenesis 19(5), 365-77.	Comparative evaluation
2003	Patlewicz G, Rodford R & Walker JD (2003). Quantitative structure-activity relationships for predicting mutagenicity and carcinogenicity. Environmental Toxicology and Chemistry 22(8), 1885-1893.	Describes Derek, TOPKAT, MCASE, ADAPT, QSAR-ES, COMPACT, COREPA
2003	Dearden JC (2003). In silico prediction of drug toxicity. Journal of Computer-	Describes TOPKAT, MCASE, Derek,

Year	Reference	Comment
	Aided Molecular Design 17(2-4), 119-127.	Oncologic, HazardExpert, COMPACT
2003	Richard A & Williams C (2003). Public sources of mutagenicity and carcinogenicity data. In Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens (R Benigni, ed), pp 151-177. CRC Press.	Public sources of mutagenicity and carcinogenicity data
2003	Passerini L (2003). QSARs for individual classes of chemical mutagens and carciongens. In Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens (R Benigni, ed), pp 81-123. CRC Press.	Literature models for congeneric classes
2003	Benigni R, Giuliani A, Gruska A & Franke R (2003). QSARs for the mutagenicity and carcinogenicity of the aromatic amines. In Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens (R Benigni, ed), pp 125-144. CRC Press.	Literature models for aromatic amines
2003	Schultz T, Cronin M & Netzeva T (2003). The present status of QSAR in toxicology. Journal of Molecular Structure-THEOCHEM 622, 23-38.	
2003	Schultz T, Cronin M, Walker J & Aptula A (2003). Quantitative structure-activity relationships in (QSARs) in toxicology: A historical perspective. Journal of Molecular Structure-THEOCHEM 622, 1-22.	
2003	White AC, Mueller RA, Gallavan RH, Aaron S & Wilson AGE (2003). A multiple in silico program approach for the prediction of mutagenicity from chemical structure. Mutation Research, Genetic Toxicology and Environmental Mutagenesis 539(1-2), 77-89.	Comparative evaluation of TOPKAT, Derek, CASETOX. Test set 1- 520 proprietary drug candidates; test set 2 - 94 commercial componds
2003	Klopman G, Chakravarti S, Harris N, Ivanov J & Saiakhov R (2003). In-silico screening of high production volume chemicals for mutagenicity using the MCASE QSAR expert system. SAR and QSAR in Environmental Research 14(2), 165-180.	Evaluation of MCASE. Test set of 2484 HPV chemicals with and without metabolic activation
2003	Toivonen H, Srinivasan A, King RD, Kramer S & Helma C (2003). Statistical evaluation of the Predictive Toxicology Challenge 2000-2001. Bioinformatics 19(10), 1183-1193.	Comparative evaluation
2003	Mekenyan O, Dimitrov S, Schmieder P & Veith G (2003). In silico modelling of hazard endpoints: current problems and perspectives. SAR and QSAR in Environmental Research 14(5-6), 361-371.	
2003	Cronin MT, Dearden J, Walker J & Worth A (2003). Quantitative structure-activity relationships for human health effects: Commonalities with other endpoints. Environmental Toxicology and Chemistry 22(8), 1829-1843.	
2003	Cronin MT, Jaworska J, Walker J, Comber M, Watts C & Worth A (2003). Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. Environmental Health Perspectives 111(10), 1391-1401.	
2003	Benigni R & Zito R (2003). Designing safer drugs: (Q)SAR-based identification of mutagens and carcinogens. Current Topics in Medicinal Chemistry 3(11), 1289-1300.	
2002	Greene N (2002). Computer systems for the prediction of toxicity: an update, Advanced Drug Delivery Reviews 54, 417–431.	Describes Derek, MCASE, TOPKAT, HazardExpert, TOXSYS, COMPACT, Oncologic
2002	Cariello NF *et al*. (2002). Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. Mutagenesis 17(4): 321-329.	Comparative evaluation
2002	Young R (2002). Genetic toxicology: Web resources. Toxicology 173(1-2), 103-121.	Review of databases
2001	Prival MJ (2001). Evaluation of the TOPKAT System for Predicting the Carcinogenicity of Chemicals." Environmental and Molecular Mutagenesis 37: 55-69.	Evaluation of TOPKAT
2000	Combes R. (2000). The use of structure-activity relationships and markers of cell toxicity to detect non-genotoxic carcinogens. Toxicology in Vitro 14(4), 387-399.	Identification of non-genotoxic carcinogens

Year	Reference	Comment
1999	Zhu X, Zhang Y, Klopman G & Rosenkranz H (1999). Thalidomide and metabolites: indications of the absence of 'genotoxic' carcinogenic potentials. Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 425(1), 153-167.	Evaluation of MCASE, META in assessment of thalidomide and metabolites
1999	Tuppurainen K (1999). Frontier orbital energies, hydrophobicity and steric factors as physical QSAR descriptors of molecular mutagenicity. A review with a case study: MX compounds. Chemosphere 38(13), 3015-3030.	
1999	Fu P, Von Tungeln L, Chiu L-H & Own Z (1999). Halogenated-polycyclic aromatic hydrocarbons: A class of genotoxic environmental pollutants. Journal of Environmental Science and Health, Part C: Environmental Carcinogenesis & Ecotoxicology Review, C17(2), 71-109.	
1998	Henry B, Grantv S, Klopman G & Rosenkranz H (1998). Induction of forward mutations at the thymidine kinase locus of mouse lymphoma cells: evidence for electrophilic and non-electrophilic mechanisms. Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 397(2), 313-335.	Evaluation of MCASE, 209 chemicals
1998	Richard A (1998). Structure-based methods for predicting mutagenicity and carcinogenicity: are we there yet? Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 400(1,2), 493-507	Describes TOPKAT, CASE/MCASE, Derek, Oncologic
1997	Dearden J *et al*. (1997). The Development and Validation of Expert Systems for Predicting Toxicity. ATLA 25, 223-252.	Describes Derek, CASE, COMPACT, TOPKAT, HazardExpert, Oncologic, REX, DTOX, PROLOG
1997	Polloth C & Mangelsdorf I (1997). Commentary on the application of (Q)SAR to the toxicological evaluation of existing chemicals, Chemosphere 35(11), 2525-2542.	Commentary, with emphasis on industrial (HPV) chemicals
1996	Marchant & Collaborative (1996). Prediction of Rodent Carcinogenicity Using the DEREK System for 30 Chemicals Currently Being Tested by the National Toxicology Program. Environmental Health Perspectives 104(5): 1065-1073.	Comparative evaluation
1996	Rosenkranz HS, Zhang YP & Klopman G (1996). Studies on the potential for genotoxic carcinogenicity of fragrances and other chemicals. Food and Chemical Toxicology 36(8), 687-696.	Comparative evaluation
1996	Zeiger E *et al*. (1996). Prediction of Salmonella mutagenicity. Mutagenesis 11(5): 471-484.	Comparative evaluation
1996	Woo Y-T, Lai D, Argus M & Arcos J (1996). Carcinogenicity of organophosphorus pesticides/compounds: An analysis of their structure-activity relationships, Environmental Carcinogenesis & Ecotoxicology Reviews, C14(1), 1-42.	Carcinogenicity of organophosphorus pesticides/compounds
1996	Lai D, Woo Y-T, Argus M & Arcos J (1996). Carcinogenic potential of organic peroxides: Prediction based on structure-activity relationships (SAR) and mechanism-based short-term tests. Environmental Carcinogenesis & Ecotoxicology Reviews, C14(1), 63-80.	Organic peroxides
1995	Rosenkranz H & Klopman G (1995). Structure-activity relationships as alternatives in the study of carcinogenesis, Alternative Methods in .Toxicology and the Life Sciences, 11 (The World Congress on Alternatives and Animal Use in the Life Sciences: Education, Research, Testing, 1993), 379-390.	
1994	Vogel E & Ashby J (1994). Structure-activity relationships: experimental approaches. SCOPE 52, 231-54.	
1994	Waters M, Richard A, Rabinowitz J, Stack F, Garrett N, Lohman P & Rosenkranz H (1994). Structure-activity relationships: computerized systems. SCOPE 52, 201-29.	
1994	Benigni R & Giuliani A (1994). Quantitative structure-activity relationship (QSAR) studies in genetic toxicology: mathematical models and the "biological activity" term of the relationship, Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 306(2), 181-6.	
1994	Klopman G & Rosenkranz H (1994). Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTI-CASE,	

Year	Reference	Comment
	Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 305(1), 33-46.	
1994	Mersch-Sundermann V, Rosenkranz H & Klopman G (1994). The structural basis of the genotoxicity of nitroarenofurans and related compounds. Mutation Research 304(2), 271-84.	Evaluation of CASE, 79 nitroarenofurans
1994	Eder E, Hoffman C, Deininger C & Scheckenbach S (1994). Risk assessment for mutagenic and carcinogenic activities of •,•-unsaturated carbonyl compounds by a screening strategy based on structure-activity relationships. Toxicology in Vitro 8(4), 707-10.	•,•-unsaturated carbonyl compounds
1992	Benigni R & Giuliani A (1992). QSAR studies in genetic toxicology: congeneric and noncongeneric chemicals. Archives of Toxicology, Supplement 15 (Med. Toxicol.), 228-237.	
1991	Tenant RW & Ashby J (1991). 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. Mutation Research 257: 209-227.	Comparative evaluation
1991	Hansch C (1991). Structure-activity relationships of chemical mutagens and carcinogens Science of the Total Environment 109-110, 17-29.	
1990	Waters M, Richard A, Rabinowitz J, Stack H & Garrett N (1990). Structure-activity relationships. Computerized systems. EPA Report EPA/600/D-90/125; Order No. PB90-263476, 75.	
1990	Rosenkranz HS & Klopman G (1990). Structural alerts to genotoxicity: the interaction of human and artificial intelligence. Mutagenesis 5(4), 333-61.	
1990	Rosenkranz H & Klopman G (1990). Evaluating the ability of CASE, an artificial intelligence structure-activity relational system, to predict structural alerts for genotoxicity. Mutagenesis 5(6), 525-527.	Evaluation of CASE, 39 chemicals
1990	Rosenkranz HS & Klopman G (1990). Structural implications of the ICPEMC method for quantifying genotoxicity data. Mutation Research 305(1), 99-116.	
1990	Blake BW, Enslein K, Gombar VK & Borgstedt HH (1990). Salmonella mutagenicity and rodent carcinogenicity: quantitative structure-activity relationships. Mutation research 241(3), 261-71.	
1989	Richard A, Rabinowitz J & Waters M (1989). Strategies for the use of computational SAR methods in assessing genotoxicity. Mutation Research, Reviews in Genetic Toxicology 221(3), 181-96.	
1989	Miertus S, Frecer V & Majekova M (1989). QSAR and mechanistic studies on the genotoxic compounds including environmental effects. International Journal of Quantum Chemistry 35(1), 153-165.	

Table 7. Primary (Q)SAR studies on genotoxicity published in the past 10 years

Reference	Chemical classes	Dataset size
Bentzien *et al*. (2010)	Primary aromatic amines	257
Pérez-Garrido *et al*. (2010)	Acrylates, methacrylates and •,•-unsaturated carbonyl compounds	220 (Ames data) 48 (mammalian cells mutagenesis data)
Carlsson *et al*. (2009)	Diverse compounds from CCRIS database	4254
Hansen *et al*. (2009)	Drugs + diverse compounds from CCRIS, VITIC, Genetox and published papers	6512
Holder *et al*. (2009)	Oxiranes (epoxides), nitrogen-containing molecules, and halogenated compounds	44
Onchoke (2009)	Triphenylene, 1-, and 2-nitrotriphenylene	3
Rothenbacher *et al*. (2009)	Polyethylene, and other substances from food packing materials	46
Toropov *et al*. (2009)	Heteroaromatic amines	95
Toropov *et al*. (2009)	Nitrated polycyclic hydro-carbons	48
Venkatapathy *et al*. (2009)	Diverse compounds from CPDB	590
Wang *et al*. (2009)	Diverse compounds from CPDB	693
Wang *et al*. (2009)	Polycyclic aromatic hydrocarbons	26
Contrera *et al*. (2008)	Pharmaceuticals, food constituents, and environmental and industrial chemicals	1254
Cunningham *et al*. (2008)	Industrial chemicals	104
Didziapetris *et al*. (2008)	Diverse compounds from CCRIS database	8000
Dorn *et al*. (2008)	Steroids and other compounds	26
Du *et al*. (2008)	Thiophene derivatives	140
Koleva *et al*. (2008)	•,•-unsaturated carbonyl compound (aldehydes and ketones)	77
Kulkarni *et al*. (2008)	Organic chemicals	11
Langham *et al*. (2008)	Diverse organic compounds from the CCRIS, Toxnet, NTP, CPDB and GAP databases	4737
Nair *et al*. (2008)	Nitroarenes	197
Papa *et al*. (2008)	Polycyclic aromatic hydrocarbons, oxo- and nitro- and unsubstituted PAHs	70
Perez-Garrido *et al*. (2008)	Haloacetic acids, nitrohaloalkanes, haloacids, haloalde- hides, halocetones, haloalcohols, haloepoxides, and haloalkanes	42
Ruiz *et al*. (2008)	Polychlorinated biphenyls	209
Singh *et al*. (2008)	Nitrated polycyclic aromatic hydrocarbons	48
Zhang *et al*. (2008)	Nitronaphthalenes and methylnitronaphthalenes	16
Benigni *et al*. (2007)	Homocyclic aromatic amines	229
Boerth (2007)	Pesticides and metabolites	N/A
Borosky (2007)	Aromatic and Heterocyclic Aromatic Amines	17
Castro *et al*. (2007)	Dental monomers	16
Chroust *et al*. (2007)	Halogenated aliphatic compounds	116
Dorn *et al*. (2007)	Lipophilic chemicals such as aliphatic hydrocarbons and alcohols, colcemid, cytochalasin B, diamide, nitrobenzene and benzonitrile, phytoestrogens genistein and daidzein, other hormonal steroids	33
Fang *et al*. (2007)	Aromatic and heteroaromatic amines	80

Reference	Chemical classes	Dataset size
Gramatica *et al.* (2007)	Nitrated Polycyclic Arom. Hydrocarbons	48
Gramatica *et al.* (2007)	Benzocyclopentaphenanthrenes/chrysenes	32
Hu *et al*. (2007)	Quinolone antibacterials	20
Mazzatorta *et al.* (2007)	Diverse compounds from the CCRIS database	5090
Sangamwar *et al.* (2007)	Antifungal compounds, mainly azoles	30
Serafimova *et al*. (2007)	Diverse compounds from NTP database and proprietary chemicals provided by BASF AG	2844
Tekiner-Gulbas *et al.* (2007)	2,5-disubstituted benzoxazole and benzimidazole derivatives	21
Toropov *et al.* (2007)	1) organic compounds 2) heteroaromatic amines 3) TIBO (tetrahydroimidazobenzodiazepinone) and HEPT (hydroxyethoxymethylphenylthiothymine) derivatives	1) 44 2) 94 3) 32
Valerio *et al*. (2007)	Small organic naturally accurring chemicals found in the human diet	120
Xiao *et al*. (2007)	Substituted phenols	29
Zhang *et al.* (2007)	Organic compounds from the CCRIS, Toxnet, NTP, CPDB and GAP databases	4555
Buttingsrud *et al.* (2006)	Aromatic and heteroaromatic nitro compounds	335
Casalegno *et al.* (2006)	Aromatic amines	100
Dobo *et al.* (2006)	Pharmaceutical actives, impurities and intermediates	272
Estrada *et al*. (2006)	Organic compounds, drugs, food additives, agrochemicals, cosmetic materials,medicinal products, and household materials	372
Hayashi *et al.* (2006)	Diverse compounds: 703 from CGX database (Kirkland *et al*., 2005) and 206 existing chemicals from Japanese ECJ database	909 (703+206)
Helma (2006)	Diverse compounds from CPDB	3895
Joshi *et al.* (2006)	Phenyltriazenes	17
Kim *et al*. (2006)	Benz[a]anthracene (BA), metabolic products - polycyclic aromatic compounds (PAH) include BA-oxides, -phenols, - quinones, -dihydrodiols and -diolepoxides	29
Knize *et al.* (2006)	Isomeric series of heterocyclic amines (amino-trimethylimidazopyridine (TMIP) isomers	11
Lanevskij *et al.* (2006)	Diverse compounds	945
Shoji *et al*. (2006)	Diverse compounds (environmental pollutants)	82
Takamura-Enya *et al.* (2006)	Aromatic nitro compounds - nitro derivatives of benzan-throne	9
Vracko (2006)	Pyriminoizodiamine isomers and aromatic/ heteroaromatic amines	12 pyriminoizodiamines and 95 aromatic/ heteroaromatic amines
Benigni *et al*. (2005)	Alpha,beta-unsaturated aldehydes	26
Bhat *et al.* (2005)	Derivatives of aromatic amines	181
Cash *et al.* (2005)	Aromatic amines	29
Cho (2005)	Mutagen X (bi-product caused by this disinfection process) and analogues	37
Contrera *et al.* (2005)	Pharmaceuticals and non-pharmaceuticals	3228
Crettaz *et al.* (2005)	Pesticides	60

Reference	Chemical classes	Dataset size
Glowienke *et al*. (2005)	Methane-, benzene- and toluene-sulfonic acid esters	19
Gonzalez *et al*. (2005)	Dental monomers	53
Hayashi *et al*. (2005)	Diverse compounds	909
Jacobs (2005)	Pharmaceuticals	N/A
Kazius *et al*. (2005)	Diverse compounds from CCRIS, Toxnet, NTP, CPDB and GAP	4872
Mahe *et al*. (2005)	1) aromatic and hetero-aromatic nitro compounds 2) Diverse compounds from CPDB	1) 230 2) 684
Mekenyan *et al*. (2005)	Pesticide active ingredients	25
Tarasov *et al*. (2005)	Diverse compounds from NTP, Gentox and GAP	105
Wang *et al*. (2005)	Nitroaromatics	219
Xiao *et al*. (2005)	Substituted benzenes compds	36
Andrews *et al*. (2004)	N-acyloxy-N-alkoxyamide analogues	41
Bang *et al*. (2004)	Mutagen X and its analogs	29
Gonzalez *et al*. (2004)	Dental monomers - cycloaliphatic epoxides	15
Gonzalez *et al*. (2004)	Dental monomers - aromatic epoxides	16
Helguera *et al*. (2004)	Dental monomers	23
Helma *et al*. (2004)	Diverse compounds from CPDB	684
Jezierska *et al*. (2004)	Aromatic and heteroaromatic amines	95
Klopman *et al*. (2004)	Organic compounds from the NTP and Gene-Tox databases	2513
Popelier *et al*. (2004)	Heteroaromatic triazenes, halogenated hydroxyfuranones (Mutagen X derivatives)	23/24
Rosenkranz (2004)	Diverse compounds from the Zeiger genotoxicity database	300
Snyder *et al*. (2004)	Pharmaceuticals	394
Valkova *et al*. (2004)	Aromatic and heteroaromatic amines	95
Votano *et al*. (2004)	Therapeutic drugs, diverse compounds from PHYSPROP, TOXNET, RTECS and drugs from the Physicians Desk Reference, version 6.0a (2003)	10,000
Votano *et al*. (2004)	Diverse compounds from TOXMET, RTECS and drugs from the Physicians Desk Reference, version 6.0a (2003)	336
Vracko *et al*. (2004)	Trimethylimidazopyridine isomers	12
Vracko *et al*. (2004)	Aromatic amines	95
Basak *et al*. (2003)	Neutral halocarbons	55
Benigni *et al*. (2003)	Simple aldehydes and •-• unsaturated. aldehydes	21
Chen *et al*. (2003)	Benzidine (BZ) and its six structural analogs	7
Cho (2003)	Mutagen X (MX), 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone and its analaogs (open and ring form)	7
Gramatica *et al*. (2003)	Aromatic amines	146
Halova *et al*. (2003)	Aromatic and heteroaromatic nitro compounds	N/A
He *et al*. (2003)	Polycyclic aromatic compounds (PACs)	277
Klopman *et al*. (2003)	High Production Volume Chemicals	2484
Lewis *et al*. (2003)	Benzidine and amino-biphenyl analogues	11
Mattioni *et al*. (2003)	Aromatic and secondary amine compounds	334
McElroy *et al*. (2003)	Diverse organic compounds from the NTP database and the Data Book of Chromosomal Aberration	297

Reference	Chemical classes	Dataset size
	Test In Vitro (Sofuni *et al.*)	
Mosier *et al*. (2003)	Thiophene derivatives	N/A
Rosenkranz (2003)	Diverse compounds (more detailed information not available)	10,000
Sztandera *et al.* (2003)	Aminoazo derivatives and their reductive cleavage products	62 aminoazo, 12 cleavage products
White *et al*. (2003)	Diverse compounds, drug candidates	94 diverse;,520 drugs
Bacha *et al*. (2002)	Diverse compounds from TOXNET and GAP	N/A
Garg *et al.* (2002)	Aminoazobenzene derivatives	43
Ivanciuc (2002)	Methylated and non-methylated polycyclic aromatic hydrocarbons (PAHs)	87
Kauffman *et al*. (2002)	Diverse secondary and aromatic amine compounds	256
Kubo *et al.* (2002)	Diverse compounds	255
Livingstone *et al*. (2002)	Diverse compounds from the literature	90
Poletti *et al*. (2002)	Cyclopentaphenanthrene derivatives	31
Roberts *et al.* (2002)	Substituted naphthoquinones	29
Rosenkranz (2002)	Diverse compounds from NTP and CPDB, as well as the Sigma, Aldrich, Lancaster Synthesis, Fluka catalogues	10,000
Toropov *et al.* (2002)	Heteroaromatic amines	73
Woo *et al*. (2002)	Disinfection Byproducts	N/A
Cash (2001)	Aromatic and heteroaromatic	95
Rosenkranz *et al.* (2001)	Diverse compounds	1500
Yourtee *et al*. (2001)	Dental monomers	54
Janzowski *et al.* (2000)	2-alkenal food relevant compds	7
Karelson *et al.* (2000)	Heteroaromatic and aromatic amines	95
Lozano *et al*. (2000)	Heterocyclic amines	12
Baeten *et al.* (1999)	Chlorinated hydrocarbons	10
Maran *et al.* (1999)	Heteroaromatic and aromatic amines	95
Tuppurainen (1999)	Halogenated hydroxyfuranones including MX	29
Zhu *et al*. (1999)	Thalidomide, and its metabolites	131

CCRIS - Carcinogenesis Research Information System database; CPDB - Carcinogenic Potency Database; GAP- Genetic Activity Profile Database developed by US EPA; NTP – National Toxicology Progam database; PHYSPROP – SRC Physical Properties Database (http://srcinc.com/); TOXNET - National Library of Medicine's Toxicology Data Network

Table 8. Primary (Q)SAR studies on carcinogenicity published in the past 10 years

Reference	Chemical classes	Dataset size
Bercu *et al.* (2010)	Diverse compounds from CPDB	694
Tanabe *et al.* (2010)	Diverse compounds	911
Fjodorova *et al.* (2010)	Diverse compounds from CPDB	805
Wang *et al.* (2009)	Diverse compounds	693
Fjodorova *et al.* (2009)	Diverse compounds	805
Massarelli *et al.* (2009)	Diverse compounds from CPDB	55
Toropov *et al.* (2009)	Diverse compounds from CPDB	401
Venkatapathy *et al.* (2009)	Diverse chemicals from CPDB	590
Bruce *et al.* (2008)	Polycyclic aromatic hydrocarbons	23
Cunningham *et al*, (2008)	Industrial chemicals	104
Fratev *et al.* (2008)	Benzene derivatives	100
Helguera *et al.* (2008)	Nitroso compounds	39
Helguera *et al.* (2008)	Nitro compounds	55
Helguera *et al.* (2008)	Nitroso compounds	26
Matthews *et al.* (2008)	Diverse compounds from FDA/CDER database	1572
Mayer *et al.* (2008)	Diverse compounds from CPDB	650
Ruiz *et al.* (2008)	Polychlorinated biphenyls	209
Vijayalakshmi *et al.* (2008)	Polycyclic aromatic hydrocarbons	17
Zhu *et al.* (2008)	Diverse compounds from NTP	384
Bull & Reckhow (2007)	Haloquinones, halogenated furans and nitrosamines	21
Contrera *et al.* (2007)	Diverse compounds, including marketed drugs and organic chemicals with pharmacologic properties, and non-therapeutics that are pesticides or industrial chemicals.	1540
Deeb *et al.* (2007)	Sulfa drugs	18
Fang *et al.* (2007)	Aromatic and heteroaromatic amines	80
Fratev *et al.* (2007)	Benzene derivatives	100
Helguera *et al.* (2007)	Nitroso compounds	35
Valerio *et al.* (2007)	Small organic naturally occurring chemicals found in the human diet	120
Dhar *et al.* (2006)	Benz[a]anthracenes and benz[c]acridine	14
Helguera *et al.* (2006)	Nitro compounds	62
Helguera *et al.* (2006)	Nitro compounds (aromatic and aliphatic)	49
Helguera *et al.* (2006)	Nitro compounds	188
Helma (2006)	Diverse compounds from CPDB	3985
Matthews *et al.* (2006a,b)	Diverse compounds	1442
Bailey *et al.* (2005)	Food contact substances from RTECS, PAFA, CPDB	N/A

Reference	Chemical classes	Dataset size
Chen *et al*. (2005)	NCTR estrogen activity data set containing 232 structurally diverse chemicals, and NCTR liver cancer database generated	232 / 996
Contrera *et al*. (2005)	Organic compounds	1072
Crettaz *et al*. (2005)	Pesticides	60
Helguera *et al*. (2005)	Diverse compounds from CPDB	189
Hemmateenejad *et al*. (2005)	Drugs	735
Lagunin *et al*. (2005)	Diverse compounds from NTP and CPDB	1602
Rallo *et al*. (2005)	Aromatic compounds containing nitrogen substituents	104
Suzuki (2005)	Diverse compounds from NTP	323
Woo *et al*. (2005)	New chemicals	N/A
Benigni *et al*. (2003)	Simple aldehydes and alpha-beta unsaturated aldehydes	8 / 21
Shen *et al*. (2003)	Aromatic amines	N/A
Toivonen *et al*. (2003)	Diverse compounds from NTP and CPDB	2419
Marino *et al*. (2002)	Methylated polycyclic aromatic hydrocarbons	49
Kabankin *et al*. (2001)	Aromatic amines	38
Gallegos *et al*. (2001)	Polycyclic aromatic hydrocarbons	78

CCRIS - Carcinogenesis Research Information System database; CPDB - Carcinogenic Potency Database; NCTR - National Center for Toxicological Research; NTP – National Toxicology Progamme database; PAFA – US FDA Priority-Based Assessment of Food Additives database

European Commission

EUR 24427 EN – Joint Research Centre – Institute for Health and Consumer Protection

Title: Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity

Author(s): Serafimova R, Fuart Gatnik M and Worth A

Luxembourg: Publications Office of the European Union

2010 – 49 pp. – 21 x 29,7 cm
EUR – Scientific and Technical Research series - ISSN 1018-5593
ISBN 978-92-79-16103-2
doi:10.2788/26123

Abstract

The review of QSARs for genotoxicity and carcinogenicity was performed in a broad sense, considering both models available in software tools and models that are published in the literature. The review considered the potential applicability of diverse models to pesticides as well as to other types of regulated chemicals and pharmaceuticals. The availability of models and information on their applicability is summarised in tables, and a range of illustrative or informative examples are described in more detail in the text. In many cases, promising models were identified but they are still at the research stage. For routine application in a regulatory setting, further efforts will be needed to explore the applicability of such models for specific purposes, and to implement them in a practically useful form (i.e. user-friendly software). It is also noted that a range of software tools are research tools suitable for model development, and these require more specialised expertise than other tools that are aimed primarily at end-users such as risk assessors. It is concluded that the most useful models are those which are implemented in software tools and associated with transparent documentation on the model development and validation process. However, it is emphasised that the assessment of model predictions requires a reasonable amount of QSAR knowledge, even if it is not necessary to be a QSAR practitioner.

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-24427-EN-N

ISBN 978-92-79-16103-2