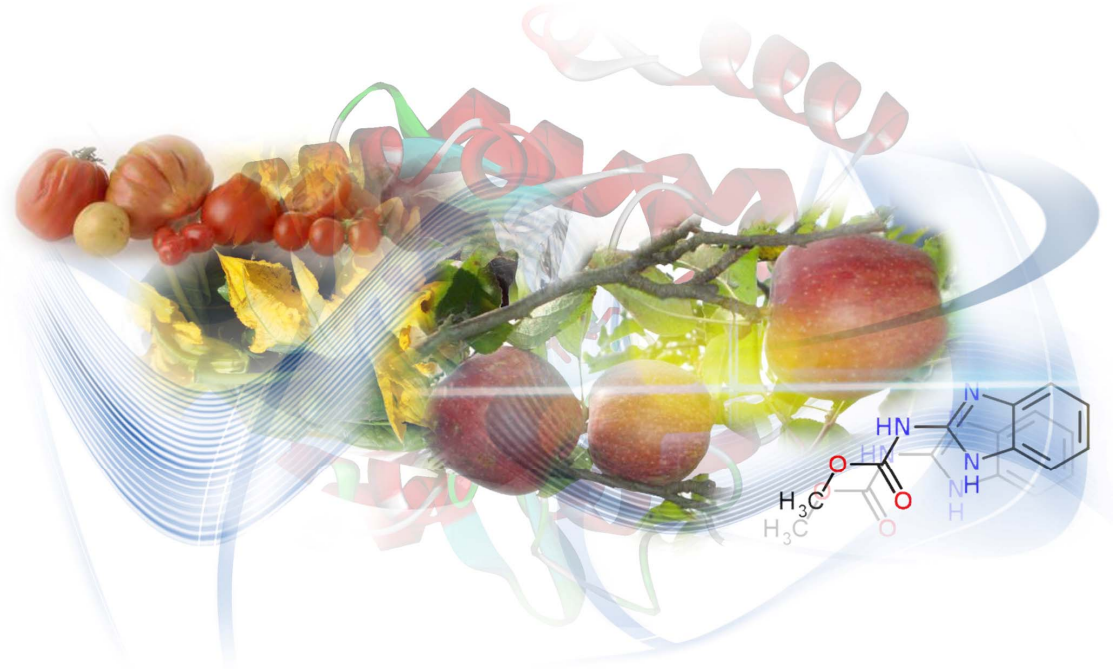
JRC Scientific and Technical Reports

# The Use of Computational Methods in the Toxicological Assessment of Chemicals in Food: Current Status and Future Prospects

**Andrew Worth, Mojca Fuart-Gatnik, Silvia Lapenna, Elena Lo Piparo, Aleksandra Mostrag-Szlichtyng and Rositsa Serafimova**



EUR 24748 EN - 2011

JRC EUROPEAN COMMISSION

ihcp Institute for Health and Consumer Protection

The mission of the JRC-IHCP is to protect the interests and health of the consumer in the framework of EU legislation on chemicals, food, and consumer products by providing scientific and technical support including risk-benefit assessment and analysis of traceability.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

**Contact information**
A. Worth
JRC - IHCP
Via Enrico Fermi, 2749 - 21027 Ispra (VA) Italy
Andrew.WORTH@ec.europa.eu
http://www.jrc.ec.europa.eu/

This publication is a Reference Report by the Joint Research Centre of the European Commission.



JRC63826

EUR 24748 EN
ISBN 978-92-79-19567-9
ISSN 1018-5593
doi:10.2788/6234

Luxembourg: Publications Office of the European Union



*Printed in Italy*

## Acknowledgements

This report is an updated and summarised version of a report prepared by the Joint Research Centre (JRC) for European Food Safety Authority (EFSA) under the terms of a Service Level Agreement between EFSA and the JRC (contract title: Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment; contract number: SLA/EFSA-JRC/2009-01).

Many collaborators, too numerous to mention by name here, contributed valuable information or advice during the EFSA project or during the preparation of this report. The authors appreciate the continuous support provided by EFSA staff (Hans Steinkellner, Luc Mohimont, Miriam Jacobs) and members of the EFSA project Steering Group (André Muller, Emilio Benfenati, Claudia Bolognesi). The authors are grateful to Matteo Stocchero (Soluzioni Informatiche, Vicenza, Italy) for support in the generation and interpretation of ToxBoxes data, as well as to Christoph Helma (*in silico* toxicology, Basel, Switzerland) for providing data from the Lazar software. Signficant contributions to the survey on the use of QSAR analysis were provided by Steven Bradbury and Mary Manibusan (US EPA), Chihae Yang (US FDA), Benoît Schilter and Paolo Mazzatorta (Nestlé Research Center, Lausanne, Switzerland). The authors also acknowledge the useful comments provided by JRC IHCP staff (Sharon Munn, Elisabet Berggren, Paolo Castello, Maurice Whelan) as well as the valuable editorial support provided by Jutta Triebe.

## Disclaimer

Any views stated in this report are those of the authors and do not constitute an official view of the Joint Research Centre or the Commission. In carrying out the reviews of computational prediction methods and databases, every effort was made to be extensive and representative. However, the explict mention of any given software tool or database does not imply any endorsement of its quality or use by the JRC or Commission; conversely, any omissions are unintentional.

## Table of Contents

## List of Abbreviations

| | |
|---|---|
| ADME | Absorption, Distribution, Metabolism and Elimination |
| ADMET | Absorption, Distribution, Metabolism, Elimination and Toxicity |
| AGES | Austrian Agency for Health and Food Safety (AGES) |
| CAS | Chemical Abstract Service |
| CRD | UK Chemicals Regulations Directorate |
| DAR | Draft Assessment Report |
| $E_{HOMO}$ | Energy of the Highest Molecular Orbital |
| $E_{LUMO}$ | Energy of the Lowest Molecular Orbital |
| EFSA | European Food Safety Authority |
| EPA | Environmental Protection Agency (EPA) |
| EU | European Union |
| FDA | Food and Drug Administration (USA) |
| JRC | Joint Research Centre |
| LOAEL | Lowest Observed Adverse Effect Level |
| MOA | Mode of Action |
| MoE | Margin of Exposure |
| MRL | Maximum Residue Level |
| MRTD | Maximum Recommended Therapeutic Dose |
| OECD | Organisation for Economic Cooperation and Development |
| PPP | Plant Protection Product |
| QMRF | QSAR Model Reporting Format |
| QPRF | QSAR Prediction Reporting Format |
| QSAR | Quantitative Structure-Activity Relationship |
| REACH | Registration, Evaluation, Authorisation and Restriction of Chemicals |
| SAR | Structure-Activity Relationship |
| SMILES | Simplified Molecular Input Line Entry System |
| TTC | Threshold of Toxicological Concern |

# EXECUTIVE SUMMARY

Chemicals present in food have various origins – they can be naturally occurring, intentionally added (additives, flavourings, preservatives, supplements), or inadvertently present (environmental pollutants, residues of pesticides and veterinary drugs, components of food contact materials, chemicals formed in food manufacturing and processing, chemicals produced by biological contaminants). Although these chemicals are generally present at very low levels, their potential for adversely affecting human health is a worldwide concern. Therefore, to protect the consumer against exposure to potentially harmful chemicals, national and international food safety standards are established on the basis of sound scientific risk assessments that define exposure levels that are considered to be "safe".

The risk assessment of chemicals present in food needs to account for the fact that in addition to the active/parent substance, which is generally well-characterised in terms of its bioavailability and toxicological properties, the consumer is also exposed to a wide range of substances resulting from metabolic and degradation processes. In the majority of cases, very limited information on the toxicological properties of metabolites and degradates is available. Since toxicological testing on animals is neither practicable nor desirable, alternative (non-animal) assessment methods are needed to support evaluations of the toxicological profile of chemicals in food, including metabolites and degradates. Computational methods that make predictions of bioavailability and toxicity on the basis of chemical structure are of particular interest, for reasons of cost-effectiveness, efficiency and animal welfare. Computational toxicology is a rapidly advancing discipline. However, the question of how to use computational methods in a reliable and practical manner for risk assessment purposes represents a considerable challenge, which is receiving increasing attention by national and international bodies, such as European Commission and the European Food Safety Authority (EFSA).

This report is based on the results obtained in the PESTISAR project, which the European Commission's Joint Research Centre (JRC) performed during 2009-2010 under contract to EFSA. The overall aim of the PESTISAR project was to evaluate the potential applicability of computational methods in the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances. Among the various types of computational estimation methods, emphasis was placed on Quantitative Structure-Activity Relationships (QSARs), Structure-Activity Relationships (SARs) and expert systems.

To address the overall aim of the PESTISAR project, the JRC performed a range of activities:

a) a survey was carried out to find out how QSAR analysis is used by national regulatory bodies and international advisory organisations in the field of food safety;

b) an extensive review was carried out of QSARs potentially useful in dietary risk assessment, focussing on toxicological endpoints (acute and repeat-dose toxicity, including organ and system-specific toxicities; genotoxicity and carcinogenicity; developmental and reproductive toxicity; immunotoxicity), and touching on endocrine-related effects (in particular nuclear hormone receptor-mediated effects);

c) an extensive review was carried out of computational models (with emphasis on QSARs and rule-based approaches) for biokinetic (ADME) properties, including oral bioavailability, human intestinal absorption, blood-brain barrier penetration, plasma protein binding, metabolism, and clearance;

d) case studies (research investigations) into the potential use of QSARs for genotoxicity and carcinogenicity, with a view to developing a conceptual framework for QSAR analysis that can be integrated with the application of the TTC concept;

e) a conceptual framework was developed for assessing the usefulness of QSAR models in terms of the practical applicability of the models and the adequacy of the predictions;

f) research and development needs were identified, leading to recommendations for further activities aimed at promoting the uptake and regulatory acceptance of computational methods in the food safety area.

This report is an update and summary of the main findings and conclusions of the PESTISAR project. Altough the PESTISAR project had a focus on pesticide risk assessment, the information presented here is broadly applicable to the risk assessment of food chemicals in general, rather than any product class in particular. As such, this report is intended to be a background document for the further development and application of computational (QSAR) methods in the food safety area.

Key words: QSAR, SAR, structural alert, expert system, *in silico*, toxicity, pesticide active substance, metabolite, dietary risk assessment, alternative method.

# 1. INTRODUCTION

The general objective of food safety policy is to protect consumer health. In the European Union (EU), Regulation (EC)178/2002 (EC, 2002) lays down the general principles and requirements of food law and procedures in matters of food safety, aiming at harmonising existing national requirements in order to ensure the free movement of food and feed throughout the EU. The regulation ensures a high level of protection of human life and health, taking into account the need to protect animal health and welfare, plant health and the environment. The risk assessment of food and feed in the EU is performed independently of risk management. The keystone of the risk assessment system for food and feed is the European Food Safety Authority (EFSA), which produces scientific opinions and advice to provide a sound foundation for EU policy and legislation. The activities of EFSA cover food and feed safety, nutrition, animal health and welfare, plant protection and plant health. As a service of the European Commission, the Joint Research Centre collaborates with EFSA in order to provide scientific support to EU policy in the food safety area.

## 1.1 Background to pesticide risk assessment and the PESTISAR project

One of the most important ways of protecting plants and plant products and of increasing agricultural yields is to use of plant protection products (PPPs). A possible consequence of their use may be the presence of pesticide residues in the treated products. It is therefore necessary to ensure that such residues should not be found in food or feed at levels presenting an unacceptable risk to humans. Maximum residue levels (MRLs) are therefore set by the European Commission at the lowest achievable level consistent with good agricultural practices to protect consumers from exposure to unacceptable levels of pesticide residues in food and feed. Regulation (EC) No 396/2005 (EC, 2005) achieves the harmonisation of pesticide MRLs, while ensuring consumer protection throughout the EU. MRLs undergo a common EU assessment to make sure that all classes of consumers, including the most vulnerable, such as children, are protected. The decision-making is science-based and a consumer intake assessment is carried out by the European Food Safety Authority (EFSA) before concluding on the safety of an MRL.

A dietary risk assessment is therefore a prerequisite for MRL setting. A major difficulty stems from the fact the only the toxicological properties of the active substance are normally directly investigated through the range of toxicological studies required according to Directive 91/414/EEC (EC 1991), which sets out uniform principles for the evaluation and authorisation of plant protection products and the active substances they contain. The new guidance document on the definition of residue (OECD, 2009), however, requires the consideration of human relevance for risk assessment of all metabolites the consumer is exposed to both in plant and animal commodities, raw or processed.

Metabolites may be produced from plant metabolism, from microbial activity in soil, or from livestock metabolism after consumption of feeding stuffs containing residues. Degradates arise from physical and chemical processes (e.g. photolysis) or from processing before the consumption (e.g. cooking) of plant and animal commodities. The consumer is therefore exposed not only to the active substance in the applied pesticide formulation, but also to a wide range of chemical compounds resulting from metabolic and degradation processes. The number and amount of distinct metabolites, defining the residue pattern, may widely differ from pesticide to pesticide depending on many parameters.

One of the outcomes of the evaluation of an application for use of an active substance on a crop is the establishment of two residue definitions, one for monitoring and one for risk assessment. The underlying rationales for these two definitions are different (OECD, 2009). While the residue definition for monitoring has regulatory purposes for the enforcement of the Maximum Residue Levels (MRLs) and must meet analytical practicalities, the residue definition for risk assessment may be wider, as its purpose is to assess consumer safety and should therefore include metabolites and degradates of toxicological relevance.

The residue definition for risk assessment should be qualitatively and quantitatively representative of the actual toxicological burden. This means that the establishment of the residue definition for risk assessment requires not only a decision on which metabolites and degradates, due to their levels, may significantly contribute to toxicological effects, but also an assessment of the toxicological endpoints of interest and related reference values (e.g. Acute Reference Dose [ARfD] and Acceptable Daily Intake [ADI]). In practice, however, very limited information on the toxicological properties of metabolites and degradates is available in the majority of cases. From the mixture (active substance, its metabolites and degradates) to which the consumer is exposed, only the toxicological properties of the active substance and their mammalian metabolites (to the extent to which they are formed in laboratory animals) are directly investigated. Furthermore, since requests for further toxicological studies are restricted as far as possible to minimise the use of animals in toxicological testing, alternative (non-animal) assessment methods are therefore needed to support the evaluation of the toxicological profile of pesticide metabolites and degradates. The information derived from such methods should reinforce the expert judgement forming the basis of the appropriate residue definition for risk assessment.

Quantitative Structure-Activity Relationship (QSAR) analysis represents a promising alternative approach, for reasons of cost-effectiveness and efficiency. For this reason, and within the framework of a collaboration agreement between the European Food Safety Authority (EFSA) and the European Commission's Joint Research Centre (JRC) a project (referred to here as PESTISAR) was initiated to evaluate the applicability of QSAR analysis in the toxicity prediction of metabolites and degradates of pesticides for dietary risk assessment (JRC, 2010).

PESTISAR was one of three projects sponsored by EFSA during 2009-2010. One of the other projects, carried out by the UK Chemicals Regulations Directorate (CRD) addressed the possible use of Threshold of Toxicological Concern (TTC) considerations in assessing metabolite/degradate toxicity (CRD, 2009), while other, carried out by the Austrian Agency for Health and Food Safety (AGES) examined the impact of metabolism and degradation on pesticide toxicity (AGES, 2010). Upon the completion of these projects, EFSA intends to pool and use the results to develop and adopt an opinion on the scientific principles for evaluating the toxicological burden related to metabolites, degradation and reaction products of active substances in food commodities. Upon adoption of this opinion, EFSA intends to develop a guidance document on the establishment of the residue definition for risk assessment in food commodities. It is foreseen that this guidance will be a practical tool to help risk assessors and regulatory authorities to adopt residue definitions based on objective criteria and weight of evidence. It could also be used for identifying cases where further information is needed.

### 1.2 Overview of work performed in the PESTISAR project

The overall aim of the PESTISAR project was to evaluate the potential applicability of computational methods in the evaluation of the toxicological relevance of metabolites and

degradates of pesticide active substances. Among the various types of computational estimation methods, emphasis was placed on Quantitative Structure-Activity Relationships (QSARs), Structure-Activity Relationships (SARs) and expert systems. The results of the PESTISAR project, completed in April 2010 (JRC, 2010), were updated throughout 2010 and published as a series of JRC Technical Reports.

The PESTISAR project involved the following activities:

- the development of a conceptual framework for assessing the usefulness of QSAR models and expert systems, in terms of their practical applicability and the adequacy of their predictions. This framework is summarised in Section 1.4 and described in detail in Worth *et al*. (2011);
- a survey of how QSAR analysis is used by national regulatory bodies and international advisory organisations in the field of food safety. This survey is summarised in Chapter 2;
- an extensive review of SARs, QSARs and expert systems for toxicological endpoints, potentially useful in dietary risk assessment. This review covered: acute and repeat-dose toxicity, including organ and system-specific toxicities (summarised in Chapter 4; described in detail in Lapenna *et al.*, 2010); genotoxicity and carcinogenicity (summarised in Chapter 5; described in detail in Serafimova *et al.*, 2010); developmental/reproductive toxicity; including endocrine-related effects (summarised in Chapter 6; described in detail in Lo Piparo *et al.*, 2010),
- an extensive review of computational models for biokinetic properties, including oral bioavailability, human intestinal absorption, blood-brain barrier penetration, plasma protein binding, metabolism, and clearance. This review is summarised in Chapter 7 and described in detail in Mostrag-Szlichtyng & Worth (2010).
- case studies (research investigations) into the potential use of QSARs for genotoxicity and carcinogenicity, with a view to developing a conceptual framework for QSAR analysis that can be integrated with the application of the TTC concept. The results of this investigation are described in detail in Worth *et al*. (2010);
- identification of research and development needs, leading to recommendations for further activities aimed at promoting the uptake and regulatory acceptance of computational methods in the food safety area. This is included in Chapter 8.

This report is an update and summary of the main findings and conclusions from the international survey and from the literature reviews.

### 1.3 Introduction to computational prediction methods

Computational prediction methods, sometimes referred to as “non-testing methods”, are based on the premise that the properties (including biological activities) of the chemical depend on its intrinsic nature and can be directly predicted from its molecular structure and inferred from the properties of similar compounds whose activities are known. These methods comprise Quantitative Structure Activity Relationship (QSAR) models as well as the less formalised approach of chemical grouping and read-across.

### 1.3.1 Quantitative Structure Activity Relationships

The term "QSAR analysis" is taken to include the development and use of Structure-Activity Relationships (SARs), Quantitative Structure Activity Relationships (QSARs), and computer-based tools (including expert systems) based on the use of one or more of these types of models.

Structure-Activity Relationships (SARs) and Quantitative Structure Activity Relationships (QSARs), collectively referred to as (Q)SARs, are theoretical models that relate the structure of chemicals to their biologic activities. (Q)SARs are used to predict the physicochemical, biological (e.g., toxicological) and fate properties of molecules from knowledge of chemical structure (Cronin, 2010).

More specifically, a SAR is a qualitative relationship between a molecular (sub)structure and the presence or absence of a given biological activity, or the capacity to modulate a biological activity imparted by another substructure. The term substructure refers to an atom, or group of adjacently connected atoms, in a molecule. A substructure associated with the presence of a biological activity is also called a structural alert. A SAR can also be based on the ensemble of steric and electronic features considered necessary to ensure the intermolecular interaction with a specific biological target molecule, which results in the manifestation of a specific biological effect. In this case, the SAR is sometimes called a 3D SAR or pharmacophore.

A QSAR is a quantitative relationship between a biological activity (e.g., toxicity), which may be categorical or quantitative, and one or more molecular descriptors that are used to predict the activity. A molecular descriptor is a structural or physicochemical property of a molecule, or part of a molecule, which specifies a particular characteristic of the molecule and is used as an independent variable in a QSAR. A comprehensive review of molecular descriptors has been published by Todeschini (Todeschini & Consonni, 2000, 2009).

### 1.3.2 Chemical grouping and read-across

In addition to the formalised approach of QSAR analysis, it is possible to estimate chemical properties and endpoints by using a less formalised approach based on the grouping and comparison of chemicals. The grouping approach can be used, for example, to support the results of QSAR analysis or to generate estimated data (and fill data gaps) in the absence of suitable QSARs. The most comprehensive guidance currently available for applying the grouping approach has been published by the OECD (OECD, 2007) and by ECHA (ECHA, 2008). The ECHA and OECD guidance documents are scientifically equivalent, except that the ECHA guidance makes additional references to REACH criteria and procedures. The concepts of grouping and read-across are further explained and illustrated by Enoch (2010).

The use of endpoint information for one chemical, called a "source chemical", to make a prediction of the same endpoint for another chemical, called a "target chemical", is termed "read-across". The source and target chemicals are considered to be similar in some way, usually on the basis of structural similarity. It is assumed that, in general, similar compounds will exhibit similar biological activity. In principle, read-across can be applied to characterise physicochemical properties, fate, human health effects and ecotoxicity, and it may be performed in a qualitative or quantitative manner, depending on the whether the data being used is categorical or numerical in nature. To estimate the properties of a given substance, read-across can be performed in a one-to-one manner (one analogue used to make the estimate) or in a many-to-one manner (two or more analogues used).

The reliability of read-across depends on the selection of appropriate analogues associated with the availability of reliable experimental data. In some cases, it is only possible to identify a limited number of suitable analogues, whereas in other cases, it is possible to build up a larger and more robust chemical group, called a chemical category. A chemical category is a group of chemicals whose physicochemical and human health and/or environmental toxicological and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristic). The presence of common behaviour or coherent trends in the chemical category is generally associated with a common underlying mechanism of action. In general, the application of read-across between analogues in a chemical category is considered to be more reliable than the application of read-across in a smaller group of analogues (in which trends are not apparent).

### 1.3.3 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 (Dearden *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 & Lai., 2005), which is freely downloadable from the US EPA website: (http://www.epa.gov/oppt/sf/pubs/oncologic.htm), Derek (Sanderson & Earnshaw, 1991; Ridings *et al*., 1996), developed by Lhasa Ltd (https://www.lhasalimited.org/), and HazardExpert (Smithing & Darvas 1992) developed by CompuDrug (http://compudrug.com/). Derek and HazardExpert can be used in conjunction with their sister programs Meteor and Metabolexpert to predict the toxicity 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 QSAR Toolbox are rule-based. Toxtree can be downloaded from the JRC (http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE) and from Sourceforge (https://sourceforge.net/projects/toxtree/). The QSAR Toolbox and guidance on its use are freely downloadable (http://www.qsartoolbox.org/).

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 http://accelrys.com/), and the publicly available Lazar (http://lazar.in-silico.de/) and CAESAR (http://www.caesar-project.eu) 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 structural alert (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*., 2007) as well as some literature-based models not implemented in software.

The advantages and disadvantages of the three main approaches are summarised in Table 1.1.

**Table 1.1 Comparison of three main approaches in expert systems**

| Approach | Advantages | Disadvantages |
|---|---|---|
| | | |
| Rule-based | • mechanistically connected to the predicted endpoint<br>• provide reasoning for the predictions<br>• in many cases support the prediction with literature references or expert knowledge | • often restricted and/or ill-defined applicability domain<br>• usually cannot explain differences of the activity within a chemical class<br>• usually have lower accuracy of the prediction than statistical models |
| Statistical | • usually have high accuracy of the predictions<br>• can be use for preliminary research when mechanism of action is unknown | • usually difficult to interpret the model predictions<br>• often do not provide mechanistically reasoning of the predictions<br>• 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 |

## 1.4 The adequacy of data generated by QSARs

The most comprehensive guidance currently available for applying QSAR analysis is provided in the REACH guidance on Information Requirements and Chemical Safety Assessment (ECHA, 2008). This guidance provides a flexible framework that can be extended, possibly with some specific adaptations, for use in the implementation of food safety legislation.

According to the framework developed for REACH, it is possible to use data from (Q)SAR models instead of experimental data if each of four main conditions is fulfilled:

- the model used is shown to be scientifically valid;
- the model used is applicable to the chemical of interest;
- the prediction (result) is relevant for the regulatory purpose; and
- appropriate documentation on the method and result is given.

Thus, multiple, overlapping conditions must be fulfilled to use a (Q)SAR prediction instead of data generated by a standard experimental test, as illustrated in Figure 1.1. The extent to which these conditions can be relaxed for indirect and supporting use of (Q)SAR data, remains to be established on the basis of experience.

The need to provide "appropriate documentation" is fulfilled by the provision of QSAR reporting formats for models and their predictions. The former type of documentation is the QSAR Model Reporting Format (QMRF) and the latter is the QSAR Prediction Reporting Format (QPRF). Information on (Q)SAR model validity, including peer-reviewed documentation, is available from various sources, including the JRC QSAR Model Database (http://qsardb.jrc.it).

The considerations necessary for demonstrating model validity, applicability and adequacy are described in detail elsewhere (Worth *et al.*, 2011). The latter report also proposes and illustrates a framework for assessing the usefulness of QSAR predictions, with reference to selected pesticides as case studies.

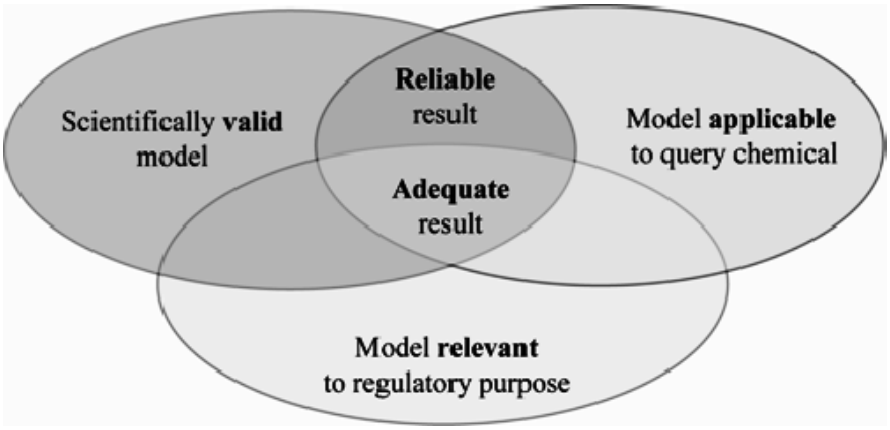



**Figure 1.1 The overlapping considerations of validity, applicability and relevance needed to demonstrate (Q)SAR adequacy**

# 2. USE OF COMPUTATIONAL METHODS ANALYSIS IN THE FIELD OF FOOD SAFETY

## 2.1 Introduction

This chapter gives an overview of how computational methods are currently used in the field of food safety by national regulatory bodies, international advisory organisations and the food industry. The results of an international survey show that currently the majority of stakeholders in the field of food safety do not apply computational methods on a routine basis, mainly because of a lack of in-house expertise. Some organisations, however, are very experienced in their use and have developed specialised in-house approaches. Despite this variable situation, computational tools are widely perceived to be a useful tool to support regulatory assessments and decision making in the field of food safety. However, there is a widespread need to develop guidance documents and software tools that will promote and harmonise the use of computational methods, together with appropriate training.

## 2.2 Method

To gain an overview of how computational methods are used internationally in the assessment of chemicals in food, a survey was carried out by EFSA and the JRC. A short and easy-to-complete questionnaire was prepared aiming at capturing major points such as which endpoints are predicted, which software and methodologies are used and how and when QSAR analysis is applied in the daily work of regulatory bodies and industry organisations.

In order to obtain as much information as possible, the questionnaire defined QSAR analysis in the broadest sense as “the use of qualitative or quantitative structure-activity relationships, chemical grouping and read-across, expert systems, or any other structure-based assessment approach.” Thus, the questionnaire was aimed at soliciting information not only on the use of (Q)SARs but also on the application of grouping and read across approaches. In addition, to obtain an overview of how (Q)SAR is perceived in the field of food safety, the questionnaire also gave room for additional comments on the practical application of (Q)SAR, and on research needs and barriers.

The questionnaire was circulated by EFSA to its European Focal Points in the Member States, and by the JRC to other organisations such as US EPA, US FDA, JECFA, Environment Canada, Health Canada and some experts in food consultancies and the food industry. Recipients of the questionnaire were encouraged to forward it to interested parties and to return the integration information, so it is not known how many recipients there were in total. The questionnaire was not circulated to Japan or other Asian countries.

To supplement the responses obtained in the survey, additional information on the use of computational methods in industry and regulatory authorities was identified from published papers (Table 2.1).

## 2.3 Main results

A total of 38 replies were received from respondents in the following countries: the Netherlands, Belgium, Bulgaria, Czech Republic, Estonia, Finland, Germany, Hungary, Latvia, Lithuania, Malta, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, USA, UK, Ireland and France. From some countries, multiple answers from were received from different institutions. A detailed description of the results is provided in the final report of the PESTISAR project (JRC, 2010).

The findings concerning software, methodology and endpoints most often used are summarised in Table 2.2, while the general results are listed as follows:

- 60% of the organisations do not use (Q)SAR analysis or other structure-based analysis for the purpose of dietary risk assessment, but 37% do (the remaining 3% didn't answer to the question).
- Of these 60% organisations that do not use (Q)SAR, 60% of them never tried to and 40% considered using it but rejected it mainly for lack of expertise.
- Only few organisations apply (Q)SAR to give insight into the mode of action and in these cases, the analysis seems to be based primarily on the use of specific structural alerts.
- Broad interest for the field was indicated by the fact that all respondents, except two, requested to receive the outcome of the project.

Concerning the use of (Q)SAR analysis, the answers received can be summarised as follows:

- (Q)SAR is used when a fast decision is necessary regarding the safety of a chemical for which no toxicity data is available, or for compounds present at very low levels.
- Read-across (SAR) is used to compare the toxicity of a compound with those of compounds with similar structures for which toxicological data exist. If the mode of action of an active substance is known, the mode of action of a related structure is supposed to be the same.
- A positive (Q)SAR result can be accepted in the absence of study results. However, a negative result is more likely to be accepted if multiple (Q)SAR models indicate that there is no alert on the molecule for toxic potential. Consensus modelling is sometimes applied based on the assumption that the predictive performance is improved when predictions from multiple types of models are combined into an overall prediction for an endpoint.
- A weight-of-evidence approach is applied to determine whether a certain metabolite should be included in the residue definition based on a toxicity evaluation. Known structural alerts for toxicity may play a role in this weight-of-evidence approach, and they may also lead to additional data requirements.

Concerning research needs and barriers to the widespread acceptance of (Q)SAR analysis, the following important observations were highlighted:

- For some endpoints (e. g. carcinogenicity and mutagenicity) too many different models have been developed and published that a user can become confused. A few relevant models for each endpoint should be identified and then guidance with a strategy on their use should be written.
- For other endpoints (e. g. developmental toxicity) there is a lack of relevant valid (Q)SAR models which, in turn, is caused by the limited availability of high-quality toxicological data for the building of models.
- One of the biggest research needs is to digitalise the data available and make them available in a high-quality public domain database.
- The application and acceptability of (Q)SAR in risk assessment should be made transparent with respect to how governmental organisations recognise and accept its

use by industry. The establishment of specific application guidance would be valuable. The selected models should preferably be freely and publicly accessible.

Other relevant points arising from the questionnaire review include:

- The need for quantitative as well as qualitative predictions. In the context of food safety, the most likely application of computational toxicology models would be in the establishment of the level of safety concern associated with the inadvertent/accidental presence of a contaminant in a food product. This requires not only qualitative information on the potential hazard (e.g. carcinogenicity) but also quantitative information (e.g. carcinogenic potency), allowing the derivation of a margin of exposure (MoE) with the estimated intake.
- Global chemical diversity. Compounds found in food and food ingredients present a wide structural diversity and complexities that may be greater than synthetic pharmaceuticals targeted for a particular purpose, and, therefore, require the development of global *in silico* models (rather than local, referring to particular classes of chemical structure).
- High reliability, relevance and transparency. Ideally, *in silico* toxicology strategies for food safety assessment should be able to predict adverse health effects in the human population. Currently, their practical application in the food sector will depend upon their potential to accurately predict biological endpoints/hazards that are used in food chemical risk assessment. This includes the need to establish confidence limits. The acceptance of these models will be possible only if the analysis is fully transparent. Therefore, the promotion of validated, freely available tools based on open-source codes is necessary and warranted. Results should be clear, concise and reproducible.

Some examples of how computational tools are used in regulatory bodies (US EPA, US FDA) and the food industry (Nestlé) are described in detail in the final report of the PESTISAR project (JRC, 2010).

### 2.4 Summary and conclusions

On the basis of the survey, it was found that the majority of key players in the food safety field either do not use (Q)SAR methodology at all or in a very limited way mainly because of a lack of expertise. When (Q)SARs are used, they are typically applied to support priority setting exercises or to fill information gaps on possible health concerns during the management of a food crisis in food industry (e.g. if a contaminant is found in food). At present, (Q)SAR is not used routinely to fill data gaps in the pre-marketing assessment of food additives, food contact substances, or pesticide and pesticide metabolite residues. However QSARs are currently being explored, developed and utilised by regulatory authorities for risk assessment purposes. Some organisations are however very experienced in the use of QSAR, notably government authorities such as the US FDA Center for Food Safety and Applied Nutrition (CFSAN) and the US EPA (OCSPP), as well as some companies (e.g. Nestlé). Other organisations lack the capacity, or use QSAR tools in a restricted manner to provide supplementary information (e.g. information on analogues).

Despite this variable situation, (Q)SAR analysis is widely perceived as a potential useful tool to support regulatory assessments in the field of food safety, and this justifies further

exploration and development. In particular to promote the further use of the methods, there is a need to perform focused substance evaluation studies that better explore the potential of QSAR analysis for application in specific situations, and there is a need to develop guidance documents and tools that will promote the harmonised use of (Q)SAR analysis in the different sectors. In addition, there is a widespread demand for training on the applications of (Q)SAR analysis in human dietary risk assessment.

## Table 2.1. Papers describing how QSARs have been developed and used by government authorities and industry

| Reference | Organisation | Key points |
|---|---|---|
| | | |
| Yang *et al*. (2009) | US Food and Drug Administration | • The Office of Food Additive Safety (OFAS) has a SAR group supporting regulatory decisions on the safety of food additives.<br>• Several 2D QSAR models have been developed and made available through commercial software via agency-approved Cooperative Research and Development Agreements (CRADA). These models have not yet been assessed by external validation.<br>• A project has been initiated to capture Agency preclinical toxicity data records in a structurally-searchable database. A data entry tool has been designed so that the toxicologists can record the data directly during the review process.<br>• Current research is investigating whether TTC values can be developed for chemical structural categories beyond the Cramer categories. The knowledgebase consists of databases, alerts and rules for modes of action.<br>• There is no regulatory guidance specific to computational toxicology methods. Many questions remain, including which models should be used, under what circumstances and how conclusions might be made from predictions generated by multiple software tools. |
| Rothenbacher *et al*. (2009) | Chemisches und Veterinäruntersuchungsamt | • Describes the use of Derek to evaluate plastic packaging materials in terms of carcinogenicity, genotoxicity, thyroid toxicity, and miscellaneous endpoints relevant to human health |
| AFSCA (2009) | Federal Agency for the Safety of the Food Chain (Belgium) | • The SAR methodology is used to compare the toxicity of a compound with those of compounds with similar structures for which toxicological data exist. |
| Jensen *et al*. (2008) | Danish National Food Institute | • Describes the use of (Q)SARs for classification and labelling using 57014 chemicals from the European Inventory of Existing Chemical Substances (EINECS) and in-house and commercial models (mainly MultiCASE), in order to identify possible reprotoxicants. |
| Mazzatorta *et al*. (2008) | Nestlé Research Center | • A multivariate chronic toxicity (LOAEL) model, using 2D descriptors, was built from a dataset of 445 different chemicals.<br>• The model reveals that the chronic toxicity effects are driven by the bioavailability of the compound that constitutes a baseline effect plus excess toxicity described by a few chemical moieties.<br>• The model predicts LOAEL with an error of 0.70. Since this error approaches the experimental error (0.64), it was concluded that the model may be used together with exposure to establish a level of safety concern of chemicals in food for which hard toxicological data are missing. |
| Maunz & Helma (2008) | Nestlé Research Center | • Describes local support vector regression models developed in-house for the prediction of Fathead Minnow Acute Toxicity (573 compounds), Maximum Recommended Therapeutic Dose (based on clinical trial data for 1215 pharmaceutical compounds). |

| Reference | Organisation | Key points |
|---|---|---|
| | | |
| Matthews *et al*. (2008) | US Food and Drug Administration | • Describes the use of four QSAR programs and an expert knowledge base system to predict the occurrence and the mode of action of carcinogenesis in rodents (weight-of-evidence).<br>• The four QSAR programs were complementary, each detecting different profiles of carcinogens. Accepting any positive prediction from two programs showed better overall performance than either of the single programs alone. |
| Contrera *et al*. (2007) | US Food and Drug Administration | • Comparison of the rodent carcinogenicity predictive performance of MC4PC and MDL-QSAR software as well as a method for combining the predictions from both programs using 1540 training set compounds.<br>• Merging MC4PC and MDL-QSAR predictions improved the overall predictive performance.<br>• Consensus rules can be tuned to reflect the priorities of the user, so that greater emphasis may be placed on predictions with high sensitivity/low false negative rates or high specificity/low false positive rates. |
| Matthews *et al*. (2007a,b) | US Food and Drug Administration | • Describes a battery of QSAR models, running in the MC4PC software, to predict reproductive and developmental (reprotox) hazards of untested chemicals.<br>• The QSARs are based on 627–2023 chemicals. |
| Valerio *et al*. (2007) | US Food and Drug Administration | • Evaluates several QSAR models for decision support in the assessment of carcinogenicity, mutagenicity and reproductive toxicity.<br>• Concludes that the *in silico* QSAR analysis is capable of identifying the rodent carcinogenic potential of naturally occurring organic molecules found in the human diet with a high degree of sensitivity. |
| Kruhlak *et al*. (2007) | US Food and Drug Administration | • Discusses some of the considerations when using computational toxicology methods for regulatory decision support of pharmaceutical impurities and degradents and gives examples of how the technology is being applied by the US FDA. |
| Tilaoui *et al*. (2007) | Nestlé Research Center | • The prediction of chronic toxicity (LOAEL) is perormed in-house by using an integrated system partly based on TOPKAT<br>• The system is used to support the prioritisation of issues in chemical food research, by establishing levels of safety concern in the absence of sufficient experimental toxicological data. |
| Mazzatorta *et al*. (2007) | Nestlé Research Center | • Describes the developmentment of a hybrid system for the prediction of Ames test mutagenicity based on a combination of a fragment-based SAR models and artificial intelligence systems. It was developed using a training set of 4337 chemicals (2401 mutagens and 1936 non-mutagens) and tested using 753 compounds (437 mutagens and 316 non-mutagens).<br>• The overall error of this system on the external test set compounds is 15%, which is quantitatively similar to the experimental error of Ames test data (average interlaboratory reproducibility determined by the National Toxicology Program). On this basis, it was concluded that the system can be applied to support early and rapid evaluation of the level of mutagenicity concern. |
| Bailey *et al*. (2005) | US Food and Drug Administration | • Describes the food contact notification (FCN) program by which the FDA reviews food contact substances (FCS) for safe use, the SAR tools available to FDA, and their use in qualitative and quantitative risk assessments of FCS |

| Reference | Organisation | Key points |
|---|---|---|
| | | |
| Tong *et al*. (2004) | US Food and Drug Administration | • Describes a stepwise "four-phase" scheme for identifying oestrogenic substances. Within each step (phase), different models were selected to work in a complementary fashion in order to minimise the rate of false negatives. The system works in a hierarchical manner to reduce the size of a dataset incrementally while increasing the accuracy of prediction. |
| Woo *et al*. (2002) | U.S. Environmental Protection Agency | • Disinfection by-products (DBPs) are formed when disinfectants react with organic and inorganic matter in water. The observations that some DBPs are carcinogenic in animal studies have raised public concern over the possible adverse health effects of DBPs. To prioritize research efforts, mechanism-based structure-activity relationship analysis was conducted to rank the carcinogenic potential of DBPs. |
| Matthews & Contrera (1998) | US Food and Drug Administration | • Describes a weight of evidence scoring method for predicting the carcinogenic potential of pharmaceuticals in rodents using MCASE QSAR-ES (Expert System) software |
| Woo *et al*. (1995) | U.S. Environmental Protection Agency | • Describes how SAR analysis has been used by the US EPA in the assessment of potential carcinogenic hazard of new chemicals for which test data are not available.<br>• Describes the major factors and rules used in Oncologic for assessing the carcinogenic potential of fibers, polymers, metals/metalloids and several major classes of organic chemicals. |

## Table 2.2 Commonly used software, methodology and main endpoints predicted

| | **Most often in all the organisations** | **by US EPA** | **by US FDA** | **by Nestlé** |
|---|---|---|---|---|
| **Software used** | Derek, MultiCase, Leadscope, TOPKAT, EPI Suite, OECD Toolbox, Toxtree. | EPI Suite, OncoLogic, Derek | MultiCase, MDL QSAR, Derek, and Leadscope. | TOPKAT, MULTICASE, Lazar, Toxtree, in-house models published and model developed under contract |
| **Methodology used** | Read across and chemical grouping | Read across and chemical grouping | Structural alerts and predictions from QSAR models | QSAR models |
| **Endpoint prediction** | Genotoxicity, carcinogenicity and chronic toxicity followed by ADME, acute toxicity and reproductive toxicity. | Mostly carcinogenicity and genotoxicity; acute and chronic non-mammalian organism toxicity; physicochemical characteristics | Mostly carcinogenicity and genotoxicity | Chronic toxicity, carcinogenicity, and genotoxicity |

# 3. SOFTWARE TOOLS FOR TOXICITY PREDICTION

## 3.1 Introduction

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

In this chapter, we provide an overview of the software packages that are commonly used in the assessment of toxicity. More-detailed information is given in Chapters 4-6 for specific (groups of endpoints). The availability of software and literature models for the prediction of biokinetic properties (including bioavailability and metabolic fate) is described in Chapter 7.

## 3.2 Software for predicting chemical toxicity

The following sections briefly describe software that is either in the public domain or commercially available. Some of the freely available software tools have been developed under the terms of open-source licensing, which means that other experts can further develop and disseminate the software. Websites for freely and commercially available tools are given in Tables 3.1 and 3.2, respectively, and their ability to predict properties and endpoints relevant to dietary toxicity assessment is highlighted in Table 3.3. A recent review is provided in Fuart Gatnik & Worth (2010). Many other reviews have also been published (Dearden *et al*., 1997; Greene *et al*., 1999; ECETOC, 2003).

### 3.2.1 Freely available software

A summary of freely available software is given in Table 3.1. The following paragraphs describe these tools in general terms.

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

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

**Lazar**: Lazar is an open-source software programme that makes predictions of toxicological endpoints (currently, mutagenicity, human liver toxicity, rodent and hamster carcinogenicity, MRDD) by analysing structural fragments in a training set (Helma, 2006; Maunz & Helma, 2008). It is based on the use of statistical algorithms for classification (k-nearest neighbours and kernel models) and regression (multi-linear regression and kernel models). In contrast to

traditional k-NN techniques, Lazar treats chemical similarities not in absolute values, but as toxicity dependent values, thereby capturing only those fragments that are relevant for the toxic endpoint under investigation. Lazar performs automatic applicability domain estimation and provides a confidence index for each prediction, and is usable without expert knowledge. Lazar runs under Linux and a web-based prototype is also freely accessible:. http://lazar.in-silico.de/

**OECD QSAR Application Toolbox**: The OECD QSAR Application Toolbox is a standalone software application for gaps in (eco)toxicity data needed for assessing the hazards of chemicals. Data gaps are filled by following a flexible workflow in which chemical categories are built and missing data are estimated by read-across or by applying local QSARs (trends within the category). The Toolbox also includes a range of profilers to quickly evaluate chemicals for common mechanisms or modes of action. In order to support read-across and trend analysis, the Toolbox contains numerous databases with results from experimental studies. The first version of the Toolbox, released in March 2008, was a proof-of-concept version. The first update (version 1.1) was released in December 2008, and the second (version 2.0) in October 2010. The release of version 3.0 is planned for October 2012. The OECD Toolbox is freely available: http://www.qsartoolbox.org/

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

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

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

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

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

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

### 3.2.2 Commercially available software

A summary of commercially available software is given in Table 3.2. The following paragraphs describe these tools in general terms.

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

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

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

**Derek**: This SAR-based system is developed by Lhasa Ltd, a non-profit company and educational charity (https://www.lhasalimited.org/). DfW contains over 50 alerts covering a wide range of toxicological endpoints in humans, other mammals and bacteria. An alert consists of a toxicophore (a substructure known or thought to be responsible for the toxicity)

and is associated with literature references, comments and examples. A key feature of DfW is the transparent reporting of the reasoning underlying each prediction.

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

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

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

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

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

QSAR within the reduced set of congeneric molecules. The modulators consist of certain substructures or physicochemical parameters that significantly enhance or diminish the activity attributable to the biophore. QSARs are then derived by incorporating the biophores and the modulators into the model. The program includes modules to predict physicochemical properties and a range of toxicological endpoints, including carcinogenicity, mutagenicity, teratogenicity, irritation, developmental toxicity, and acute toxicity. For the endpoints, the software uses it own toxicity scale, from 0 to 100 CASE units, to cover the range from inactive, marginally active and active. In many cases, it is difficult to relate these CASE units to traditional measures of toxicity.

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

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

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

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

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

**Table 3.1 Some commonly used freely available software tools**

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

**Table 3.2 Some commonly used commercial software tools**

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

**Table 3.3 Software capable of predicting toxicological endpoints relevant to dietary risk assessment**

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

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

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

# 4. PREDICTION OF ACUTE AND SYSTEMIC TOXICTY

Systemic toxicity studies aim at investigating the effects of chemicals in laboratory animals exposed to various dosage regiments for different durations. Exposure is generally through the oral, dermal or inhalation routes. The information from systems toxicity studies is used in hazard and risk assessment of chemicals occurring in food, industrial chemicals, biocides, and cosmetics. In this chapter, we give an overview of the software packages used in the assessment of acute systemic toxicity, chronic systemic toxicity and organ- and system-specific toxicity, as well as the databases available for obtaining such data. Reviews on QSAR studies published in the literature are available elsewhere (Devillers & Devillers, 2009; Lapenna *et al*., 2010; Tsakovska *et al*., 2008).

## 4.1 Acute systemic toxicity

Acute toxicity describes the adverse effects caused by either a single exposure to a chemical substance or multiple exposures within 24 hours.. The acute lethal dose to 50% of the treated animals ($LD_{50}$ value) is the basis for the hazard assessment and classification of chemicals and is widely used for regulatory purposes. However, the $LD_{50}$ value presents some drawbacks when used for QSAR modelling. First, acute toxicity effects may result from a wide spectrum of biokinetic, cellular and molecular events. Converting the complex, whole-body phenomena related to acute toxicity into a simple number necessarily leads to a loss of information. Second, available data are highly variable, having been generated by different laboratories, protocols, animal species and strains. This undermines the reliability and repeatability of acute toxicity measurements. These facts complicate the modelling process and may explain why there are relatively few (Q)SAR models and expert systems for predicting oral acute toxicity, in comparison with other endpoints.

### 4.1.1 Software for predicting acute systemic toxicity

Software tools capable of predicting endpoints related to systemic toxicity are listed in Table 4.1.

The commercial software ACD/Tox Suite (now developed and marketed by Advanced Chemistry Development [ACD/Labs] and formerly by Pharma Algorithms as ToxBoxes) predicts toxicity in both the mouse and rat for various administration routes, including oral, as either quantitative $LD_{50}$ values or classification into the five GHS categories.

The statistically-based programs TOPKAT and MCASE use multiple QSARs on small and homogenous sets of data. The rat oral $LD_{50}$ module in TOPKAT comprises 19 regression analyses developed using experimental values of approx. 4000 chemicals from RTECS, including pesticides and industrial chemicals. The rat oral $LD_{50}$ module in MCASE (named A56) is based on and comprises data for 7920 chemicals from the FDA, WHO and NTP datasets. Tunkel and coworkers (Tunkel *et al*., 2005) compared the performance of the TOPKAT and MCASE rat $LD_{50}$ modules against an external test set of 73 organic compounds covering 32 chemical categories retrieved from submissions to the EPA High Production Volume (HPV) Challenge Program (http://www.epa.gov/chemrtk/). The predictive accuracy of each software tool was assessed by applying the EPA's New Chemical classification approach (http://www.epa.gov/oppt/newchems/index.htm), from the low-concern class (>2000 mg/kg) to the high-concern class (<15 mg/kg). While neither model was able to classify all 73 compounds, TOPKAT correctly classified 67% of the chemicals, while

MCASE classified 70% correctly. However, it should be noted that the test set used was significantly skewed toward "low concern" chemicals, which both models predicted correctly with a high degree of accuracy (82% and 100% correct for TOPKAT and MCASE, respectively). Moreover, a high degree of false negatives was found for moderate and high concern HPV chemicals (TOPKAT, 72%; MCASE, 100%), suggesting that these programs are less reliable for the identification of more toxic compounds. The authors also compared the model outputs against the GHS five-tier scheme for classification of rat oral acute toxicants (<5, 5-50, 50-300, 300-2000, and 2000-5000 mg/kg), which is similar to the one adopted by EPA (<15, 15-50, 50-500, 500-2000, >2000 mg/kg). When compared against the GHS scheme, the ability of TOPKAT and MCASE to produce correct classifications was 73% and 70%, respectively, for the HPV test set chemicals, thereby changing slightly with respect to the EPA scheme, albeit enough to invert the rank order of these models. Overall, these results support the usefulness of the TOPKAT and MCASE tools when used for hazard classification.

Other software tools available for predicting acute toxicity ($LD_{50}$) to rat/mouse, are also available, such as MDL QSAR and TerraQSAR. The TerraQSAR software, based on neural network methodology, includes models for predicting both oral and intravenous LD50 values in mice and rats (http://www.terrabase-inc.com/).

### 4.1.2 Databases containing information on acute systemic toxicity

Sources of rat $LD_{50}$ values which may be suitable for the development of QSARs, the application of read-across, and the evaluation of high-throughput *in vitro* methods, are listed in Table 4.2. In particular, Acutoxbase has been developed in the context of the EU FP6 project 'A-Cute-Tox' (http://www.acutetox.org), which aims to optimise and "pre-validate" an *in vitro* testing strategy for predicting acute human toxicity. At present, Acutoxbase is not publicly accessible. However, parts of the data have been published in the literature (Kinsner-Ovaskainen *et al*., 2009).

In order to be useful for QSAR development, datasets should be first curated, i.e. the accuracy of the structures should be verified and the quality of biological data should be reviewed. In addition, inorganic and organometallic compounds, salts, and compound mixtures are often removed from the analysis. For the development of QSARs, $LD_{50}$ values should be converted to log[1/(mol/kg)] (if originally expressed as mol/kg). Finally, approximate $LD_{50}$ values should be converted to discrete values, and multiple $LD_{50}$ values from different labs/experiments should be converted to a single value. The ChemIDplus and ZEBET databases have been recently employed as data sources for QSAR analyses (Zhu *et al*., 2009a,b).

## Table 4.1. Software tools for systemic toxicity endpoints

| SOFTWARE (AND DEVELOPER) | AVAILABILITY | ENDPOINT | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Acute (oral) toxicity | Chronic (oral) toxicity | Hepatotoxicity | Nephrotoxicity (+ urinary tract toxicity) | Neurotoxicity | Cytotoxicity | Immunotoxicity (1) |
| ACD/Tox Suite (ToxBoxes) | Commercial | ● | | | | | | |
| ADMET Predictor (Simulations Plus Inc.) | Commercial | | | ● | | | | |
| ADME/Tox WEB | Freely available | ● | | | | | | |
| BioEpisteme | Commercial | | | ● | ● | | | |
| CAESAR models (Mario Negri Institute) | Freely available | | | | | | | |
| Derek (Lhasa Ltd) | Commercial | | | ● | ● | ● | | ● |
| HazardExpert (CompuDrug) | Commercial | | | | | ● | | ● |
| Lazar (*In silico* Toxicology; Freiburg university) | Freely available | | ● | ● | | | | |
| Leadscope (Leadscope) | Commercial | | | ● | ● | ● | | |
| MCASE/MC4PC (MultiCASE) | Commercial | ● | | ● | ● | | ● | |
| MDL QSAR (Symyx) | Commercial | ● | | ● | ● | | | |
| Molcode Toolbox ( Molcode Ltd) | Commercial | | ● | | | | ● | |
| OASIS-TIMES (Laboratory of Mathematical Chemistry, Bourgas University) | Commercial | | | | | | | |
| OncoLogic (US EPA) | Freely available | | | | | | | |
| Pallas Suite including ToxAlert, Cytotoxicity (CompuDrug) | Commercial | | | | | ● | ● | |
| TerraQSAR (TerraBase) | Commercial | ● | | | | | | |
| TOPKAT (Accelrys) | Commercial | ● | ● | | | | | |
| Toxtree (JRC) | Freely available | | | | | | | |

(1) immunotoxicity other than skin sensitisaton.

## Table 4.2 Databases containing acute toxicity information

| Database | Availability | Information |
|---|---|---|
| | | |
| Acutoxbase, linked to the EU FP6 project 'A-Cute-Tox';<br>https://acubase.amwaw.edu.pl | Access through the internet, currently restricted to project partners | The following data are available for 97 reference chemicals (i.e. 52% drugs, 31% industrial chemicals, 12% pesticides, 5% others):<br>*in vitro*: approx. 100 *in vitro* assays including general acute cytotoxicity, metabolism-mediated toxicity, biokinetics, and organ-specific toxicity.<br>*in vivo*: Over 2200 LD50 values in rodents (rat and mouse) and other animals (e.g. guinea pig, dog) with various administration routes (oral, intravenous, etc.) compiled from published literature.<br>For 86 reference chemicals, human acute poisoning cases from clinical/forensic reports are also available. |
| ChemIDplus, developed by the US NLM;<br>http://chem.sis.nlm.nih.gov/chemidplus/ | Freely available through the Internet, structure-searchable | Toxicity data for over 139,000 records, retrieved from TOXNET® (TOXicology Data NETwork; http://toxnet.nlm.nih.gov) which includes HSDB (Hazardous Substances Data Bank). The HSDB is an older subset of the RTECS database. A search for rat and mouse oral $LD_{50}$ values found 13,548 and 28,033 records, respectively. |
| CEBS, developed by the US NIEHS;<br>http://cebs.niehs.nih.gov/ | Freely available through the Internet | *In vivo* study data and acute dose of a small number of known hepatotoxicants to rat. |
| RTECS, originally compiled and maintained (until 2001) by the US NIOSH and currently maintained by Symyx Technologies. Structure-searchable through the Symyx Toxicity Database:<br>http://www.symyx.com/products/databases/bioactivity/rtecs/index.jsp<br>Also searchable via the Leadscope Toxicity Database (http://www.leadscope.com/databases/) | Commercial | Rat acute oral toxicity ($LD_{50}$) and acute inhalation toxicity ($LC_{50}$) data compiled from the open scientific literature for approx. 7,000 compounds (organic, inorganic and mixtures), including approx. 4000 organic compounds. |
| ZEBET, compiled by BfR ZEBET;<br>http://www.dimdi.de | Freely searchable through the DIMDI website. Published in a report by ICCVAM (ICCVAM *et al.*, 2001) | Includes rat or mouse $LD_{50}$ values (from the RTECS database) and cytotoxicity ($IC_{50}$) data for 347 compounds compiled from the open literature. |

Abbreviations: CEBS, Chemical Effects in Biological Systems; DIMDI, German Institute for Medical Documentation and Information; ICCVAM, Interagency Coordinating Committee on the Validation of Alternative Methods; RTECS, Registry of Toxic Effects of Chemical Substances; US NLM, US National Library of Medicine; US NIEHS, US National Institute of Environmental Health Sciences; US NIOSH, US National Institute of Occupational Safety and Health; BfR ZEBET, Centre for Documentation and Evaluation of Alternatives to Animal Experiments of the German Federal Institute for Risk Assessment.

### 4.1.3 Conclusions on the ability to predict acute systemic toxicity

Some currently available software tools (e.g. TOPKAT and MCASE) are useful for predicting acute toxicity in categorical terms (e.g. in terms of GHS classifications). However, these tools should be further investigated in relation to apparently high degree of false negatives generated, since this would be undesirable in the regulatory assessment of pesticides. The performance of other software tools in predicting acute toxicity should also be investigated. It is recommended that targeted studies are carried out to explore the usefulness of these software tools not only for classifying chemicals but also for making quantitative predictions of LD50 values for chemical inventories of regulatory importance (e.g. pesticides).

In the scientific literature, QSAR models have been generated for sets of congeneric compounds (organophosphates, aromatic amines, anilines, etc.) and are scattered over many original publications. Some of these studies have also explored the use of *in vitro* data as aadditional descriptors in the derivation of so-called quantitative structure activity-activity relationships (Lessigiarska *et al*., 2006). Despite their limited applicability when taken individually, these local models might be usefully combined into an expert system for toxicity predictions. Further research and development in this area is therefore encouraged. In addition, several recent research studies (Zhu *et al*, 2009a,b; Raevsky *et al*., 2009) have demonstrated the ability to make reasonable quantitative predictions for structurally diverse datasets, especially when high throughput bioactivity data are used in combination with traditional QSAR descriptors. These approaches should be explored further with a view to practical implementation. In this respect, the future availability of the models developed by Zhu and coworkers for use as $LD_{50}$ predictors via the EPA website and the ChemBench web portal (Zhu *et al*., 2009a,b) are promising initiatives.

## 4.2 Chronic systemic toxicity

Chronic (repeated dose) toxicity refers to the general toxicological effects in mammals occurring as a result of prolonged and repeated (oral, dermal or inhalation) exposure to a substance. The general toxicity includes a wide range of possible adverse effects including changes in morphology, physiology, growth, development or life span which result in impaired functional capacity, impaired capacity to compensate for additional stress or increased susceptibility to the harmful effects of other environmental influences.

The most commonly performed animal tests are the subacute (28-day) and subchronic (90-day) oral toxicity tests in rodents. Testing is sometimes performed with a longer testing period (12 months or more), and sometimes with in a non-rodent species (e.g. dogs, primates). The studies are used to identify adverse effects on various organs and tissues (e.g. liver, kidney, central nervous system, reproductive organs, immune system, and the endocrine system), and to establish a dose metric for risk assessment - the lowest dose that induces an adverse effect (Lowest Observed Adverse Effect Level; LOAEL) or the highest dose with no biologically or statistically significant adverse effects (No Observed Effect Level; NOEL). In this assessment, all toxicological responses are taken into account and the critical (most sensitive) effect is identified. The results of repeated-dose testing can also be used to classify chemicals on the basis of systemic toxicity. Within the Globally Harmonised Classification System (GHS) for chemicals, the results of repeated dose studies can be used, in a weight-of-evidence approach, to place systemic toxicants in two hazard categories.

Thus, chronic toxicity is not really a single endpoint, but a common term for a multitude of biological effects that have different mechanisms, occur in different tissues and organs and over different time scales. This presents a challenge for QSAR modelling, which should

ideally focus on groups of chemicals with a common mode of action. Perhaps for this reason, there have been few attempts to develop QSAR models for chronic toxicity in mammals.

### 4.2.1 Software for predicting repeated dose toxicity

Software tools capable of predicting repeated dose toxicity are given in Table 4.1. At present, the best known is probably TOPKAT, which predicts oral rat chronic LOAEL values. The model includes five regression-based models for five classes of chemicals (acyclics, alicyclics, heteroaromatics, single benzenes and multiple benzenes), developed on the basis of 393 chemicals from various sources (EPA and National Cancer Institute/National Toxicology Program (NCI/NTP) databases; FDA drug applications reports; and the open literature). The paper describing the original model development (Mumtaz *et al*., 1995), based on 234 structurally-diverse chemicals for which chronic data (12 months or more) were available from the above-mentioned sources, provides a transparent description of the model – it is mulitilinear regression QSAR based on 44 structural descriptors. In contrast, the algorithm for the updated TOPKAT model, based on five regression models and an extended dataset of 393 chemicals, has not been published.

In a model assessment study by Venkatapathy *et al*. (2004), the predictive performance of TOPKAT was tested against 343 chemicals from the EPA's Office of Pesticide Programs (OPP) database. After removal of compounds that TOPKAT could not recognise or which generated various types of warnings, the percentages of chemicals in TOPKAT's database that had a LOAEL predicted within a factor of 2, 5 and 10 of the experimental LOAEL were 65%, 83%, and 91%, respectively. When testing against chemicals not already in TOPKAT's database (i.e. an external validation), the corresponding percentages were 34%, 57% and 72%. Similar statistics were obtained when the TOPKAT predictions were compared against 313 chemical in the "IHP database", so-called because it was derived the Integrated Risk Information System (IRIS), Health Effects Assessment Summary Tables (HEAST), and Provisional Toxicity Value (PTV) databases. If prediction within a factor of 2 is taken as the criterion for "correct classification", this implies a misclassification rate of 35-66%; and if a factor of 10 is adopted, the corresponding misclassification rate would be 9-28%.

In another assessment, Tilaoui *et al*. (2007) investigated the ability of TOPKAT to predict the LOAELs of substances typically occurring in food, on the basis of 607 substances taken from Munro *et al*. (1996). After excluding the 267 substances in the TOPKAT training set, the number of validation substances was reduced to 340. Of those 340 molecules, 287 had predicted LOAELs with the model applicability domain (OPS), of which 86% were predicted within a factor of 2.

In addition to providing point estimates of chronic toxicity, the similarity search capacity of TOPKAT can be used to identify analogues in the TOPKAT database for use in read-across assessments. For example, in order to predict the LOAEL of dichlorobenzophenone (DCBP), which is a metabolite of chlorobenzilate, dichlorodiphenyltrichloroethane, and dicofol, Mougdal *et al*. (2003) identified 47 potential analogues in the TOPKAT database, of which five were selected on the basis that there were toxicity data in an EPA database (IRIS, HEAST or PTV). Among the five potential surrogates, chlorobenzilate was chosen as a surrogate for DCBP, since it had the most conservative chronic oral reference dose (RfD). The RfD is the US EPA's maximum acceptable oral dose of a toxic substance, obtained by dividing the NOEL or LOAEL by various uncertainty factors.

The other main software tool capable of predicting LOAELs, is a module of the recently developed MolCode Toolbox. A QMRF for this model is available in the JRC QSAR Model Database.

### 4.2.2 Databases containing information on repeated dose toxicity

There are two main databases suitable for the development and assessment of (Q)SARs for repeat-dose toxicity (Table 4.3). The RepDose database developed by the Fraunhofer Institute (Bitsch *et al*., 2006) contains NOELs and LOAELs for over 650 industrial chemicals, but is not made publicly available. A database of human Maximum Recommended Therapeutic Dose (MRTD) values has been compiled and made publicly available by the US FDA (Matthews *et al*., 2004b).

In addition to these databases, there are several datasets in the published literature. Munro *et al*. (1996) developed a database of 612 structurally well-defined organic chemicals, divided into the three structural Cramer classes (Cramer *et al*., 1978) and associated with 2944 (subchronic and chronic) NOELs derived from non-carcinogenic endpoints in oral rodent or rabbit studies. This database has provided the basis of the TTC concept. Oral NOELs for 45 consumer product ingredients (not in the Munro database) have been published by Blackburn *et al*. (2005).

### 4.2.3 Conclusions on the ability to predict repeated dose toxicity

The availability of (Q)SAR models for chronic toxicity endpoints is currently very limited. Since a large number of potential targets and mechanisms are associated with repeated dose effects, it is unlikely that any single model or software tool will be capable of making reliable predictions for all chemicals of interest to dietary risk assessment. The most commonly used software tool at present is TOPKAT, and despite the lack of transparency in its predictions, several studies have shown that it gives reasonable predictions for a range of chemicals (including pesticides, industrial chemicals). Another more recently developed tool is a module of MolCode Toolboxes. Predictions from such tools could be used in a weight-of-evidence approach along with additional data. Additional research investigations into the applicability of TOPKAT and MolCode Toolboxes across a wide range of food chemicals would be worthwhile. In addition, a transparent expert system or battery of (Q)SAR models needs to be developed for this endpoint. The studies performed by Garcia-Domenech and co-workers, using the same data as used for TOPKAT, have shown that simple, transparent regression and classification models can be developed, with an equivalent performance to TOPKAT. Thus, it is recommended that the predictive abilities of these models are compared, and refinements of the literature models explored.

A useful alternative to QSAR when limited data are available is to estimate the toxicity of a chemical of interest by reading across from the corresponding data for suitable analogues. Thus, read across provides an alternative or additional approach to the use QSAR in the estimation of chronic toxicity. Several studies have demonstrated the usefulness of reading across chronic toxicity data, and at least one freely available software tool is available to automate the task in the case of human MRTDs (Lazar). In view of the limited availability of QSARs and predictive software for chronic toxicity effects, the read-across approach merits further investigation, and automated software should be developed further.

**Table 4.3. Databases containing repeated dose toxicity information**

| Database | Availability | Information |
|---|---|---|
| US FDA Maximum Recommended Therapeutic Dose (MRTD) Database<br><br>http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm092199.htm<br>http://www.epa.gov/ncct/dsstox/sdf_fdamdd.html | Freely available | MRTD values for 1215 pharmaceuticals from clinical trials, mostly by oral administration and daily treatments, usually for 3-12 months. (with 5% of the pharmaceuticals being administered intravenously and/or intramuscularly). Includes structures. Available from FDA and EPA DSSTOX |
| RepDose database developed by Fraunhofer Institute of Toxicology and Experimental Medicine<br><br>http://www.fraunhofer-repdose.de/ | Freely available for online searching | Subacute to chronic, oral and inhalation NOELs and LOAELs and for 655 industrial chemicals (version 2009); publicly available rat, mouse and dog studies; includes structures, physicochemical properties and study designs |
| Mazzatorta *et al.* (2008)<br>http://pubs.acs.org/doi/suppl/10.1021/ci8001974 | Freely available as MS Excel file | molecular structures (encoded as canonical SMILES strings) with LOAEL values for 445 unique chemicals |
| The Munro and Cramer datasets:<br>http://apps.ideaconsult.net:8080/ambit2/dataset/26538?max=100<br>http://apps.ideaconsult.net:8080/ambit2/dataset?search=Cramer<br><br>http://www.efsa.europa.eu/en/scdocs.htm | Available from AMBIT website.<br><br>Also expected to be available from EFSA in 2011 | Munro database contains 612 structurally well-defined organic chemicals and aassociated NOELs<br>Cramer dataset contains 83 structures (no toxicological data) |

### 4.3 Organ-specific and system-specific toxicity

In addition to models for acute and repeated dose toxicity at the *in vivo* level, a limited number of models have been developed for predicting toxicities at the cellular, tissue and organ levels. For example, models have been developed for hepatic and urinary tract toxicities (e.g. Matthews *et al*. 2009b), nephrotoxicity and neurotoxicity, as reviewed by Lapenna *et al*. (2010). Some of these models are based on the concept of reactivity-based toxicity. The covalent binding of reactive electrophiles to cellular targets (i.e. nucleophilic sites of macromolecules) has the potential to initiate a chain of biological effects (e.g. depletion of glutathione and protein thiols) resulting in specific organ and and system toxicities.

Among the commonly used software tools, Derek for Windows v.12 estimates neurotoxicity using the following structural alerts: γ-diketone or precursor, acrylamide or glycidamide, nitroimidazole, carbon disulphide or precursor, pyrethroid, 1-methyl-1,2,3,6-tetrahydropyridine, lead or lead compound and organophosphorus ester.

In general, the modelling of organ-specific and system-specific effects represents an underdeveloped field, ripe for future research but far from regulatory applications. Future research initiative could include, for example, re-examination of the datasets for hepatobiliary and urinary tract toxicities of drugs with a view to developing more accessible models and assessing their applicability to chemicals other than pharmaceuticals. In addition, the concept of reactivity-based toxicity, now established as a plausible mechanism for hepatocyte toxicity, could be further exploited using data from hepatocyte cultures and cell lines. In some areas, such as immunotoxicity, short-term progress seems unlikely. The complexity of such effects probably means that alternative (e.g. systems biology) approaches will need to be investigated in the longer term. Ultimately, it seems unlikely that QSAR models for organ-specific and system-specific effects will be used directly for regulatory purposes, where the focus is on the assessment of apical endpoints. However, these models could become a useful contribution to priority setting exercises, and provide means of providing supporting information, such as on the mechanisms of toxicity.

### 4.4 The Threshold of Toxicological Concern approach

Chronic systemic toxicity studies after oral exposure have been used to develop the Threshold of Toxicological Concern (TTC) concept. The TTC is a generic human exposure level for chemicals below which there is low probability of risk to human health, assuming lifetime exposure. The principle of TTC is built on the premise that a safe level of exposure can be identified for chemicals present at low concentrations in the diet, even for those with unknown toxicity, on the basis of their chemical structure (Kroes *et al.*, 2004). As such it can be used to support preliminary hazard characterisation and to set priorities in toxicity testing (Barlow, 2005).

The idea that toxicologically insignificant exposure levels to chemicals exist was proposed by Frawley due to an increasing demand for toxicity testing (Frawley, 1967). Although his estimation of a threshold level was based on limited systemic toxicity studies, the concept became broadly accepted (Safford, 2008).

The first toxicological threshold level for chemicals migrating from food packaging was developed by a probabilistic assessment of the distribution of carcinogenic potency data, from rodent lifetime studies (Rulis, 1992). Rulis proposed a level of exposure of 0.5 ppb equivalent to an intake of 1.5µg/day/adult (Safford, 2008) which would be protective for known and unknown carcinogens. The cut-off value was then accepted by the US Food and Drug

Administration (FDA) as a Threshold of Regulation (ToR), which meant that no further testing was required for substances migrating from packaging into food below this level of exposure. This was the first use of TTC concept for regulatory purposes.

Further development of the TTC concept was carried out by Cheeseman and colleagues, who confirmed that the threshold level of 1.5µg/day, proposed by Rulis, is valid for most carcinogens, and that the dose would be protective also against other toxic endpoints (Cheeseman *et al.*, 1999). They also proposed higher exposure threshold levels for chemicals lacking structural alerts for carcinogenicity, chemicals that were negative in genotoxicity testing and having acute toxicity (LD50) values above 1000mg/kg.

The TTC approach was subsequently refined by different authors with the aim of providing a tiered approach based mostly on chemical structure and oral systemic toxicity data. Munro and colleagues developed a generic threshold for chemicals where non-carcinogenic toxic effects are expected, by evaluating the impact of chemical structure on toxicity. For this purpose they applied the Cramer decision tree, which places chemicals into three structural classes according to the level of concern based on systemic toxicity. The Cramer decision tree approach uses the knowledge on structure activity relationships, metabolism, chemical reactivity, human exposure levels and other relevant information (Cramer *et al.*, 1978). The decision tree consists of 33 questions. Each question can be answered as yes or no, leading to the final classification of a chemical into one of three classes, reflecting the presumption of low, moderate and high toxicity. As a result substances are classified into one of three classes.

- Class I (Low) contains substances of simple chemical structure with known metabolic pathways and innocuous end products which suggest a low order of oral toxicity.
- Class II (Intermediate) contains substances that are intermediate. They possess structures that are less innocuous than those in Class 1 but they do not contain structural features that are suggestive of toxicity like those in Class 3.
- Class III (High) contains substances with a chemical structure that permits no strong initial impression of safety and may even suggest a significant toxicity

The Cramer scheme (and its Toxtree implementation, see 5.1) is applicable to organic molecules and their salts. Polymers, oligomers and inorganics cannot be classified by the decision tree.

Munro *et al.* (1996) proposed human exposure thresholds of 1800, 540 and 90µg/person/day for classes II and II, respectively. To further evaluate the thresholds proposed by Munro, an expert group was established by International Life Sciences Institute (ILSI) Europe. The group concluded that adverse effects on the nervous system, immune system, endocrine system and development were covered by the thresholds previously proposed by Munro for the three Cramer classes. An exception was identified for organophosphates, which are more toxic. For this group of substances, a specific TTC of 18 µg/person/day was derived (Kroes *et al.*, 2004).

The so-called “cohort of concern” was identified. This includes aflatoxin-like, azoxy- and nitroso- compounds, which are genotoxic, and TCDD (2,3,7,8-dibenzo-p-dioxin and its analogues) and steroids, which are endocrine disruptors. Since these groups of compounds were considered to result in the highest risks if present at very low concentrations in the diet, they were excluded from the TTC approach. Other exclusions from the TTC approach include polyhalogenated dibenzodioxins/dibenzofurans/biphenyls and heavy metals, all of which are known to accumulate in the body; and proteins, because of their allergenic potential. For

chemicals having structural alerts for genotoxicity but which do not belong to the cohort of concern, a TTC of 0.15 μg/day was recommended (Kroes *et al.*, 2004).

The ILSI expert group also proposed a decision tree to act as guidance on how and when the TTC principle could be applied as a preliminary step in safety evaluation of chemicals (Kroes *et al.*, 2004). The decision tree is intended for use on chemicals with known structure and low molecular mass. Data on total human exposure are relevant for the successful application of the TTC approach.

So far, the TTC approach has been successfully applied in the safety assessment of food contaminants migrating from packaging by the US FDA, as well as flavouring agents by the Joint FAO/WHO Expert Committee on Food Additives (JECFA). The European Food Safety Authority (EFSA) uses the TTC approach to evaluate flavouring substances, and the European Medicines Agency (EMA) uses it in support of marketing applications for genotoxic impurities in pharmaceutical preparations, and recommends a TTC of 1.5 μg per day for all but highly potent subset of compounds (EMEA, 2006). The US FDA also issued a (draft) guidance document on recommended approaches for genotoxic and carcinogenic impurities in drug products enumerating acceptable TTC values, e.g. 1.5 μg per day for both marketing applications and grater than 1-year clinical trials (FDA, 2008). The application of the TTC approach has also been explored for its applicability to consumer products (Safford 2008; Felter *et al.*, 2009). It has also been proposed that the TTC could be adapted for environmental risk assessment (Barlow, 2005).

The scientific basis of the Cramer TTC scheme and its applicability in different regulatory areas has been assessed by various researchers (Phillips *et al.*, 1987) and institutions. For example, an EFSA opinion on the applicability of TTC in the food and feed areas is currently being developed and will be published in 2011.

### 4.4.1 Databases underlying the derivation of toxicological threshold values

The main databases that have been used to develop the TTC concept and to derive structure-based threshold values (as described in the above-mentioned studies) are summarised in Table 4.4.

**Table 4.4. Summary of Threshold of Toxicological Concern datasets**

| Author | Database (no of substances) | Evaluated experimental data | Conclusions |
|---|---|---|---|
| | | | |
| Rulis (1986) | CPDB carcinogens (343) | Chronic long term exposure | Proposed ToR of 0.5 ppb equivalent to 1.5µg/day adult intake |
| Munro (1996) | JECFA, US EPA IRIS, non tumour from NTP, DART, literature (611) | Oral toxicity data from chronic, sub-chronic, reproductive, teratology studies | Proposed TTC for the three Cramer classes: 1880 µg/day for Class I; 540 µg/day for Class II; 90 µg/day for Class III |
| Cheeseman (1999) | CPDB carcinogens (709) | Short-term toxicity data, genotoxicity testing | Confirmation of the validity of 1.5µg/day for subsets of potent and non potent carcinogens |
| | RTECS (3306)<br>RTECS (2542) | Oral reproductive toxicity data<br>Data from other repeat-dose toxicity tests | Confirmation of the validity of 1.5µg/day for other toxic effects |
| ILSI working group (2000, 2004) | Munro DB<br>JECFA, US EPA IRIS, non tumour from NTP, DART, literature (611) | Subchronic neurotoxicity data (45)<br>Acute neurotoxicitytoxicity data (37)<br>Developmental neurotoxicity (52)<br>Immunotoxicity (37)<br>Developmental (81) | Confirmation of TTC proposed for the three Cramer classes, also for other toxic endpoints<br>Lower TTC of 18 µg/day for organophosphates |
| | Cheesman's CPDB carcinogens (709) extended (730) | | Identified 5 groups of chemicals of highest concern "cohort of concern": 3 groups of genotoxic compounds (aflatoxin-like compounds, azoxy-compounds, nitroso-compounds) and 2 groups of endocrine disruptors (TCDD, steroids) |

### 4.4.2 Software to support the derivation of toxicological threshold values

One of the best known software tools for supporting TTC estimations is the JRC's Toxtree software. Toxtree is a freely available open source software tool that estimates toxic hazard by applying a decision tree approach. It was developed by Ideaconsult Ltd (Bulgaria) under the terms of a JRC contract. It is designed to be user-friendly and flexible, being capable of extensions and revisions to its rulebases (plug-ins). Since it is licensed under the General Public License (GPL), any user has the right to modify and redistribute the software in accordance with the GPL licensing conditions. Toxtree can be downloaded from the JRC (http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE) and from Sourceforge (https://sourceforge.net/projects/toxtree/)

The current version of Toxtree (v2.1.0, June 2010), includes the following plug-ins (rulebases) related to TTC assessment

1) the original Cramer rulebase (Cramer *et al.*, 1978; Figure 4.1)

   The Toxtree implementation of the original Cramer decision has been evaluated by Patlewicz *et al*. (2008).

2) the Cramer rulebase with extensions (Figure 4.2)

   This rulebase (first available in v1.60, July 2009) works by assigning compounds to Class I, II, or III, according to the rules from Cramer, and some extra ones. Several compounds were classified by Munro as Class I or Class II compounds according to the Cramer rules, even though Munro reported low NOAEL values upon oral administration (indicating relatively high toxicity). To overcome such misclassifications, five rules were introduced to capture the possible toxicity of these compounds. This plug-in was developed by Curious-IT, The Netherlands, on behalf of JRC.

3) the TTC decision tree of Kroes *et al.* (2004).

   This rulebase (first available in v2.1.0) results in three possible outcomes: a) substance would not be expected to be a safety concern; b) negligible risk (low probability of a life-time cancer risk greater than 1 in $10^6$); and c) risk assessment requires compound-specific data. It incorporates the Benigni/Bossa rules for the identification of genotoxic carcinogens (developed earlier by ISS, Italy on behalf of the JRC), and requires the user to input the estimated daily intake.

### 4.4.3 Summary and conclusions on the TTC approach

The TTC approach has been applied successfully in the food safety area (especially in the evaluation of flavourings, food contact substances, and pesticide metabolites in groundwater). It has also been evaluated in terms of its applicability in other areas, including pesticide metabolites and degradation products (CRD, 2010), drinking water contaminants, and genotoxic consituents in herbal substances and preparations. There are ongoing discussions at the EU and international levels on how to harmonise the TTC approach across different food sectors, how to provide more detailed guidance on its application, and how to improve the

scientific basis of individual building blocks (steps), such as the Cramer classification scheme (which is based on rules developed in 1978). Ongoing activities include the work of the EFSA Working Group on the Threshold of Toxicological Concern and the ILSI Europe Task Force on Risk Assessment of Chemicals in Food, which has recently established a working group on Chemical Risk assessment in Absence of adequate Toxicological Information.

To take recent scientific advances into account, there is a short-term need (within 3 years) to clarify and refine some of the rules in the Cramer classification scheme. This could involve the rescoping, addition, deletion and reordering of one or more rules. In the medium term (3 years and beyond), there is an opportunity to completely rebuild TTC assessment schemes based on newly developed methodologies, including toxicity and ADME prediction tools, as well as advanced *in vitro* test methods.

In general, it is concluded that the use of currently available (Q)SAR prediction tools in food safety assessment will be most effective in the context of the TTC approach. This will require further research to identify the most suitable models/tools and to develop appropriate ways of interpreting and integrating the predictions they generate.

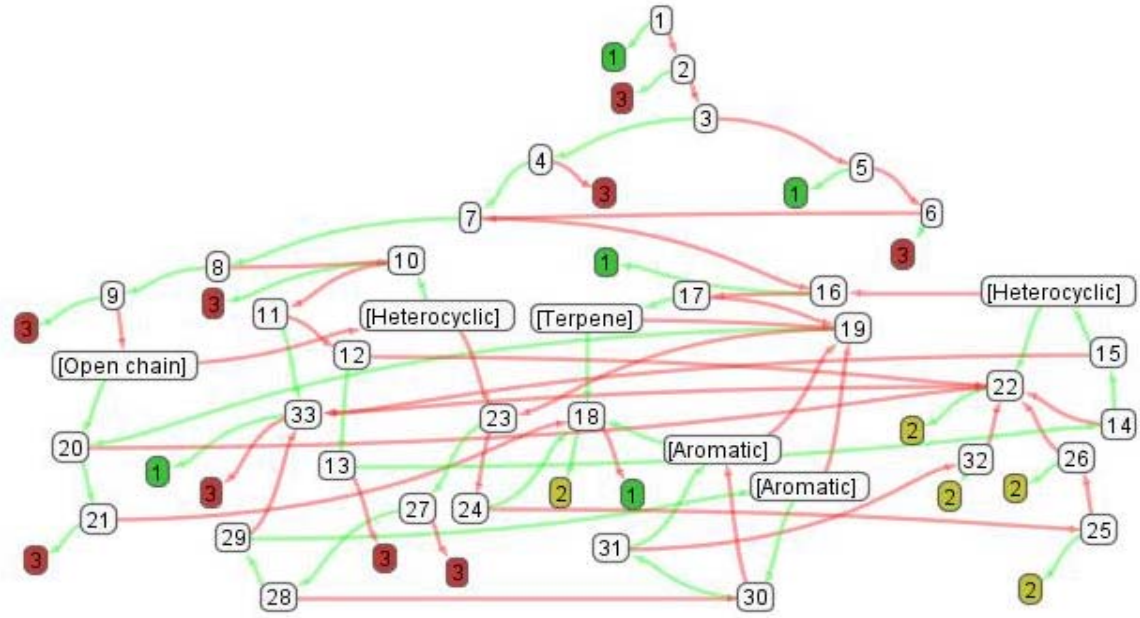



**Figure 4.1. Cramer scheme (original). Yes branch in green. No branch in red. Terminal nodes (labelled 1, 2 & 3) refer to Cramer classifications I, II and III.**

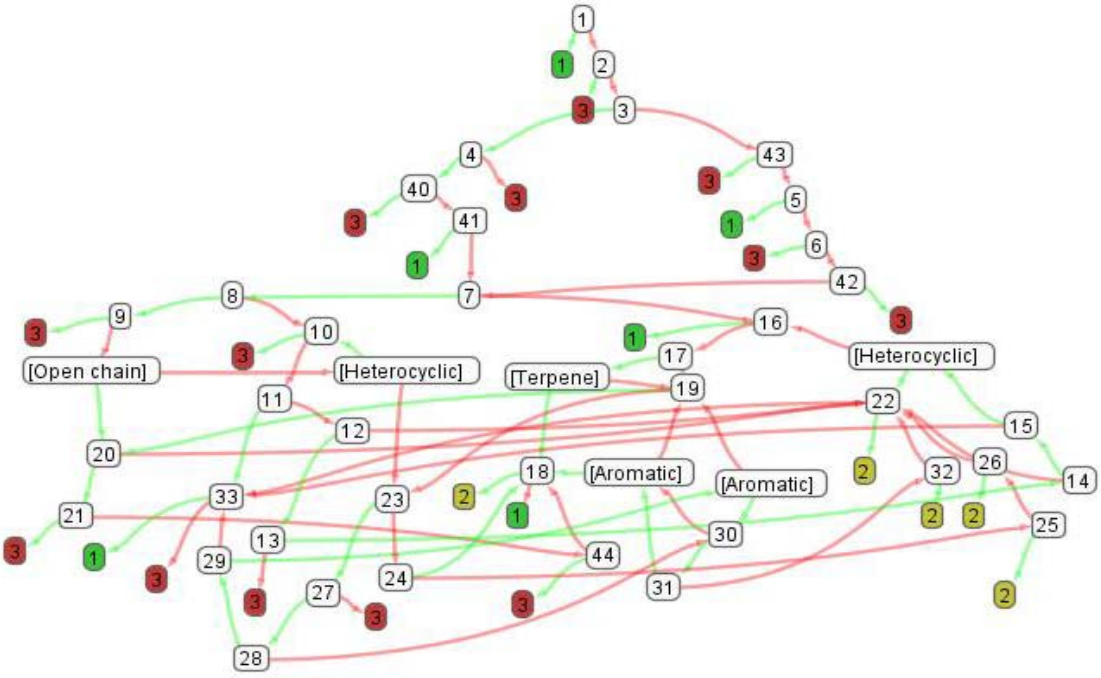



**Figure 4.2. Cramer scheme with extensions decision tree. Yes branch in green. No branch in red. Terminal nodes (labelled 1, 2 & 3) refer to Cramer classifications I, II and III.**

# 5. PREDICTION OF GENOTOXICITY AND CARCINOGENICITY

## 5.1 Introduction

To date, hundreds of (Q)SAR models have been published in the literature for predicting genotoxicity and carcinogenicity, and there are numerous software packages implementing such models. The most commonly modelled endpoint for genotoxicity has been Ames test mutagenicity, whereas carcinogenicity models have focused mostly on the rodent bioassay. This chapter describes the background biology, the various methodologies used, and summarises some of the key conclusions from an extensive literature review concerning the predictivity and applicability of existing models (Serafimova *et al.*, 2010).

## 5.2 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 genotoxicity tests is given in Table 5.1.

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 (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" and mechanistically based approach for the identification of genotoxic potential. The mechanistic chemistry concerning the structural alerts associated with covalent DNA binding has been reviewed in detail by Enoch *et al*. (2010). 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. Other studies have focussed on the development of decision tree approaches (e.g. Purdy, 1996).

## Table 5.1 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<br>OECD 471 |
| *In vitro* mammalian cell gene mutation test – *hprt* test | Mutagenicity: gene mutations | EU B.17<br>OECD 476 |
| *In vitro* mammalian cell gene mutation test – Mouse lymphoma assay | Mutagenicity: gene mutations and structural chromosome aberrations | EU B.17<br>OECD 476 |
| *In vitro* mammalian chromosome aberration test | Mutagenicity: structural and numerical chromosome aberrations | EU B.10<br>OECD 473 |
| *In vitro* micronucleus test | Mutagenicity: structural and numerical chromosome aberrations | EU (none)<br>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<br>OECD 475 |
| *In vivo* mammalian erythrocyte micronucleus test | Mutagenicity: structural and numerical chromosome aberrations | EU B.12<br>OECD 474 |
| Unscheduled DNA synthesis (UDS) test in mammalian liver cells *in vivo* | Genotoxicity: DNA repair | EU B.39<br>OECD 486 |
| Transgenic animal models | Mutagenicity: gene mutations | EU (none)<br>OECD (none) |
| *In vivo* alkaline single-cell gel electrophoresis assay for DNA strand breaks (Comet assay) | Genotoxicity: DNA strand breaks | EU (none)<br>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<br>OECD 483 |
| Rodent dominant lethal test | Mutagenicity: structural and numerical chromosome aberrations | EU B.22<br>OECD 478 |
| Transgenic animal models | Mutagenicity: gene mutations | EU none<br>OECD none |
| *In vivo* alkaline single-cell gel electrophoresis assay for DNA strand breaks (Comet assay) | Genotoxicity: DNA strand breaks | EU none<br>OECD none |
| Unscheduled DNA synthesis (UDS) test in testicular cells *in vivo* | Genotoxicity: DNA repair | |

### 5.3 Databases ontaining information on genotoxicity and carcinogenicity

A number of web-based databases provide access to experimental data for genotoxicity and 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 5.2.

**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/) and the Technical University of Denmark (DTU) website (http://130.226.165.14/). 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 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.

**Table 5.2 Public databases for genotoxicity and carcinogenicity**

| Database (name and link) | Information |
|---|---|
| Benchmark Data Set for *In silico* Prediction of Ames Mutagenicity<br>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)<br>http://potency.berkeley.edu/cpdb.html | Contains of the results of 6540 chronic, long-term animal cancer tests on 1547 chemicals |
| Danish QSAR database<br>DTU site: http://130.226.165.14/<br>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<br>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:<br>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)<br>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)<br>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<br>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<br>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)<br>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)<br>http://toxnet.nlm.nih.gov/ | CCRIS contains over 9000 chemical records with animal carcinogenicity, mutagenicity, tumour promotion, and tumor inhibition test results.<br>GENE-TOX contains genetic toxicology (mutagenicity) test data, resulting from expert peer review of the open scientific literature, on over 3000 chemicals |

## 5.4 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 & Tennant, 1988; Tennant & Ashby, 1991). 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 5.1).

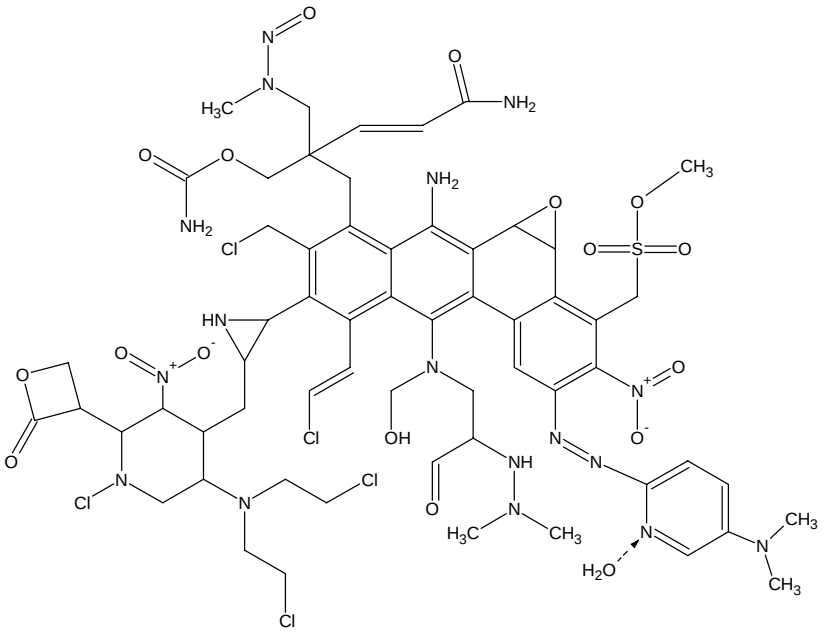


**Figure 5.1 Ashby's poly-carcinogen model, modified after Ashby and Tennant (1988) and Tennant and Ashby (1991)**

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*. (1996).

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 QSAR Toolbox (see below).

The relationships (overlaps) between the different lists of SAs are illustrated in Figure 5.2. 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).

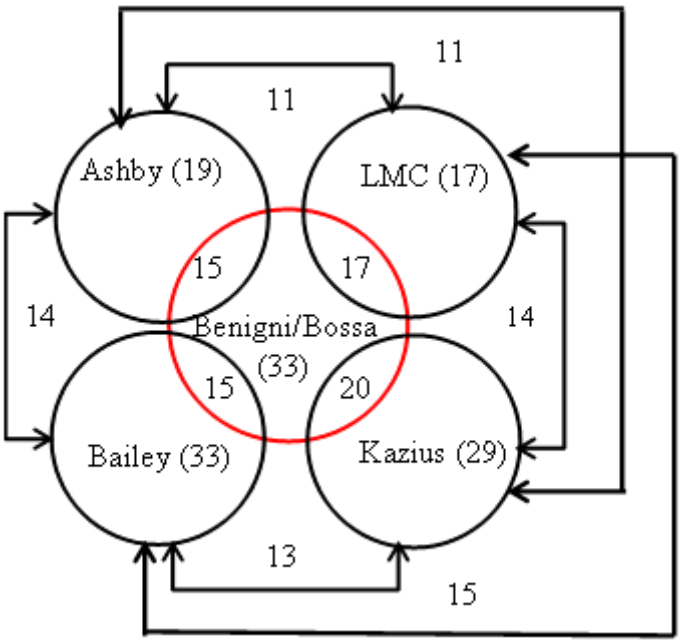



**Figure 5.2 Comparison between the Ashby, Benigni-Bossa, LMC, Bailey and Kazius lists of SAs**

### 5.5 Software for predicting genotoxicity and carcinogenicity

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 5.3.

**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 (Sanderson & Earnshaw, 1991). 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 Derek 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 Derek mutagenicity model is available in the JRC QSAR Model Database.

Crettaz and Benigni (2005) assessed the ability of Derek 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 Derek. 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), Derek was used to predict the genotoxicity and carcinogenicity of 100 randomly selected pesticide active substances (CRD, 2009). It was concluded that Derek 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 Derek and other (Q)SAR programs.

**HazardExpert:** The HazardExpert models (Smithing & 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.

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

**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 DTU website http://130.226.165.14/) as well the as JRC website (http://ecbqsar.jrc.it/). 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; http://www.qsartoolbox.org/) implements five so-called "profilers" connected with genotoxicity and carcinogenicity. Two are general mechanistic profilers: DNA binding by OECD (OECD, 2010) and DNA binding by OASIS (Serafimova *et al*., 2007); and three are endpoint-specific: micronucleus alerts by Benigni/Bossa (Benigni *et al*., 2009); mutagenicity/carcinogenicity alerts by Benigni/Bossa (Benigni *et al*., 2008b) and Oncologic Primary classification (see below).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 CPDB database – 1536 chemicals containing data for Ames mutagenicity and carcinogenicity; c) the OASIS Genotox database – 7500 chemicals with data for Ames mutagenicity and chromosomal aberrations as well as data for metabolism; d) the Toxicity Japan MHLW database – 252 chemicals containing data for Ames mutagenicity e) the Micronucleus ISSMIC database – 142 chemicals with data from *in vivo* Micronucleus assay; f) the Micronucleus OASIS database – 557 chemicals with data from *in vivo* Micronucleus assay. 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/sf/pubs/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 one study (Enslein *et al.*, 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).

## Table 5.3 Software for genotoxicity and carcinogenicity

| Software | Availability | Comments (endpoints predicted, applicability and performance) |
|---|---|---|
| CAESAR<br>http://www.caesar-project.eu/ | Freely available | Mutagenicity, carcinogenicity |
| Derek (Lhasa Ltd.)<br>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<br>http://www.compudrug.com | Commercial | Mutagenicity, oncogenicity |
| Lazar<br>http://lazar.in-silico.de | Freely available | Ames mutagenicity, carcinogenicity |
| MDL-QSAR<br>http://www.symyx.com/ | Commercial | Carcinogenicity |
| MolCode Toolbox<br>http://molcode.com/ | Commercial | Ames mutagenicity, carcinogenicity |
| Multicase (MCASE/MC4PC)<br>MultiCASE Inc<br>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<br>http://www.oasis-lmc.org | Commercial | Ames mutagenicity, chromosomal aberrations |
| OECD Toolbox<br>http://toolbox.oasis-lmc.org<br>http://www.qsartoolbox.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™<br>http://www.epa.gov/oppt/newchems/tools/oncologic.htm | Freely available | Carcinogenicity |
| PASS<br>Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Moscow<br>http://ibmc.p450.ru/PASS// | Commercial | Classification models giving probability of mutagenic effects. There are two models, one for Ames mutagenicity, and another covered multiple*in vitro* and *in vivo* mutagnicity endpoints in mammals. |
| TOPKAT (Accelrys)<br>http://www.accelrys.com | Commercial | Ames mutagenicity, carcinogenicity |
| Toxtree<br>http://ecb.jrc.ec.europa.eu/qsar/ | Freely available | Includes modules for mutagenicity, carcinogenicity, and the *in vivo* micronucleus assay |

## 5.6 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 published since 2000 is given in Table 5.4. 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 range of "multi-model evaluation studies" have been 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 mutagenicity. The three structure-based systems produced equivalent results (71-76% concordance), whereas the physicochemical system produced a lower (61%) concordance. Similar results for Derek and TOPKAT were reported by Cariello *et al*. (2002) - the accuracy of prediction of Ames mutagenicity by Derek 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, Derek, 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 & Zito, 2004).

An informative survey was performed by Benigni and Bossa (2008a). 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 Derek. 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. The observations for TOPKAT and MultiCase were similar to those for Derek. 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.

### 5.7 Conclusions on the ability to predict genotoxicity and carcinogenicity

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.

## Table 5.4 Reviews and model evaluation studies on (Q)SARs for genotoxicity and carcinogenicity (since 2000)

| Year | Reference |
|---|---|
| 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. |
| 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 |
| 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. |
| 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 |
| 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. |
| 2008 | Mayer JM *et al*. (2008). Structure–activity relationship analysis tools: Validation and applicability in predicting carcinogens. Regulatory Toxicology and Pharmacology 50, 50-58. |
| 2008 | Custer LL & Sweder KS (2008). The role of genetic toxicology in drug discovery and optimization. Current Drug Metabolism 9, 978-985. |
| 2008 | Benigni R & Bossa C (2008). Predictivity of QSAR. Journal of Chemical Information and Modeling 48, 971-980. |
| 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 |
| 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. |
| 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. |
| 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-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. |
| 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. |
| 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. |

| Year | Reference |
|---|---|
| 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. |
| 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. |
| 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), 18-23. |
| 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. |
| 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. |
| 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. |
| 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. |
| 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. |
| 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. |

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

| Year | Reference |
|---|---|
| 2002 | Cariello NF *et al*. (2002). Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. Mutagenesis 17(4): 321-329. |
| 2002 | Young R (2002). Genetic toxicology: Web resources. Toxicology 173(1-2), 103-121. |
| 2001 | Prival MJ (2001). Evaluation of the TOPKAT System for Predicting the Carcinogenicity of Chemicals." Environmental and Molecular Mutagenesis 37, 55-69. |
| 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. |

# 6. PREDICTION OF REPRODUCTIVE TOXICITY

Reproductive and developmental toxicity (referred to collectively here as reprotoxicity) studies are used to identify the adverse effects a chemical may have on sexual function and fertility in adult males and females, developmental toxicity in the offspring, as well as effects on, or mediated via, lactation. Thus, reproductive toxicity refers to a range of endpoints relating to the impairment of male and female reproductive capacity (fertility) and the induction of non-heritable harmful effects on the progeny (developmental toxicity). The variety of observable effects are brought about by a plethora of mechanisms of action, many of which are unknown or only partially understood at the molecular and cellular level. Along with carcinogenicity studies, reprotoxicity studies are among the most costly and time-consuming experimental procedures. Furthermore, reprotoxicity testing requires the highest number of test animals. For all these reasons, the development of alternative (non-animal) methods for reprotoxicity assessment is a high political priority.

There are relatively few (Q)SARs for reproductive toxicity, which is partly due to the complexity of the endpoint (many of the underlying mechanisms of action are unknown or only partially understood), and partly due to the paucity of high quality data suitable for model development (Cronin & Worth, 2008). A detailed review of available software and literature models is given in Lo Piparo *et al*. (2010).

## 6.1 Databases

To improve the availability of (Q)SARs and other *in silico* methods for reprotoxicity endpoints, there is a need to develop reprotoxicity databases of high quality and high resolution, in terms of capturing the wide variety of adverse effects and underlying mechanisms of action. Currently available databases are summarised in Table 6.1.

This need has been acknowledged by the International Life Sciences Institute Risk Science Institute (ILSI RSI), who convened a working group to review methodology used to construct statistically based SAR systems for developmental toxicity (Julien *et al*., 2004). It was concluded that an improved process is needed for utilizing developmental toxicity data in the construction of statistically based SAR models. As result of the ILSI RSI report (Julien *et al*., 2004), ILSI is developing a QSAR-ready and peer-reviewed database (http://www.ilsi.org/Lists/Activities/AllItems.aspx) with the assistance of Leadscope Inc. (Columbus, Ohio, USA), and with data contributions coming from a range of governmental and academic organisations, as well as contract research laboratories and major pharmaceutical companies. At the time of writing, this database is not yet available.

To support the establishment of interoperable databases and more consistent risk assessment practices, the DevTox project is developing an internationally harmonised terminology for endpoints in developmental toxicity studies (Paumgartten *et al.*, 2009), including a controlled vocabulary of terms and a defined set of hierarchical relationships between the terms (ontology). An online database (http://www.devtox.org/index.htm) is freely accessible and contains supporting data in the form of image files and text descriptions of tissue anomalies that can be used as the basid for assigning malformations. The DevTox project was initiated by the German Federal Institute for Risk Assessment (BfR), and sponsored by the German Federal Ministry of the Environment, Nature Conservation and Nuclear Safety under the auspices of the International Programme on Chemical Safety (IPCS).

The US FDA has developed and made publicly available the ICSAS Reprotox database (named after the developer research unit, the Informatics and Computational Safety Analysis Staff [ICSAS]), as reported by Matthews *et al*. (2007a,b). The majority of the data were taken from five publicly available sources: Reproductive Toxicology Center System (REPROTOX), Shepard's Catalog of Teratogenic Agents, Teratogen Information System (TERIS), The Registry of Toxic Effects of Chemical Substances (RTECS), and The Physicians' Desk Reference (PDR). In addition, a small portion of internal FDA reprotoxicity data was included. A review of the many duplicate records provided an opportunity to investigate the consistency of information that was reported in the different public databases but extracted from the same original source. This investigation revealed a consistent interpretation of the data from the original sources with the exception of RTECS, indicating in a lesser reliability of this database. The reprotoxicity data were classified into seven general classes (male reproductive toxicity, female reproductive toxicity, fetal dysmorphogenesis, functional toxicity, mortality, growth, and newborn behavioural toxicity), and 90 specific categories. Each specific category contained over 500 chemicals, but the percentage of active chemicals is low, generally only 0.1–10%. In total, the database contains 51,724 study records from over 10,000 individual reprotoxicity studies in which each record is linked to the test chemical structure. The majority of reprotoxicity studies were conducted in rats, mice and rabbits. The majority of test substances were pharmaceuticals, with a relatively limited number of industrial chemicals. The chemical structures are represented as "mol" files and as SMILES (Simplified Molecular Input Line Entry System) codes. The database contains 2134 organic chemicals that are suitable for QSAR modelling. In the QSAR-ready database, built for QSAR analysis, the inorganics, organometallics, high molecular weight polymers, and mixtures of organic chemicals, were excluded.

In support of the ToxCast predictive toxicology effort (Dix *et al*., 2007) the US EPA has developed and made publicly available the Toxicity Reference Database (ToxRefDB) for capturing information from publicly available *in vivo* toxicity studies. This database contains standard toxicity test results for pesticides and other environmental chemicals. It includes the Developmental Toxicity Endpoints dataset (Knudsen *et al*., 2009) resulting from 383 rat and 368 rabbit prenatal studies on 387 chemicals, mostly pesticides; and the Reproductive Toxicity Endpoints dataset (Martin *et al*., 2009) results from multigeneration reproductive toxicity studies on 316 chemicals. The multigeneration reproductive toxicity data set includes assessment of gonadal function, the oestrous cycle, mating behaviour, conception, gestation, parturition, lactation, weaning, and on the growth and development of the offspring. The information in the ToxRefDB is well structured, searchable and downloadable, which makes it a potentially useful resource for QSAR modelling and other developments in predictive toxicology. In order to develop models capable of supporting risk assessment, dose-response data will need to be added.

## Table 6.1 Databases for reproductive toxicity (including receptor binding)

| Database | Availability | Information |
|---|---|---|
| Toxicology Data Network (TOXNET) Developmental and Reproductive Toxicology Database (DART)<br>http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?DARTETIC. | Freely available | Bibliographic database containing over 200,000 references to literature published since 1965. It covers teratology and other aspects of developmental and reproductive toxicology. Users can search by subject terms, title words, chemical name, Chemical Abstracts Service Registry Number (RN), and author. |
| Endocrine Disruptor Knowledge Base (EDKB) database (US FDA)<br>http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm | Freely available | Biological activity database including *in vitro* and *in vivo* experimental data for more than 3,000 chemicals and chemical-structure search capabilities. It includes two datasets: Estrogen Receptor (ER) binding dataset (containing 131 ER binders and 101 non-ER binders), and Androgen Receptor (AR) bataset (containing 146 AR binders and 56 non-AR binders). Searchable by assay type and by structure; provides a search ranking based on a structure similarity index. |
| Endocrine Active Substances Portal (JRC) | Under development | Searchable database giving information on chemical identity (e.g. CAS number), chemical structure, toxicity (both to humans and wildlife), physicochemical properties, mode and mechanism of action, for about 520 chemicals, including those on the EU priority list of substances<br>(http://ec.europa.eu/environment/endocrine/strategy/substances_en.htm) |
| ICSAS Reprotox Database (US FDA)<br>http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm092217.htm | Freely available | Weight-of-Evidence values for 2134 organic chemicals (most of them pharmaceuticals; plus limited numbers of industrial chemicals). SMILES and mol files available. |
| ILSI Developmental Toxicity database | Underdevelopment | Will be available in downloadable format from the ILSI website (http://www.ilsi.org/Lists/Activities/AllItems.aspx) and via the DSSTox website (http://www.epa.gov/ncct/dsstox/) |
| NureXbase<br>http://nurexbase.prabi.fr | Freely available | Information on endocrine-active compunds linked to their receptor targets. Sequence, expression and 3D structures data are linked. |
| NURSA (Nuclear Receptor Signaling Atlas)<br>http://www.nursa.org/ | Freely available | Information on chemical structure, crystal structure, SMILES, physical descriptors, nuclear receptors and mechanism of endocrine action. |

| Database | Availability | Information |
|---|---|---|
| | | |
| OECD (Q)SAR Toolbox<br>http://www.qsartoolbox.org/ | Freely available | Although primarily a tool for chemical categories and read-acros, it also includes several databases, including reprotoxicity data: 166,072 ER binding data from Danish EPA (pre-generated predictions, not experimental values) as well as 1606 experimental ER binding affinity values from the OASIS commercial database. |
| REDIPED (Relational Database of Information on Potential Endocrine Disrupters) developed by the Institute for Environment & Health, University of Leicester, Leicester, UK.<br>http://www.cranfield.ac.uk/health/researchareas/environmenthealth/ | Commercial | Includes references and data on chemical identity, physical properties, production volumes, uses, regulations, sources of exposure, exposure assessment, environmental fate & transport (i.e. accumulation, degradation, fate), and biological activity (*in vitro* and *in vivo* activity, binding abilities, relative activity, and general toxic effects). |
| ToxRefDB<br>http://www.epa.gov/NCCT/toxrefdb/ | Freely available | Standard toxicity test results for pesticides and other environmental chemicals including developmental toxicity (387 chemicals) and multigeneration reproductive toxicity (316 chemicals). |

## 6.2 Software

A number of computer programs generate structure-based predictions of reprotoxicity endpoints, as summarised in Table 6.2, and reviewed briefly below. Some of these models are classification models, amking categorical predictions, whereas others make quantitative predictions.

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

**ADMET Predictor:** This commercial program is designed to estimate certain ADMET (Absorption, Distribution, Metabolism, Elimination, and Toxicity) properties of a drug-like chemical from its molecular structure (see Chapter 7). It includes a qualitative assessment of oestrogen receptor toxicity in rats (TOX_ER_filter), together with a quantitative measure of oestrogen receptor toxicity in rats (TOX_ER (IC50(estrogen)) that is applied only for compounds classified as 'Toxic' by the previous model.

**CAESAR**: The freely accessible CAESAR model for developmental toxicity was built using 292 compounds. Two models were developed, one using WEKA (Waikato Environment for

Knowledge Analysis) and Random Forest, and the other using the Adaptive Fuzzy Partition (AFP) classification model.

**Derek**: This rule-based system includes structural alerts for three specific endpoints: developmental toxicity (3 alerts), teratogenicity (5 alerts), testicular toxicity (1 alert) and oestrogenicity (4 alerts).

Pearl *et al*. (2001) conducted a small validation study with 34 chemicals, and reported 100% specificity (equivalent to 0% false positives) and 72% sensitivity (28% false negatives). However, due to the small size of the dataset, it is difficult to draw general conclusions from these results.

The Dutch National Institute for Public Health and the Environment (RIVM) published a study (Hulzebos & Posthumus, 2003) where Derek predictions for the reproductive toxicity effects of 60 substances were compared with experimental data. The authors concluded that reprotoxicity is poorly predicted by this software. A further study by the RIVM (Maslankiewicz, 2005) reached the same conclusion. The study examined the ability to correctly predict the developmental toxicities of 108 industrial chemicals by using Derek and by applying the chemical categories developed by the US EPA to support the implementation of the Toxic Substances Control Act (TSCA; http://www.epa.gov/compliance/civil/tsca/tscaenfstatreq.html). The conclusion was based on the observation that Derek only recognised 10% of substances which may cause impaired fertility, and only 19% of chemicals which may harm the foetus (on the basis of the harmonised EU classifications of chemicals in Annex I of the Dangerous Substances Directive). However, this conclusion is unfair to the extent that it ignores the fact that Derek is only designed to identify positives and does not make negative predictions – the absence of a prediction simply means there are no rules identifying chemical features of toxicological concern, and does not necessarily reflect the absence of toxicity. For the same reason, use of the ten TSCA categories also revealed low sensitivities (percentage of correctly predicted positive substances) – 19% and 18% for fertility and teratogenicity effects, respectively. The authors also noted that Derek and TSCA had one structural alert in common for the studied chemicals and thus the applicability domain is different for the two predictive approaches. For this reason, it would be worthwhile to build on the RIVM study by investigating the combined use of prediction based on the use of TSCA categories and Derek.

**Endocrine Disruptor Knowledge Base (EDKB)**: This online database, developed and made publicly available by the US FDA's National Center for Toxicological Research (NCTR), contains computer-based predictive models to predict the binding affinity of compounds to the oestrogen and androgen nuclear receptor proteins

**Leadscope**: The Leadscope software has a module containing QSAR models for predicting the developmental toxicity of the rodent foetus, including dysmorphogenesis (structural and visceral birth defects), developmental toxicity (foetal growth retardation and weight decrease), and foetal survival (foetal death, post-implantation loss, and preimplantation loss). The Leadscope QSAR models for reproductive toxicity include rodent male reproductive, rodent male sperm, female reproductive.

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

**MultiCASE**: The US FDA have applied MultiCASE methodology (the MC4PC software) to the above-mentioned FDA database to develop a battery of QSAR models for reproductive and developmental toxicity hazard identification (Matthews *et al*. 2007a, 2007b). Their models were designed to predict seven general reprotoxicity classes: male and female reproductive toxicity, foetal dysmorphogenesis, functional toxicity, mortality, growth, and newborn behavioural toxicity. These are different to the models included in the marketed version of the software. The QSARs were derived from weighted reproductive toxicity findings, in order to incorporate a WoE paradigm based on data from as many as three mammalian species (rats, mice, and rabbits) and to identify trans-species reprotoxicants with a high probability of being reprotoxic in humans. The authors reported a good predictive performance for the majority of the QSARs in this battery: high specificity (>80%), low false positive rate (<20%), and high database coverage (>80%). Because of the large size of the training sets (containing 627 to 2023 chemicals) and the diversity of molecular structures they represent, the authors argue that the QSARs to have a wide applicability domain. However, the models are not documented in sufficient detail to be reproduced and they are not readily transferable. Therefore, in order to use the models, it would be necessary to purchase the MC4PC software and redevelop the models using the same dataset. In conclusion, these studies provide support for the ability to model specific reprotoxicity endpoints, but they are of limited practical usefulness.

**OECD QSAR Application Toolbox**: This freely available software (http://www.qsartoolbox.org/) includes a profiler for predicting ER binding potential, based on a decision tree developed by the US EPA described below (OECD, 2009).

**PASS:** The PASS (Prediction of Activity Spectra for Substances) is developed and marketed by the Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences. Chemicals structures are presented in mol format and used to generate Multilevel Neighbourhood of Atoms (MNA) descriptors (Filimonov *et al*., 1999). A Bayesian algorithm is used to predict various biological activities in terms of the probabilities of presence (Pa) and absence (Pi) of each particular activity (Filimonov & Poroikov, 2008; Poroikov *et al*., 2007). Further information is available at: http://195.178.207.233/PASS/

**PharmMapper Server**: This is a freell accessible web service (http://59.78.96.61/pharmmapper/) which identifies potential target proteins for small molecules, using a pharmacophore mapping approach (Liu *et al*., 2010). Over 7,000 receptor-based pharmacophore models (covering 1,627 protein targets) are accessible. Protein targets include ER, thyroid and progesterone receptors.

**TerraQSAR - E2-RBA**: This suite of software modules, developed and marketed by TerraBase Inc (http://www.terrabase-inc.com/) includes the TerraQSAR - E2-RBA programme, which applies a probabilistic neural network for the computation of estrogen receptor binding affinity (RBA; %) values, relative to that of 17*beta*-estradiol, for organic substances.

**T.E.S.T.:** The Toxicity Estimation Software Tool is an open-source application developed by the US EPA. It estimates the toxicity of a compound by applying several QSAR methodologies thus allowing the user to have greater confidence in predicted toxicities. Among other toxicities it predicts developmental toxicity. The tool is freely downloadable from the EPA website (http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST). The models are well documented and the training set is made available as a structure (sdf) file.

**TIMES**: TIssue MEtabolism Simulator is a heuristic algorithm to generate metabolic maps from a library of biotransformations and abiotic reactions. It allows prioritization of chemicals

according to toxicity of their metabolites. The TIMES platform is also used to predict different endpoints including receptor mediated endpoints for oestrogen, androgen and aryl hydrocarbon binding affinity. They are based on the Common Reactivity Pattern (COREPA) approach developed by the Laboratory of Mathematical Chemistry at the Bourgas University, Bulgaria. The COREPA approach is a probabilistic classification method which assesses the impact of molecular flexibility on stereo electronic properties of chemicals. Similarity between chemicals is analysed by comparing their conformational distributions, and the system automatically identifies the parameter that best discriminate chemicals in groups. A Bayesian decision tree is then developed for classifying untested chemicals. The use of COREPA to predict oestrogenicity has been well described elsewhere (Mekenyan *et al*., 2003a, b; Schmieder *et al*., 2003).

**TOPKAT**: The Developmental Toxicity Potential (DTP) module of the TOPKAT software was developed from experimental studies selected after review of literature citations on rat oral data. TOPKAT comprises three QSAR models, each applicable to a specific class of chemicals. The output is the probability of a submitted chemical structure being a developmental toxicant in the rat. A probability below 0.3 indicates no potential for developmental toxicity (NEG), whereas a probability above 0.7 signifies developmental toxicity potential (POS). The probability range between 0.3 and 0.7 refers to the "indeterminate" zone (IND). The TOPKAT model automatically determines whether the submitted structure belongs to the Optimum Prediction Space (OPS) of the model in order to evaluate the reliability of prediction. The original models were published by Enslein *et al*. (1983) and by Gombar *et al*. (1995), although it is not clear whether the models now implemented in the software are the same as, or refinements of, the original models.

**Toxmatch**: This freely available software (http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/) does not in itself generate predictions of reprotoxicity endpoints, but it can be used to develop categories and support read-across assessments. This has been demonstrated in a study by Enoch *et al*. (2009). This study illustrates the use of 2D similarity indices within Toxmatch to form categories for 57 query chemicals. The underlying hypothesis is that chemicals selected as being similar should act via a single mechanism of action, even if that mechanism is unknown. Read-across predictions were performed for the 17 query chemicals for which a category could be formed. The authors concluded that 2D similarity methods offer a useful method for building chemical categories for reproductive toxicity in which a priori mechanistic knowledge is limited. Although the categories proposed are limited in terms of their applicability (40 query chemicals were not allocated to categories), the results form a good basis for further investigations.

**VirtualToxLab**: This is a commercial tool for predicting endocrine disrupting potential by simulating and quantifying the interactions with aryl hydrocarbon, oestrogen alpha/beta, androgen, thyroid alpha/beta, glucocorticoid, liver X, mineralocorticoid and peroxisome proliferator-activated receptor gamma (Vedani *et al*., 2009; Vedani & Smiesko 2009). It also includes metabolic considerations by simulating interactions with the enzymes CYP450 3A4 and 2A13. The tool is based on the combined use of automated flexible docking with multi-dimensional QSAR (mQSAR).

## Table 6.2 Software for reproductive toxicity (including receptor binding)

| Software | Availability | Applicability |
|---|---|---|
| ACD ToxSuite (ToxBoxes);<br>http://www.acdlabs.com/products/admet/tox/ | Commercial<br>Free web application: http://www.pharma-algorithms.com/webboxes/ | ER binding affinity prediction. Identifies and visualises specific structural toxicophores. Identifies analogues from its training set. Algorithms and datasets not disclosed. Predictions associated with confidence intervals and probabilities, providing prediction reliability. |
| ADMET Predictor<br>http://www.simulations-plus.com/ | Commercial | Qualitative and quantitative prediction of oestrogen receptor toxicity in rats. Based on two models: a qualitative model and, if toxic, the quantitative ratio of IC50 estradiol/IC50 compound. |
| CAESAR<br>http://www.caesar-project.eu/ | Freely available | Two classification models for developmental toxicity based on the dataset of Arena *et al.* (2004) including 292 compounds. |
| Derek<br>http://www.lhasalimited.org/ | Commercial | Classification models (different levels of likelihood) based on 23 alerts for developmental toxicity; 4 alerts for oestrogenicity. |
| Endocrine Disruptor Knowledge Base (EDKB) database (US FDA)<br>http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm | Freely available | Quantitative models to predict the binding affinity of compounds to the estrogen and androgen nuclear receptor proteins. |
| Leadscope<br>http://www.leadscope.com/ | Commercial | Classification models for developmental toxicity in the rodent fetus: dysmorphogenesis (structural and visceral birth defects), developmental toxicity (fetal growth retardation and weight decrease), and fetal survival (fetal death, post-implantation loss, and preimplantation loss). Models of reproductive toxicity: rodent male reproductive, rodent male sperm, female reproductive. |
| MolCode Toolbox<br>http://molcode.com/ | Commercial | Quantitative prediction of rat ER binding affinity and AhR binding affinity |
| MultiCASE (MC4PC)<br>http://www.multicase.com/ | Commercial | Classifcation models for developmental toxicity associated with a variety of datasets, mainly drugs. The marketed software includes modules for predicting mammal sperm toxicity, developmental toxicity, developmental fetal growth retardation, development fetal weight decrease and survival fetal death. |

| Software | Availability | Applicability |
|---|---|---|
| OSIRIS property explorer<br>http://www.organic-chemistry.org/prog/peo/ | Freely available | Classification model which predicts "undesirable" effects (mutagenicity, tumorigenicity, irritating effects and reproductive effects), mainly based on the RTECS database of >3500 compounds. |
| PASS<br>Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Moscow<br>http://195.178.207.233/PASS/ | Commercial, with free internet service, and downloadable demo | Classification models giving probability of reprotoxic effects. The embryotoxicity model predicts the probability that a substance crosses the placental membrane and causes any toxic effect (e.g. fetal bradycardia, low birth weight) or death of an embryo. The teratogenicity model predicts the probability that a substance crosses the placental membrane and causes abnormal development of one or more body systems in the embryo. |
| PharmMapper Server<br>http://59.78.96.61/pharmmapper/ | Free web service | Identifies potential target proteins for small molecules, using a pharmacophore mapping approach. Over 7,000 receptor-based pharmacophore models (covering 1,627 protein targets) are accessible. Protein targets include ER, thyroid and progesterone receptors. |
| TerraQSAR – E2 RBA<br>http://www.terrabase-inc.com/ | Commercial | Neural network model that computes the estrogen receptor binding affinity (RBA; %), relative to that of 17*beta*-estradiol (E2), for organic chemicals. |
| T.E.S.T.: The Toxicity Estimation Software Tool http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST) | Freely available | Developmental toxicity estimation. The prediction is done by applying several QSAR methodologies resulting in a greater confidence of the results. |
| TIMES (COREPA)<br>Laboratory of Mathematical Chemistry, Bourgas University<br>http://oasis-lmc.org/ | Commercial | Classification models for the prediction of estrogen, androgen and aryl hydrocarbon binding.The chemical is predicted to fall in one of several activity bins (ranges of binding affinity). |
| TOPKAT (Accelrys)<br>http://www.accelrys.com | Commercial | Classification model for developmental toxicity of pesticides, industrial chemicals. |
| ToxBoxes<br>Pharma Algorithms<br>http://pharma-algorithms.com/tox_boxes.htm | Commercial | Classification model for the prediction of ER binding. |
| VirtualToxLab<br>http://www.biograf.ch | Commercial | Classification model for endocrine-disruptiong potential based on simulations of the interactions towards aryl hydrocarbon, estrogen α/β, androgen, thyroid α/β, glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated receptor γ, as well as the enzymes CYP450 3A4 and 2A13. |

## 6.3 Endocrine-related effects

### 6.3.1 Endocrine Active Substances and potential Endocrine Disruptors

Endocrine Active Substances (EAS) are chemicals having the potential to interfere with the endocrine systems, as judged from *in vitro* or *in vivo* tests. Such chemicals may be regarded as endocrine disruptors (EDs) if there is evidence that the substance causes adverse health effects in an intact organism, or its progeny, secondary to changes in endocrine function. In particular, EDs have been associated with reprotoxicity, as well as cancer, diabetes and obesity. Numerous mechanisms of action have been associated with endocrine disruption, and a wide variety of *in vitro* tests have been developed to identify chemicals acting via these mechanisms. The status of these *in vitro* tests has been reviewed by Jacobs *et al*. (2008), who also emphasise the need to incorporate metabolic considerations into the assessment of EAS. While endocrine disruption is not a defined endpoint in the framework of EU legislation on chemicals or pesticides, chemicals with ED potential are of particular concern for human health and the environment, especially if their potential adverse effects are not detected by other endpoint assays. In REACH, EDs are considered to be Substances of Equivalent Concern as other Substances of Very High Concern.

EAS act via a range of mechanisms with the result of enhancing or suppressing normal hormone responses, including homeostatic and feedback mechanisms. In many cases, EAS act by binding to nuclear hormone receptors (NRs), which are ligand-inducible transcription factors involved in the regulation of specific target genes and of critical cellular processes such as cell growth, differentiation and metabolic processes. Members of the NR superfamily include receptors for various steroid hormones oestrogen (ER), androgen (AR), progesterone (PR), several corticosteroids, retinoic acid, thyroid hormones, vitamin D, and dietary lipids (the peroxisome proliferator activated receptor; PPAR).

The largest and best studied group of NRs is the Oestrogen Receptor (ER) family. The ER is a ligand-dependent transcription factor - when a hormone binds to the ligand binding domain (LBD), it induces a conformational change in the receptor that initiates a series of events that culminate in the activation or repression of responsive genes (Anstead *et al*., 1997). The crystallographic structures available for the ER have provided insights into mechanisms of action and have given an input to the development of highly specific *in silico* models. The mobility and plasticity of the ER ligand-binding cavity have been identified as important factors allowing the binding of compounds of different structural types to the receptor site (Pike *et al*., 1999). In absence of the ligand, ERs are in an inactive conformation in the target cell nuclei. The binding of an agonist switches the ER into an active conformation, while the binding of an antagonist blocks agonist access. A third category of ligands, termed selective ER modulators (SERMS), have the ability to act as both agonists and antagonists, depending on the cellular and promoter context.

### 6.3.2 *In silico* modelling of endocrine-related effects

There is an extensive literature on the modelling of NR binding and endocrine activity, including studies based on traditional QSAR, molecular modelling, and decision tree approaches. This section reviews, with illustrative examples, these main types of *in silico* methods that have been developed to support the identification of EDs. Strictly, these should not be regarded as *in silico* models for endocrine disruption, since they do not in themselves provide sufficient information to determine whether adverse effects are produced secondary

to changes in endocrine function. However, they could be regarded as models for the identification of EAS.

A number of literature studies were reviewed in the context of a JRC-funded study entitled the "Validation of non-commercial (Q)SAR models for ER and AR binding", which was performed by Mario Negri Institute (Benfenati *et al*., 2005). In this study, non-proprietary models for ER and AR binding activity were reviewed in order to identify interesting publications related to ER and AR endpoints. A scheme for scoring each model/publication was based on the availability of key information (experimental biological data, structures, descriptors, chemical domain and models). A total of 158 models (published until 2005) were scored. Additional studies (published after 2005) are reviewed in Lo Piparo *et al.* (2010)

Several studies (e.g. Netzeva *et al*., 2006; Gallegos Saliner *et al*., 2006) have reported decision trees for categorising chemicals based on the NR binding potential. These are potentially useful for regulatory applications, due to their simplicity, transparency, reproducibility and transferability

A decision tree for predicting ER binding has been developed by the US EPA and included as a "profiler" in the OECD (Q)SAR Application Toolbox, thereby making it freely available and readily applicable (OECD, 2009). The decision tree is based on the hypothesis that the structural domain of chemicals that can bind to the ER is determined by the energy and steric constraints of the ER itself. Based on experimental data available in literature, the nature of the chemical interactions in the various "subpockets" within the ER-binding domain(s) was hypothesised. Three primary ER binding subpockets were identified, having different requirements for hydrogen bonding. The decision tree described uses basic structural features and simple properties to match chemicals with "similar" chemical groups. The system examines each chemical and places them into groups of inactive chemicals, "drug-like" chemicals (which have the potential for strong ER binding affinity), or groups of chemicals which may have weak-to-moderate binding affinity, depending on specific properties or structural features.

When the 3D structure of the protein receptor is known, *in silico* approaches such as molecular docking can be applied. Docking is used to find the best match between a biological macromolecule and a ligand. The ligand is placed inside the receptor pocket and the free energy of binding of the molecular complex is estimated computationally. The receptor structure needs to available from experimental studies, usually X-ray crystallography or NMR. In the case of ERs, several crystal structures of the receptor with different ligands (both agonists and antagonists) are available from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do).

Another *in silico* approach often used for ER affinity prediction is 3D-QSAR based on so-called field-based descriptors that describe the micro-environment surrounding the (ligand) molecules (molecular electrostatic and steric potential and Van der Waals volume). For example, Comparative Molecular Field Analysis (CoMFA) is a modelling method that examines molecules in three-dimensional detail, describing the magnitude and directions of electronic and steric interactions (Cramer *et al*., 2002). CoMFA produces an imaginary 3D box around the ligand, consisting of steric and electrostatic interaction energies at each grid point. These values become the descriptors for QSAR analysis. The main advantages of CoMFA methods are: a) the crystal structure of the protein target is not needed, since the analysis is derived entirely from the ligand; and b) by describing properties in terms of 3D fields, it is possible to visualise areas within the 3D space around the ligand that are positively or negatively related to the activity. The main disadvantage of CoMFA is the need

to align (superimpose) numerous 3D structures, which makes it difficult to study heterogeneous datasets. CoMFA is a research tool that requires considerable expertise to implement. It is useful for investigations into mechanisms of binding and in the development of QSARs, but is not suited for the routine assessment of chemicals by non-specialists.

An alternative to CoMFA, which avoids alignment difficulties, is to use VolSurf (Cruciani *et al*., 2000) and ALMOND (Pastor *et al*., 2000), which are commercially available 3D-QSAR methods developed by Molecular Discovery (http://www.moldiscovery.com/index.php). These are sophisticated yet easy-to-handle computational procedures that can be used to explore the physicochemical property space of a molecule, using a simple molecular input such as SMILES. There is no need to use and manipulate 3D structures since these operations are automatically performed by the software. VolSurf automatically generates 3D maps and compresses the information into numerical descriptors. ALMOND generates and handles alignment-independent descriptors called GRIND (GRid INdependent Descriptors). These are a new generation of 3D-molecular descriptors - being alignment independent, they are quickly and automatically computed. These methodologies are promising research tools for future QSAR development.

A more recent development is VirtualToxLab, developed by Vedani and colleagues (Vedani *et al*., 2009; Vedani & Smiesko 2009). This is an *in silico* tool for predicting the endocrine-disrupting potential of compounds by simulating their interactions towards a series of proteins known to trigger adverse effects. It is based on a fully automated protocol, calculating the binding affinity of a molecule towards a series of proteins and estimating the resulting toxic potential. Currently, 12 protein targets are included: the androgen, aryl hydrocarbon, oestrogen alpha/beta, glucocorticoid, mineralocorticoid, thyroid alpha/beta liver X and the peroxisome proliferator-activated receptor gamma (PPAR-γ), as well as the enzymes cytochrome P450 3A4 and 2A13. Toxic potential is estimated automatically by simulating the interactions with the macromolecular targets, by quantifying these interactions in terms individual binding affinities and combining the flexible docking routine with multidimensional QSAR. The technology is accessible over the Internet (http://www.biograf.ch/).

### 6.4 Regulatory use of *in silico* predictions

*In silico* models for reprotoxicity endpoints and NR binding have mainly been used for setting priorities for testing, rather than to fill data gaps for hazard and risk assessment.

An example of how (Q)SARs can be used in classification and labelling has been reported by the Danish National Food Institute in Denmark (Jensen *et al*., 2008). They performed a screening exercise of 57, 014 European Inventory of Existing Chemical Substances (EINECS) chemicals by using in-house and commercial QSAR models (mainly MultiCASE) in order to identify possible reprotoxicants. Three QSAR models were used for reproductive toxicity for the endpoints teratogenic risk to humans, dominant lethal effect in rodents and *Drosophila melanogaster* sex-linked recessive lethal effect. In addition, the chemicals were also screened by using three models for endocrine activity. Chemicals were considered predicted positive for reproductive toxicity if a positive prediction was obtained in any of the models within the applicability domain. On this basis, 5240 EINECS chemicals (9.2% of the chemicals screened) were predicted as reprotoxicants by one or more of the models. The authors also interpreted the model outputs in terms of EU classifications for reproductive toxicity - category Xn (Harmful) and R63 (Possible risk of harm to the unborn child). The list of chemicals with EU classifications suggested on the basis of QSAR, have been submitted to

the Danish EPA to support a future update of the advisory classification list (which industry can use to support the self-classification of chemicals).

### 6.5 Conclusions

At present, the availability of (Q)SARs for reprotoxicity endpoints (excluding models related to endocrine activity) is limited as a result of the diversity and biological complexity of the endpoints, and the paucity of data suitable for modelling. Available models are potentially useful as a means of supporting hazard identification and priority setting, but not yet for the establishment of toxic potencies for use in risk assessment.

Given the nature of the reprotoxicity endpoints, it is unlikely that an entirely structure-based approach will be capable of fully describing and predicting the *in vivo* effects. Thus, available models should not be used in isolation but to contribute to WoE assessments, and to guide experimental testing, where necessary. Batteries of models and *in vitro* tests will need to be developed, and this has been the aim of an ongoing EU-funded Reprotect project (http://www.reprotect.eu/). This project has included the development of QSARs for predicting passage across the placental barrier (Hewitt *et al*., 2007).

At the current state of development, it is not possible to give clear recommendations on how to use the results of models for reprotoxicity endpoints. For short-term progress (next 3 years), it is recommended that further research on the regulatory applicability of current models is performed, for example along the lines of the Danish EPA study (Jensen *et al*., 2008). In additional to traditional QSAR approaches, the grouping and read-across approach has also been found to be a promising means of making predictions, especially when mechanistic insights are lacking (Fabjan *et al*., 2006). Further work will also need to be aimed at the development and assessment of integrated strategies including *in vitro* data as well as *in silico* models (Hewitt *et al.*, 2010).

The future development of (Q)SAR models and databases will also depend on the development of a standardised vocabulary for describing the plethora of reprotoxic effects at different levels of biological organisation. ILSI and Leadscope have already started such an initiative. In relation to databases, an important achievement has been the construction, from publicly available information sources, of the US FDA's weight-of-evidence (WoE) reprotox database suitable for QSAR modelling (Matthews *et al*., 2007a, 2007b).

In the longer term (5 years and more), the development of systems biology approaches incorporating "omic" and HTS data is likely to become increasingly important. Preliminary investigations have started, for example in connection with the US ToxCast initiative (Martin *et al*., 2009; Knudsen *et al*., 2009). It is too early to judge whether this approach, which reflects a shift from modelling apical endpoints to toxicity pathways, will ultimately be useful in the routine regulatory assessment of chemicals.

In contrast to reprotoxicity, there is an extensive and growing range of software and literature models for predicting endocrine-related activities, and especially binding to the ER and AR receptors. In many cases, these models are at the research stage and require specialised expertise to recreate them in molecular modelling software. However, there are a number of potentially useful models, including simple decision tree approaches (e.g. OECD, 2009) as well as commercial models (e.g. the VirtualToxLab approach; Vedani *et al*., 2009). One of the main challenges here is to develop agreed approaches for interpreting model results for regulatory applications other than priority setting.

# 7. PREDICTION OF BIOKINETIC (ADME) PROPERTIES

## 7.1 Introduction

The term ADME refers to **A**bsorption, **D**istribution, **M**etabolism and **E**xcretion, the four processes related to the toxicokinetic (pharmacokinetic) profile of the chemicals interacting with living organisms. Collectively, these processes determine the fate of the substance inside the body. The term ADMET is sometimes also used, especially in the pharmacological area, the express the overall profiling of ADME properties and **T**oxicological effects of a substance.

The development of methods for determining ADME properties, including *in silico* methods, is a large and rapidly evolving field. This chapter provides an introduction to the background biology, and reviews the current status of available databases, software tools and literature models relevant to ADME prediction. The *in silico* methods cover a range of approaches, including but not limited to (Q)SAR models. A more detailed review is provided by Mostrag-Szlichtyng & Worth (2010).

## 7.2 Background biology

**Absorption** is a complicated process governed by a wide variety of factors, including not only the intrinsic properties of the substance (molecular size, solubility ($logS_{aq}$), ionization constant (pKa) and octanol/water partition coefficient (logP) values), but also physiological conditions inside the organism (local pH, absorptive surface area), and activities of enzymes, transporters and carriers along the gastrointestinal (GI) tract. Absorption in the upper GI tract (in mouth and stomach) is minimal and occurs as a result of passive diffusion. Substances absorbed in mouth (despite enzymatic degradation processes) enter directly the systemic circulation; substances absorbed in the stomach (despite hydrolysis and biotransformation processes) go to the liver first and their actual bioavailability is usually limited by first-pass metabolism. The most intensive absorption takes place in the lower GI tract, especially via large mucous surface of small intestine. The predominant absorption mechanism there is passive diffusion, although large molecules may be taken up by pinocytosis. In the large intestine absorption is less efficient and occurs by passive diffusion or active transport (in case of electrolytes). The activity of gut microflora, enzymatic degradation processes and hepatic first-pass metabolism usually diminish the amount of parent molecule that enters systemic circulation.

Human intestinal absorption (HIA) is usually measured as the percentage of the dose that reaches the portal vein after passing the intestinal wall (%HIA) and is a basis of most *in silico* absorption models. The percentage of the dose that remains after absorption and first-pass hepatic metabolism is defined as the oral bioavailability (F) of the compound. In other words, bioavailability describes the passage of a substance from the site of absorption into the systemic circulation and is usually not equivalent to the amount of a substance absorbed.

Once a compound enters the systemic circulation, it is distributed inside the body. This distribution process is governed by two main factors, namely the permeability of a substance between blood and particular tissues and the affinity of a substance to bind with tissues and plasma proteins.

One of the most important tissue/blood partitioning coefficients is blood/brain (BB) partition coefficient, usually expressed as logBB and defined as the ratio of substance concentration in

blood to its concentration in brain. The passage of compounds across the blood/brain barrier (BBB), an important determinant of neurotoxicity, is based mainly on passive diffusion across the BBB membrane. However active transport also may be important. For nutrients and endogenous compounds, such as amino acids, monocarboxylic acids, amines, hexoses, thyroid hormones, purine bases and nucleosides, several transport systems regulating the entry of the respective compound classes into the brain have been identified. In addition, there is evidence that active efflux pumps like the multidrug transporter P-glycoprotein (P-gp) on the luminal membrane of the brain capillary endothelial cells serve to impede the entry of hydrophobic compounds into the brain.

Compounds in the blood may exist in bound or unbound form. The protein binding of a substance influences the half-life inside the body and the bound fraction often serves as a reservoir from which the substance is slowly released to the unbound form. Unbound substances cross membrane barriers more readily, and may be metabolised and/or excreted. Hence the percentage of plasma protein binding (%PPB) is one of the key determinants in distribution. The most abundant protein in blood plasma is human serum albumin (HSA) accounting for about 60% of the total plasma protein. Since HSA is capable of binding diverse molecules, it significantly affects the overall %PPB.

**Metabolism** (**biotransformation**) is one of the main factors influencing the fate and toxicity of a chemical. Metabolism includes a set of chemical reactions (so-called metabolic pathways) inside the organism, which generally convert xenobiotics into more polar and more easily excreted (i.e. less toxic) forms. However, in some cases metabolism may lead to the formation of toxic metabolites or/and intermediates. Traditionally biotransformation is divided into two main phases - phase I and phase II. Phase I, the so-called functionalisation phase, has a major impact on lipophilic molecules, rendering them more polar and more readily excretable. In phase II, often referred to as detoxification, such functionalised moieties are subsequently conjugated with highly polar molecules before they are excreted. Both phases are catalysed by specific enzymes which are either membrane-bound (microsomal proteins) or present in the cytosol (cytosolic or soluble enzymes). The superfamily of cytochrome P450 (CYP450; also termed heme-thiolate protein P450) enzymes, including more than 70 families of proteins, catalyses the oxidative (and sometimes reductive) phase I metabolic reactions of diverse compounds. Phase II metabolism is governed by various enzymes acting on different types of molecules. The most significant among them are glutathione S-transferase (GST), methyltransferase (MT), N-acetyltransferase (NAT), sulfotransferase (SULT) and UDP-glucuronosyltransferase (UGT). Besides phase I and phase II metabolism, the liver causes specific pre-systemic (first-pass) effects, especially following the oral intake. In addition, phase III metabolism refers to the excretion of metabolites from cells with efflux transporters.

**Excretion** is the process of eliminating waste metabolic products, the major route of which is renal (urinary) excretion via the kidneys. The major non-metabolic routes of clearance ($CL_{tot}$) include bile and urinary elimination of unchanged compounds. The excretion with sweat, faeces and expired air as well as the ability of compounds to be excreted into breast milk and transferred to neonates may also be significant.

### 7.3 Literature reviews on the modelling of ADME properties

Despite difficulties in the modelling of ADME (e.g. low availability and/or quality of experimental data, complexity of physiological mechanisms inside the organisms), a large number of *in silico* prediction models and tools have been developed for ADME and ADME-related properties. As an illustration of the vastness of the ADME literature, Table 7.1 lists the

major reviews and expert opinions that have been published only during the last five years (2005-2010). A recent literature review (Mostrag-Szlichtyng & Worth, 2010) describes in detail a range of models for human intestinal absorption, human oral bioavailability, blood/brain barrier permeability, plasma protein binding, metabolism and excretion

## Table 7.1. Recent (2005-2010) reviews/expert opinions concerning *in silico* studies in ADME and ADME-related endpoints

| Year | Reference |
|---|---|
| 2010 | Mostrag-Szlichtyng & Worth (2010). Review of QSAR Models and Software Tools for predicting Biokinetic Properties. JRC Technical Report EUR 24377 EN |
| 2010 | Madden (2010). *In silico* approaches for predicting ADME properties |
| 2010 | Veselovsky *et al*. (2010). Computer-based substrate specificity prediction for cytochrome P450 |
| 2010 | Wang & Skolnik (2010). Mitigating permeability-mediated risks in drug discovery |
| 2010 | Kortagere & Ekins (2010). Troubleshooting computational methods in drug discovery |
| 2010 | Cross & Cruciani (2010). Molecular fields in drug discovery: getting old or reaching maturity? |
| 2010 | Sprous *et al*. (2010). QSAR in the pharmaceutical research setting: QSAR models for broad, large problems |
| 2010 | Kharkar (2010). Two-Dimensional (2D) *in silico* models for Absorption, Distribution, Metabolism, Excretion and Toxicity (ADME/T) in drug discovery |
| 2010 | Ekins & Williams (2010). Precompetitive preclinical ADME/Tox data: set it free on the web to facilitate computational model building and assist drug development |
| 2009 | Cruciani *et al*. (2009). ChemInform abstract: *In silico* pKa prediction and ADME profiling |
| 2009 | Franklin (2009). *In silico* studies in ADME/Tox: caveat emptor |
| 2009 | Livingstone & van de Waterbeemd (2009). *In silico* prediction of human bioavailability |
| 2009 | Vastag & Keserü (2009). Current *in vitro* and *in silico* models of BBB penetration: a practical view |
| 2008 | Chohan *et al*. (2008). Advancements in predictive *in silico* models for ADME |
| 2008 | Hou & Wang (2008). Structure-ADME relationship: still a long way to go? |
| 2008 | Jacobs *et al*. (2008). The use of metabolising systems for *in vitro* testing of endocrine disruptors |
| 2008 | Li *et al*. (2008). Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction |
| 2007 | Clark (2007). *In silico* ADMET tools: a dawn of a new generation? |
| 2007 | Dearden & Worth (2007). *In silico* prediction of physicochemical properties |
| 2007 | Dearden (2007). *In silico* prediction of ADMET properties: how far have we come? |
| 2007 | Khan & Sylte (2007). Predictive QSAR modeling for the successful predictions of the ADMET properties of candidate drug molecules |
| 2007 | Mohan *et al*. (2007). Computer-assisted methods in chemical toxicity prediction |
| 2007 | Al-Fahemi *et al*. (2007). Investigating the utility of momentum-space descriptors for predicting BBB penetration |
| 2007 | Ekins *et al*. (2007a). *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling |
| 2007 | Ekins *et al*. (2007b). *In silico* pharmacology for drug discovery: applications to targets and beyond |
| 2007 | Ekins *et al*. (2007c). Novel applications of kernel-PLS to modeling a comprehensive array of properties for drug discovery |
| 2007 | Trainor (2007). The importance of plasma protein binding in drug discovery |

| Year | Reference |
|---|---|
| 2006 | Hou *et al*. (2006). Recent advances in computational prediction of drug bbsorption and permeability in drug discovery |
| 2006 | Chohan *et al*. (2006). Quantitative Structure Activity Relationships in drug metabolism |
| 2006 | Crivori & Pogessi (2006). Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs |
| 2006 | Fox & Kriegl (2006). Machine learning techniques for *in silico* modeling of drug metabolism |
| 2006 | Norinder & Bergström (2006). Prediction of ADMET properties |
| 2006 | Gola *et al*. (2006). ADMET property prediction: The state of the art and current challenges |
| 2006 | Schuster *et al*. (2006). Predicting drug metabolism induction *in silico* |
| 2006 | Tetko *et al*. (2006). Can we estimate the accuracy of ADME-Tox predictions? |
| 2006 | Wan & Ulander (2006). High-throughput pKa screening and prediction amenable for ADME profiling |
| 2006 | Segall *et al*. (2006). Focus on success: using a probabilistic approach to achieve an optimal balance of compound properties in drug discovery |
| 2006 | Hyland *et al*. (2006). Utility of human/human-derived reagents in drug discovery and development: An industrial perspective |
| 2006 | Luco & Marchevsky (2006). QSAR studies on blood-brain barrier permeation |
| 2006 | Allen & Geldenhuys (2006). Molecular modeling of blood–brain barrier nutrient transporters: *In silico* basis for evaluation of potential drug delivery to the central nervous system |
| 2006 | Cianchetta *et al*. (2006). Molecular Interaction Fields in ADME and safety |
| 2005 | Colmenarejo (2005). *In silico* ADME prediction: Data sets and models |
| 2005 | De Graaf *et al*. (2005). Cytochrome P450 *in silico*: an integrative modeling approach |
| 2005 | Delisle *et al*. (2005). Computational ADME/Tox modeling: aiding understanding and enhancing decision making in drug design |
| 2005 | Goodwin & Clark (2005). *In silico* predictions of BBB penetration: considerations to "keep in mind" |
| 2005 | Ekins *et al*. (2005). Computational prediction of human drug metabolism |
| 2005 | Ekins *et al*. (2005). Techniques: Application of systems biology to absorption, distribution, metabolism, excretion and toxicity |
| 2005 | Kaznessis (2005). A review of methods for computational prediction of BB partitioning |
| 2005 | Otagiri (2005). A molecular functional study on the interactions of drugs with plasma proteins |
| 2005 | Testa *et al*. (2005b). Musings on ADME predictions and structure-activity relations |
| 2005 | Votano (2005). Recent uses of topological indices in the development of *in silico* ADMET models |

## 7.4 Databases and literature datasets

Although a wide range of diverse molecules have been screened in terms of their ADME properties, mainly to satisfy the needs of the pharmaceutical industry, relatively few data are publicly available. The majority of information on drug candidates are proprietary. Furthermore, ADME data for other types of chemicals (e.g. food additives, environmental pollutants, industrial chemicals, pesticides, etc.) are scarce. Thus, for the purpose of developing new ADME models, limited information is available. It is also unclear whether models developed for pharmaceuticals are applicable to a broader range of compounds, since pharmaceuticals are designed to be bioavalabnle and bioactive.

A list of available databases suitable for the development of QSARs for ADME properties is given in Table 7.2. One of them is **WOMBAT-PK 2009**, the clinical pharmacokinetics database of top selling drugs, provided by Sunset Molecular (http://www.sunsetmolecular.com/). It includes information about over 13,000 clinical

pharmacokinetic measurements for 1230 molecules (1230 unique SMILES) and is being constantly expanded (over 100 drugs are planned to be added in 2010). All WOMBAT-PK 2009 drugs are represented (if possible) in neutral species. The searchable categories of WOMBAT-PK 2009 database include, among others, percentage oral bioavailability (for 818 drugs), percentage plasma protein binding (for 1006 drugs), percentage urinary excretion (for 811), qualitative blood brain barrier permeability (for 519 drugs) and phase I metabolizing enzymes (for 511 drugs). The **Metabolism & Transport Drug Interaction Database** (DIDB) has been developed by the University of Washington scientists (http://www.druginteractioninfo.org/). It contains *in vitro* and *in vivo* information on drug interactions in humans and provides pharmacokinetic profiles of drugs. The **Metabolite**$^{TM}$ Database provided by Symyx (http://www.symyx.com/) indexes paths and schemes of biotransformation for xenobiotics and medicinal drugs and collects experimental data from *in vivo* and *in vitro* studies. **ADME DB**, a database provided by Fujitsu (http://www.fqs.pl/), contains data on interactions of substances with drug metabolizing enzymes and drug transporters. It includes information on ADME properties (e.g. CYP and other phase I and phase II enzymes) as well as interactions between drugs.

Among freely available databases (Table 7.2), two are of importance. The **ADME-AP** database developed by Bio Info & Drug Design (http://xin.cz3.nus.edu.sg/group/admeap/admeap.asp/) (Sun *et al*., 2002), provides data on diverse ADME-associated proteins including physiological function of each protein, pharmacokinetic effects, ADME classification, direction and driving force of disposition, location and tissue distribution, substrates, synonyms, gene name and protein availability in other species. The **PK/DB database** (http://www.pkdb.ifsc.usp.br/) includes 1203 compounds with respect to 2973 pharmacokinetic measurements (Moda *et al*., 2008). This database also includes five models for *in silico* ADME prediction (human intestinal absorption, human oral bioavailability, plasma protein binding, blood/brain barrier permeability and water solubility).

Numerous datasets published recently in the literature are also of importance as far as the modeling of ADME properties is concerned (Table 7.3). They can be used for a wide range of predictive purposes, e.g. for human intestinal absorption, human oral bioavailability, plasma protein binding, blood brain barrier permeation and metabolic pathway modelling.

## Table 7.2. Databases for ADME

| Database (developer and availability) | Database size/ chemical classes | Provided properties | Details |
|---|---|---|---|
| ADME INDEX™ DATABASE<br>Bio-Rad Laboratories<br>http://www.bio-rad.com/<br>(commercial; hosted by Bio-Rad Lab KnowItAll) | FDA-approved drugs and non-approved compounds | ADME | Experimental *in vitro* ADME data generated by Lighthouse Data Solutions (LDS) Laboratory |
| ADME DB<br>Fujitsu<br>http://www.fqs.pl/<br>(commercial, available online) | Drugs | Drug metabolizing enzymes, kinetic metabolism, transporters | Protein information about enzymes and transporters, metabolic reactions, types of drug-drug interactions, structures of drugs and metabolites, kinetic information |
| ADME-associated proteins (ADME-AP) Database<br>Bio Info & Drug Design<br>(Sun *et al.*, 2002)<br>http://xin.cz3.nus.edu.sg/group/admeap/admeap.asp/<br>(freely available online) | 321 proteins and 964 substrates | ADME | Drug ADME associated proteins, functions, similarities, substrates/ligands, and tissue distributions |
| AurSCOPE® ADME/DDI<br>Aureus Pharma<br>http://www.aureus-pharma.com/<br>(commercial) | 7000 compounds | ADME<br>Drug-drug interactions | Biological and chemical information on metabolic properties of drugs |
| BioPath Database<br>Molecular Networks<br>http://www.molecular-networks.com/<br>(trial version freely available online)<br>(commercial full version) | Endogenous compounds<br>1175 chemical structures in free online version<br>2074 chemical structures in commercial version | 1545 biochemical transformations (in free online version)<br>2881 biochemical transformations (in commercial version) | Biochemical pathways (metabolic transformations and cellular regulations).<br>Covered organisms: prokaryotes, plants, yeasts and animals<br>Subcellular localisation of pathways including: cytosol, chloroplasts, mitochondria, endoplasmatic reticulum, peroxysomes, endothelium of blood vessels, vascular muscle cell, animal extracellular matrix, nucleus, animal cell membrane, plant cell wall |

| Database (developer and availability) | Database size/ chemical classes | Provided properties | Details |
|---|---|---|---|
| BioPrint®<br>CEREP<br>http://www.cerep.fr/<br>(commercial) | 2500 compounds | Pharmacology and ADME database | Chemical descriptors (structures, 2D and 3D); *in vitro* profiles; *in vivo* effects. Enzyme/solubility/absorption assays |
| KEGG (Kyoto Encyclopaedia of Genes and Genomes) Database<br>Kanehisa Laboratories (Kyoto University & University of Tokyo)<br>http://www.genome.jp/kegg/<br>(freely available for academic use only; for other purposes available commercially under license agreement with Pathway Solutions Inc.,<br>http://www.pathway.jp/licensing/commercial.html) | 16 databases including 344 metabolic pathway maps, 9150 drugs, 1231 organisms, 16083 metabolites and other small molecules, 8064 biochemical reactions and many others | Metabolism | KEGG metabolism information includes (among others) the following aspects: carbohydrate/ energy/ lipid/ nucleotide/ amino acid/ metabolism; biosynthesis of secondary metabolites; xenobiotic biodegradation and metabolism |
| Metabolism Database<br>Accelrys<br>http://accelrys.com/<br>(commercial) | Drugs, agrochemicals, food additives and industrial & environmental chemicals (69,241 records) | Metabolism | Metabolism data for vertebrates, invertebrates and plants; data on pathways and related compounds |
| Metabolism & Transport Drug Interaction Database (DIDB)<br>University of Washington<br>http://www.druginteractioninfo.org/<br>(commercial) | Drugs | Pharmacokinetic data;<br>Enzyme/transporter interactions | Drug interactions in humans, pharmacokinetic profiles of drugs |
| Metabolite™<br>Symyx<br>http://www.symyx.com/<br>(commercial) | Xenobiotics and drugs | Metabolism | Metabolic paths and schemes; experimental data |

| Database (developer and availability) | Database size/ chemical classes | Provided properties | Details |
|---|---|---|---|
| PharmGKB Database<br>Stanford University<br>http://www.pharmgkb.org/<br>(freely available for research purposes) | Drugs, genes, pathways, diseases, information about people who have participated in pharmcogenomics research studies | Pharmacokinetic data | Clinical and basic pharmacokinetic and pharmacogenomic research in the cardiovascular, pulmonary, cancer, pathways, metabolic and transporter domains |
| PharmaPendium™ Database<br>Elsevier<br>https://www.pharmapendium.com/<br>(commercial) | Data from the FDA freedom of information documents and EMEA "EPAR" approval documents (structure/substructure searchable) | Pharmacokinetic data | Data on efficacy, indications and dosage, safety, pharmacokinetics, pharmacology and mode of action, preclinical and clinical toxicity (extracted from documents), adverse effects (extracted from documents), general product information |
| PK/DB Database<br>(Moda *et al*., 2008)<br>http://www.pkdb.ifsc.usp.br/<br>(freely available online) | 1203 compounds | Pharmacokinetic data | Human intestinal absorption, human oral bioavailability, plasma protein binding, blood/brain barrier penetration |
| Prous Ensemble ® Database<br>Prous Science<br>http://www.prous.com/<br>(commercial) | 127 000 bioactive compounds<br>275 000 references | Pharmacokinetic and metabolism data | Drug monographs containing information on the synthesis, pharmacological actions, pharmacokinetics and metabolism, toxicity, clinical studies, manufacturers and references |
| Symcyp<br>http://www.simcyp.com/<br>(commercial) | 47 drugs -experimental data from *in vitro* enzyme and cellular systems, physicochemical properties and dosage forms | ADME, pharmacokinetic profiles, drug-drug interactions | Population-based PBPK simulator for modelling ADME and drug-drug interactions in virtual patient populations. |
| WOMBAT-PK 2009<br>Sunset Molecular<br>http://www.sunsetmolecular.com/<br>(commercial) | 1230 drugs | Pharmacokinetic data | Percentage oral bioavailability, percentage plasma protein binding, qualitative blood/brain barrier permeability, phase 1 metabolizing enzymes |

**Table 7.3. Literature datasets for ADME**

| Dataset (reference) | Dataset size/chemical class(es) | *ADME* and *ADME*-related properties provided | Information available |
|---|---|---|---|
| Hou *et al*. (2007b) | 648 compounds | Human intestinal absorption (HIA) | HIA |
| Hou *et al*. (2007a) | 768 compounds | Oral bioavailability (F) | Oral bioavailability |
| Moda *et al*. (2007a) | 302 drugs | | |
| Sietsema *et al*. (1989) | (Dataset & 550 references) | | |
| Konovalov *et al*. (2007) | 328 compounds | Blood/brain barrier (BBB) penetration | LogBB |
| Zhao *et al*. (2007) | 1593 compounds | | Binary classification (BBB+/BBB-) |
| Abraham *et al*. (2006) | 328 drugs and organic compounds | | Blood/plasma/serum to rat brain distribution coefficients |
| Li *et al*. (2005) | 415 compounds | | Binary classification (BBB+/BBB-) |
| Hollósy *et al*. (2006) | 179 drugs | Plasma protein binding (PPB) | Percentage PPB, urinary excretion and other ADME data |
| Votano *et al*. (2006) | 1008 compounds | | Percentage human plasma protein binding |
| Turner *et al*. (2004b) | 62 drugs | | Human plasma protein binding; total and renal clearance |
| Thummel & Shen (2001) | 320 drugs | | Percentage PPB, urinary excretion and other ADME data |
| Kalgutkar *et al*. (2005) | (464 references) | Metabolic pathways | Structural alerts |
| Manga *et al*. (2005) | 147 drugs | CYP metabolism | CYP isoforms predominantly responsible for their metabolism (CYP3A4/2D6/2C9) |
| Yap *et al*. (2006) | 503 compounds | Clearance ($CL_{tot}$) | Total clearance in humans |

## 7.5 Software for predicting ADME properties

Tables 7.4 and 7.5 indicate the extensive range of software tools for the purpose of ADME and ADME-related predictions. The vast majority of available software tools are commercial. The tools differ greatly in terms of their capabilities and applications.

Some software, e.g. ACD/PhysChem Suite, ASTER, EPISUITE, ClogP (Table 7.4) were designed to perform the predictions of basic physicochemical properties (e.g. ionizstion constant pKa, octanol/water partition coefficient logP, distribution coefficient logD or aqueous solubility $logS_{aq}$). The best accuracy attained in physicochemical property prediction is close to that of measured data. The only approach that promises to improve the predictive

accuracy of such models seems to be consensus modelling, in which the results of multiple models are combined.

The importance of physicochemical property prediction is that the estimated data often serve as inputs to models of key ADME properties, such as gastrointestinal absorption, BBB permeability, oral bioavailability and plasma protein binding. Software tools such as Know-it-All, ADME Boxes, and ADMET Predictor (Table 7.5) generate physicochemical property predictions and use them in further ADME modeling.

In addition to structure-based models, there is a trend towards developing more sophisticated, mathematical PBPK models (Table 7.5). In these tools, *in vitro* and/or *in vivo* ADME data are integrated with the results of QSAR/QSPR models (e.g. for percentage plasma protein binding or blood/brain barrier penetration) for organism-based ADME modelling. Examples of such software tools include GastroPlus and Cloe which mimic the processes inside living organisms.

Simcyp (http://www.simcyp.com/) is a proprietary PBPK simulator that provides a platform for modelling the ADME properties of drugs and their metabolites, as well as drug-drug interactions, in virtual patient populations (Jamei *et al*., 2009). By predicting inter-individual variability, it can be used to identify people at the extreme risks arising from both oral and non-parenteral routes (lungs and skin) of drug administration / exposure. The populations included are: Healthy Volunteers, North European Caucasians, Japanese, Cirrhotic (different degree), Renal Impairment (different degrees), Obese (different levels), and all paediatric age groups. A Bayesian based parameter estimation module can be used to predict individual as well as population parameters. Various QSAR-based predictors are included to predict ADME parameters if measured data are not available. Symcyp is based on and includes a database of demographic, physiological, genomic and *in vitro* biochemical data. It has been developed by a consortium of pharmaceutical companies, academic institutes and regulatory authorities. In addition, as a module to the Simcyp Population-based ADME Simulator, Simcyp Rat is a 'virtual animal' for predicting drug kinetics in rats. Simcyp is a unique and comprehensive tool, and although is has been developed to support the safety assessment of drugs and their metabolites, it would be worth investigating for its applicability in dietary risk assessment.

**Table 7.4. Software tools for predicting physicochemical properties useful as input data for ADME modelling**

| SOFTWARE (COMPANY) | AVAILABILITY | PROPERTY | | | |
|---|---|---|---|---|---|
| | | Ionization constant (pKa) | Octanol/water partition coefficient (logP) | Distribution coefficient (logD) | Aqueous solubility ($logSa_q$) |
| ACD/PhysChem Suite/Batch (ACD Labs) http://www.acdlabs.com/ | Commercial | ● | ● | ● | ● |
| ASTER (U.S. EPA) http://www.epa.gov/med/Prods_Pubs/aster.htm/ | Not publicly available | ● | ● | | ● |
| ChemOffice (CambridgeSoft) http://www.cambridgesoft.com/ | Commercial | | ● | | |
| ChemProp (Helmholtz Centre for Environmental Research, UFZ) http://www.ufz.de/ | Commercial | | ● | | ● |
| ClogP (DAYLIGHT) http://www.daylight.com/ | Commercial | | ● | | |
| EPISUITE (U.S. EPA) http://www.epa.gov/oppt/exposure/pubs/episuite.htm/ | Freely downloadable | | ● | | ● |
| JAGUAR (Schrödinger) http://www.schrodinger.com/ | Commercial | ● | | | |
| Molecular Modeling Pro (ChemSW) http://www.chemsw.com/molecularmodeling.htm/ | Commercial | | ● | | ● |
| MoKa (Molecular Discovery) http://www.moldiscovery.com/ | Commercial | ● | | | |
| Pipeline Pilot (Accelrys Scitegic) http://accelrys.com/ | Commercial | ● | ● | | ● |
| SPARC (U.S. EPA) http://ibmlc2.chem.uga.edu/sparc/ | Free on-line application | ● | ● | | ● |
| TSAR (Accelrys) http://accelrys.com/ | Commercial | | ● | | |
| VCCLAB (Virtual Computational Chemistry Lab) http://www.vcclab.org/ | Free on-line application | ● | ● | | ● |

**Table 7.5. Software tools for physicochemical-based and organism-based ADME predictions**

| SOFTWARE (COMPANY) | AVAILABILITY | PROPERTY | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ionization constant (pKa) | Octanol/water partition coefficient (logP) | Distribution coefficient (logD) | Aqueous solubility ($logSa_q$) | GI absorption | Caco-2 | Oral bioavailability | BBB | PPB | Metabolism | Transporters | Others (e.g. excretion) |
| ACD/ADME Suite with AbSolv module (ACD Labs) http://www.acdlabs.com/ | Commercial | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | | ● |
| Accord for Excel with ADME/Tox Add-on (Accelrys) http://accelrys.com/ | Commercial | ● | ● | ● | ● | ● | | | ● | ● | ● | | |
| ADME Batches[1] (Pharma Algorithms) – now included in ACD/ADME Suite | Commercial | | | | ● | ● | | | | | | | |
| ADME Boxes[1] (Pharma Algorithms) – now included in ACD/ADME Suite | Commercial | ● | ● | ● | ● | ● | ● | ● | | ● | | ● | |
| DISCOVERY STUDIO including Cerius2 (Accelrys) http://accelrys.com/ | Commercial | | | | ● | ● | | | ● | ● | ● | | ● |
| ADMENSA[1] (Inpharmatica) | Commercial | | ● | | ● | ● | | ● | | ● | ● | | |
| ADMET Predictor (Simulations Plus Inc.) http://www.simulations-plus.com/ | Commercial | ● | ● | ● | ● | ● | | | ● | ● | | | ● |
| ADMETox/Pallas including MetabolExpert, MEXAlert, pKalc, PrologD, TPSA, RetroMEX, RuleOf5, PrologP, ToxAlert, Cytotoxicity (CompuDrug) http://www.compudrug.com/ | Commercial | ● | ● | ● | ● | | | | | | ● | | ● |
| ADMEWORKS including Predictor and ModelBuilder (Fujitsu) http://www.fqs.pl/ | Commercial | | ● | | ● | ● | | | ● | | ● | | ● |

| SOFTWARE (COMPANY) | AVAILABILITY | PROPERTY | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ionization constant (pKa) | Octanol/water partition coefficient (logP) | Distribution coefficient (logD) | Aqueous solubility (log$Sa_q$) | GI absorption | Caco-2 | Oral bioavailability | BBB | PPB | Metabolism | Transporters | Others (e.g. excretion) |
| BioFrontier/P450 (Fujitsu)<br>http://www.fqs.pl/ | Commercial | | | | | | | | | | ● | | |
| ChemDBsoft with MOLPRO Package including SLIPPER (ChemDBsoft)<br>http://www.chemdbsoft.com/ | Commercial | ● | ● | ● | ● | ● | | | | | | | |
| ChemSilico Predictors, i.e. CS LogWS/D/P, CS BBB/PB/HIA (ChemSilico)<br>http://chemsilico.com/ | Commercial | | ● | ● | ● | ● | | | ● | ● | | | |
| Cloe® including Cloe PK, Cloe PredictHIA (Cyprotex)*<br>http://www.cyprotex.com/ | Commercial | | | | | ● | | | | | ● | | ● |
| COMPACT (Computer-Optimised Molecular Parametric Analysis of Chemical Toxicity), University of Surrey, Guildford, UK<br>Lewis *et al*. (1996, 2001) | Neither commercial nor public | | | | | | | | | | ● | | |
| GastroPlus (Simulations Plus Inc.)*<br>http://www.simulations-plus.com/ | Commercial | | | | ● | ● | | ● | | | ● | | |
| iDEA ADME[1] (Lion Biosciences) | Commercial | | | | ● | ● | ● | ● | ● | | ● | | |
| iDEA PKexpress[1] (Lion Biosciences) | Commercial | | | | | ● | | | | | ● | | |
| Jchem with Calculator Plugins (ChemAxon)<br>http://www.chemaxon.com/ | Commercial | ● | ● | ● | | | | | | | | | ● |
| KnowItAll ADME/Tox (Bio-Rad Laboratories)<br>http://www.bio-rad.com/ | Commercial | ● | ● | ● | ● | ● | | ● | ● | ● | | | ● |

| SOFTWARE (COMPANY) | AVAILABILITY | PROPERTY | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ionization constant (pKa) | Octanol/water partition coefficient (logP) | Distribution coefficient (logD) | Aqueous solubility (log$Sa_q$) | GI absorption | Caco-2 | Oral bioavailability | BBB | PPB | Metabolism | Transporters | Others (e.g. excretion) |
| META/METAPC/ MCASE ADME Module (MultiCASE)<br>Klopman *et al*. (1994, 1997, 1999), Talafous *et al*. (1994)<br>http://www.multicase.com/ | Commercial | | | | | | | ● | | ● | ● | | ● |
| MetaDrug™ (Genego)<br>http://www.genego.com/ | Commercial | | ● | ● | ● | | | | ● | ● | ● | ● | ● |
| MetaSite (Molecular Discovery)<br>Cruciani *et al*. (2005)<br>http://www.moldiscovery.com/ | Commercial | | | | | | | | | | ● | | |
| METEOR (Lhasa Ltd.)<br>Testa *et al*. (2005a)<br>http://www.lhasalimited.org/ | Commercial | | | | | | | | | | ● | | |
| MolCode ToolBox (MolCode)<br>http://www.molcode.com/ | Commercial | | | | | | ● | | ● | ● | ● | | |
| NorayMet ADME (Noray Bioinformatics)<br>http://www.noraybio.com/ | Commercial | ● | ● | ● | ● | ● | ● | | | ● | ● | | ● |
| OraSpotter[1] (ZyxBio) | Commercial | | | | ● | ● | | | | | | ● | |
| PK SiM (Bayer Technology Services)<br>http://www.systems-biology.com/ | Commercial | | | | | ● | | ● | | | ● | ● | |
| ProPred (CAPEC)<br>http://www.capec.kt.dtu.dk/ | Commercial | | ● | | ● | | | | | | | | |

| SOFTWARE (COMPANY) | AVAILABILITY | PROPERTY | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ionization constant (pKa) | Octanol/water partition coefficient (logP) | Distribution coefficient (logD) | Aqueous solubility ($logSa_q$) | GI absorption | Caco-2 | Oral bioavailability | BBB | PPB | Metabolism | Transporters | Others (e.g. excretion) |
| PreADME (Bioinformatics and Molecular Design Research Centre) | Commercial | | ● | | ● | ● | ● | | ● | ● | | | ● |
| PreADMET web-based application (BMDRC) http://www.bmdrc.org/ | | | | | | | | | | | | | |
| q-ADME (Quantum Lead) http://www.q-lead.com/ | | | | | | ● | ● | | | ● | | | |
| QikProp (Schrödinger) http://www.schrodinger.com/ | Commercial | | ● | | ● | | ● | | ● | ● | | | ● |
| QMPRPlus[1] (Simulations Plus Inc.) http://www.simulations-plus.com/ | Commercial | | ● | ● | ● | ● | | | ● | | | | |
| StarDrop (BioFocus DPI) http://www.scientific-computing.com/ | Commercial | ● | ● | ● | ● | ● | | | ● | ● | ● | ● | ● |
| Simcyp® (SimCYP)* http://www.simcyp.com/ | Commercial | | | | | ● | | | | | ● | | |
| OASIS-TIMES (Laboratory of Mathematical Chemistry, Bourgas University) http://www.oasis-lmc.org/ | Commercial | | | | | | | | | | ● | | |
| TruPK[1] (Strand Genomics), now a part of KnowItAll platform from Bio-Rad Labs | Commercial | | | | | ● | | | | ● | ● | | |
| VolSurf/VolSurf+ (Molecular Discovery & Tripos) http://www.moldiscovery.com/ | Commercial | | ● | ● | ● | ● | ● | | ● | ● | ● | | |

[1] Former software, not commercially available now, but often cited and still possibly in use

## 7.6 Types of *in silico* modelling approaches

Thousands of ADME models have been published in the scientific literature during the last ten years. These models can be divided into a few categories of modelling approaches. The selection of the most useful approach depends on the aims of investigation and is usually driven by the availability of necessary input data, as well as by the level of information needed as an output (e.g. high-throughput screening of numerous compounds or detailed analysis of particular metabolic reaction).

The simplest approach is based on rules-of-thumb and structural alerts. Their main advantages, i.e. simplicity and transparent interpretability, make them very useful for fast screening of large datasets. As far as ADME-related endpoints are concerned, several rules-of-thumb have been developed (Table 7.6), especially for assessing the likelihood of human intestinal absorption, blood/brain barrier penetration and plasma protein binding. Structural alerts have been identified mainly for metabolism-related issues, but also for human intestinal absorption (Raevsky *et al*., 2002). Models in this category are suitable for routine assessments by non-specialists, especially when rough approximations are sufficient.

Another approach to ADME modelling is data-based modelling. This includes conventional QSAR/QSPR and the application of different statistical algorithms, from relatively simple linear multivariate methods, such as Multiple Linear Regression (MLR), Partial Least Squares (PLS) and Linear Discriminant Analysis (LDA) to sophisticated nonlinear ones, such as Artificial Neural Networks (ANN). They are usually combined with learning methods such as Genetic Algorithms (GAs), Support Vector Machines (SVMs), Inductive Logic Programming (ILP), Bayesian Modelling (BM) and Self-Organizing Maps (SOMs). In addition, the approach of Hologram Quantitative-Structure Property Relationship (HQSAR) has been applied to ADME modelling. This technique is based on the arrangement of molecular fragments in a molecular hologram which allows three-dimensional information to be obtained from two-dimensional input structures (Wang *et al*., 2006; Moda *et al*., 2007a). Models in this category may be suitable for routine assessments by non-specialists, provided that a software implementation of the model is available.

QSARs for ADME properties tend to be local models, i.e. are based on small, homogenous data sets, with reliable predictions being obtained for the compounds falling within the model's applicability domain. Relatively few models have been developed on structurally diverse datasets containing more than 100 compounds. However, the accuracy of predictions across structurally diverse datasets can be improved by the application of consensus modelling, which transfers the strengths of multiple single models to a final consensus one. This approach has been demonstrated, for example in the modelling of blood/brain barrier penetration (Zhang *et al*., 2008a) and total clearance (Yap *et al*., 2006).

To obtain detailed information on the mechanisms of interaction between molecules, similarity-based molecular modelling may be useful. The methods within this category are used mainly in metabolism-related studies, especially for assessing the role of cytochrome P450 or identifying reaction sites (atoms) on particular enzyme substrates. Such methods include 3D-QSAR (e.g. Comparative Molecular Field Analysis, CoMFA); quantitative molecular similarity analysis (QMSA), based on experimental data or computed molecular descriptors; pharmacophore modelling and docking. Models in this category tend to require highly specialised modelling expertise, and as such are not suitable for routine assessments by non-specialists.

This review of literature-based ADME models given below focuses on the conventional QSAR/QSPR approach (data-based modelling category), which could be useful for dietary risk

assessment purposes. Given the large number of QSAR studies published and the wide variety of ADME properties, the description is limited to a few illustrative examples, focusing on key ADME properties: human intestinal absorption (predicted as percentage fractional absorption, [%FA] or percentage human intestinal absorption [%HSA]), oral bioavailability (classification models), blood/brain barrier permeability (logBB and classification models), plasma protein binding (human serum albumin [HSA] binding or percentage plasma protein binding [%PPB]) and excretion (total clearance [$CL_{tot}$] renal clearance). Since conventional QSARs for metabolism prediction are highly limited in terms of their applicability, other types of biotransformation models (e.g. 3D-QSAR, pharmacophore modelling, docking) are also included. Additional detail is given in Mostrag-Szlichtyng & Worth (2010).

## Table 7.6. Rules-of-thumb for ADME evaluation

| Reference | ADME property | Rules-of-thumb details |
|---|---|---|
| Gleeson (2008) | Solubility<br>Bioavailability<br>PPB<br>Brain/tissue binding<br>CYP1A2/2C9/2C1/2D6/ 3A4 inhibition | The influence of molecular weight (MW), ionization state (pKa) and calculated octanol/water partition coefficient (ClogP) on various ADME properties was discussed, e.g:<br>• **Solubility** increases as: MW decreases and ClogP decreases. In terms of pKa: zwitterionic molecules containing both an acidic and basic functional group are the most highly soluble, while neutral molecules are the least soluble; acidic molecules are more soluble than basic molecules;<br>• **Bioavailability** increases as MW decreases; ClogP does not have a significant influence. In terms of pKa: bioavailabilities for neutral, basic and zwitterionic molecules are quite similar;<br>• **Plasma protein binding** increases as: MW increases and ClogP increases. In terms of pKa, PPB follows the trend: acids > neutrals > zwitterions > bases;<br>• **Brain/tissue binding** increases as MW increases and ClogP increases. In terms of pKa, no significant relationships have been observed |
| Lobell *et al.*.(2006) | GI absorption | Good GI absorption is characteristic for reasonably soluble, not too liphophilic, large, polar or flexible compounds. The combined calculated values of physicochemical properties determining these factors, i.e. aqueous solubility ($logS_{aq}$), octanol/water partition coefficient (ClogP), molecular weight (MW), polar surface area (PSA) and the number of rotatable bonds (RotB) give a "**traffic light" (TL) scheme for absorption**, as follows:<br>• **Green**: $logS_{aq} \geqq 50$; ClogP ≤ 3; MW ≤ 400; PSA ≤ 120; RotB ≤ 7;<br>• **Yellow**: $logS_{aq}$: 10-50; ClogP: 3-5; MW: 400-500; PSA: 120-140 RotB: 8-10;<br>• **Red**: $logS_{aq} < 10$; ClogP > 5; MW > 500; PSA > 140; RotB ≧ 11 |
| Zmuidinavicius *et al.* (2003) | Human intestinal absorption | • Compounds with quaternary nitrogens or biphosphonate moieties are poorly absorbed;<br>• Compounds with molecular weight < 255 have good absorption;<br>• Compounds with molecular weight between 255 and 580, polar surface area < 154 $Å^2$ and one of two following conditions hold: logP > 0 or hydrogen bond acidity < 1.3 display good absorption;<br>• Compounds with molecular weight > 580, polar surface area < 291 $Å^2$ and logP > 0 are well absorbed |
| Norinder & Haberlein (2002) | BBB penetration | • The molecule has a high chance of entering the brain if the number of nitrogen and oxygen atoms (N+O) atoms is ≤ 5;<br>• LogBB is positive if [logP-(N+O)] is positive |

| Reference | ADME property | Rules-of-thumb details |
|---|---|---|
| Veber *et al*. (2002) | Oral bioavailability | High probability of good oral bioavailability for compounds with:<br>• ≤ 10 rotatable bonds;<br>• Polar surface area ≤ 140 $A^2$ or<br>• The sum of hydrogen bond donors and acceptors ≤ 12 |
| Kelder *et al*. (1999) | BBB penetration | • The upper limit for the polar surface area (PSA) for a molecule that has a high chance of entering the brain is < 60-70 Å |
| Van der Waterbeemd *et al*. (1998) | BBB penetration | • The upper limit for the polar surface area (PSA) for a molecule that has a high chance of entering the brain is around 90 Å<br>• The molecular weight (MW) of such molecule should be not larger than 450 g/mol |
| Lipinski *et al*. (1997, 2001) | Absorption | "Rule of 5", indicating that a molecule is prone to poor absorption if:<br>• Molecular weight > 500;<br>• Sum of OH and NH hydrogen bond donors > 5;<br>• Sum of O and N hydrogen bond donors > 10;<br>• ClogP > 5 |

## 7.7 Current status of *in silico* models for key ADME properties

### 7.7.1 Human intestinal absorption models

The majority of published models for human intestinal absorption have been developed using datasets including drugs and drug-like molecules, what creates a significant shortcoming as far as their applicability to different classes of chemicals is concerned. Furthermore, the published models are at the research stage, and not yet implemented into software suitable for the routine assessment of chemicals.

Nevertheless, some general findings have been identified that may be useful in further studies. The most significant descriptors for HIA are related to hydrogen bonding, molecular size, lipophilicity and surface polarity. Moreover, some generic functional groups which have detrimental impact on HIA have been identified, e.g. quaternary nitrogens and biphosphonates. The datasets used in modelling procedures should include, if possible, chemicals covering the whole range of %HIA values in order to avoid biases towards poorly/highly absorbed compounds. Some compounds (usually those actively transported, insoluble or acting as P-glycoprotein substrates) appear as outliers or rule contradictors in HIA models – in such cases a set of preliminary models for active transport/solubility/P-gp binding could be developed before HIA prediction in order to identify the outliers and avoid final prediction errors. In addition, future research efforts should investigate ways of incorporating metabolic effects into QSAR models.

### 7.7.2 Bioavailability models

Bioavailability is a very challenging property to model, due to the diversity of the underlying determinants, some of which (e.g. first-pass metabolism) are very difficult to model. One of the handicaps in bioavailability modelling is the paucity of data publicly available to the scientific community and the fact that the majority of the data available concerns mainly drugs and drug-like molecules.

Despite these difficulties, several attempts have been made to model human oral bioavailability, generally in categorical terms (e.g. high *vs* low bioavailability). These studies have resulted in a reasonable or good ability to identify high bioavailability compounds, but a relatively poor ability to identify low bioavailability compounds. Available studies also show that modelling strategies based on whole-molecule descriptors of diverse structures is not sufficient, as it does not allow to effectively characterise the first-pass metabolism. The more successful models employ well-defined substructures, which are probably related to different metabolism pathways.

### 7.7.3 Blood-brain barrier models

There is a wealth of BBB permeability information published in the literature and available databases, which could potentially be applied by researchers to develop *in silico* models of brain penetration. However, the major shortcomings of existing data sets is that they tend to be relatively small (less than 100 compounds), they come from a variety of sources and may not be sufficiently consistent for modelling purposes. Other datasets were compiled specifically for drugs. Very few models have been proposed for determination of logBB for pollutants. Hence, one of the most urgent needs is the generation of larger and more diverse datasets with accurate measurements of logBB values. Nevertheless, the majority of recently developed QSAR models based on logBB data represent good predictivity as determined by both internal validation against the training set and external validation against test sets. There are a number of *in silico*

models yielding logBB predictions of around 0.35-0.45 log units that could be used for screening purposes. By examining the wide variety of potentially useful molecular descriptors that have been reported, and some important generalisations for further modelling studies can made. Generally it is possible to distinguish two categories of descriptors. The first includes descriptors of size (i.e. molar refraction, connectivity and topological indices, molecular mass, surface area) while the second includes descriptors of polarity (i.e. polar surface area, partial charges, functions of hydrogen bond acid or hydrogen bond base groups). The descriptors from the first class are important predictors for the partitioning of non-polar compounds in the brain, whereas the descriptors from the second category express the features of polar molecules which are determine their tendency to partition in the blood.

### 7.7.4 Models for plasma protein binding

The relatively small number of studies performed for plasma protein binding is a result of complexity of factors influencing the binding process on the one hand and the paucity of PPB human data on the other. Large differences between data obtained from various species put into question the utility of models developed on non-human plasma proteins to predict human plasma protein binding. The majority of available human PPB models are based on data for drug molecules and tend to have a local character with applicability domains limited to small sets of structurally similar molecules. Although such models are relatively simple (they are based on relatively small number of descriptors, with lipophilicity being the most significant one) and probably easily reproducible and transferable, they cannot be applied to sets of structurally diverse compounds. However, a few investigations (discussed above) were based on broader datasets. Based on these studies, it can be concluded that lipophilicity alone is not important but not sufficient to model PPB processes, especially in the case of large and diverse datasets are concerned. It is necessary to use additional descriptors of various types (e.g. structural, topological, quantum mechanical) to obtain more complex and reliable human PPB models, and the use of non-linear modelling technques may also be necessary. However, this is usually connected with a decreased transparency and reproducibility of the models.

### 7.7.5 Metabolic fate models

The utility of conventional QSARs predicting the metabolic fate of chemicals is highly limited. However, computer-based expert systems (COMPACT, META, MetabolExpert, METEOR, TIMES; see Table 7.5) have a much broader applicability.

A few QSAR models in the literature have provided some promising results for further research studies (discussed above). Most of these were designed to predict the phase I metabolism, with CYP450 isoforms playing a predominant role in the biotransformation of human drugs and xenobiotics. The modelling of phase II metabolism has not received as much attention; in most cases, these models have been developed for GST-catalyzed biotransformation.

Although progress is being made in the development of QSARs for metabolism, currently available models are typically derived from small data sets (only few of them are based on more than 100 compounds) and thus show poor predictivity for heterogenous sets of compounds. Most of the available QSARs have been developed for the purposes of drugs discovery and development. Furthermore, the model-building methodology, underlying training sets and model algorithms are often not transparent, which is an impediment to interpretation and reproducibility. A major bottleneck is the paucity of high quality and relevant experimental (*in vitro* or *in vivo*) data for use in model building and validation. Thus, it is difficult to make clear recommendations about which currently available literature models could be used in the dietary

risk assessment of chemicals other than drugs. To make progress in this respect, more transparent descriptions of the applied approaches and training datasets are needed.

As far as modelling of CYP inhibition is concerned, literature QSARs are at an early stage of development as they usually give poor predictions when tested on the external sets of compounds. Much better results can be obtained from the models predicting the site of the metabolism (predictivity of 80% or more). The most challenging task seems to be modelling the rates of metabolism.

Significant improvement could probably be obtained by combining multiple *in silico* models for metabolism prediction (consensus modelling) along with physiologically based pharmacokinetic (PBPK) modelling utilising the data from different sources (*in silico*, *in vivo* and *in vitro*). However, this represents a long-term research effort.

### 7.7.6 Excretion (clearance) models

The complexity of excretion processes and paucity of experimental data have hindered the development of models for excretion. Some efforts to model human total, urinary and (to a lesser extent) biliary clearance have been made only recently. These studies have identified some important trends governing the clearance processes, which form a useful basis for further research and model development. Most of the models are based on non-linear relationships and utilise large numbers of molecular descriptors in order to capture the multiple features affecting the clearance process. These models tend to be less transparently documented and thus of low reproducibility. However, if encoded into software tools, they could be practically useful. From the available literature, it seems that the software-based VolSurf approach (Cruciani *et al.*, 2000), shown to be successful for modelling human intestinal absorption, oral bioavailability and blood/brain barrier penetration modelling, also works well for renal clearance prediction. Given the emphasis of published studies on drugs, the applicability of these approaches to other types of chemicals significant in dietary risk assessment, would require further investigation.

# 8. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

## 8.1 Survey on the use of computational methods

The survey of regulatory use resulted in responses from 38 organisations around the world, including government authorities and industry. The results indicate that the majority of stakeholders in the food safety field do not currently apply QSAR analysis on a routine basis, mainly through lack of in-house expertise. A few organisations, however, have developed specialised in-house expertise and approaches, mainly for priority setting and filling data gaps in emergency situations. In general, the respondents were in favour of further using QSAR analysis (including chemical grouping and read-across), and requested additional guidance and training.

## 8.2 A conceptual framework for assessing QSAR predictions

A framework for assessing the usefulness of QSARs was proposed (Worth *et al.*, 2011), building on guidance already adopted for the REACH regulation, including international (OECD) guidance on the validation and documentation of QSAR models for regulatory purposes. Assessing the usefulness of a model, in terms of both its practical applicability and the adequacy (relevance, reliability and completeness) of its predictions, is not a trivial exercise, and needs to be performed on a case-by-case basis. It is therefore difficult to make generalisations and provide firm guidance on the assessment of model usefulness. Since REACH is still in an early phase of implementation, the current guidance has not been tried and tested to any significant extent, and it leaves many issues open to judgement, such as the level of detail needed in the reportining formats for QSAR models (QMRFs) and their predictions (QPRFs), and the criteria for deciding when a given estimate is adequate for a specific regulatory purpose. There is therefore a need to develop further this guidance, and also adapt it for regulatory purposes in the food safety field. With this eventual goal in mind, a checklist of questions is proposed that focus on issues that could be reasonably considered by the risk assessor. Other issues can only be assessed by drawing upon specialised QSAR expertise. The application of the checklist to a range of software models for gentoxicity prediction was illustrated with several examples using two pesticide actives as case studies (Worth *et al.*, 2011).

## 8.3 The availability of models for toxicity prediction

The review of (Q)SARs for toxicological endpoints was performed in a broad sense, considering both models available in software tools and models that are published in the literature. The review identified numerous (Q)SAR models for toxicological endpoints relevant to dietary risk assessment. The models are based on a wide variety of approaches, including models that are mechanistically-based or at least mechanistically plausible, and models that have no apparent mechanistic basis. Literature-based models tend to be “local” models, applicable only to specific groups of chemicals, whereas software tools tend to be “global” models, with a wider (but not universal) applicability. Literature-based models are not as accessible to the end-user, unless they have also been encoded in the form of a software tool.

The review also revealed considerable differences in the availability of models depending on the endpoint. At one extreme, there is a huge literature and range of software tools for predicting genotoxicity and carcinogencity, and at the other extreme, there are few or no models for organ and system-specific toxicities. The quality of the models also varies

depending on the endpoint: in general, models for acute toxicity are more reliable than "complex" endpoints which comprise a large number of partially understood mechanisms, such as chronic toxicity, systemic toxicities, and reproductive toxicity. For mutagenicity and carcinogenicity, there is a relative abundance of reliable models, mainly based on the fact that these toxic effects are driven by chemical reactivity (electrophilic binding to DNA). 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 was also noted that a range of software tools are research tools suitable for model development, but 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 which are associated with transparent documentation on the model development and validation process. While transparency is much valued, this does not necessarily mean that all aspects of model development and validation need to be completely open for a model to be useful in a regulatory context. It could be argued that commercial software tools do not need to document their predictive algorithms, provided that information on model applicability is provided, and the user obtains an indication of the reliability of individual predictions, ideally with supporting data for one or more analogues. Similarly, it could be argued that the training set does not need to be completely transparent, provided that a celar definition of the applicabity domain is provided. In the case of models that are published in the literature however, a greater degree of transparency is needed in order to ensure reproducibility and transferability. Ultimately, the degree of information needed to support the regulatory use of models should be a policy decision, and this will most likely be context-dependent, and even decided on a case-by-case basis.

At present, it is difficult to give firm guidance on how to use available models for specific groups of food chemicals and for specific purposes in dietary risk assessment. This would require focussed research investigations. In general, most models are not sufficiently validated (across a diverse range of chemical groups) or documented to promote confident use. However, it may be possible to use such models in dietary risk assessment, on a case-by-case basis, in a weight-of-evidence approach in which the predictions are substantiated by other available information, e.g. experimental data on close analogues. In addition, it is often recommended to combine the use of multiple models for a given endpoint, to improve the reliability of prediction, although clear guidance on how to combine the results of multiple models (e.g. in the form of model batteries or decision tree approaches) still needs to be developed.

### 8.4 The availability of computational models for ADME prediction

The review of computational models for ADME properties was also performed in a broad sense, with emphasis on QSARs and rule-based approaches. This revealed a vast and rapidly growing literature and software range, especially for the prediction of certain ADME properties (e.g. blood/brain barrier permeability, human intestinal absorption). While it is difficult to give firm conclusions on the applicability of such tools, it is clear that many have been developed with pharmaceutical applications in mind, and as such may not be applicable to other types of chemicals (this would require further research investigation). Most of the available data sets are thus skewed toward drug molecules. If these models are applied to other classes of chemicals, the predictions may be unreliable, and in many cases the user will

not be able to judge on this, since the applicability domains have not been explicitly defined and in many cases the training sets are confidential. On the other hand, a range of predictive methodologies have been explored and found promising, so there is merit in pursuing their applicability in the field of food safety. Many of the software tools are not transparent in terms of their predictive algorithms or underlying datasets. However, the literature identifies a set of commonly used and easily-interpreted descriptors that have been found useful in ADME prediction, so further research and model development activities could be based on such studies.

To promote the wider use of *in silico* models for ADME properties in the risk assessment of food chemicals, various significant research initiatives would need to be undertaken: a) it will be necessary to generate high-quality experimental datasets for relevant classes of chemicals other than drugs (e.g. pollutants, food additives, food contact materials, pesticides); and b) the applicability of each model would have to be determined, on a case-by-case basis, by comparing its predictions with experimental data for chemical inventories of interest.

### 8.5 The applicability of models for genotoxicity and carcinogenicity prediction

The case studies (JRC, 2010; Worth *et al.*, 2010) focussed on the applicability of several software tools for predicting genotoxicity and carcinogenicity endpoints. This was identified as particularly important in terms of incorporating such tools into TTC assessment schemes. To assess the predictive performance of software models, some conclusions could be drawn from a large and structurally diverse dataset containing more than 700 chemicals. The abilities of several software tools to predict genotoxicity and carcinogenicity were comparable to previously published evaluations.

The ability of individual models to identify carcinogens was found to be moderately better than chance (typical sensitivities of 66-71%, typical false negative rates of 29-33%), which might not be considered adequate. In contrast, several tools were good identifiers of Ames mutagenicity (typical sensitivities of 80-93%; typical false negative rates of 7-20%). The boundaries of these ranges (93% sensitivity and 7% false negatives) are likely to represent extreme values of predictivity, since in the case of statistical models, a defined but variable percentage of the test chemicals are also present in the model training sets. Furthermore, some of these tools were good identifers of classified mutagens (highest sensitivities of 73-87%; lowest false negative rates of 13-27%). Pairwise combinations of these tools could increase the overall sensitivity (to about 90%) and reduce the false negative rate (to about 10%). Such tools could be employed to identify potential genotoxicants. In the context of a TTC assessment, such chemicals could either be excluded from the TTC scheme, and therefore assessed on a case-by-case casis, or subjected to a lower threshold of toxicological concern. For example, following the proposal of Kroes *et al*. (2004), the dietary intake of compounds predicted to be genotoxic should be compared with a TTC of 0.15μg/person/day (as opposed to thresholds of 1800, 540 and 90 μg/person/day for chemicals in Cramer classes I, II and III, respectively). Chemicals that are predicted to be classified (*in vivo*) mutagens (or both genotoxic and carcinogenic) could perhaps be excluded from the TTC scheme and assessed on a case-by-basis using experimental data. The performances of the various software tools were assessed by applying transparent schemes for interpreting the predicted data. It principle, it is possible to modify these schemes in order to optimise the prediction of positives (usually at the expense of correctly predicting negatives) or *vice versa*. Thus, future research could focus on optimising the data interpretation schemes with a particular purpose in mind (e.g. improving the ability to identify positives). Furthermore, models that are good at predicting positives can be used alongside models that are good at predicting negatives, to

produce an optimised model battery. Further research is needed to investigate how to develop optimised model batteries against fixed and pre-defined criteria (which may include, for example, criteria based on sensitivity, specificity and rates of indeterminate/equivocal predictions).

### 8.6 The use of computational models in dietary risk assessment

Based on the reviews of QSARs and other *in silico* models, it is not possible to give strong conclusions and provide firm guidance on which models should be used in the routine assessment of chemicals in food. In general, however, it can be concluded that:

a) model predictions should be substantiated with additional information, e.g. a comparison of model predictions with experimental data for one or more analogues, thereby adding confidence to the prediction for the untested chemical;

b) predictions from multiple models are sometimes superior to individual model predictions; however, identification of optimal model combinations (batteries) still requires considerable research;

c) individual models have their own strengths and weaknesses, for example in terms of their applicability to different areas of chemical space; however, building a detailed understanding, endpoint-by-endpoint, of where individual models perform more or less reliably will require a considerable research effort.

d) In cases where user-friendly QSAR models / expert systems are either not available or reliable, the grouping and read-across approaches could be useful in order to predict toxicological endpoints and ADME properties. Various computational tools exist to support grouping and read-across approaches. However, the application of these methodologies requires expert judgement drawing on QSAR, chemistry and toxicology, so they are not as stratighforward as some of the conventional QSAR software packages.

e) For the purposes of dietary risk assessment, computational models will be used most profitably in the overall context of the TTC approach, which has been shown to be valid for assessing. Further research is needed to optimise TTC approaches for food safety applications.

A summary of the state-of-the art of software models is given in Table 8.1, along with some observations that may guide the user. These observations are based on the reviews and the experience of the authors, and should therefore not be taken as absolute guidance.

### 8.7 Recommendations

To achieve a wider and more judicious use of QSARs and related computational methods in the food safety area, efforts will be needed to provide tailor-made training courses and materials, to raise awareness through improved communication and outreach activities, and by investing in focussed research strategies.

Research strategies should partly focus on available models, investigating their applicability, endpoint-by-endpoint, to chemical inventories of relevance in the food safety area (e.g. pesticides, food contact materials, food additives). In parallel, efforts will be needed to develop tailor-made models using relevant and quality-assured databases. Given the scale of undertaking these approaches, international collaboration and coordination will be essential.

Once a better understanding has been obtained concerning the performances of various software tools in predicting toxicological and ADME endpoints relevant to food safety assessment, international agencies such as EFSA should consider developing their own criteria for acceptance of model estimates according to specific regulatory applications. For example, in assessing the toxicological relevance of pesticide residues for possible inclusion in the residue definition, it is desirable to minimise the number of false negative predictions for critical endpoints such as genotoxicity. In order to establish the usefulness of different models in this context, it is necessary to set clear criteria for the generation of false negatives, as well as sensitivity, and perhaps also in terms of percentage coverage of the pesticides (PPP) inventory. Models or combinations of models can then be assessed and optimised against these criteria. For example, a battery of models could be developed, although it is by no means guaranteed *a priori* that a given set of criteria will be achievable.

Recommendations for further activities, both in the short-term and long-term, are summarised in Table 8.2.

## Table 8.1. Summary of the status of software models and observations on their potential applicability

| Endpoint | Comments |
|---|---|
| **Acute oral toxicity** | A range of models are available for estimating rodent LD50 balues or for classifying chemical on the basis of acute toxicity. Literature models tend to be local models with limited applicability domains.<br>Freely available and user-friendly tools, suitable for the non-specialist user, are almost completely lacking: an online ToxBoxes application provides LD50 estimations; ChemBench web portal is in development.<br>Commercially available and user-friendly tools, suitable for the non-specialist user, include: ACD/Tox Suite, MCASE/MC4PC, MDL QSAR, TerraQSAR and TOPKAT. |
| **Repeat dose oral toxicity** | Very few models are available.<br>User-friendly commercial tools include TOPKAT and MolCode Toolbox, which predict chronic (12 month or more) LOAELs.<br>User-friendly and freely available tools include Lazar, which predicts MRTDs in humans; and Toxtree, which predicts Cramer classifications, reflecting the level of concern based on oral toxicity. |
| **Organ and system-specific toxicities** | No software was identified as potentially useful for routine assessment purposes. The development of such tools would provide a useful means of supplementing the use of models for apical effects.<br>A few models are available for effects such as hepatotoxicity, nephrotoxicity and (developmental) neurotoxicity; these are restricted to literature models with limited applicability domains.<br>Several software packages (CASE/MC4PC, MDL-QSAR, BioEpisteme, and Leadscope Predictive Data Miner) have been reported as useful research tools in the development of models for hepatic and urinary tract toxicities. |
| **Genotoxicity** | A vast range of models are available, mostly for chemicals that are electrophilic and DNA-reactive. The majority of models predict Ames mutagenicity. Many models predict genotoxic effects in general withour reference to a specific endpoint.<br>Freely available and user-friendly tools, suitable for the non-specialist user, are limited: online ToxBoxes application, Lazar, Toxtree (Benigni-Bossa and *In vivo* Micronucleus), CAESAR, OECD QSAR Toolbox (DNA-binding profilers).<br>Based on the research investigation, various gentoxicity models are considered sufficiently predictive for the identification of potential genotoxic carcinogens in the context of a TTC assessment. |

| Endpoint | Comments |
|---|---|
| **Carcinogenicity** | A vast range of classification models are available for predicting carcinogenic potential, mostly in rodents. Models for predicting carcinogenic potency are lacking.<br>Various compilations of structural alerts have been found useful (e.g. Ashby, Bailey, Kazius, Benigni-Bossa)<br>User-friendly and freely available tools, suitable for the non-specialist user, are limited: Lazar, Oncologic, Toxtree.<br>User-friendly commercial tools include: Derek, TOPKAT, MolCode Toolbox, HazardExpert, OASIS TIMES.<br>Based on the research investigation, current carcinogenicity models are not considered sufficiently predictive for use in isolation. However, they may have some value when used to supplement the use of genotoxicity models in the context of a TTC assessment.<br>Several software packages have been reported as useful research tools for model building: CASE/MC4PC, MDL-QSAR. |
| **Developmental and reproductive toxicity** | Very few models are available, and these are mostly classification models. Literature models tend to be local models with limited applicability domains.<br>User-friendly and freely available tools, suitable for the non-specialist user, are almost non-existent: CAESAR (developmental toxicity)<br>User-friendly and commercial software tools, suitable for the non-specialist end-user, include: TOPKAT (developmental toxicity), Derek, Leadscope.<br>Several software packages have been reported as useful research tools for model building: CASE/MC4PC, Toxmatch |
| **Endocrine activity** | Many models are available for predicting nuclear hormone receptor binding or receptor-mediated effects, especially for the oestrogen, androgen and aryl hormone receptors. In many cases, the models are the results of research investigations and not suitable for routine use.<br>A range of user-friendly software tools, suitable for the non-specialist end-user, are commercially available, including: ADMET Predictor, ToxBoxes, MolCode Toolbox, TerraQSAR, TIMES and VirtualToxLab.<br>Simple and freely available decision tree approaches have been found useful for screening and priority setting, especially the US EPA decision tree which is implemented in the OECD QSAR Toolbox.<br>Important challenges are to develop an understanding of how to use these model results for regulatory purposes, and how to integrate metabolic information into the assessment. |

| Endpoint | Comments |
|---|---|
| **Physicochemical properties useful in ADME prediction** | A wide range of software tools are available for predicting key physicochemical properties such as partition coefficients (logP), distribution coefficients (logD), ionisation constants (pKa) and solubility; these are generally perceived to be highly reliable within their applicability domains.<br>User-friendly and freely available and tools, suitable for the non-specialist user, include: EpiSuite, SPARC, VCCLab and ADMEBoxes online application.<br>Commercially available and user-friendly tools, suitable for the non-specialist user, include: Accord for Excel with ADME/Tox Add-on , ACD ADMEBoxes, |
| **ADME properties** | A vast and rapidly growing range of models are available, although in many cases they have been developed for pharmaceuticals., which raises questions about their applicability to other types of chemicals. Literature models tend to be local models with limited applicability domains.<br>A wide range of research tools are available (e.g, for modelling enzyme interactions and metabolism), which are not suitable for the routine assessment of chemicals by non-specialists.<br>For routine application in the assessment of chemicals, a number of simple rules-of-thumb have been found useful for approximate estimations and screening purposes.<br>User-friendly and freely available tools, suitable for the non-specialist user, are almost completely lacking: an online version of ADMEBoxes provides bioavailability estimation, volume of distribution and P-gp binding.<br>User-friendly and commercially available tools, suitable for the non-specialist user, include: ACD/ADME Suite, MetabolExpert, Meteor, Accord for Excel with ADME/Tox Add-on, Symcyp. |

The comments in this table are based on the outcome of the literature review and on the experience of the authors. It is not intended to provide complete and definitive guidance on the applicability of available models for specific purposes. A reasonable amount of QSAR expertise is always required to interpret model predictions.

## Table 8.2. Recommendations for further activities to promote the regulatory use of QSARs in the food safety field

| Short term (1-3 years) | Long-term (>3 years) |
|---|---|
| | |
| **General** | |
| 1) Need to investigate the applicability (predictivity and scope) of different software tools on an endpoint-by-endpoint basis<br>2) Need to explore the advantanges of combining the use of multiple tools in model batteries and Integrated Testing Strategies<br>3) Need to take policy decisions on how much information is needed to support the regulatory use of models<br>4) Need to establish criteria for model acceptability according to the regulatory purpose and context (e.g. mimising false negatives in the identification of genotoxicants in the context of a TTC scheme)<br>5) Need to investigate and develop detailed guidance on how to use the outputs of models for defined regulatory purposes<br>6) Need for training on how to use software tools, and interpret their outputs in a regulatory context<br>7) Need to clarify and refine the Cramer classification scheme | 1) The development of publicly accessible, structured and searchable databases will be vital for further model development and validation<br>2) Focussed research and development activities are needed to implement potentially useful literature models in the form of software tools. Ideally, a sufficient range of models will be made freely available<br>3) Agencies such as EFSA could consider developing purpose-built models based on their own databases, e.g. pesticides<br>4) Opportunity to replace the Cramer classification scheme with a new TTX assessment scheme incorporating the latest scientific developments in *in silico* and *in vitro* toxicology |
| **Prediction of acute oral toxicity** | |
| 1) Need to further explore the combined use of structural descriptors and *in vitro* data via quantitative structure-activity-activity analysis (QSAAR)<br>2) Need to increase availability of models in the public domain. Several promising model-building methodologies have been identified. | |

| Short term (1-3 years) | Long-term (>3 years) |
|---|---|
| | |
| **Prediction of repeat-dose toxicity** | |
| 1) Need to compare the applicability (predictivity and scope) of TOPKAT and MolCode Toolbox. | 1) Considerable scope for new model development, but this will depend on increased availability of high quality data. Given the biological complexity of the endpoint, systems biology approaches could be worth pursuing in addition to traditional QSAR analysis. |
| **Prediction of organ and system-specific toxicities** | |
| 1) Need to build public databases suitable for modelling organ and system toxicities | 1) Considerable scope for new model development, but this will depend on increased availability of high quality data.<br>2) Models could be integrated with models for acute and chronic apical effects. |
| **Prediction of genotoxicity and carcinogenicity** | |
| 1) Need to add or refine rules in current expert systems in order to reduce false positive predictions of genotoxicity and carcinogenicity<br>2) Hybrid approaches, based on both statistical algorithms and mechanistic knowledge, represent a promising way forward in model development.<br>3) Need to build models for specific endpoints rather than genotoxicity or carcinogenicity in general.<br>4) Need to explore alternative ways of interpreting model predictions with a view to optimising positive or negative predictivity.<br>5) Need to further explore use of model batteries. | 1) Need to incorporate ADME into current models / expert systems.<br>2) Need to increase availability of models for non-genotoxic carcinogenicity. |

| Short term (1-3 years) | Long-term (>3 years) |
|---|---|
| | |
| **Prediction of developmental and reproductive toxicity** | |
| 1) Need to develop ublic databases suitable for modelling reprotoxic effects (e.g. ILSI initiative) | 1) Need to expand current ontolgies linking adverse effects with underlying changes at the molecular and cellular and tissue levels<br>2) Considerable scope for new model development, but this will depend on increased availability of high quality data. Need to build models based on point estimates (e.g. LOAELs) as well as classification models. Given the biological complexity of the endpoint, systems biology approaches could be worth pursuing in addition to traditional QSAR analysis. |
| **Prediction of endocrine activity** | |
| 1) Need to develop models for a wider range receptors (other than ER, AR, ArH)<br>2) Need to better understand how to use data from receptor models | 1) Need to incorporate ADME considerations into model |

| Short term (1-3 years) | Long-term (>3 years) |
|---|---|
| | |
| **Prediction of ADME properties** | |
| 1) Need to increase availability of models in the public domain. A range of promising model-building methodologies have been identified.<br>2) Need to assess available software tools in terms of their applicability to chemicals other than pharmacuticals | 1) A wide range of ADME databases are commercially available. Substantial efforts are needed to bring more information into the public domain.<br>2) Traditional QSAR approaches should be supplemented by mathematically and physiologically based models - PBBK models. |

# 9. REFERENCES

## Chapter 1 references


AGES (2010). Impact of metabolic and degradation processes on the toxicological properties of residues of pesticides in food commodities. Report from the Austrian Austrian Agency for Health and Food Safety (AGES) to the European Food Safety Authority (EFSA). Available at: http://www.efsa.europa.eu/en/scdocs/scdoc/49e.htm

Cronin M (2010). Quantitative structure-activity relationships (QSARs) – applications and methodology. Chapter 10 in *Recent Advances in QSAR Studies: Methods and Applications.* T Puzyn, J Leszczynski & MTD Cronin (Eds). Springer, Heidelberg, Germany, pp. 3-11.

CRD (2010) Applicability of thresholds of toxicological concern in the dietary risk assessment of metabolites, degradation and reaction products of pesticides. Report from the UK Chemicals Regulation Directorate (CRD) to the European Food Safety Authority (EFSA). Available at: http://www.efsa.europa.eu/en/scdocs/scdoc/44e.htm

Dearden JC, Cronin MTD & Kaiser KLE (2009). How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR and QSAR in Environmental Research 20, 241-266.

EC (1991). Council Directive 91/414/EEC of 15 July 1991 concerning the placing of plant protection products on the market. Official Journal of the European Union, L 230/1 of 19.08.1991. Office for Official Publications of the European Communities (OPOCE), Luxembourg.

EC (2002). Regulation (EC) No 178/2008 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal L 031, 1-24. 1 February 2002.

EC (2005). Regulation (EC) No 396/2005 of the European Parliament and of the Council of 23 February 2005 on maximum residue levels of pesticides in or on food and feed of plant and animal origin and amending Council Directive 91/414/EEC. Official Journal L070, 1-16. 16 March 2005.

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

Enoch S (2010). Chemical category formation and read-across for the prediction of toxicity. Chapter 10 in *Recent Advances in QSAR Studies: Methods and Applications.* T Puzyn, J Leszczynski & MTD Cronin (Eds). Springer, Heidelberg, Germany, pp. 209-219.

JRC (2010). Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Report from the European Commission's Joint Research Centre (JRC) to the European Food Safety Authority (EFSA). Available at: http://www.efsa.europa.eu/en/scdocs/scdoc/50e.htm

Lapenna S, Fuart-Gatnik M & Worth A (2010). Review of QSAR Models and Software Tools for predicting Acute and Chronic Systemic Toxicity. JRC Technical Report EUR 24639 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

Lo Piparo E & Worth A (2010). Review of QSAR Models and Software Tools for predicting Developmental and Reproductive Toxicity. JRC Technical Report EUR 24522 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

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.

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

OECD (2007). Guidance on Grouping of Chemicals. Series on Testing and Assessment Number 80. ENV/JM/MONO(2007)28. Organisation for Economic Cooperation and Development, Paris, France. Available at: http://www.oecd.org/

OECD (2009). Guidance Document on the Definition of Residue. Series on Testing and Assessment Number 63, Series on Pesticides Number 31. Organisation for Economic Cooperation and Development, Paris, France. Available at: http://www.oecd.org/

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 struc 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.

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

Serafimova R, Fuart Gatnik M & Worth A (2010). Review of QSAR Models and Software Tools for predicting Genotoxicity and Carcinogenicity. JRC Technical Report EUR 24427 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

Todeschini R & Consonni V (2000). Handbook of Molecular Descriptors. In *Series of Methods and Principles in Medicinal Chemistry* (Mannhold R., Kubinyi H, Timmerman H, Eds), Volume 11, Wiley-VCH.

Todeschini R & Consonni V (2009). Molecular Descriptors for Chemoinformatics. In *Series of Methods and Principles in Medicinal Chemistry* (Mannhold R., Kubinyi H, Timmerman H, Eds), Volume 41, Wiley-VCH.

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

Worth A, Lapenna S, Lo Piparo E, Mostrag-Szlichtyng A & Serafimova R (2010). The Applicability of Software Tools for Genotoxicity and Carcinogenicity Prediction: Case Studies relevant to the Assessment of Pesticides. JRC Technical Report EUR 24640 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

Worth A, Lapenna S, Lo Piparo E, Mostrag-Szlichtyng A & Serafimova R (2011). A Framework for assessing *in silico* Toxicity Predictions: Case Studies with selected Pesticides. JRC Technical Report EUR 24705 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

## Chapter 2 references

AFSCA (2009). Comitè Scientifique de l'Agence Fédérale pour la Sécurité de la Chaine Alimentaire (conseil urgent 2009). Migration de 4-méthylbenzophénone de l'emballage en carton imprimé vers les céréales de petit déjeuner. Available at : http://www.favv-afsca.fgov.be/comitescientifique/avis/_documents/CONSEILurgent_05-2009_FR_DOSSIER2009-05.pdf

Arvidson KB Chanderbhan R, Muldoon-Jacobs K, Mayer J & Ogungbesan A (2010). Regulatory use of computational toxicology tools and databases at the United States Food and Drug Administration's Office of Food Additive Safety. Expert Opinion on Drug Metabolism & Toxicology 6, 793-796.

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.

Benfenati E, Benigni R, Demarini DM, Helma C, Kirkland D, Martin TM, Mazzatorta P, Ouédraogo-Arras G, Richard AM, Schilter B, Schoonen WG, Snyder RD & Yang C. (2009). Predictive models for carcinogenicity and mutagenicity: frameworks, state-of-the-art, and perspectives. Journal of Environmental Science and Health, Part C Environmental Carcinogenesis and Ecotoxicology Reviews 27, 57-90.

Contrera JF, Matthews EJ, Kruhlak NL & Benz RD (2004). Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modelling of the human maximum recommended daily dose. Regulatory Toxicology and Pharmacology 40, 185-2006.

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.

FDA/OFAS (2002). Guidance for Industry –Preparation of Notifications for Substamces: Toxicology Recommendations. Available at: http://fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/ucm081825.htm

Jensen GE, Niemelä JR, Wedebye EB & Nikolov NG (2008). QSAR models for reproductive toxicity and endocrine disruption in regulatory use – a preliminary investigation. SAR and QSAR in Environmental Research 19(7), 631 - 641.

JRC (2010). Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Report from the European Commission's Joint Research Centre (JRC) to the European Food Safety Authority (EFSA). Available \t: http://www.efsa.europa.eu/en/scdocs/scdoc/50e.htm

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

Kruhlak NL, Contrera JF, Benz RD & Matthews EJ (2007). Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. Advanced Drug Delivery Reviews **59**(1), 43-55

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, Benz RD & Contrera JF (2004). Assessment of the health effects of chemicals in humans: I. QSAR estimation of the maximum recommended therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data. Current Drug Discovery Technologies 1(1), 61-76.

Matthews EJ, Kruhlak NL, Benz RD & Contrera JF (2007a). A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. Regulatory Toxicology and Pharmacology 47(2), 115-135.

Matthews EJ, Kruhlak NL, Benz RD, Ivanov J, Klopman G & Contrera JF (2007b). A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. Regulatory Toxicology and Pharmacology 47(2), 136-155.

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.Mazzatorta P, Dominguez Estevez M, Coulet M & Schilter B (2008). Modelling oral rat chronic toxicity. Journal of Chemical Information and Modeling 48, 1949-1954.

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, 34-38.

Mazzatorta P, Dominguez Estevez M, Coulet M & Schilter B (2008). Modelling oral rat chronic toxicity. Journal of Chemical Information and Modeling 48, 1949-1954.

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

OECD (2006). Guidance Document on the Definition of Residue. Series on Texting and Assessment No. 63/Series on Pesticides no.31. 10 October 2006. Available at: http://appli1.oecd.org/olis/2006doc.nsf/linkto/env-jm-mono(2006)31

OECD (2009). Guidance Document on the Definition of Residue. Series on Testing and Assessment No. 63 and Series on Pesticides no.31. 28 July 2009. Available at: http://www.oecd.org/

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.

Tilaoui L, Schilter B, Tran L-A, Mazzatorta P & Grigorov M (2007). Integrated Computational Methods for Prediction of the Lowest Observable Adverse Effect Level of Food-Borne Molecules. QSAR & Combinatorial Science 26, 102-108.

Tong W, Xie Q, Hong H, Shi L, Fang H & Perkins R (2004). Assessment of Prediction Confidence and Domain Extrapolation of Two Structure–Activity Relationship Models for Predicting Estrogen Receptor Binding Activity. Environmental Health Perspectives 112, 1249-1254.

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.

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, 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

Yang C, Valerio LG Jr & Arvidson KB (2009). Computational toxicology approaches at the US Food and Drug Administration. Alternatives to Laboratory Animals 37(5), 523-531.

## Chapter 3 references

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

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

Enoch SJ, Madden JC & Cronin MT (2008). Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. SAR and QSAR in Environmental Research 19(5-6), 555-578.

Fuart-Gatnik M & Worth AP (2010). Review of Software Tools for Toxicity Prediction. JRC Technical Report EUR 24489 EN. Publications Office of the European Union. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

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

Klopman G & Rosenkranz H (1994). Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTI-CASE, Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis 305(1), 33-46.

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG & Würtzen G (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food & Chemical Toxicology 42(1), 65-83.

Maunz A & Helma C (2008). Prediction of chemical toxicity with local support vector regression and activity-specific kernels. SAR and QSAR in Environmental Research **19**(5-6), 413-431.

Rydberg P Gloriam DE & Olsen L (2010). The SMARTCyp cytochrome P450 metabolism prediction server. Bioinformatics 26, 2988-2989.

## Chapter 4 references

Barlow S (2005). Threshold of toxicological concern (TTC) - A tool for assessing substances of unknown toxicity present at low levels in the diet. ILSI Europe's Concise Monographs and Report, 34.

Bitsch A, Jacobi S, Melber C, Wahnschaffe U, Simetska N & Mangelsdorf I (2006). REPDOSE: A database on repeated dose toxicity studies of commercial chemicals - A multifunctional tool. Regulatory Toxicology and Pharmacology **46**(3), 202-210.

Blackburn K, Stickney JA, Carlson-Lynch HL, McGinnis PM, Chappell L & Felter SP (2005). Application of the threshold of toxicological concern approach to ingredients in personal and household care products. Regulatory Toxicology and Pharmacology **43**(3), 249-259.

Cheeseman MA, Machuga EJ & Bailey AB (1999). A tiered approach to threshold of regulation. Food and Chemical Toxicology 37, 387-412.

Cramer GM, Ford RA & Hall RL (1978). Estimation of toxic hazard - a decision tree approach. Food and Cosmetics Toxicology 16, 255–276.

CRD (2010) Applicability of thresholds of toxicological concern in the dietary risk assessment of metabolites, degradation and reaction products of pesticides. Report from the UK Chemicals Regulation Directorate (CRD) to the European Food Safety Authority (EFSA). Available at: http://www.efsa.europa.eu/en/scdocs/scdoc/44e.htm

Devillers J & Devillers H (2009). Prediction of acute mammalian toxicity from QSARs and interspecies correlations, SAR & QSAR in Environmental Research 20, 467-500.

FDA (2008). Guidance for Industry. Genotoxic and Carcinogenic Impurities in Drug Substances and Products: Recommended Approaches. Draft document: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm079235.pdf

Felter S, Lane RW, Latulippe ME, Llewellyn GC, Olin SS, Scimeca JA & Trautman TD (2009). Refining the threshold of toxicological concern (TTC) for risk prioritization of trace chemicals in food. Food & Chemical Toxicology 47(9), 2236-2245.

Frawley JP (1967). Scientific evidence and common sense as a basis for food packaging regulations. Food and Cosmetics Toxicology 5, 293-308.

Kinsner-Ovaskainen A, Rzepka R, Rudowski R, Coecke S, Cole T & Prieto P (2009). Acutoxbase, an innovative database for in vitro acute toxicity studies. Toxicology in Vitro 23(3), 476-485.

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG & Würtzen G (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food & Chemical Toxicology 42(1), 65-83.

Lapenna S, Fuart-Gatnik M & Worth A (2010). Review of QSAR Models and Software Tools for predicting Acute and Chronic Systemic Toxicity. JRC Technical Report EUR 24639 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

Lessigiarska I, Worth AP, Netzeva TI, Dearden JC & Cronin MTD (2006). Quantitative structure-activity-activity and quantitative structure-activity investigations of human and rodent toxicity. Chemosphere 65(10), 1878-1887.

Matthews EJ, Ursem CJ, Kruhlak NL, Benz RD, Sabaté DA, Yang C, Klopman G & Contrera JF (2009b). Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: Part B. Use of (Q)SAR systems for early detection of drug-induced hepatobiliary and urinary tract toxicities. Regulatory Toxicology and Pharmacology 54(1), 23-42.

Matthews EJ, Kruhlak NL, Weaver JL, Benz RD & Contrera JF (2004b). Assessment of the health effects of chemicals in humans: II. Construction of an adverse effects database for QSAR modeling. Current Drug Discovery Technologies **1**(4), 243-254.

Mazzatorta P, Estevez MD, Coulet M & Schilter B (2008). Modeling oral rat chronic toxicity. Journal of Chemical Information and Modeling **48**(10), 1949-1954.

Mumtaz MM, Knauf LA, Reisman DJ, Peirano WB, DeRosa CT, Gombar VK, Enslein K, Carter JR, Blake BW, Huque KI & *et al*. (1995). Assessment of effect levels of chemicals from quantitative structure-activity relationship (QSAR) models. I. Chronic lowest-observed-adverse-effect level (LOAEL). Toxicology Letters **79**(1-3), 131-143.

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

Raevsky OA, Grigor'ev VJ, Modina AE & Worth AP (2010). Prediction of acute toxicity to mice by the arithmetic mean toxicity (AMT) modelling approach. SAR and QSAR in Environmental Research 21(1), 265-275.

Rulis AM (1992). Threshold of Regulation: Options for handling minimal risk situations. In: Food Safety Assessment, Chapter 14, Eds Finlay JW, Robinson SF, Armstrong DJ American Chemical Society Symposium Series No 484. American Chemical Society, Washington DC. pp 132-139.

Phillips JC, Purchase R, Watts P & Gangolli SD (1987). An evaluation of the decision tree approach for assessing priorities for safety testing of food additives. Food Additives & Contaminants 4, 109- 123.

Safford RJ (2008). The Dermal Sensitisation Threshold - A TTC approach for allergic contact dermatitis. Regulatory Toxicology and Pharmacology 51, 195-200.

Tilaoui L, Schilter B, Tran L-A, Mazzatorta P & Grigorov M (2007). Integrated Computational Methods for Prediction of the Lowest Observable Adverse Effect Level of Food-Borne Molecules. QSAR & Combinatorial Science 26, 102-108.

Tsakovska I, Lessigiarska I, Netzeva T & Worth Andrew P (2008). A Mini Review of Mammalian Toxicity (Q)SAR Models. QSAR & Combinatorial Science 27(1), 41-48.

Tunkel J, Mayo K, Austin C, Hickerson A & Howard P (2005). Practical Considerations on the Use of Predictive Models for Regulatory Purposes. Environmental Science & Technology 39(7), 2188-2199.

Venkatapathy R, Moudgal CJ & Bruce RM (2004). Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. Journal of Chemical Information and Computer Sciences **44**(5), 1623-1629.

Zhu H, Martin TM, Ye L, Sedykh A, Young DM & Tropsha A (2009a). Quantitative Structure−Activity Relationship Modeling of Rat Acute Toxicity by Oral Exposure. Chemical Research in Toxicology 22(12), 1913–1921.

Zhu H, Ye L, Richard A, Golbraikh A, Wright FA, Rusyn I & Tropsha A (2009b). A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents. Environmental Health Perspectives 117(8), 1257-1264.

## Chapter 5 references

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

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.

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

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). 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 (2008a). Predictivity of QSAR. Journal of Chemical Information and Modeling 48, 971-980.

Benigni R & Bossa C (2008b). 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. JRC Technical Report EUR 23241 EN. Publications Office of the European Union, Luxembourg. Available at: 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. JRC Technical Report EUR 23844 EN. Publications Office of the European Union, Luxembourg. Available at: 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 25,335-341.

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

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.

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.

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 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.

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.

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.

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

Enoch SJ & Cronin MTD (2010). A review of the electrophilic reaction chemistry involved in covalent DNA binding. Critical Reviews in Toxicology 40, 728-748.

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.

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.

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.

Hansen K, Mika S, Schroeter T, Sutter A, Laak At, Steger-Hartmann T, Heinrich N & Muller 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.

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.

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.

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.

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

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

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, 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

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.

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.

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.

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.

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

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

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.

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.

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.

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.

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.

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.

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.

Serafimova R, Fuart Gatnik M & Worth A (2010). Review of QSAR Models and Software Tools for predicting Genotoxicity and Carcinogenicity. JRC Technical Report EUR 24427 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

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 (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 & 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.

Tennant 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.

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.

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

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.

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

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.

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

## Chapter 6 references


Anstead GM, Carlson KE & Katzenellenbogen JA (1997). The estradiol pharmacophore: Ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. Steroids 62(3), 268-303.

Arena VC, Sussman NB, Mazumdar S, Yu S & Macina OT (2004). The Utility of Structure–Activity Relationship (SAR) Models for Prediction and Covariate Selection in Developmental Toxicity: Comparative Analysis of Logistic Regression and Decision Tree Models. SAR and QSAR in Environmental Research 15(1), 1 - 18.

Benfenati E, Roncaglioni A, Boriani E, Porcelli C, Spreafico M & Lo Piparo E (2005). Validation of selected, non-commercial (Q)SAR models for ER and AR binding. Final report of JRC contract CCR.IHCP.C430414.X0.

Cramer RD, Patterson DE & Bunce JD (2002). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society 110(18), 5959-5967.

Cronin MTD & Worth AP (2008). (Q)SARs for Predicting Effects Relating to Reproductive Toxicity. QSAR & Combinatorial Science 27(1), 91-100.

Cruciani G, Crivori P, Carrupt PA & Testa B (2000). Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. Journal of Molecular Structure: THEOCHEM 503(1-2), 17-30.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW & Kavlock RJ (2007). The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. Toxicological Sciences 95(1), 5-12.

Enoch SJ, Cronin MTD, Madden JC & Hewitt M (2009). Formation of Structural Categories to Allow for Read-Across for Teratogenicity. QSAR & Combinatorial Science 28(6-7), 696-708.

Enslein K, Lander TR & Strange JR (1983). Teratogenesis: a statistical structure-activity model. Teratogenesis, Carcinogenesis, and Mutagenesis 3(3), 289-309.

Fabjan E, Hulzebos E, Mennes W & Piersma AH (2006). A Category Approach for Reproductive Effects of Phthalates. Critical Reviews in Toxicology 36(9), 695-726.

Filimonov D & Poroikov V (2008). Probabilistic Approaches in Activity Prediction. In: Chemoinformatics Approaches to Virtual Screening (Alexandre Varnek and Alexander Tropsha eds.). Cambridge, RSC Publishing, 182-216.

Filimonov D, Poroikov V, Borodina Yu & Gloriozova T (1999). Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. Journal of Chemical Information and Computer Sciences 39(4), 666-670.

Gallegos Saliner A, Netzeva TI & Worth AP (2006). Prediction of Estrogenicity: Validation of a Classification Model. SAR and QSAR in Environmental Research 17, 195-223.

Gombar VK, Enslein K & Blake BW (1995). Assessment of developmental toxicity potential of chemicals by quantitative structure-toxicity relationship models. Chemosphere 31(1), 2499-2510.

Hewitt M, Madden JC, Rowe PH & Cronin MTD (2007). Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier. SAR and QSAR in Environmental Research 18(1), 57 - 76.

Hewitt M, Ellison CM, Enoch SJ, Madden JC & Cronin MTD (2010). Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. Reproductive Toxicology 30(1), 147-160.

Hulzebos EM & Posthumus R (2003). (Q)SARs: gatekeepers against risk on chemicals? SAR and QSAR in Environmental Research 14(4), 285-316.

Jacobs MN, Janssens W, Bernauer U, Brandon E, Coecke S, Combes R, Edwards P, Freidig A, Freyberger A, Kolanczyk R, Mc Ardle C, Mekenyan O, Schmieder P, Schrader T, Takeyoshi M & van der Burg B (2008). The use of metabolising systems for in vitro testing of endocrine disruptors. Current Drug Metabolism 9(8), 796-826.

Jensen GE, Niemelä JR, Wedebye EB & Nikolov NG (2008). QSAR models for reproductive toxicity and endocrine disruption in regulatory use – a preliminary investigation. SAR and QSAR in Environmental Research 19(7), 631 - 641.

Julien E, Willhite CC, Richard AM, DeSesso JM & Group IRW (2004). Challenges in constructing statistically based structure-activity relationship models for developmental toxicity. Birth Defects Research Part A: Clinical and Molecular Teratology 70(12), 902-911.

Knudsen TB, Martin MT, Kavlock RJ, Judson RS, Dix DJ & Singh AV (2009). Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB. Reproductive Toxicology 28(2), 209-219.

Liu X, Ouyang S, Yu B, Huang K, Liu Y, Gong J, Zheng S, Li Z, Li H & Jiang H (2010). PharmMapper Server: a web server for potential drug target identification via pharmacophore mapping approach. Nucleic Acids Research 38, W609-W614.

Lo Piparo E & Worth A (2010). Review of QSAR Models and Software Tools for predicting Developmental and Reproductive Toxicity. JRC Technical Report EUR 24522 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

Martin MT, Mendez E, Corum DG, Judson RS, Kavlock RJ, Rotroff DM & Dix DJ (2009). Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database. Toxicological Sciences 110(1), 181-190.

Maslankiewicz L HE, Vermeire TG, Muller JJA, Piersma AH (2005). Can chemical structure predict reproductive toxicity? Available at: http://hdl.handle.net/10029/7374

Matthews EJ, Kruhlak NL, Daniel Benz R & Contrera JF (2007a). A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. Regulatory Toxicology and Pharmacology 47(2), 115-135.

Matthews EJ, Kruhlak NL, Daniel Benz R, Ivanov J, Klopman G & Contrera JF (2007b). A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. Regulatory Toxicology and Pharmacology 47(2), 136-155.

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

Mekenyan O, Nikolova N & Schmieder P (2003b). Dynamic 3D QSAR techniques: applications in toxicology Journal of Molecular Structure: THEOCHEM 622, 147-165.

Netzeva TI, Gallegos Saliner A & Worth AP (2006). Comparison of the Applicability Domain of a QSAR for Estrogenicity with a Large Chemical Inventory. Environmental Toxicology & Chemistry 25, 1223-1230.

OECD (2009). (Quantitative) Structure Activity Relationships [(Q)SARs] of Estrogen binding affinity to support prioritization of heterogenous chemicals within defined inventories for screening and testing. Schmieder P, Kolanczyk R, Aladjov H, Hornung M, Tapper M, Denny J & Sheedy B. Organisation for Economic Cooperation and Development. Available at: http://www.olis.oecd.org/olis/2009doc.nsf/LinkTo/NT00004CD6/$FILE/JT03268747.PDF

Pastor M, Cruciani G, McLay I, Pickett S & Clementi S (2000). GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. Journal of Medicinal Chemistry 43(17), 3233-3243.

Paumgartten F, Solecki R, Buschmann J, Clark R, Grote K, Rauch M & Chahoud I (2009). Harmonization of terminology in developmental toxicology: The quest for a more precise description and a harmonized classification of fetal observations. Reproductive Toxicology 27(1), 8-13.

Pearl GM, Livingston-Carr S, Durham S (2001). Integration of computational analysis as a sentinel tool in toxicological assessments. Current Topics in Medicinal Chemistry 1, 247-255.

Poroikov V., Filimonov D., Lagunin A., Gloriozova T & Zakharov A (2007). PASS: identification of probable targets and mechanisms of toxicity. SAR and QSAR in Environmental Research 18(1), 101-110.

Pike ACW, Brzozowski AM, Hubbard RE, Bonn T, Thorsell A-G, Engstrom O, Ljunggren J, Gustafsson J-A & Carlquist M (1999). Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. EMBO Journal 18(17), 4608-4618.

Schmieder P, Mekenyan O, Bradbury S & Veith G (2003). QSAR prioritization of chemical inventories for endocrine disruptor testing. Pure and Applied Chemistry 75, 2389-2396.

Vedani A, Dobler M & Lill MA (2005). Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor†. Journal of Medicinal Chemistry 48(11), 3700-3703.

Vedani A & Smiesko M (2009). In silico toxicology in drug discovery - concepts based on three-dimensional models. Alternatives to Laboratory Animals 37(5), 477-496.

Vedani A, Smiesko M, Spreafico M, Peristera O & Dobler M (2009). VirtualToxLab - in silico prediction of the toxic (endocrine-disrupting) potential of drugs, chemicals and natural products. Two years and 2,000 compounds of experience: a progress report. ALTEX 26(3), 167-176.

## Chapter 7 references

Abraham MH, Ibrahim A, Zhao Y & Acree WE Jr (2006). A data base for partition of volatile organic compounds and drugs from blood/plasma/serum to brain, and an LFER analysis of the data. Journal of Pharmaceutical Sciences 95(10), 2091-2100.

Al-Fahemi JHA, Cooper DL & Allan NL (2007). Investigating the utility of momentum-space descriptors for predicting blood-brain barrier penetration. Journal of Molecular Graphics and Modelling 26(3), 607-612.

Allen DD & Geldenhuys WJ (2006). Molecular modeling of blood-brain barrier nutrient transporters: In silico basis for evaluation of potential drug delivery to the central nervous system. Life Sciences 78(10), 1029-1033.

Chohan KK, Paine SW & Waters NJ (2006). Quantitative Structure Activity Relationships in Drug Metabolism. Current Topics in Medicinal Chemistry 6, 1569-1578.

Chohan KK, Paine SW & Waters NJ (2008). Advancements in predictive in silico models for ADME. Current Chemical Biology 2, 215-228.

Cianchetta G, Yi L, Singleton R, Zhang M,Wildgoose M, Rampe D, Kang J & Vaz RJ (2006). Molecular Interaction Fields in ADME and safety, in: Molecular Interaction Fields (G. Cruciani ed). Wiley-VCH, Weinheim, pp. 197-218.

Clark DE (2007). In silico ADMET tools: a dawn of a new generation? Expert Opinion on Drug Discovery 2, 1423-1429.

Colmenarejo G (2005). In Silico ADME Prediction: Data Sets and Models. Current Computer-Aided Drug Design 1, 365-376.

Crivori P & Poggesi I (2006). Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs. European Journal of Medicinal Chemistry 41(7), 795-808.

Cross S & Cruciani G (2010). Molecular fields in drug discovery: getting old or reaching maturity? Drug Discovery Today 15(1-2), 23-32.

Cruciani G , Milletti F, Storchi L, Sforna G & Goracci L (2009). *In silico* pKa prediction and ADME profiling. Chemistry & Biodiversity 6(11), 1812-1821.

Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T & Vianello R (2005). MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. Journal of Medicinal Chemistry 48(22), 6970-6979.

Cruciani G, Pastor M & Guba W (2000). VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. European Journal of Pharmaceutical Sciences 11(Supplement 2), S29-S39.

Dearden J & Worth A (2007). In Silico Prediction of Physicochemical Properties. JRC Technical Report EUR 23051 EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

Dearden JC (2007). In silico prediction of ADMET properties: how far have we come? Expert Opinion on Drug Metabolism & Toxicology 3(5), 635-639.

de Graaf C, Vermeulen NPE & Feenstra KA (2005). Cytochrome P450 in Silico: An Integrative Modeling Approach. Journal of Medicinal Chemistry 48(8), 2725-2755.

Delisle RK, Lowrie JF, Hobbs DW & Diller DJ (2005). Computational ADME/Tox modeling: aiding understanding and enhancing decision making in drug design. Current Computer-Aided Drug Design 1, 325-345.

Ekins S, Andreyev S, Ryabov A, Kirillov E, Rakhmatulin EA, Bugrim A & Nikolskaya T (2005). Computational prediction of human drug metabolism. Expert Opinion on Drug Metabolism & Toxicology 1(2), 303-324.

Ekins S, Nikolsky Y & Nikolskaya T (2005). Techniques: Application of systems biology to absorption, distribution, metabolism, excretion and toxicity. Trends in Pharmacological Sciences 26(4), 202-209.

Ekins S & Williams AJ (2010). Precompetitive preclinical ADME/Tox data: set it free on the web to facilitate computational model building and assist drug development. Lab on a Chip.

Ekins S, Mestres J & Testa B (2007a). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. British Journal of Pharmacology 152(1), 9-20.

Ekins S, Mestres J & Testa B (2007b). In silico pharmacology for drug discovery: applications to targets and beyond. British Journal of Pharmacology 152(1), 21-37.

Ekins S, Embrechts MJ, Breneman CM, Jim K & Wery J-P (2007c). Novel Applications of Kernel-Partial Least Squares to Modeling a Comprehensive Array of Properties for Drug Discovery. In Computational Toxicology (S Ekins, ed), pp. 403-432.

Fox T & Kriegl JM (2006). Machine Learning Techniques for In Silico Modeling of Drug Metabolism. Current Topics in Medicinal Chemistry 6, 1579-1591.

Franklin RB (2009). In silico studies in ADME/Tox: caveat emptor. Current Computer-Aided Drug Design 5, 128-138. However beautiful the strategy, you should occasionally look at the results. Winston Churchill (1874-1965). British Statesman.

Gleeson MP (2008). Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. Journal of Medicinal Chemistry 51(4), 817-834.

Gola J, Obrezanova O, Champness E & Segall M (2006). ADMET Property Prediction: The State of the Art and Current Challenges. QSAR & Combinatorial Science 25(12), 1172-1180.

Goodwin JT & Clark DE (2005). In Silico Predictions of Blood-Brain Barrier Penetration: Considerations to "Keep in Mind". Journal of Pharmacology and Experimental Therapeutics 315(2), 477-483.

Greene N, Judson P, Langowski J & Marchant C (1999). Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR and QSAR in Environmental Research 10, 299-314.

Hollosy F, Valko K, Hersey A, Nunhuck S, Keri G & Bevan C (2006). Estimation of Volume of Distribution in Humans from High Throughput HPLC-Based Measurements of Human Serum Albumin Binding and Immobilized Artificial Membrane Partitioning. Journal of Medicinal Chemistry 49(24), 6958-6971.

Hou T, Wang J, Zhang W, Wang W & Xu X (2006). Recent Advances in Computational Prediction of Drug Absorption and Permeability in Drug Discovery. Current Medicinal Chemistry 13, 2653-2667.

Hou T, Wang J, Zhang W & Xu X (2007a). ADME Evaluation in Drug Discovery. 6. Can Oral Bioavailability in Humans Be Effectively Predicted by Simple Molecular Property-Based Rules? Journal of Chemical Information and Modeling 47(2), 460-463.

Hou T, Wang J, Zhang W & Xu X (2007b). ADME Evaluation in Drug Discovery. 7. Prediction of Oral Absorption by Correlation and Classification. Journal of Chemical Information and Modeling 47(1), 208-218.

Hou T & Wang J (2008). Structure – ADME relationship: still a long way to go? Expert Opinion on Drug Metabolism & Toxicology 4(6), 759-770.

Hyland R, Jones B & van de Waterbeemd H (2006). Utility of human/human-derived reagents in drug discovery and development: An industrial perspective. Environmental Toxicology and Pharmacology 21(2), 179-183.

Jamei M, Marciniak S, Feng K, Barnett A, Tucker G & Rostami-Hodjegan A (2009). The Simcyp® Population-based ADME Simulator. Expert Opinion on Drug Metabolism & Toxicology 5(2), 211-223.

Jacobs MN, Janssens W, Bernauer U, Brandon E, Coecke S, Combes R, Edwards P, Freidig A, Freyberger A, Kolanczyk R, Mc Ardle C, Mekenyan O, Schmieder P, Schrader T, Takeyoshi M & Burg Bvd (2008). The Use of Metabolising Systems for In Vitro Testing of Endocrine Disruptors. Current Drug Metabolism 9, 796-826.

Khan & Sylte (2007). Predictive QSAR modeling for the successful predictions of the ADMET properties of candidate drug molecules

Kalgutkar AS, Gardner I, Obach RS, Shaffer CL, Callegari E, Henne KR, Mutlib AE, Dalvie DK, Lee JS, Nakai Y, O'Donnell JP, Boer J & Harriman SP (2005). A Comprehensive Listing of Bioactivation Pathways of Organic Functional Groups. Current Drug Metabolism 6, 161-225.

Kaznessis YN (2005). A Review of Methods for Computational Prediction of Blood-Brain Partitioning. Current Medicinal Chemistry - Central Nervous System Agents 5, 185-191.

Kharkar (2010). Two-Dimensional (2D) *in silico* models for Absorption, Distribution, Metabolism, Excretion and Toxicity (ADME/T) in drug discovery

Klopman G, Dimayuga M & Talafous J (1994). META. 1. A Program for the Evaluation of Metabolic Transformation of Chemicals. Journal of Chemical Information and Computer Sciences 34(6), 1320-1325.

Klopman G, Tu M & Fan BT (1999). META 4. Prediction of the metabolism of polycyclic aromatic hydrocarbons. Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta) 102(1), 33-38.

Klopman G, Tu M & Talafous J (1997). META. 3. A Genetic Algorithm for Metabolic Transform Priorities Optimization. Journal of Chemical Information and Computer Sciences 37(2), 329-334.

Konovalov DA, Coomans D, Deconinck E & Vander Heyden Y (2007). Benchmarking of QSAR Models for Blood-Brain Barrier Permeation. Journal of Chemical Information and Modeling 47(4), 1648-1656.

Kortagere & Ekins (2010). Troubleshooting computational methods in drug discovery. Journal of Pharmacological and Toxicological Methods 61(2), 67-75.

Lewis DF, Ioannides C & Parke DV (1996). COMPACT and molecular structure in toxicity assessment: a prospective evaluation of 30 chemicals currently being tested for rodent carcinogenicity by the NCI/NTP. Environmental Health Perspectives 104, 1011–1016.

Lewis DFV (2001). COMPACT: a structural approach to the modelling of cytochromes P450 and their interactions with xenobiotics. Journal of Chemical Technology & Biotechnology 76(3), 237-244.

Li H, Sun J, Fan X, Sui X, Zhang L, Wang Y & He Z (2008). Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. Journal of Computer-Aided Molecular Design 22(11), 843-855.

Li H, Yap CW, Ung CY, Xue Y, Cao ZW & Chen YZ (2005). Effect of Selection of Molecular Descriptors on the Prediction of Blood−Brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods. Journal of Chemical Information and Modeling 45(5), 1376-1384.

Lipinski C, Lombardo F, Dominy B & Feeney P (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews 46, 3-26.

Lipinski CA, Lombardo F, Dominy BW & Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews 23(1-3), 3-25.

Livingstone DJ & van de Waterbeemnd H (2009). In silico prediction of human bioavailability, in Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability (Methods and

Principles in Medicinal Chemistry 40). H van de Waterbeemd, B Testa, R Mannhold, H Kubinyi, G Folkers (Eds). pp. 433-451.

Luco JM & Marchevsky E (2006). QSAR Studies on Blood-Brain Barrier Permeation. Current Computer-Aided Drug Design 2, 31-55.

Madden (2010). In silico approaches for predicting ADME properties. Chapter 10 in *Recent Advances in QSAR Studies: Methods and Applications.* T Puzyn, J Leszczynski & MTD Cronin (Eds). Springer, Heidelberg, Germany, pp. 283-304.

Manga N, Duffy JC, Rowe, PH & Cronin, MT (2005). Structure-based methods for the prediction of the dominant P450 enzyme in human drug biotransformation: consideration of CYP3A4, CYP2C9, CYP2D6. SAR and QSAR in Environmental Research 16(1-2), 43-61.

Moda TL, Montanari CA & Andricopulo AD (2007a). Hologram QSAR model for the prediction of human oral bioavailability. Bioorganic & Medicinal Chemistry 15(24), 7738-7745.

Moda TL, Torres LG, Carrara AE & Andricopulo AD (2008). PK/DB: database for pharmacokinetic properties and predictive in silico ADME models. Bioinformatics 24(19), 2270-2271.

Mohan C, Ghandi T, Garg D & Shinde R (2007). Computer-assisted methods in chemical toxicity prediction. Mini Reviews in Medicinal Chemistry 7(5), 499-507.

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

Norinder U & Bergström CAS (2006). Prediction of ADMET Properties. ChemMedChem 1(9), 920-937.

Norinder U & Haeberlein M (2002). Computational approaches to the prediction of the blood-brain distribution. Advanced Drug Delivery Reviews 54(3), 291-313.

Otagiri (2005). A molecular functional study on the interactions of drugs with plasma proteins. Drug Metabolism and Pharmacokinetics 20(5), 309-323.

Raevsky OA, Schaper K-J, Artursson P & McFarland JW (2002). A Novel Approach for Prediction of Intestinal Absorption of Drugs in Humans based on Hydrogen Bond Descriptors and Structural Similarity. Quantitative Structure-Activity Relationships 20(5-6), 402-413.

Schuster D, Steindl TM & Langer T (2006). Predicting Drug Metabolism Induction In Silico. Current Topics in Medicinal Chemistry 6, 1627-1640.

Segall MD, Beresford AP, Gola JM, Hawksley D & Tarbit MH (2006). Focus on success: using a probabilistic approach to achieve an optimal balance of compound properties in drug discovery. Expert Opinion on Drug Metabolism & Toxicology 2(2), 325-337.

Sietsema WK (1989). The absolute oral bioavailability of selected drugs International Journal of Clinical Pharmacology, Therapy and Toxicology 27(4), 179-211.

Sprous *et al*. (2010). QSAR in the pharmaceutical research setting: QSAR models for broad, large problems

Sun LZ, Ji ZL, Chen X, Wang JF & Chen YZ (2002). ADME-AP: a database of ADME associated proteins. Bioinformatics 18(12), 1699-1700.

Talafous J, Sayre L, Mieyal J & Klopman G (1994). META. 2. A dictionary model of mammalian xenobiotic metabolism. Journal of Chemical Information and Computer Sciences 34, 1326-1333.

Testa B, Balmat A-L, Long A & Judson P (2005a). Predicting Drug Metabolism - An Evaluation of the Expert System METEOR. Chemistry & Biodiversity 2(7), 872-885.

Testa B, Vistoli G & Pedretti A (2005b). Musings on ADME Predictions and Structure-Activity Relations. Chemistry & Biodiversity 2(11), 1411-1427.

Tetko IV, Bruneau P, Mewes HW, Rohrer DC & Poda GI (2006). Can we estimate the accuracy of ADME-Tox predictions? Drug Discovery Today 11(15-16), 700-707.

Thummel KE & Shen DD (2001). Design and optimization of dosage regimens: pharmacokinetic data. In: Goodman and Gilman's the pharmacological basis of therapeutics (Hardman JG, Limbird LE, Gilman AG, eds). McGraw-Hill, New York, pp. 1917-2023.

Trainor GL (2007). The importance of plasma protein binding in drug discovery Expert Opinion on Drug Discovery 2(1), 51-64.

Turner JV, Maddalena DJ & Cutler DJ (2004b). Pharmacokinetic parameter prediction from drug structure using artificial neural networks. International Journal of Pharmaceutics 270(1-2), 209-219.

Waterbeemd H, Camenisch G, Folkers G, Chretien JR & Raevsky OA (1998). Estimation of blood–brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. Journal of Drug Targeting 6, 151-165.

Vastag M & Keseru GM (2009). Current in vitro and in silico models of blood-brain barrier penetration: a practical view. Current Opinion in Drug Discovery & Development 12(1), 115-124.

Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW & Kopple KD (2002). Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. Journal of Medicinal Chemistry 45(12), 2615-2623.

Veselovsky *et al*. (2010). Computer-based substrate specificity prediction for cytochrome P450

Votano J (2005). Recent uses of topological indices in the development of in silico ADMET models.Votano JR. Current Opinion in Drug Discovery & Development 8(1), 32-37.

Votano JR, Parham M, Hall LM, Hall LH, Kier LB, Oloff S & Tropsha A (2006). QSAR Modeling of Human Serum Protein Binding with Several Modeling Techniques Utilizing Structure−Information Representation. Journal of Medicinal Chemistry 49(24), 7169-7181.

Wang J, Krudy G, Xie X-Q, Wu C & Holland G (2006). Genetic Algorithm-Optimized QSPR Models for Bioavailability, Protein Binding, and Urinary Excretion. Journal of Chemical Information and Modeling 46(6), 2674-2683.

Wang J & Skolnik S (2010). Mitigating permeability-mediated risks in drug discovery. Expert Opinion on Drug Metabolism and Toxicology 6(2), 171-187.

Wan H & Ulander J (2006). High-throughput pKa screening and prediction amenable for ADME profiling. Expert Opinion on Drug Metabolism and Toxicology 2(1), 139-155.

Yap CW, Li ZR & Chen YZ (2006). Quantitative structure-pharmacokinetic relationships for drug clearance by using statistical learning methods. Journal of Molecular Graphics and Modelling 24(5), 383-395.

Zhang L, Zhu H, Oprea T, Golbraikh A & Tropsha A (2008a). QSAR Modeling of the Blood–Brain Barrier Permeability for Diverse Organic Compounds. Pharmaceutical Research 25(8), 1902-1914.

Zhao YH, Abraham MH, Ibrahim A, Fish PV, Cole S, Lewis ML, de Groot MJ & Reynolds DP (2007). Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes. Journal of Chemical Information and Modeling 47(1), 170-175.

Zmuidinavicius D, Didziapetris R, Japertas P, Avdeef A & Petrauskas A (2003). Classification structure-activity relations (C-SAR) in prediction of human intestinal absorption. Journal of Pharmaceutical Sciences 92(3), 621-633.

## Chapter 8 references

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG & Würtzen G (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food & Chemical Toxicology 42(1), 65-83.

JRC (2010). Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Report from the European Commission's Joint Research Centre (JRC) to the European Food Safety Authority (EFSA). Available at: http://www.efsa.europa.eu/en/scdocs/scdoc/50e.htm

Worth A, Lapenna S, Lo Piparo E, Mostrag-Szlichtyng A & Serafimova R (2010a). A Framework for assessing *in silico* Toxicity Predictions: Case Studies with selected Pesticides. JRC Technical Report EUR 246XX EN. Publications Office of the European Union, Luxembourg. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/

European Commission

**EUR 24748 EN – Joint Research Centre – Institute for Health and Consumer Protection**

Title: The Use of Computational Methods in the Toxicological Assessment of Chemicals in Food: Current Status and Future Prospects

Author(s): Andrew Worth, Mojca Fuart-Gatnik, Silvia Lapenna, Elena Lo Piparo, Aleksandra Mostrag-Szlichtyng and Rositsa Serafimova

Luxembourg: Publications Office of the European Union

2011 – 137 pp. – 21 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593

ISBN 978-92-79-19567-9

doi:10.2788/6234

**Abstract**

A wide range of chemicals are intentionally added to, or unintentially found in, food products, often in very small amounts. Depending on the situation, the experimental data needed to complete a dietary risk assessment, which is the scientific basis for protecting human health, may not be available or obtainable, for reasons of cost, time and animal welfare. For example, toxicity data are often lacking for the metabolites and degradation products of pesticide active ingredients. There is therefore an interest in the development and application of efficient and effective non-animal methods for assessing chemical toxicity, including Quantitative Structure-Activity Relationship (QSAR) models and related computational methods.

This report gives an overview of how computational methods are currently used in the field of food safety by national regulatory bodies, international advisory organisations and the food industry. On the basis of an international survey, a comprehensive literature review and a detailed QSAR analysis, a range of recommendations are made with the long-term aim of promoting the judicious use of suitable QSAR methods. The current status of QSAR methods is reviewed not only for toxicological endpoints relevant to dietary risk assessment, but also for Absorption, Distribution, Metabolism and Excretion (ADME) properties, which are often important in discriminating between the toxicological profiles of parent compounds and their reaction products. By referring to the concept of the Threshold of Toxicological Concern (TTC), the risk assessment context in which QSAR methods can be expected to be used is also discussed.

This report provides a summary and update of the findings obtained in a study carried out by the JRC under the terms of a contract awarded by the European Food Safety Authority (EFSA).

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.


LB-NA-24748-EN-C