IGHRC

The Interdepartmental Group
on Health Risks from Chemicals

Predictive approaches to **chemical hazard identification and characterisation:** Current use by UK Government Departments and Agencies

cr16

The Interdepartmental Group
on Health Risks from Chemicals

Predictive approaches to **chemical hazard identification and characterisation:** Current use by UK Government Departments and Agencies

cr16

The Interdepartmental Group on Health Risks from Chemicals aims to stimulate the development of new, improved approaches to the assessment of risks to human health from chemicals.

The Steering Committee of the Interdepartmental Group on Health Risks from Chemicals comprises participants from the Department for Environment, Food and Rural Affairs, the Department of Health, (including Public Health England), the Department for Business, Enterprise and Regulatory Reform, the Home Office, the Environment Agency (EA), the Health and Safety Executive (including Chemicals Regulation Directorate), the Food Standards Agency, the Medicines and Healthcare Products Regulatory Agency, the Veterinary Medicines Directorate, the Biotechnology and Biological Sciences Research Council, the Medical Research Council and the Natural Environment Research Council.

The Secretariat is based at the Institute of Environment and Health, Cranfield University.

Acknowledgements: This document is a result of the considerable input provided by IGHRC member departments and agencies, supervised by the HSE.

Contributors include: Susy Brescia, Helen McGarry, Ian Dewhurst and Helen Smith (HSE); Sue Fraser (HSL); Henry Stemplewski (MHRA); David Gott, Diane Benford and David Parker (FSA); Philippa Edwards and Steve Robjohns (HPA)

Please cite as: IGHRC (2013) Predictive approaches to chemical hazard identification and characterisation: Current use by UK Government Departments and Agencies (cr16). Institute of Environment and Health, Cranfield University, UK.

Published by the Institute of Environment and Health, 2013

ISBN 978-1-899110-50-6

Institute of Environment and Health
Vincent Building
Cranfield University
Cranfield
Bedfordshire MK43 0AL
UK

Telephone +44(0) 1234 758279
Facsimile +44(0) 1234 758280
Website http://ieh.cranfield.ac.uk/ighrc/ighrc.html

Foreword

This document has been produced by the Interdepartmental Group on Health Risks from Chemicals (IGHRC) as part of its Phase 3 work programme (October 2007 to May 2013). It is about using knowledge of the chemical structure and physicochemical properties of a substance to make predictions about its toxicity. Such predictions can be used in a variety of ways: to screen for alerts, to supplement other forms of data, to give a basis for decision making when nothing else is available; and sometimes to obviate the need for experimental testing.

The document shows what tools and approaches are available and the circumstances in which these predictions are currently used in chemicals regulation - the predictions being made or (if made by others) accepted by UK government departments. There is clearly an appetite and a role for predictive approaches. With this in mind, the document recommends how the basis for such predictions should be presented to UK government departments to maximise the chances of acceptance, and suggests initiatives that could further advance the development and application of predictive techniques.

Cases are described that exemplify how things stand at present. An important inference is that, currently, predictive techniques are not being used significantly in ways outside of those described, probably because they are not deemed to be sufficiently reliable for the purpose in question. If any reader of this document thinks that the current UK government perspective is too limited, the IGHRC Secretariat will be happy to receive worked examples that aim to show how predictivity could be applied more widely and the advantages that could arise from so doing.

Dr Steve Fairhurst
Chairman of IGHRC
Chemicals Regulation Directorate, Health and Safety Executive

Contents

1 Predictive Approaches: Current use by UK Government Department and Agencies

1.1 Introduction

There is a growing interest in, and increasing imperatives for, the development of alternative methods for toxicological evaluations that reduce, refine or replace (3Rs) the use of animals. The drive for designing alternative means to characterise the toxicological properties of substances, including the development of different test methods, is being fuelled by a combination of societal expectations, legislative pressure, economic costs and scientific opportunity (Kimber *et al.,* 2011). In addition, in the UK there has been a recent pledge (May 2010) by the current coalition government to reduce animal experimentation[1].

The vision for the safety assessment of chemicals in the 21st century seeks to utilise the increasing power of modern computational chemistry in combination with advanced systems' biology, high-throughput *in vitro* screening and 'omic' technologies to predict the toxic potential and to prioritise chemicals with respect to the need for detailed toxicological evaluation.

The focus of this document is on the recent advances in computational chemistry and the increasing understanding of the relationship between chemical structure and biological activity as a means to predict the toxicity of chemicals (human health hazard assessment). The predictive approaches considered in this report are those non-testing techniques based on structure–activity relationships of chemicals. These can be computer based, such as (quantitative) structure–activity relationship (Q)SAR models (*in silico* methods) or based on less-formalised analyses of information, such as read-across and grouping of chemicals (categories).

There are many promising benefits to these techniques, including their cost effectiveness, speed compared with traditional testing and reduction in animal use.

(Q)SAR and other predictive techniques have been widely used by the pharmaceutical industry for many years to predict drug activity and identify new candidate drugs for further development. Generally these models, which are designed for the prediction of the desired pharmacological action rather than the unwanted toxic activity of a substance, are proprietary and confidential, and thus tend not to be available to the wider scientific community. In addition, (Q)SAR tools developed for the prediction of drug activity (computational pharmacology) are not generally suitable for toxicity prediction for several reasons. Drugs cover a relatively narrow chemical range (i.e. a limited number of potential structures), are developed with discrete biological targets in mind, have relatively high target molecule affinities, conform to a limited range of physicochemical properties, have well-understood metabolic profiles and also have well-known and quantified patterns of use. Many of the tools used in computational toxicology by the pharmaceutical industry were developed with these aspects in mind. They have had to be adjusted to the broader structural universe (chemical space) of environmental chemicals to become of some use for toxicity prediction. Environmental chemicals may not have discrete intended biological targets (with the exception of pesticides and biocides), usually do not exhibit high-affinity interactions with molecular targets, often have largely unknown metabolic patterns and can have highly variable patterns of use that may result in highly variable exposures.

Although progress in the use and further development of *in silico* techniques applicable to toxicology has been more gradual than that of models applicable to pharmacology, new European legislation, such as

---

[1] www.cabinetoffice.gov.uk/sites/default/files/resources/coalition_programme_for_government.pdf (p. 18; accessed May 2013)

REACH and the Cosmetics Directive, has led to increased interest in, and improved dissemination of, the principles of these methods, resulting in an enhancement of specific tools for toxicity prediction. This is because these pieces of legislation promote the use of alternatives to laboratory animals to provide estimates of toxicity for hazard and risk assessment purposes.

The use of predictive approaches can be important in human health hazard assessment of chemicals. Predictive approaches are of value when specific toxicity data are not available on an individual chemical or a mixture of chemicals.

- The registration requirements of REACH call for toxicological information for a large number of chemicals, and the ability to utilise existing data on related chemicals in ways that avoid the need for new vertebrate animal testing could result in significant savings in resources and animals.
- Even if the chemical of concern is covered by a prior authorisation regulatory scheme (e.g. pesticides, biocides or food additives) demanding extensive packages of test data before authorisation, there might be opportunities to reduce animal use as well as cost and time implications to industry.
- In some instances, such as the finding of contamination in food or environmental samples at relatively low levels, there can be a need to quickly make some assessment of the possible threat before new data can be generated, in order to provide the basis for decision making aimed at reducing exposures (e.g. removal of an item from sale).

In addition to the use in formal hazard assessment, it is also possible to use predictive approaches in other ways.

- To prioritise future testing and/or detailed assessment, based on the level of concern identified in the prediction
- To eliminate potentially toxic compounds before they might be synthesised as commercially useful substances
- To provide support for, or the basis of a challenge to, data from a non-standard investigation
- To identify components in a mixture that present the greatest level of concern
- To form part of a weight-of-evidence (WoE) approach, alongside test data and exposure estimation

Predictive approaches can lead to conclusions that are qualitative/categorical (e.g. chemical X has an alert for mutagenicity), quantitative/continuous (e.g. chemical X has a 4h-LC_{50} value of 25 mg m^{-3}), or a combination of the two (e.g. chemical X is likely to be a weak sensitiser). The type of prediction used can be driven by the required output.

Recent developments in computing power, the ability to create extensive databases and the use of the internet to compile, organise and distribute information have increased the capability to investigate relationships between chemical structure and biological activity. A wide range of (Q)SAR programs that use a range of approaches to predict a chemical's toxicity profile are now available. As test data become available on a wider range of chemicals, the potential increases to develop the means to extrapolate or read-across with confidence from the data to a structurally closely related compound.

While the increases in knowledge of the relationships between chemical structure and biological activity have enhanced the ability to predict the toxicity of untested chemicals, they have also demonstrated that subtle changes in chemical structure can have a significant impact on biological activity, especially if the toxicity is mediated by binding to a receptor.

For this reason, certain minimum criteria need to be taken into account when performing hazard assessment with predictive approaches. It is therefore important to be aware of the limitations of predictive approaches and the basic requirements associated with their use in human health hazard assessment.

This document describes some of the predictive approaches available for human health hazard assessment of chemicals and how they have been used by UK government departments and agencies. It does not aim to be a definitive guide, but is more a compilation of currently used approaches together with some case studies.

As part of the development of this document, a questionnaire was circulated to UK departments and agencies to determine their current uses of predictive approaches. The main results of the survey are outlined below, by department. The tabulated results of the questionnaire are presented in Appendix 1. In addition, a workshop to explore in more depth the perception and use of predictive approaches across UK government regulatory bodies was held in January 2012; findings from that workshop have been incorporated into this document.

Overall, the survey has shown that read-across from structurally related chemicals and grouping approaches are used routinely by UK government for different types of chemical within various pieces of legislation, in different contexts and for different purposes or assessments. (Q)SAR models are used much less commonly, mainly because of a lack of in-house expertise; they are usually utilised in specific situations (e.g. for minor impurities, residues, by-products or metabolites) requiring only a limited toxicological assessment. They also tend to be applied in a weight-of-evidence (WoE) approach together with other strands of information. If a consensus prediction is obtained from several different models, then the prediction is more likely to be accepted. In addition, a positive (Q)SAR result tends to be accepted, in the absence of study results, more readily than a negative result.

In many instances, there appear to be no formal guidance documents on the use of such predictive approaches and no freely available software tools that, together with appropriate training, can help promote and harmonise the use of these techniques. Another obstacle to the use of computational methods is the need to ensure IT security within government departments. However, despite all these barriers, predictive toxicology is widely perceived by UK government to be a useful tool to support regulatory assessments and decision making.

The survey has also shown that the reason for applying these techniques varies, depending on the context. There are situations where urgency and/or speed of decision making is the driver because there is no time for generating new data (e.g. in chemical contamination of food incidents). There are also circumstances where the application of predictive methods is the only available option because there is no legislative instrument to request test data (e.g. in an environmental contaminant with no data). On some occasions, the use of predictive tools is deemed to be acceptable because it is only to assist in prioritising chemicals for further detailed assessment or because the decision-making process does not lead to major regulatory consequences (e.g. in voluntary, industry-driven evaluation schemes such as the OECD HPV programme). In many instances, the main factor driving the implementation and acceptability of these approaches is the reduction of animal testing, costs and regulatory burden (e.g. in use of read-across for some endpoints for pesticides, biocides, food-contact materials and food additives).

There are also situations where the legislation is not prescriptive in requiring animal testing but instead offers some flexibility in the type of methods that can be selected to perform the hazard assessment (e.g. for impurities, residues, metabolites, by-products, co-formulants). In these instances, the application of predictive toxicology is considered acceptable either because, even if the prediction is wrong, it is not as critical to the overall evaluation (e.g. where an impurity is present at very low levels); or a pragmatic approach is adopted because it would be impractical, unsustainable and unworkable to request full toxicological packages of test data on a potentially very high number of different but related chemicals (e.g. for residues, metabolites and by-products of pesticides and biocides).

It should be noted that although these approaches are applied always with caution and only if it is deemed that they are adequate and scientifically valid, the general view remains that they are considered to be less reliable than animal testing.

1.2 Current use of predictive approaches by UK government regulatory departments and agencies

1.2.1 Health and Safety Executive: Chemicals' Regulation Directorate

The Chemicals' Regulation Directorate (CRD) in HSE is responsible for the implementation of several pieces of EU chemical legislation such as Plant Protection Products Regulation (PPPR), Biocidal Products Directive (BPD; to be replaced by the Biocidal Products Regulation (BPR)), Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) and Classification, Labelling and Packaging Regulation (CLP).

Plant protection products (pesticides) and biocides

For the evaluation of active substances under the PPPR, predictive approaches are rarely used by the industry duty holders, as submission of extensive packages of test data is prescribed by the legislation. Only very occasionally, read-across arguments from very closely related analogues have been proposed for some endpoints by duty holders. The situation with the evaluation of active substances under the BPD and BPR is similar. Read-across from structurally related substances is proposed from time to time by the industry duty holders in relation to a wide range of different toxicological endpoints; in these cases, the use of the read-across approach is proposed by industry to fill data gaps for chronic toxicity, carcinogenicity and reprotoxicity (endpoints requiring the conduct of very expensive studies with high levels of animal usage). (Q)SAR model predictions (usually the expert system Derek Nexus) are sometimes utilised by duty holders to support read-across proposals. Under these regulatory schemes, the validity of such approaches is thoroughly assessed by HSE to determine whether to accept, reject or modify them. Subsequently, if accepted, HSE has to defend them to the other Member State regulatory bodies and EU institutions involved. On occasion, HSE may also choose to perform its own computational prediction (Derek Nexus is used) to expedite the evaluation process.

(Q)SAR model predictions tend to be used on a more regular basis for the toxicological evaluation of impurities in new sources of pesticide active substances or for the toxicological assessment of products/residues and groundwater metabolites arising from metabolism of the pesticides by plants and soil organisms. In some instances, these *in silico* predictions are generated and put forward by industry (on its own accord or at the request of HSE) and subsequently assessed and verified by HSE, but on occasion HSE chooses to perform its own computational predictions (again, Derek Nexus is used). These (Q)SAR predictions are never used completely in isolation in such assessments, but always in a WoE approach, together with some test data (*in vitro* and *in vivo*) and consideration of other factors such as potential exposure levels, structural similarity to the parent chemical or metabolites produced in experimental rodent studies and expert predictions of kinetic behaviour.

For the evaluation of products under the PPPR and the BPD/BPR, read-across is used on a routine basis to extrapolate acute toxicity data and dermal absorption values from one formulation to another very similar formulation (same active substance and similar composition of co-formulants). Read-across arguments are generally developed by the industry duty holders and subsequently evaluated for their validity/adequacy by HSE.

REACH (chemicals in general)

For the assembly of information profiles of general chemicals under REACH, predictive approaches play a pivotal role; in fact, the legal text states that one of the purposes of the regulation is ‘the promotion of alternative methods for assessment of hazards of substances ...’. Consequently, non-testing methods should be used ‘whenever possible’ and the use of vertebrate animal tests to generate new toxicity data is regarded as a last resort. For registration purposes, read-across and grouping approaches have been used extensively by industry to fill data gaps. For some endpoints, up to approximately 30% of the registrations submitted to ECHA before the deadline of December 2010 included a read-across argument. A comprehensive picture of the acceptability of these arguments to the regulatory bodies involved, including the UK REACH Competent Authority in HSE (CRD), has not been produced yet, as the compliance check of selected registration dossiers is still ongoing. So far, the appraisal of the first batch of dossiers by ECHA shows that the quality of the proposed read-across arguments is rather poor: they tend to lack sufficient and adequate justification.

(Q)SAR model predictions can also be utilised. However, it appears that only 0.1–0.2% of the registrations submitted by industry to ECHA before the deadline of December 2010 included the use of *in silico* techniques as stand-alone methods. From the available statistics produced by ECHA, it is difficult to determine whether (Q)SAR methods have been used as supporting evidence in conjunction with other data in a higher number of registrations, since one of the categories used in the analysis was ‘weight of evidence approach’. Generally, (Q)SAR does not yet have regulatory acceptance as a stand-alone method. Detailed guidance documents are available to support the use of these techniques for the purposes of REACH.

Classification and labelling

For the purposes of harmonised classification and labelling (C&L) of substances, read-across and grouping approaches, which are normally put forward by industry and considered or evaluated for their validity by the relevant regulatory authorities, have been used on an *ad hoc* basis. (Q)SAR model predictions are rarely used and never as stand-alone methods. It is also significant that C&L decisions have generally been based on the data available on a substance, which has not provided a stimulus to fill data gaps. However, for some endpoints, the presence of some substructures (structural alerts) has been sufficient for a positive classification in the absence of *in vivo* data (e.g. respiratory sensitisation for isocyanates, skin/eye irritation for organic peroxides).

OECD high production volume chemicals’ programme

CRD (together with the Health Protection Agency and the Environment Agency) also plays a role in the OECD-HPV (high production volume) chemicals’ programme, which is a voluntary collaboration between sponsor countries and industry to establish agreed hazard profiles on high production volume substances. Within this programme, industry has made extensive use of the category approach to minimise the testing of individual chemicals. Many of the categories proposed have been accepted by the regulatory bodies involved. However, it should be noted that within this programme the conclusions of the evaluation do not lead to major regulatory consequences; in general, there is a more relaxed approach and an acceptance of less exactitude where there are no such implications.

1.2.2 Current use of predictive approaches by the Food Standards Agency (FSA)

In the food area, the basis for the evaluation of flavouring compounds is grouping approaches and read-across from data on related substances. This approach was incorporated into the European regulatory framework for flavouring compounds and was based on an opinion from the Scientific Committee on Food in 1999[2]. Flavourings impart taste and/or smell to food. Any particular flavouring is a complex mixture of a number of individual chemicals which in combination produce the desired taste or smell. When combined, structurally closely related compounds may modify the perceived taste or smell. As a consequence a large number of compounds may be used in flavourings often in relatively small quantities. Test data on every individual chemical are usually limited. However, because compounds can be grouped, it is possible to apply a stepwise approach that integrates information on intake from current uses, structure–activity relationships, metabolism and, when needed, toxicity. The flavourings are first subdivided into three structural classes (I, II, III). Then it is predicted whether they will be metabolised to innocuous products. If they are, then intakes are compared to the thresholds specified for the structural classes below which exposures are not considered to present a safety concern. If exposures are greater than the threshold then, unless the compound is endogenous, under intended-use conditions an adequate margin of safety must exist from the NOAEL for the substance or a structurally related substance (allowing for any perceived difference in toxicity between the substance and the related substances). Although a similar system for flavouring evaluations is used by the Joint FAO/WHO Expert Committee on Food Additives (JECFA), the European Food Safety Authority (EFSA) places greater emphasis on absence of possible genotoxicity and uses two exposure approaches.

Comparison of the relative bioavailability of many nutrients from nutrient sources has relied on estimates comparing the relative dissolution data for each particular nutrient source with bioavailability information on a known nutrient source. This approach was considered proportionate and pragmatic, since the legislation is based on comparison of nutrient sources rather than being product specific, therefore the influence of other formulation components on the actual bioavailability can be ignored.

Read-across from structurally related chemicals is also used on an *ad hoc* basis in the evaluation of food contact materials and food additives. Components of food contact materials, such as printing inks and adhesives, are often compositionally similar to materials intended for other uses (e.g. in medical devices) sold under the same trade name. Some of these other uses may require toxicological data on that specific material. Where data exist on such a material and sufficient comparative compositional information is available, then data would be read-across to the food contact material. Although in the case of food additives toxicological studies are required, it has been recognised that there are some categories, for example botanical extracts, where testing each individual product would be neither scientifically nor ethically justifiable. Although different extraction processes produce extracts of differing composition, by testing representative extracts (ideally from both extremes and the centre of the potential compositions) potential differences in biological effect should be observed. It is possible to extrapolate the observed findings to the untested extracts within the tested domain and determine whether or not these would be acceptable. A good example of this approach is in the opinion on rosemary extracts from the EFSA Panel on Food Additives, Flavourings, Processing Aids and Materials in Contact with Food (EFSA, 2008[3]). In this example toxicological and compositional data were available for five different extraction methods, provided the specifications included allowable ranges of some principal components were defined these data could be extrapolated to other rosemary extracts.

Predictive approaches including *in silico* methods are applied extensively to the evaluation of food incidents as there is often a need for an urgent conclusion or decision to be taken, and no time to wait for new information to be generated via testing.

(Q)SAR model predictions (Derek Nexus, Toxtree, Lazar and Osiris) have only been used occasionally in conjunction with genotoxicity test data in the assessment of impurities present in veterinary medicinal products and in the toxicological evaluation of animal feed additives, either of which may lead to residues in meat and milk.

In 2012 the EFSA Scientific Committee adopted an opinion endorsing the use of the threshold of toxicological concern (TTC) under defined circumstances. From a scientific perspective the TTC approach could, in principle, be applied to any substances for which exposures are low and toxicity data are sparse. However, in the context of the EU there is a legislative requirement to submit toxicity

---

[2] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31999D0217:EN:NOT [accessed May 2013]

[3] www.efsa.europa.eu/en/efsajournal/pub/721.htm [accessed March 2013]

data in several areas (e.g. the technically active substances in pesticides, food additives and feed additives). The TTC approach was recommended as a useful screening tool either for the setting of priorities for data needs and for risk management action or for deciding whether exposure is so low that the probability of adverse health effects is low and consequently further data are not needed. In the case of chemicals subject to legislative requirements the latter could be applied to unavoidable residuals and contaminants, degradation products and metabolites[4]. A number of chemical groups where the TTC approach should not be used are high potency carcinogens (i.e. aflatoxin-like, azoxy- or *N*-nitroso-compounds, benzidines, hydrazines), inorganic substances, metals and organometallics, proteins, steroids, substances that are known or predicted to bioaccumulate, nanomaterials, radioactive substances and mixtures of substances containing unknown chemical structures.

1.2.3 Health Protection Agency

The Health Protection Agency (HPA) is a non-regulatory body that is responsible for providing toxicological advice to other government departments and agencies in relation to environmental contaminants in drinking water, soil, air and waste processes.

Read-across from data on related substances and grouping approaches are often used for the assessment of environmental contaminants, as test data can be limited for the chemicals of interest and there is no regulatory option available for requiring a specific duty holder to perform further testing.

In the case of drinking water, the HPA provides toxicological advice to the Drinking Water Inspectorate (DWI) regarding its regulatory approval scheme for products permitted for use in contact with the public water supply. Migration tests are conducted on some products. Some chemicals can be detected leaching into the test water at relatively low levels. These chemicals are usually not product ingredients, but degradation products, reaction by-products or impurities etc. For the majority of these chemicals, there are no or very limited toxicity data. In such cases, the HPA may use a read-across approach where relevant toxicity data are identified for very similar compounds. When adopting this approach, the uncertainty is clearly acknowledged. A pragmatic approach has to be adopted. Structural alerts for mutagenicity may also be used for chemicals with no or very limited toxicity data. For example, the HPA has advised that chemicals such as nitrosamines, primary aromatic amines and oxiranes should be regarded as potentially mutagenic and carcinogenic and thus, exposure should be as low as reasonably practicable (ALARP).

The HPA also provides toxicological advice on test results that have been conducted for public and private water supplies. These can be carried out in response to taste and odour problems. Hydrocarbon compounds originating from a petroleum product source (e.g. heating oil spill) are often detected at relatively low levels. Toxicity data are not available for many individual hydrocarbon chemicals. However, a risk assessment can be conducted by using oral reference doses (RfD) for various hydrocarbon fractions containing very similar compounds. This approach is recommended by the World Health Organization (WHO) Guidelines on Drinking Water Quality[5] and based on an approach derived by the Total Petroleum Hydrocarbon Criteria Group (TPHCWG, 1997). The RfD are derived from individual chemicals or hydrocarbon mixtures considered to be sufficiently representative of the toxicity of certain hydrocarbon fractions (each fraction contains a large number of individual chemicals within a specific carbon number range).

A similar approach has been adopted when the HPA has been asked to provide toxicological advice on exposure to vapours from petroleum hydrocarbon products, such as that arising from an indoor leak or spill of heating oil (kerosene). In such cases exposure is likely to occur to a mixture of volatile hydrocarbon compounds for which there are limited or no individual inhalation toxicity data. Comparisons can be made with the various inhalation reference concentrations (RfC) derived by the TPHCWG (1997). For example, the concentration of the detected hydrocarbon fraction in air can be compared with TPHCWG (1997) chronic (lifetime) RfC of $1\,\mathrm{mg\,m^{-3}}$ for USA JP8 jet fuel (C8–C16), which is considered to be sufficiently similar to heating oil vapour.

For dibenzo-*p*-dioxins and related compounds produced by combustion processes, that can be present in soils, internationally agreed toxic equivalent factors (TEF) are normally used. The application of TEF is based on extrapolation, or read-across from data available on some congeners to structurally related congeners lacking such data.

Where no genotoxicity data are available, (Q)SAR model predictions can be used on an *ad hoc* basis for the assessment of mutagens and genotoxic

---

[4] www.efsa.europa.eu/en/efsajournal/doc/2750.pdf [accessed March 2013]

[5] www.who.int/water_sanitation_health/publications/2011/dwq_guidelines/en/index.html [accessed March 2013]

carcinogens. The HPA acting as secretariat to the Committee on Mutagenicity (COM) has undertaken some initial investigations using a combination of CAESAR and Toxtree. These two models are freely available and combine a statistical model (CAESAR) with a knowledge-based model (Toxtree). This proposal is currently under review by the COM.

1.2.4 Medicines and Healthcare Products Regulatory Agency

The Medicines and Healthcare Products Regulatory Agency (MHRA) is a government body. Its functions include the regulation of human medicines and medical devices and equipment used in healthcare and the investigation of harmful incidents. The principal aim of the agency is to safeguard public health.

Generally, for new active substances at the marketing authorisation stage a read-across or category approach is not used. Each compound is considered on a case-by-case basis. The rationale for this approach is that small changes in the chemical or molecular structure of the drug substance can have a significant impact on biological activity. Also, changes in the drug product, such as changes in the excipients or formulation, may impact on biological activity.

In addition, predictive approaches are not generally used since an extensive package of test data is required by the legislation. If a non-clinical dataset (which generally includes data generated *in vitro* and in animal studies on the pharmacodynamics, safety pharmacology, pharmacokinetics and toxicity of the compound) is not complete or the studies are not of the required quality, then the relevant data would be requested from the applicant. This may involve further experimental studies. It is the responsibility of the applicant to submit an appropriate non-clinical package of studies. These data are considered to be commercial-in-confidence property of the applicant.

Regarding the potential impact of human medicines on the environment, QSAR and read-across from other structurally related substances may be used to help interpret the data, but not to replace completely the need for guideline studies.

Predictive approaches may be used in certain circumstances and some endpoints. Several examples are given below.

Concerning the genotoxicity of a compound, if a chemical is mutagenic *in vivo*, it is considered to be potentially carcinogenic and there is no need for carcinogenicity studies. If a compound is genotoxic to somatic cells, it is considered to be also genotoxic to germ cells and there is no need for germ cell mutagenicity assays. If a compound is positive *in vitro*, it is considered to be an *in vivo* genotoxin and *in vivo* studies are generally not required. These are examples of activity–activity relationship (AAR) rather than SAR where certain toxicological properties are predicted on the basis of the presence of hazards which share similar mechanisms of action.

If a substance belongs to a specific class or group of chemicals it is possible to apply QSAR in its simplest form. For example, certain anti-cancer compounds such as anthracyclines are considered to be genotoxic, carcinogenic and toxic to reproduction without the need for experimental verification. Compounds that belong to a class that has been well characterised as causing developmental toxicity do not require reproductive toxicity studies. If a chemical is genotoxic and targets rapidly dividing cells, reproductive toxicity studies are not required.

Another example of the application of predictive approaches is the use of read-across in the authorisation process of generic medicines. If a medicinal product is considered to be essentially similar to a currently marketed product, additional non-clinical studies are generally not required. An overview of the toxicological profile of the product based on information on existing products (usually from the published literature) is acceptable. However, it is important to emphasise that this is not a generic position which is always applicable. For example, for drugs comprising a liposomal formulation, specific data may be required as it has been shown that even small changes in various components of the formulation, including liposome morphology, may affect the biological properties of the product.

Predictive approaches are often used in clinical trials in the initial trial dose setting. Computer-aided prediction of drug toxicity through the application of several software models (e.g. Derek Nexus, MCASE, TOPKAT) together with information on target receptor occupancy (based on *in vitro* studies) play an important role in the determination of the maximum recommended starting dose (MRSD) in humans. In addition, for high risk medicinal products, the human minimal anticipated biological effect level (MABEL) is also established. The MABEL is estimated through the application of a pharmacodynamic/pharmacokinetic (PD/PK) modelling approach which integrates information from (i) QSAR prediction of drug toxicity, (ii) receptor binding and receptor occupancy studies in target cells from human and relevant animal species and *in vivo*, (iii) concentration–response curves *in vitro* in target cells from human and relevant animal species and *in vivo* dose responses in relevant animal species, and (iv) exposure at the pharmacologically active dose in the relevant species.

1.3 EU survey on the use of (Q)SAR models and *in silico* methods

An EU-wide survey similar to the one reported here has been conducted as part of the ORCHESTRA project; this is a project which is funded by the European Commission with the intention of conducting and presenting the findings of research in the area of *in silico* methods for toxicity testing. The aim of this project is to promote a wider understanding, awareness and appropriate use of *in silico* methods to regulators, industry users and academics.

For the ORCHESTRA project survey, data was collected between September 2010 and April 2011. The results are available on the project website[6] and open-access literature (Mays *et al.,* 2012). Thirty-three stakeholders from within the EU participated in the survey. Of the respondents, 13 were academics or consultants (having no direct stake in chemical manufacturing, nor a mission to protect or regulate public health), 12 were regulators (having a regulatory mission) and eight were from industry (chemical manufacturers or their organisations, having a direct stake in the outcome of specific REACH dossiers).

The majority of respondents (20/33) stated that they had used (Q)SAR or *in silico* methods and intended to continue this use in the future. For those that stated that they had not previously used *in silico* techniques, it was highlighted that these respondents were open to their use in the future, if their roles were to allow them to do so.

The stakeholder respondents cited a total of 28 *in silico* models when asked which *in silico* models, methods or software they had previously applied. Of the 28 models cited, the OECD (Q)SAR Toolbox and EPIsuite were the most frequently utilised across each stakeholder group, with CAESAR being the only other tool which was cited by all groups. Additional methods which were cited in more than one stakeholder group included ECOSAR, SPARC and Toxtree. The remainder of the citations indicated a great disparity in the selection of models used within and across stakeholder groups (e.g. 11 different models were cited by a single respondent).

Respondents were also asked which domains (physicochemical properties, human toxicology, environmental fate properties or ecotoxicology) they most commonly used in their *in silico* assessments. The academics and consultants stated that they most frequently applied the physicochemical domain, whilst industry seemed to use domains for human toxicology and environmental fate properties. Within the stakeholder groups it was noted that few of the regulator and industry organisations had used all domains when applying *in silico* methods.

Regarding the functions for which respondents had applied *in silico* methods, academics and consultants adopted *in silico* methods as a time-saving device and for research, whereas industry and regulatory agencies primarily used *in silico* methods to generate supporting information to fulfil regulatory requirements. Several of the regulatory respondents stated that, although they had previously used (Q)SAR tools (primarily the OECD (Q)SAR Toolbox), they were not sure about the pertinent functions, flagging an unfamiliarity with the available tools as an issue. There was a consensus across stakeholder groups that in the future *in silico* methods will be useful in the prioritisation of chemicals of concern and for assessing thousands of chemicals simultaneously.

In addition to the current uses of *in silico* methods by stakeholder groups, the ORCHESTRA survey also tried to identify the barriers that prevent the use of *in silico* methods and ways in which their use might be increased. In general, regulatory bodies tended to highlight problems with the accessibility of these computational tools, such as cost, required expertise to correctly interpret the model outputs, knowledge and training. Many emphasised the lack of documentation, clarity, simplicity and transparency of these *in silico* models (see Mays *et al*., 2012 for further details).

---

[6] www.orchestra-qsar.eu [accessed March 2013]

2 Available Techniques

There are many regulatory programmes to assess the risks that chemical substances pose to humans or to the environment. For reasons of animal welfare and cost, it is necessary to limit the number of tests conducted where this is scientifically justifiable (OECD, 2007a).

Non-testing data can be generated by three main approaches.

- (Quantitative) structure–activity relationships ((Q)SAR) including the use of expert systems
- Read-across
- Chemical category formation

These approaches are based on the theory that similar chemicals should have similar biological activities (‘similarity principle’) and will be discussed further in this section.

2.1 Guidance on (Q)SAR

2.1.1 Explanation of the (Q)SAR concept and expert systems

(Q)SAR is the collective term for SAR and QSAR; these are theoretical models that can be used to predict the physical, biological or environmental properties of a chemical (ECHA, 2008). They are defined as follows.

Structure–activity relationship

A structure–activity relationship (SAR) is a qualitative relationship (i.e. an association) between a particular molecular structure or substructure and the presence or absence of a biological activity, or the capacity to modulate a biological activity imparted by another substructure. A substructure associated with the presence of a biological activity may be referred to as a structural alert (ECHA, 2008).

A SAR can also be based on the ensemble of steric and electronic features (biophore or toxicophore) considered necessary to ensure the intermolecular interaction with a specific biological target molecule, which results in the manifestation of a specific biological effect.

Similarly, a biophobe or toxicophobe refers to the features that are necessary to ensure the optimal intermolecular interactions with a specific biological target molecule that result in the absence of a specific toxic effect (OECD, 2004[7]).

Quantitative structure–activity relationship

A quantitative structure–activity relationship (QSAR) is a quantitative relationship between a biological activity (e.g. toxicity) and one or more molecular descriptors that are used to predict the activity (OECD, 2004); that is, it is a mathematical or statistical model that correlates quantitative parameters from the chemical structure (molecular properties) to a quantitative measure of a particular activity or property (ECHA, 2008). That is,

$$B = f\,(p_1,\ p_2,\ \ldots\ p_n)$$

where B is the level of biological activity, and p_1 to p_n are molecular properties.

Often, the relationship is linear, so we have

$$B = K_1p_1 + K_2p_2 + \ldots\ K_np_n$$

where k_1 to k_n are constants to be determined by statistical analysis.

The level of biological activity is often a measure of toxicity, such as a half-maximal effective

---

[7] http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2004)24 [accessed March 2013]

concentration (EC_{50}) or a minimum effective dose. Molecular properties can either be measured properties such as molecular weight, boiling point or ionisation energy; or they can be calculated properties such as molecular volume, hydration energy or polarisability. In statistical terminology, the biological activity is the dependent variable, and the molecular properties are the independent variables.

Mathematical or statistical QSAR models have often no mechanistic basis; hence, they tend to perform well with toxicity endpoints (e.g. irritation) with a single, well-defined molecular initiating event (MIE), but usually fail when modelling toxicity endpoints with multiple, competing MIE (e.g. repeated dose toxicity, carcinogenicity, reproductive toxicity). A MIE is the initial chemical-induced perturbation of some biological system.

Mathematical or statistical QSAR include local and global models. A QSAR model needs to have a balance between a wide applicability domain (see Section 2.1.2) and its predictive accuracy. A global QSAR model produces its prediction from a large, diverse dataset, covering a wide range of chemicals. Although the use of a global model extends the applicability domain, the predictive accuracy of these models is often weakened. Local QSAR models are applicable to fewer chemicals representing a narrower area of chemical space and produce their predictions from a narrower chemical subset, usually based on analogous chemicals. Global QSAR models have the advantage of being generally applicable (although detailed information on interactions may be lost), whereas local QSAR models may provide more insight into the mechanisms involved in the process (at the cost of being applicable to fewer chemicals). Several studies have shown that the use of a smaller dataset of structurally similar molecules improves models' predictive performance. For example, Bergstrom *et al.* (2004) divided a global QSAR database into several subgroups according to functional groups (e.g. acids, bases and ampholytes) to assess aqueous solubility and found that the use of two of the subgroups (bases and ampholytes) improved the model's predictive accuracy compared with the use of the global database.

Expert systems

An expert system is any formalised system (not necessarily computer-based) that enables a user to make rational predictions about the properties or activities of chemicals. All expert systems for the prediction of chemical toxicity or activities are built upon experimental data representing one or more effects of chemicals in biological systems (the database) and rules derived from such data (the rule-base; OECD, 2004). Expert systems are based on mechanistic descriptors (e.g. functional groups, structural alerts) or have a more mechanistic base. To be developed they require a detailed analysis of the toxicological endpoint data available in the scientific literature such that rules that relate structural features to toxicity can be derived. This is a slow and time-consuming process. They can be used for more heterogeneous groups of compounds than can straightforward SAR and QSAR, and they are suitable to identify active chemicals for a wide range of endpoints, including those toxicity endpoints involving multiple, competing MIE (e.g. repeated dose toxicity, carcinogenicity, reproductive toxicity).

The rule-base can be of two main types.

1 Rules based on statistics, where the models have been derived from the statistical analysis of data, using pattern recognition techniques – these are typically quantitative in nature.

2 Rules based on existing knowledge and expert judgement (based on knowledge of reactive chemistry) – these are typically qualitative in nature.

Therefore, expert systems are characterised into two main groups according to the rules employed. Automated rule induction systems primarily use statistically induced rules; examples include the commercially available models TOPKAT and MULTICASE and the freely available software tool Lazar. Knowledge-based systems primarily use rules derived from expert opinion (akin to asking a group of experts for their opinion). Examples of the latter include Derek Nexus, based on knowledge of structure–toxicity relationships and mechanisms of action, Toxtree, OncoLogic and HazardExpert. A simple example of such a rule is: 'A primary aromatic amine is probably carcinogenic'.

A third type of expert system, employed by some models, combines both knowledge-based and statistically based rules in a hybrid system. Hybrid expert systems use algorithms to fragment chemicals into potential structural alerts. Statistical methods are then used to select the structural alerts that can distinguish active chemicals from inactive chemicals. An example of this system is TIMES-SS, which is used to construct (Q)SAR models to estimate skin sensitisation potency.

The advantages and disadvantages of the three main types of expert system are summarised in Table 2.1.

The predictive ability of a knowledge-based expert system can be assessed in terms of:

Table 2.1 Comparison of the three main approaches in expert systems

Approach	Advantages	Disadvantages
Knowledge based	Mechanistically connected to the predicted endpoint Provides reasoning for the predictions In many cases supports the prediction with literature references or expert knowledge No need to invoke statistics to rationalise prediction	Often restricted and/or ill-defined applicability domain Usually cannot explain differences of the activity within a chemical class (is qualitative in nature) Usually has lower accuracy of prediction than statistical models Negative prediction could mean inactive chemical or insufficient knowledge in the expert system
Statistical	Usually has high accuracy of prediction Can be used for preliminary research when mechanism of action is unknown	Usually difficult to interpret the model predictions Often does not provide mechanistic reasoning of predictions Often non-transparent to the end user
Hybrid	Combines advantages of knowledge-based and statistical approaches, including mechanistic interpretability and overall accuracy	Likely to have restricted applicability domain

Source: JRC Report, 2011

- its sensitivity (the ratio of correctly predicted toxic chemicals to the total number of toxic chemicals tested)
- its specificity (the ratio of correctly predicted non-toxic compounds to the total number of non-toxic compounds tested)
- its concordance (the ratio of all correctly predicted compounds to the total number of compounds tested).

In the regulatory environment, high sensitivity is to be preferred in order to reduce the risk of false negative predictions. However, most knowledge-based expert systems currently have only moderate sensitivity. One reason for this is that, because a given type of toxicity probably covers only a small region of toxicological space (i.e. non-toxic chemicals usually occupy a much larger region of chemical space than do toxic compounds), it is easier to predict correctly that a chemical will be non-toxic than that it will be toxic (Cronin & Madden, 2010).

2.1.2 Validity, applicability and acceptance of (Q)SAR

In order for (Q)SAR results to be used in regulatory assessment as an acceptable alternative to experimental data the models should be scientifically valid. The (Q)SAR model should be assessed in terms of its validity with reference to the internationally agreed OECD principles for the validation of (Q)SAR. The OECD member countries and the EU adopted these principles in 2004 (ECHA, 2008).

Since non-testing data will be used in different ways depending on the context, fixed validation criteria are difficult to define. Therefore the validation principles for (Q)SAR are not criteria for the regulatory acceptance of (Q)SAR, but they identify useful information for their assessment (ECHA, 2008). The information on each substance should therefore be analysed as to whether it is adequate for regulation on a case-by-case basis (ECHA, 2008).

Validation principles

Introduction to validation: Definition of validation for (Q)SAR, validation process, and application of validation principles

The OECD member countries and the European Commission agreed a set of five principles of validation for (Q)SAR models for regulatory purposes. This agreement was made at the 37th Joint Meeting of the Chemicals' Committee and the working party of chemicals, pesticides and biotechnology in November 2004 (OECD, 2007a; ECHA, 2008). These principles should be read in conjunction with the associated explanatory notes, which were agreed at the same meeting, and a checklist, which was designed to provide guidance on the application of these principles, as shown in pp. 94–98 of the Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] Models (OECD, 2007a).

The agreed OECD principles of validation are as follows (OECD, 2007a; ECHA, 2008). A note identifying those responsible for applying these principles is given in parentheses.

To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information

1 A defined endpoint

A (Q)SAR model should be associated with a 'defined endpoint'. In this context an endpoint refers to any physicochemical, biological or environmental effect that can be measured and therefore modelled. This principle is intended to ensure transparency in the endpoint being predicted by the model, as otherwise different experimental protocols or conditions could give the same endpoint. (The developer)

2 An unambiguous algorithm

A (Q)SAR model should be expressed in the form of an 'unambiguous algorithm'. This principle is intended to ensure transparency of the model algorithm; however, it is recognised that in the case of commercial models, this information is not always publicly available. (The developer)

3 A defined domain of applicability

A (Q)SAR model should be associated with a 'defined domain of applicability'. This is because in terms of chemical structures, properties and mechanisms of action, a (Q)SAR is greatly simplified and is therefore subject to limitations. (Established by the developer but can be modified by the user)

4 Appropriate measures of goodness-of-fit, robustness and predictivity

A (Q)SAR model should be associated with 'appropriate measures of goodness-of-fit, robustness and predictivity'. Two types of information are therefore needed to fulfil this principle: a) the internal performance of the model (from the goodness-of-fit and robustness); and b) the predictivity of the model (from using an appropriate test set). (Mainly the developer)

5 A mechanistic interpretation, if possible

A (Q)SAR should be associated with a 'mechanistic interpretation' where this is possible. Providing a mechanistic interpretation is not always possible, but this principle is intended to ensure that the mechanistic associations between the descriptors in the model and the endpoint are assessed and, where possible, documented. (Expert systems tend to have an inherent mechanistic interpretation. Mathematical (Q)SAR do not tend to have a mechanistic interpretation, but if the developer tries to explain a mechanistic basis, the prediction is more robust. If there is no mechanistic basis, the user should check whether the prediction is linked to a mechanistic interpretation.)

The OECD Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (OECD, 2005) defines the term 'validation' as follows.

> "…the process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose" (OECD, 2005; ECHA, 2008)

This definition is difficult to interpret in the context of (Q)SAR and the validation principles, so an alternative definition has been proposed for the practical validation of (Q)SAR models intended for use in regulatory assessment.

> "The validation of a (Q)SAR is the process by which the performance and mechanistic interpretation of the model are assessed for a particular purpose" (ECHA, 2008)

In this context 'performance' of the model refers to goodness-of-fit, robustness and predictive ability. 'Purpose' of the (Q)SAR is expressed by the defined endpoint and applicability domain. 'Mechanistic interpretation' could be regarded as the scientific relevance of the model (ECHA, 2008). This definition therefore captures the principles of validation and these collectively would determine the validity (reliability and relevance) of the model (ECHA, 2008).

The term 'validation process' refers to any procedure during which the principles of validation are applied to a model or set of models (OECD, 2007a). This should include a judgement of the validity of the model (based on model performance, the endpoint and the chemical domain) and the relevance of the model for the particular regulatory purpose for which it is intended (OECD, 2007a). A dossier should be produced from the validation process and this should be structured according to the (Q)SAR validation principles. It may not be possible to fulfil all the principles of validation because (Q)SAR models may not have been designed with these in mind, or they may not have been reported with the details needed to fulfil them. Therefore the dossier should be as complete as possible and regulators will need to be flexible when considering the acceptability of the (Q)SAR, taking into consideration its constraints (OECD, 2007a).

These principles of validation are intended for use with a diverse range of model types, including SAR, (Q)SAR, decision trees, neural network models and expert systems. For systems that are based on multiple models, the principles of validation should be applied to the smallest component that functions independently (OECD, 2007a).

The use of (Q)SAR for regulatory assessment is highly dependent on the context. Users should be aware of features of the model that could contribute to a high or low performance for a particular situation (OECD, 2007a). Information on the model should be provided to the assessor of the prediction (e.g. the regulator) in the form of a reporting format (see Section 2.1.4). Examples of the sort of information that can be expected in one of these reports are given in the OECD's guidance on the validation of (Q)SAR models (OECD, 2007b) and includes the type of model, the definition of the model (including algorithms and applicability domains), the development of the model (including the availability of the training set), the validation and the applications of the model (OECD, 2007a).

Principle of defined endpoints

In this principle, the term 'endpoint' is wide ranging and could mean any of the following.

- A specific biological effect, for example increased serum levels of aminotransferases
- A variety of toxicological effects that could range from observations (clinical signs, e.g. change in body weight) to haematology, urinalysis or pathology findings (also called 'test protocol endpoints'; OECD, 2007a)
- Results of several tests conducted with different protocols, or the same protocol in different organisms (also called 'regulatory endpoints'; OECD, 2007a)

As a general rule, the closer the match between the endpoint predicted by the (Q)SAR and the endpoint required for the regulatory assessment, the more reliable the prediction made (OECD, 2007a). How reliable a particular (Q)SAR will be in a particular context should be assessed on a case-by-case basis. In any case, this principle is designed to ensure that detailed information is used as a basis for this decision (OECD, 2007a).

Defining the endpoint for which a model has been built determines the nature of the prediction that can be made. In some cases, the endpoint may be obvious, for example, a mouse oral LD_{50} measured at a specific time point. However, consider the case where the endpoint to be modelled is mutagenicity. Within this regulatory endpoint it is important to define what exactly has been modelled. Is it bacterial mutagenicity (Ames test), or chromosomal damage in mammalian cells, or point mutation in mammalian cells, or micronuclei in mice, or something else?

When considering the use of a (Q)SAR, the experimental protocol and data that were used to develop it should be assessed. This includes whether the data used were from a single experimental protocol or if data from several protocols have been combined. The design of the assay, the methods used and how the assay was evaluated should also be examined. The experimental protocol used will affect the variability of the dataset and therefore the quality of the model. In the ideal world (Q)SAR should be developed using a single protocol, but this is not always possible. Another issue facing regulators and companies using (Q)SAR is that this information is not always publicly available (OECD, 2007a). Some endpoints are extremely complex and cannot be accommodated by a single model. In these instances, different datasets may be used to evaluate the substance effectively (OECD, 2007a).

(Q)SAR are often grouped according to the defined endpoints that are associated with the OECD test guidelines. Table 2.2 shows the common regulatory endpoints associated with the OECD (2007a) guidelines.

Table 2.2 Common regulatory endpoints associated with OECD test guidelines

Type of endpoint or effect	Common regulatory endpoints
Physicochemical properties	Melting point
	Boiling point
	Vapour pressure
	K octanol/water
	K organic C/water
	Water solubility
Human health effects	Acute oral toxicity
	Acute inhalation toxicity
	Acute dermal toxicity
	Skin irritation/corrosion
	Eye irritation/corrosion
	Skin sensitisation
	Repeated dose toxicity
	Genotoxicity (*in vitro*)
	Genotoxicity (*in vitro*, non-bacterial)
	Genotoxicity (*in vivo*)
	Developmental toxicity
	Carcinogenicity
	Organ toxicity (e.g. hepatotoxicity, cardiotoxicity, nephrotoxicity)

If the same endpoint can result from several different mechanisms, the (Q)SAR must be:

- either developed for each specific mechanism and then applied to a specific class of chemicals;

- or a general (Q)SAR which models the same observed toxic effect for several classes of chemicals and toxicity mechanisms (OECD, 2007a). This relationship does not give an explanation to why a particular substance produces that particular endpoint, and in order to gain a better understanding of the mechanisms involved either multiple (Q)SAR models with different domains of applicability should be used for the same endpoint, or instead a statistical method capable of modelling across multiple mechanisms (OECD, 2007a).

A defined endpoint should therefore have the following features (OECD, 2007a).

1. Detailed information should be provided about the test protocols used to generate the test data, with special reference to factors that impact variability, uncertainties and deviations from the standard test guidelines.
2. Experimental protocols that have differences from the test protocol should not lead to an endpoint value that is markedly different.
3. Differences within the protocol should not lead to differences in the endpoint that cannot be rationalised.
4. The (Q)SAR should not be used for substances outside the chemical domain of the test protocol.
5. The endpoint predicted by the (Q)SAR and the endpoint measured by the test protocol should be the same.
6. A well-defined endpoint should reflect differences between chemical structures.

Principle of unambiguous algorithms

An important factor in the transparency of a model is a description of how the (Q)SAR estimates were produced. This allows others to reproduce the model in the future (OECD, 2007a).

The ‘algorithm’ of a model could be a mathematical equation between descriptor variables and response (endpoint) variables (in case of QSAR) or a knowledge-based rule (in case of SAR) that forms the relationship between the chemical structure/substructure or other molecular descriptors of the substance and the predicted endpoint (OECD, 2007a).

Molecular descriptors are used to describe different features of a chemical. They provide a means of representing molecular structures in a numerical form. The number may be a theoretical attribute (e.g. relating to size or shape) or measurable property (e.g. log P_{ow} or aqueous solubility). The most simple and commonly used molecular descriptors are molecular weight, atomic composition indices and atomic count descriptors. Other commonly used descriptors are those representing structural fragments (2-D descriptors). More complex descriptors are the topological descriptors which tend to represent structural features of the molecule such as size, shape, symmetry, branching, cyclicity and bond multiplicity. Equally complex are the electronic descriptors, which include the ability to cross biological membranes or bind to macromolecules (correlated with hydrogen bond donating/accepting ability or dipole interactions) and chemical reactivity associated with covalent binding (usually estimated from frontier molecular orbital information). Even more complex descriptors are the geometrical 3-D descriptors. These require knowledge of the relative positions of the atoms in 3-D space and tend to be applicable to mechanisms of action where toxicity is induced by a complex interaction between the chemical and the biological molecule (e.g. an enzyme or a receptor).

The majority of models contain algorithms that are from an unambiguous source and have been evaluated by peer review (Table 2.3). However, some algorithms do not describe how the estimate is produced, and for many commercial models the algorithm is not always publicly available (OECD, 2007a). For endpoints that are binary in nature, the (Q)SAR is primarily a classification model and the algorithm describes the presence or absence of a chemical substructure (OECD, 2007a). Some of the more exploratory algorithms are inherently ambiguous and these are not recommended for use in regulatory applications.

It is vital that (Q)SAR results are explained in such a way that they can be reproduced by other members of the scientific community. The algorithm used to relate the descriptors of the chemical structure to an endpoint is an extremely important part of this explanation, and (Q)SAR models which do not have this explanation would be less acceptable for regulatory use. An unambiguous algorithm enables the model to be tested and the user to explain how the (Q)SAR estimate was made (OECD, 2007a).

Any conclusions drawn from the model are related to and defined by the dataset used to develop the model. This is especially relevant for small datasets that used only a limited number of chemicals, or where the variability is large (OECD, 2007a).

When assessing the algorithm the following elements should be considered.

- The dataset of chemicals, endpoint values and molecular descriptor values

- How the descriptors were derived and measured
- The test dataset and training dataset, and any reasoning for the removal of outliers
- The mathematical model used
- Statistical parameters showing how the model performs
- The parameters and their values in the (Q)SAR (OECD, 2007a)

The training dataset includes a set of chemical substances with their molecular descriptors and *measured* endpoint values used to develop a (Q)SAR model.

Linear regression - Linear regression fits a straight line through the scatterplot of the molecular properties (or other explanatory variables) such that the distance from this line to the points is minimised (it is a 'line of best fit'). The distance is usually calculated as the sum of the squares of the distance from the points to the line, in which case it is sometimes called an ordinary least squares regression. Where there is more than one explanatory variable it is sometimes called multiple regression. It is possible to perform linear regression on transformed (e.g. log based) molecular properties.

Linear regression is well understood, computationally cheap and relatively easy to interpret. If there is reason to suspect that the biological behaviour can be explained by a small number of known molecular properties then this is the best option. Unfortunately, the analytical workflow becomes cumbersome where there are more than a few tens of descriptor variables, or if their effects may be non-linear.

Linear regression and genetic algorithms - In the case of many independent variables (molecular properties), rather than manually inspecting the results of a linear regression model, updating it and repeating it, genetic algorithms can be used to evolve automatically the 'best' model. This involves running a population of different models and calculating which ones best fit the data. Poorly fitting models are

Table 2.3 Some common algorithms used in (Q)SAR

Algorithm	Advantages	Disadvantages	Notes	Suitability
Univariate regression (linear and non-linear)	Describes a simple relationship between the descriptor and the endpoint	Non-linear models can be over fitted	Can be linear or non-linear	Univariate linear regression (ULR) is generally unambiguous
Multiple linear regression (MLR)	Applied when the endpoint needs more than one descriptor to be modelled	Possible over fitting Correlated variables can be an issue	Example of MLR: ordinary least squares	Produces a transparent and reproducible algorithm
Principal component analysis (PCA) & principal component regression (PCR)	Compresses data Enables users to identify patterns in data of high dimension	Some components are neglected, these may or may not be relevant Noise can remain in the model Can give too much weight to outliers	Has poor predictive ability	Suitable for dimension reduction by neglecting some components
Partial least squares	Useful when have colinearity in the descriptors	Outliers are difficult to identify Introduction of outliers afterwards also affects the model	Is a combination of MLR and PCR	Suitable to explain the variance in the independent variable
Artificial neural nets	These models are very flexible Useful in pattern recognition, process analysis and non-linear modelling	Demand a large amount of data	Learns from examples in the data	Ambiguous
Genetic algorithms	Not limited by statistical functions Very flexible	Strongly influenced by decisions of programmer Best suited to development of (Q)SAR (rather than use)	Contains a learning process	Ambiguous?

discontinued, while the better fitting ones spawn a set of offspring that have been randomly altered (mutated) from the parent model by adding or removing independent variables. After several generations the fit will be improved to some optimal value.

Genetic algorithms solve the limitations of linear regressions, allowing the developer to consider hundreds of molecular properties, and not needing to specify their effect beforehand. Running many models for many generations is more computationally expensive, though not prohibitively so. A larger problem is that the algorithms are not guaranteed to converge on a global optimum.

Principal component analysis - Principal component analysis (PCA) creates a new set of independent variables ('predictor variables') from the existing ones by using an orthogonal transformation. These predictor variables have the property of not being correlated with each other. As a bonus, the new variables can be scored by their contribution to the variance of the data. This allows the developer to reduce the number of variables in the model by discarding those with a score below some cut-off threshold. Principal component regression (PCR) takes the new set of independent variables and uses them in a linear regression. PCR is computationally cheap and allows the developer to simplify models from many descriptors down to a small number. However, the new variables (which are combinations of some of the original variables) may be hard to interpret, and partial least squares regression (see below) is widely considered to be superior in most cases.

Partial least squares - Partial least squares (PLSR) is a modern variation of PCR. Rather than choosing predictor variables that maximise the contribution to the variance of the dependent variable, the independent variables are chosen to maximise the covariance between each other. This reduces the impact of large but irrelevant predictor variables. PLSR shares many advantages and disadvantages with PCR, but should be preferred to PCR.

Principle of a defined domain of applicability

All (Q)SAR models have limitations based on the structure, physicochemical characteristics and activity of the chemicals used in the model training dataset, and this principle therefore establishes the scope of the model. A (Q)SAR can only be expected to be reliable for chemicals that are similar to those used to develop the model. Any predictions outside of this scope are less likely to be reliable (OECD, 2007a).

The scope of the (Q)SAR together with its limitations is referred to as the 'applicability domain' and when using a (Q)SAR it should be stated whether or not it is being used within its applicability domain. For some (Q)SAR models this may be a simple statement that the chemical is in or out of the domain. For other more quantitative assessments this may include a confidence interval that expresses the degree of similarity between the chemical of interest and the model training dataset (OECD, 2007a).

It is important to stress that the applicability domain is not a measure of performance. Making a prediction outside a model's applicability domain does not automatically mean that prediction is not accurate. It simply indicates that the model used to make that prediction is not designed to do so, and is being operated outside its intended area of relevance (i.e. prediction via extrapolation). However, in many cases, predicting outside a model's applicability domain does result in increased prediction error.

The applicability domain of a (Q)SAR model is defined as *"the response and chemical structure space in which the model makes predictions with a given reliability"* (OECD, 2007a).

In the case of a SAR, the applicability domain is a description of any limits on its applicability (e.g. inclusion and/or exclusion rules regarding the chemical classes to which the substructure is applicable). In the case of a QSAR, the applicability domain is a description of the ranges of the descriptor and response variables for which the QSAR makes reliable estimates.

'Chemical structure' could be information on physicochemical properties and/or structural fragments. 'Response' refers to any physicochemical, biological or environmental effect that is being predicted (i.e. the defined endpoint; OECD, 2007a). Information about the applicability domain helps the user of the model to judge whether the prediction will be appropriate for the chemical in question. The chemical would be assessed in terms of its similarity to the dataset used to develop the model (OECD, 2007a).

An applicability domain should be specific to that particular model and should be derived from the chemicals in the dataset, their molecular descriptors and the statistical approach used to develop the model. The applicability domain should be defined and documented by the developer of the model and include the following information (OECD, 2007a).

- A statement of the unambiguous algorithm
- Details of the dataset (chemical used, descriptors and endpoint values)

- Details of the statistical method to develop the model
- Structural requirements of chemicals to which the model applies

Validation of the applicability domain is not an absolute boundary for a given model. If the domain is very constrained the model will not be able to predict the endpoint for a larger number of chemicals; conversely if it is less constrained the (Q)SAR will be able to make predictions for more chemicals but the predictions may be less reliable (OECD, 2007a).

It should be borne in mind that using the (Q)SAR within its applicability domain does not guarantee that it will be reliable. Even if a chemical lies within the established scope of the applicability domain, it could act by a different mechanism of action that is not captured within the model. Conversely, if the applicability domain is defined in mechanistic terms, the (Q)SAR may be able reliably to predict beyond the defined dataset or applicability domain. Therefore, in addition to the physicochemical and structural domains, the mechanism of action of the chemicals used to develop the model is a useful addition to the applicability domain (OECD, 2007a).

The identification of mechanistic domains relies heavily on expert judgement at present; however some software tools are available which can assist, as in these examples (OECD, 2007a).

- Derek Nexus applies knowledge-based rules for toxicity prediction.
- HazardExpert issues an alert if a toxic fragment is found in the query molecule.
- Multicase identifies substructural molecular fragments and molecular descriptors that are correlated with specific toxicological activities.
- MDL–(Q)SAR identifies molecular descriptors that are correlated with specific types of toxicological activity.
- TOPKAT classifies substances into chemical classes and applies quantitative models for toxicity prediction.
- CATABOL, META, MetabolExpert and METEOR are tools for metabolism prediction and can be used to predict the metabolites of the substance.

Ideally when assessing the applicability domain, information on the query molecule and the (Q)SAR dataset should be provided in terms of structural and physicochemical similarity. This can be expressed either qualitatively or quantitatively. In addition, the mechanism by which the query chemical acts should also be compared with the mechanism of the chemicals in the (Q)SAR dataset. If this assessment is not possible, the similarity may also be expressed statistically (OECD, 2007a).

Some software tools, such as MultiCASE, Leadscope and Prediction Model Builder, analyse the query molecule by looking at the molecular fragments. They check whether these fragments are represented within the dataset of the (Q)SAR. The higher the incidence of the fragments occurring in the dataset, the more confidence exists that the query chemical will be predicted reliably (OECD, 2007a).

Other software tools, such as Ambit Disclosure software, Leadscope and MDL–(Q)SAR, use algorithms to measure similarity (OECD, 2007a).

The applicability domain should also be endpoint-specific because different classes of chemicals may still behave in the same way for a different endpoint.

Commercial (Q)SAR tools vary in extent of available information concerning the applicability domain. For example, Derek Nexus will fire an alert if a substance has a particular structural feature, but expert judgement is needed to decide if this is relevant in the required context. TOPKAT enables the user to assess whether the substance falls within the applicability domain in terms of the fragment and descriptors. It also tells the user if the substance is within its own database and retrieves similar chemicals and their data (ECHA, 2008).

If multiple (Q)SAR are being used to predict the same endpoint there are two approaches to validating the applicability domain.

1 When the query chemical falls within both (or several) applicability domains, the confidence can be found by averaging the individual predictions. If the prediction is the same for two or more models, confidence should be higher than for a single model. The common applicability domain is narrow for multiple models, so this restricts the number of chemicals that can be predicted (OECD, 2007a).

2 Multiple (Q)SAR are chosen that use different algorithms, but have high specificity for the endpoint. The collective applicability domain is wider because the (Q)SAR do not necessarily overlap. This approach is used to gain insight into certain molecular properties that are correlated with the same endpoint or toxic effect (OECD, 2007a).

A stepwise approach has been proposed to determine the applicability domain of a model, as follows.

1. Identify whether the query chemical falls within the range of the physicochemical properties for the model
2. Define the structural similarity between the query chemical and those within the dataset of the model
3. Check that the query chemical contains the reactive groups that are thought to cause the desired effect (mechanistic check)
4. Check that the query chemical will be metabolically active, as chemicals that undergo metabolic transformation may provide an inaccurate prediction

These four stages should be applied in a sequential manner and those that satisfy all four conditions have an increased reliability in their prediction. However, this rigorous approach also reduces the number of chemicals that can be predicted (OECD, 2007a).

Although an understanding of the applicability domain of the model can either increase or decrease confidence in a (Q)SAR estimate, it should be noted that the applicability domain can never provide absolute certainty about the estimate. Even if the query chemical is well within the defined applicability domain, the prediction may still be unreliable. The reverse can also be true where the query chemical is outside the applicability domain and yet the prediction is reliable. Applicability domains should therefore be used to support the decision made by expert judgement (OECD, 2007a).

The identification of special atom arrangements that cause certain types of toxicity (toxicophores) provides a way of defining mechanistic domains. Within these mechanistic-based domains, chemicals that contain multiple functional groups deserve special attention as their toxicological activity can be modulated by the presence of additional functional groups.

Principle of measures of goodness-of-fit, robustness and predictivity

This principle is required in order to establish the performance of the model, including both internal model performance (goodness-of-fit and robustness) and external model performance (predictivity). Assessment of the performance of the model is also called ‘statistical validation’. This statistical validation should be considered in combination with knowledge of the applicability domain of the model (OECD, 2007a).

‘Goodness-of-fit’ is a measure of how well the model accounts for the variance of the data it is based on (OECD, 2007a). The goodness-of-fit of a model to its original training dataset is the absolute minimum of information needed to assess the performance of the model. It shows whether a model is statistically significant and how much the molecular descriptors account for the variation in the training dataset (OECD, 2007a).

The ‘robustness’ of a model refers to the stability of the prediction when one or more data points change (OECD, 2007a). It demonstrates how sensitive the parameters and predictions are to changes in the training dataset. If a model is not robust then it is unlikely to be useful for predictive purposes (OECD, 2007a).

The ‘predictive ability’ of the model is a measure of how well the model can predict new data (OECD, 2007a). Models that are too complex (over fitted) will not predict data as reliably as the internal validation may indicate; therefore it is important to evaluate the predictivity of the model as well. External validation is generally regarded as the most rigorous assessment of predictivity, since predictions are made for chemicals not used in model development.

Models with high statistical performance are likely to have the following features (OECD, 2007a).

- A high power of prediction with a minimum number of variables
- Low correlation between the predictor variables

Models with low statistical performance are likely to have the following features (OECD, 2007a).

- A lack of one or more relevant variables (i.e. insufficient fitting capability)
- A marked difference between goodness-of-fit and predictivity
- One or more irrelevant variables correlated with the endpoint by chance alone
- A high correlation between the predictor variables

The goodness-of-fit and robustness of (Q)SAR models have been evaluated for a variety of common algorithms, as shown in Appendix 2.

One of the most important factors when choosing a (Q)SAR model is its ability to accurately predict the desired endpoint with a substance that was not used in its development. In order to assess this predictive power, the (Q)SAR should be validated with chemicals that were outside the original dataset and thus were not used to develop it. External validation should be used to supplement the internal validation procedures, so that a model is shown to be both robust and predictive (OECD, 2007a).

Owing to the practical difficulties in obtaining new, experimentally tested chemicals, in terms of cost, time and animal welfare, the dataset available to develop a model is usually split into two; one is used to develop the (Q)SAR model whilst the other is used to validate it. When dividing the dataset in this way the chemicals should be split into groups according to pre-defined and suitable criteria, such as the experimental design used to generate the data.

In assessing the relevance of a negative prediction one needs to be clear about what a 'negative' actually means in terms of the particular model used. For example, a result of 'nothing to report' given by Derek Nexus has two possible explanations: either the substance class has not been studied for inclusion, or the substance class has been studied but there is no evidence of toxicity. The user has to determine the correct interpretation.

In addition to the data-driven (Q)SAR models described above that are based on mathematical algorithms, there are knowledge-driven (Q)SAR. Knowledge-driven, mechanistic-based or rule-based (Q)SAR assign chemicals to classes or groups before attempting to predict the endpoint.

One of the challenges in using knowledge-based systems is how to assess the predictive ability of the model and the significance of a certain chemical property or mechanism assigned to the target chemical, compared with the conventional statistical methods outlined previously (OECD, 2007a). Derek Nexus provides the following likelihood levels (i.e. an indication of probability) for its predictions, which are based on reasoning.

- Certain: there is proof that the proposition is true.
- Probable: there is at least one strong argument that the proposition is true and there are no arguments against it.
- Plausible: the WoE supports the proposition.
- Equivocal: there is an equal WoE.
- Doubted: the WoE opposes the proposition.
- Improbable: there is at least one strong argument that the proposition is false and there are no arguments that it is true.
- Impossible: there is proof that the proposition is false.

Principle of mechanistic interpretation

The credibility and acceptance of a model is increased when the validation process is consistent with existing theories and knowledge from chemistry or toxicology. This is because it enables the user to explain how and why an estimated value was produced. A mechanistic interpretation of a (Q)SAR model can add to the understanding of the statistical validity and applicability domain (OECD, 2007a).

The principle includes the wording "if possible" because the evolution of a (Q)SAR model is an iterative process that involves the statistical explanation of data, hypothesis generation and hypothesis testing. The iterative process generally leads to a series of refinements to the training set, both in terms of chemicals included and their molecular descriptors.

The basis of a mechanistic interpretation is that the biological properties of a chemical are inherently linked to the molecular structure and its attributes. In (Q)SAR models, the structural attributes of a chemical are represented in a mathematical form by the so-called molecular descriptors. If the molecular descriptors cannot be associated with changes in the hydrophobic, electronic and steric attributes of the chemical, then it would be very unlikely that there is a mechanistic basis for the model (OECD, 2007a).

There is a general drive to extend the domain of applicability of (Q)SAR, but this will generally reduce the mechanistic relevance of the model. Combining (Q)SAR into an expert system is a long-term solution to this problem as it improves the overall performance of (Q)SAR predictions (OECD, 2007a).

There are two types of parameter or descriptor used in (Q)SAR models.

1 Those derived from a measurable property of the molecule, for example vapour pressure, partition coefficient, dissociation constant

2 Those used to quantify important attributes of the chemical structure, also called molecular descriptors

Molecular descriptors are usually derived using a computational method. They are a very important starting point in the development of (Q)SAR, so it is essential that the method of calculating the molecular descriptors is also available to the user. It is also important that, when selecting descriptors, the role that these descriptors play either in the way the chemical behaves or in the way the endpoint is expressed should be known.

An example of a mechanistic-based molecular descriptor is a substructure within the whole molecule that indicates a potential for biological activity (e.g. a functional unit). (Q)SAR models based on substructures have the advantage that the query chemical can be assessed against an extensive number of substructures. However, if the query chemical does not possess any substructure present in the original dataset, no prediction can be made. Also, some interactions between substructures may be difficult to anticipate (OECD, 2007a).

Other examples of mechanistic-based descriptors are the molecular connectivity indices (MCI). These descriptors are based on topological information and are related to molecular volume, surface area, presence of branching and potential steric hindrance, presence of rings and their substituents.

More recently, another type of mechanistic-based descriptor has been developed, called electronic descriptors. Examples of electronic descriptors are the dipole moment, molecular polarisability, solvent accessible surface area, atomic charge on an atom, and nucleophilic and electrophilic bonds.

Expert systems are based on mechanistic descriptors or have a mechanistic base. The expert knowledge incorporated into Derek Nexus is based on a series of structural alerts or toxicophores associated with certain types of toxic activity. When a query structure is processed the alerts that match are displayed and the endpoints that are triggered by these alerts are indicated. In addition, the bibliographic references associated with the matched alerts and the triggered endpoints are shown (OECD, 2007a).

The mechanistic interpretation of a (Q)SAR can be established in two ways.

- The descriptors are selected before modelling on the basis of their known or anticipated role in driving the response (called *a priori*).
- The descriptors are selected on the basis of statistical fit alone (called *a posteriori*) with their mechanistic rationale being formalised after modelling.

For (Q)SAR with continuous descriptors, the mechanistic interpretation can be based on the physicochemical interpretation of each descriptor and its association with a mechanism of action (OECD, 2007a).

For SAR, the mechanistic interpretation can be based on the chemical reactivity or interaction of the substructure of interest (OECD, 2007a).

For expert systems, the mechanistic interpretation is based on expert knowledge and learned rules.

Reliability of (Q)SAR prediction

Once a model has been assessed as valid it is then necessary to check that the (Q)SAR is appropriate for the query chemical and to evaluate whether the prediction is reliable. This assessment can be divided into the following stages.

1. Check that the query chemical is within the applicability domain of the model.
2. Check that the applicability domain is suitable for the required regulatory purpose. The reliability of (Q)SAR predictions is a relative concept and depends upon the context of application; a greater or lesser degree of reliability would be acceptable for different applications. The applicability domain can therefore be defined to suit the particular regulatory context (ECHA, 2008). Most of the currently available models were not developed with regulatory assessment in mind. They also contain biases within the model, which may or may not be useful, depending on the use required. For example, some models are biased towards certain types of chemical for which there is no regulatory requirement, and others are biased towards a certain type of prediction, for example optimised to correctly identify positives at the expenses of correctly identifying negatives. Such biases within the model do not necessarily affect its validity, but they may affect its applicability. So it is necessary to check that there is a good match between the applicability of the (Q)SAR and the regulatory purpose (ECHA, 2008).
3. Check how well the model predicts chemicals similar to the query chemical. This is performed to ensure a model is appropriate and should be done with a chemical analogous to the substance of interest and for which measured values exist. This is effectively performing read-across to support the use of the model (read-across is described in more detail in Section 2.2; ECHA, 2008).
4. Check that the model estimate is reasonable according to other information on the query chemical. This stage of the process requires expert judgement. One approach is to compare the calculated values for the query chemical and its analogues to measured data.

Adequacy of (Q)SAR prediction

If a model is valid and considered reliable it should lastly be evaluated in terms of the extent that it fulfils the necessary regulatory requirements. Computer-based tools are available for this process, but expert judgement is also important (ECHA, 2008).

In order for a (Q)SAR to be adequate for a regulatory purpose it must fulfil the following four conditions (ECHA, 2008).

1. The model used is shown to be scientifically valid.
2. The model used is applicable to the query chemical with the necessary level of reliability (i.e. the compound of interest falls within the applicability domain in terms of structure, physicochemical properties, metabolism and mechanism of action).
3. The prediction is relevant for the regulatory purpose.
4. Appropriate documentation on the model and prediction is given.

During assessment of a chemical it is also necessary both to consider the completeness of the overall information on the chemical and document the adequacy of the (Q)SAR results using the appropriate reporting format.

Only limited guidance is available on the regulatory acceptance of (Q)SAR predictions. However, in general, the following principles should be taken into account in decision making for regulatory purposes.

- The principle of proportionality: the amount of information needed is dependent on the importance of the decision that will result from it.
- The principle of caution: the amount of information needed is dependent on the risk the substance poses: the more severe the consequences, the more conservative the approach.

These principles specify that the level of validity required from a model can change depending on the regulatory decision being made (ECHA, 2008).

2.1.3 Regulatory use of (Q)SAR

Current experience

(Q)SAR have been used quite widely in EU regulatory programmes; however, there is little documentation available to describe why a particular approach was taken (ECHA, 2008).

Some ways that (Q)SAR have been used in the past are listed and summarised below.

- To provide data when testing is not technically possible
- To provide data when it is not available for a substance that is not a priority
- To assess the reliability of measured data
- To estimate properties for a range of components in a multi-constituent substance
- To provide information on environmental effects and estimate environmental fate data
- To argue against the need for other tests to be done; the absence of a reactive substructure can be used to justify the omission of some tests
- To argue for other more unusual testing to be performed
- To provide information about mode of action and sensitivity for ecotoxicity tests
- To provide supporting information for mode of uptake or toxicokinetics

Ways that (Q)SAR have been used in classification and labelling include the following.

- Self-classification: A number of industry sectors have published guidance for the self- classification of chemicals. To support the self-classification process, the Danish Environmental Protection Agency (EPA) has published an advisory list for self-classification of dangerous substances that was developed using the Danish EPA (Q)SAR database (ECHA, 2008).
- EU classification according to Directive 67/548/EEC: Classifications in Annex I of the directive (now in Annex VI of the Regulation on classification, labelling and packaging of substances and mixtures – CLP) have been agreed by EU member states and are legally binding. The criteria by which decisions on classification and labelling are made under this directive are largely based on experimental results but, where appropriate, validated (Q)SAR can be used (ECHA, 2008). Likewise, under CLP, a WoE approach can be used, which can include (Q)SAR.

Ways that (Q)SAR have been used in persistence, bioaccumulation and toxicity (PBT) assessment and very persistent and very bioaccumulative (vPvB) assessment include the following.

- In combination with experimental data
- For the selection of PBT candidates where experimental data did not exist or were considered unreliable
- To confirm or negate PBT status
- As an initial screening exercise
- To deselect PBT candidates from further consideration
- For short-term aquatic toxicity to algae, fish and *Daphnia*
- To evaluate constituents of multi-component mixtures

The performance of (Q)SAR analysis is required for two aspects of the evaluation of pesticides in the EU.

1 Where a new impurity is present in a new source of a pesticide active substance, or the levels of an existing impurity are increased significantly, a (Q)SAR should be performed and any alerts compared with alerts produced by the active substance itself (EC, 2011).
2 (Q)SAR is cited as a technique of potential value in the evaluation of pesticide metabolites that can leach into groundwater at levels above 0.1 µg l^{-1} (EC, 2003).

(Q)SAR has also been proposed as one of the tools of potential value for use in other areas of pesticide assessment where there are limited data available, for example in assessment of metabolites present in edible parts of treated crops.

(Q)SAR analysis is generally performed as a standalone approach, or in combination with other techniques in the hazard assessment of environmental contaminants and food flavourings, as test data are rarely available for these chemicals and/or there are no legal instruments to enforce testing.

Since (Q)SAR are mathematical models based on an underlying dataset, they have an inherent degree of uncertainty. This is caused by two main factors.

- The variability of the original data
- A model only represents part of the real situation; it cannot model all types of mechanism or chemical.

2.1.4 Reporting formats

When using (Q)SAR instead of test data, it is important that adequate and reliable documentation is provided. Several (Q)SAR reporting formats have been developed that demonstrate the documentation required for (Q)SAR. These are designed to provide a standard framework for the key information about (Q)SAR and to ensure the same information is available to industry and regulators alike (ECHA, 2008).

There are two different reporting formats, designed to capture different types of information.

1 The (Q)SAR model reporting format (QMRF) includes a description of the algorithm, the model development and validation according to the OECD principles and evaluation studies performed with the model (ECHA, 2008). The structure of the QMRF is designed to evaluate the (Q)SAR model according to the OECD validation principles and it should be used to facilitate the regulatory process. The QMRF has been agreed at OECD level. The QMRF may also contain information about the original dataset to develop the model and the test set, that is chemical name, identifying numbers (e.g. Chemical Abstracts Service (CAS) number), simplified molecular-input line-entry system (SMILES), IUPAC International Chemical Identifier (InChl), format used to encode molecular structures (mol file), structural formula, values for the dependant variable and for the descriptors (ECHA, 2008).
2 The (Q)SAR prediction reporting format (QPRF) explains how the estimate was derived by applying a specific model to a specific substance. It should include information of the model prediction, the endpoint, the substance modelled, the relationship between the substance and the applicability domain, analogues of the substance and the relationship between the predicted endpoint and the regulatory endpoint (ECHA, 2008). In other words, it is a framework to evaluate the specific substance by the chosen model. The prediction made for the query chemical should be presented and assessed for reliability, making reference to the QMRF.

These reporting formats collectively provide a comprehensive description of the use of the (Q)SAR and other information required for the assessment of the substance. The structure of the formats is not meant to be fixed, and it is expected that they will evolve as experience is gained, but they should be designed to ensure transparency, consistency and acceptability (ECHA, 2008).

Some examples of reporting formats are given in the OECD's guidance on the validation of (Q)SAR models (OECD, 2007a).

2.1.5 Computational tools for applying (Q)SAR

Introduction: (Q)SAR and knowledge-based tools

There is a wide variety of publicly available and commercial tools for the development and application of (Q)SAR. These include tools for data management, data mining, descriptor generation, molecular similarity analysis, analogue searching and hazard assessment (ECHA, 2008).

If any (Q)SAR model is intended as a replacement for experimental data, the model should be validated according to the OECD principles and documented using the appropriate reporting formats, as described in previous sections (ECHA, 2008).

Summary tables of computational tools

Some computational tools can be regarded primarily as databases used for data management, data mining and analogue searching. Other computational tools are primarily (Q)SAR applications for toxicity and/or endpoint prediction. Some of the most commonly used of these are presented in Table 2.4 and Table 2.5 below.

It has been pointed out that most of the publications by software developers claim good predictions, often with better than 90% concordance. However, valid comparisons of performance need to be carried out independently of software developers, using a given test set. Of course, no test set gives perfect coverage of chemical or biological space, and different test sets will almost always give different results (Cronin & Madden, 2010).

Improved predictivity can often be achieved by combining the predictions from two or more expert systems and/or (Q)SAR; this is most probably because individual prediction errors are to some extent averaged out. A consensus prediction can simply be the arithmetic mean of all predictions, or it can involve using a leverage-weighted mean, in which the most predictive model has the greatest contribution. For classification endpoints, using consensus-positive predictions from any two or more programs can yield improved predictivity (Cronin & Madden, 2010).

There are now several *in silico* tools for the prediction of a range of toxicity endpoints. All depend on the quality and availability of experimental data; these are to some extent still lacking and are, in some cases, of dubious accuracy. Hence, a primary ongoing need for *in silico* prediction is more and better experimental data available in the public domain.

There has been a rapid increase in recent years in the number of software models available for toxicity prediction. Many of these meet the OECD Principles for the Validation of (Q)SAR, but some do not – or at least it is not clear from the limited information on company websites whether or not they do so. It is recommended that all developers and marketers of toxicity prediction software publish this information.

Table 2.4 Summary table of the most commonly used databases

Name	Overview, uses	Details, what it does	Search options
Ambit & Ambit Discovery Developed by CEFIC	Data management, searchable databases, tools for grouping & applicability domain assessment	**Ambit** stores chemical structures, identifiers (e.g. CAS), attributes (e.g. molecular descriptors), experimental data, test descriptors and literature references; software generates 2-D & 3-D molecular descriptors **Ambit Discovery** performs grouping, assesses applicability domain with statistical methods or using mechanistic understanding or structural similarity	Search by name, CAS, SMILES, substructure, structure-based similarity & chemical identifier
Danish (Q)SAR database Developed for JRC by the Danish EPA	Database of (Q)SAR predictions for 166 000 organic chemicals & a range of endpoints	Contains approximately 60 predictions for each chemical including a yes/no statement for MultiCase predictions for assessment of the applicability domain Has endpoints for physiochemical properties, fate, ecotoxicity, absorption, metabolism & toxicity	Search by structure, (substructure/exact match), identifiers (e.g. name, CAS) & endpoint
JRC (Q)SAR Model Database Currently being developed by JRC	Searchable tool for linking chemicals to (Q)SAR	(Q)SAR summaries being compiled in the standard (Q)SAR model reporting format	Search by chemical (CAS, EC number, structure), endpoint, descriptors & model author

Source: ECHA (2008)
CEFIC, Conseil Européen des Fédérations de l'Industrie Chimique; Danish EPA, Danish Environmental Protection Agency;
JRC, Joint Research Council
Sources of these databases are given in Appendix 3

Table 2.5 Summary table of the most commonly used computational (Q)SAR applications for endpoint prediction

Name	Overview, uses	Details, what it does
Toxtree Developed for JRC by Ideaconsult Ltd	Estimates different types of toxic hazard using structural rules Includes options for applying the Cramer decision tree	Cramer scheme: uses chemical structures to estimate the threshold of toxicological concern Uses recognised pathways for metabolic deactivation/activation, toxicity, presence in food and endogenous metabolites to categorise substances into three classes based on their predicted toxicity, Class 3 being the most toxic
(Q)SAR Toolbox Developed by OECD	Can be used to implement stepwise approaches for analogue and categories	Summarises information about the validation results of each model according to the OECD validation principles to enable user to decide which are appropriate for that particular regulatory purpose Lists analogues with their estimates Gives information about metabolite activation/detoxification
OncoLogic®	Expert system that assesses potential of chemicals to cause cancer	Uses rules from SAR, mechanisms of action and epidemiology studies Produces assessment of potential carcinogenicity and detailed reasoning for this outcome from information about the compound structure Consists of four subsystems that evaluate fibres, metals, polymers and organic chemicals
Derek Nexus Developed by Lhasa Ltd	Knowledge-based expert system Uses knowledge of structure–toxicity relationships, mechanisms of action & metabolism	Has over 504 alerts covering wide range of toxicological endpoints (mutagenicity, carcinogenicity, organ toxicity, reproductive toxicity, skin irritation and skin sensitisation) Alerts consist of a toxicophore (substructure thought to be responsible for toxicity) and literature references, comments and examples Main strengths are in mutagenicity, carcinogenicity and skin sensitisation Rules are based on mechanisms of action, chemical class or empirical relationships Has nine levels of confidence for result: certain, probable, plausible, equivocal, doubted, improbably, impossible, open and contradicted Has been found to comply well with the OECD Principles for the Validation of (Q)SAR Concordances of 60%, 60% and 53% have been reported for skin sensitisation, mutagenicity and carcinogenicity, respectively (Cronin & Madden, 2010)
HazardExpert Developed by CompuDrug Ltd	Predicts toxicity, bioavailability & bioaccumulation of compounds	Calculates log P and pK_a Default knowledge base is based on information from US EPA and information collected by CompuDrug Ltd Rule-based system allows the user to understand, expand, modify or optimise the data Covers the following endpoints: oncogenicity, mutagenicity, teratogenicity, membrane irritation, sensitisation, immunotoxicity & neurotoxicity Has a link to MetabolExpert system
TOPKAT Developed by Accelrys Inc.	Statistical system consisting of suite of (Q)SAR for a range of endpoints	16 models for the following endpoints: aerobic biodegradability, Ames mutagenicity, *Daphnia magna* EC_{50}, developmental toxicity, fathead minnow LC_{50}, FDA rodent carcinogenicity, NTP rodent carcinogenicity, ocular irritancy, log K_{ow}, rabbit skin irritancy, rat chronic LOAEL, rat inhalation toxicity LC_{50}, rat MTD, rat oral LD_{50}, skin sensitisation & WoE rodent carcinogenicity Models are typically based on large toxicological datasets

Name	Overview, uses	Details, what it does
The CASE family of methods Developed by Klopman & Rosenkranz	A variety of models for a multitude of endpoints and hardware platforms, including CASE, MULTICASE, MCASE, CASETOX & TOXALERT	Uses fragment-based technology and statistical analysis of a database of chemicals and their toxicity data The program checks the chemical substructures for biophores (substructures thought to be responsible for activity) and orders the chemicals according to statistical significance (i.e. the top biophore is the one responsible for the largest number of active molecules); next the program looks for ‘modulators’, substructures that significantly alter the activity of the biophores This information is then used to predict the biological activity of new chemicals that are not in the dataset It covers the following endpoints: carcinogenicity, mutagenicity, teratogenicity, irritation, developmental toxicity, acute toxicity & biodegredation
TIMES (tissue metabolism simulator) Developed by Laboratory of Mathematical Chemistry (LMC)	Integrates on the same platform metabolic simulators and (Q)SAR models for predicting toxicity of selected metabolites	(Q)SAR models for predicting toxicity of metabolites Can predict skin sensitisation, mutagenicity, chromosomal aberration & ER (oestrogen)/AR (androgen) binding affinities of chemicals, while also accounting for metabolic activation Can also predict toxicity to aquatic species Models toxicity for two types of toxicochemical domain: reversible (non-covalent) and irreversible covalent bioreactive chemicals Models metabolism using a combination of rules and a library of biotransformations and abiotic reactions User can calibrate probability of metabolites according to the situation, e.g. to limit the metabolites to just the likely ones Consideration can be given to physicochemical properties, e.g. water solubility or log K_{ow}
Meteor Developed by Lhasa Ltd	Predicts metabolic fate of a query chemical using knowledge of structure–metabolism rules	Based on biotransformation rules Has an integrated reasoning model which allows the system to evaluate the likelihood of each biotransformation taking place; potential biotransformations can be compared User can analyse the query at different search levels, depending on likelihood of biotransformations User can also add own rules, search metabolic tree and view/graph/use individual biotransformations
CATABOL	Simulates biodegradation pathways and predicts physicochemical & toxic endpoints	Gives a quantitative assessment by using a mechanistic approach Biodegradability simulator includes a library of transformations, ordered by a hierarchy; these transformations are adjusted to reproduce the degradation pathways for over 500 chemicals

Source: ECHA (2008)
MTD, maximum tolerated dose
Sources of these applications are given in Appendix 3

Faced with the need to make a toxicity prediction, how should one decide which model to use? There are several considerations to take into account. First, of course, is the selection of one or more models that offer the requisite endpoint. Then comes the key aspect of cost. A few models (e.g. CAESAR, OncoLogic, TIMES, Toxtree) are freely available. The price of others varies greatly (Cronin & Madden, 2010).

2.1.6 Computational (Q)SAR tools for human health endpoints

Tools and databases for prediction of toxicokinetic properties

As the overall toxicity elicited by a chemical may be determined as much by its kinetic properties as its intrinsic toxicity, prediction of biokinetic properties is essential in assessing the overall toxicity of a substance.

Public availability of large, high-quality datasets of toxicokinetic properties is limited, although increasing awareness of the importance of such properties is leading to more data being published. Much of the available toxicokinetic data and models relate to pharmaceutical chemicals. This presents two difficulties. First, the available data are skewed towards those chemicals which are pharmacokinetically viable. Secondly, the chemical space covered by such models may not be representative of the chemical space of environmentally relevant chemicals. Indeed the lack of kinetic information for non-pharmaceuticals presents a real challenge in the risk assessment of industrial chemicals. Another important consideration in the modelling of toxicokinetics is that, whereas toxicity can be related to specific interactions, kinetic characteristics are often associated with more global properties of the molecule such as the influence of the log P_{ow} of the compound on its ability to penetrate membranes (Cronin & Madden, 2010).

Despite difficulties in the modelling of toxicokinetics, a large number of *in silico* prediction tools have been developed, including models for human intestinal absorption, human oral bioavailability, blood–brain barrier permeability, plasma protein binding, metabolism and excretion or clearance.

Metabolism is of particular relevance to toxicity prediction because of the potential to generate toxic metabolites from innocuous parent compounds. For this reason, methods to predict metabolism have generated much interest. Many of these models are based on knowledge of general reaction chemistry or specific, known metabolic pathways relevant to the functional group identified within the molecule (Cronin & Madden, 2010).

The vast majority of available software tools are commercial. The tools differ greatly in terms of their capabilities and applications (see Table 2.6 below).

QSAR for toxicokinetic properties tend to be local models; that is, they are based on small, homogeneous datasets, 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 (JRC Report, 2011).

Tools and databases for acute toxicity prediction

There are only a few (Q)SAR and expert systems that are capable of modelling acute toxicity; this is because the endpoint itself is very complex. Acute toxicity could be a result of several mechanisms acting on the whole body resulting in a wide spectrum of biokinetic, cellular and molecular events, making it difficult to model. Converting the complex, whole-body phenomena related to acute toxicity into a simple number necessarily leads to a loss of information. Also, 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 (JRC Report, 2011).

Most literature-based models are restricted to single classes of chemicals, for example phenols, alcohols or anilines (ECHA, 2008). Some examples of the use of (Q)SAR to predict acute toxicity are listed below.

- Acute inhalation toxicity – Some regression models have been developed to predict inhalation toxicity for volatile substances; these typically use vapour pressure and boiling point as parameters. For example, the estimation of acute neurotoxicity by Cronin (1996).
- LD_{50} prediction – There are a few sets of models that predict the LD_{50} for small sets of compounds. For example, Hansch & Kurup (2003) used data from Cope (1939) to develop a (Q)SAR to predict the LD_{50} of barbiturates in female white mice.
- Human toxicity – The human LD_{100} values of various drugs have been estimated using data from King (1985).

Table 2.6 Commercially available computational tools to predict toxicokinetics

Software	Provider	Predicted toxicokinetic properties
Discovery Studio ADMET	Accelrys	Absorption, plasma protein binding, CYP2D6 binding
MetabolExpert	CompuDrug	Metabolic fate of compound
Meteor	Lhasa	Metabolic fate of compound
META/METAPC (from the CASE family of models)	MultiCASE	Metabolic transformations
ADME Boxes	ACD Labs	Absorption, bioavailability, plasma protein binding, volume of distribution

Sources of these tools are given in Appendix 3

- Prediction of *in vitro* effects – A number of (Q)SAR for cytotoxicity have been developed for *in vitro* testing; these are not currently directly relevant to the assessment of acute toxicity for regulatory purposes, but since reliable (Q)SAR can be developed for cytotoxicity of defined groups of chemicals *in vitro*, these methods look promising for the future (ECHA, 2008).

Some information about the acute toxic effects and corrosivity of a substance can be gained from its physicochemical properties. The vapour pressure and log K_{ow} are examples of properties that determine the feasibility of exposure and the possible distribution in the airways. Particle size is also important when looking at deposition of the substance in the respiratory tract and potential toxicity. Some properties are also important when assessing potential dermal exposure; these include log K_{ow}, molecular weight, degree of hydrogen bonding and melting point (ECHA, 2008).

Some (Q)SAR can be used to predict dermal penetration or metabolic pathways. However, these have not been extensively validated with experimental data, and so these predictions can only be used for hazard identification or risk assessment as part of a WoE approach (ECHA, 2008).

Some currently available software tools (e.g. TOPKAT and MCASE) are useful for predicting acute toxicity in categorical terms (e.g. in terms of Globally Harmonised Systems – GHS – classification). However, these tools should be further investigated in relation to the apparently high degree of false negatives generated, since this would be undesirable in a regulatory setting (JRC Report, 2011).

Databases containing information on acute toxicity which may be suitable for the development of (Q)SAR include Acutoxbase, ChemIDplus, CEBS and RTECS (JRC Report, 2011).

Tools and databases for repeated dose toxicity prediction

Chronic repeated dose toxicity is currently assessed in rodents, and an important challenge for the future is how to measure the health risks from repeat exposure without the use of animals. Such a task is very difficult because of the wide range of effects that could result from prolonged and repeated exposure (Kimber *et al.,* 2011). Repeated dose 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 represents a challenge for (Q)SAR modelling, which should ideally focus on groups of chemicals with a common mode of action. The use of (Q)SAR in assessing chronic toxicity is very limited. However, the JRC is currently building an inventory of evaluated (Q)SAR models for regulatory use (the JRC (Q)SAR Model Database, see Section 2.1.5 for details; ECHA, 2008).

Before any *in vivo* testing is conducted for repeated dose toxicity, the physicochemical properties of the substance should be considered. These properties can be used to determine whether the substance can be absorbed following oral, dermal or inhalation exposure and whether the substance itself or its metabolites are likely to reach the target organ or tissues. They are also used to decide on the appropriate administration route for any *in vivo* testing and whether such testing is technically possible. Substances that are, for example, highly volatile, highly reactive or unstable may be impossible to test (ECHA, 2008).

For repeated dose toxicity, (Q)SAR are used to give an indication of the mechanisms which may occur and possible organ or systemic toxicity. They are currently not well validated for use in repeated dose toxicity because there is a large number of potential targets and mechanisms that may be associated with a substance. Therefore at present there are no recommendations regarding the use of (Q)SAR for routine use in this area. When considered alongside other data, (Q)SAR can be used as part of a WoE approach and they can be used to support the use of read-across or grouping (ECHA, 2008).

At present, the most commonly used software tool for repeated dose toxicity prediction is TOPKAT, and despite the lack of transparency in its predictions, several studies (Venkatapathy *et al.,* 2004; Tilaoui *et al.,* 2007) have shown that it gives reasonable predictions for a range of chemicals (including pesticides and industrial chemicals). Another more recently developed tool is a module of MolCode Toolboxes. Predictions from such tools could be used in a WoE approach along with additional data. Additional research investigations into the applicability of TOPKAT and MolCode Toolboxes across a wide range of chemicals would be worthwhile. In addition, a transparent expert system or battery of (Q)SAR models needs to be developed for this endpoint (JRC Report, 2011).

There are three main databases or datasets suitable for the development and assessment of (Q)SAR for repeated dose toxicity. The RepDose database developed by the Fraunhofer Institute[8] contains NOAEL and LOAEL for over 650 industrial chemicals, but has not been made publicly available.

---

[8] www.fraunhofer-repdose.de [accessed March 2013]

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) NOAEL derived from non-carcinogenic endpoints in oral rodent or rabbit studies. This database has provided the basis of the TTC concept (JRC Report, 2011).

In addition to models for repeated dose toxicity, a limited number of *in silico* tools have been developed for predicting organ-specific and system-specific toxicity. For example, among the commonly used software tools, Derek Nexus estimates neurotoxicity using the following structural alerts: γ-diketone or precursor, acrylamide or glycidamide, nitroimidazole, carbon disulfide or precursor, pyrethroid, 1-methyl-1,2,3,6-tetrahydropyridine, lead or lead compounds and organophosphorus ester (JRC Report, 2011).

Tools and databases for prediction of irritant or corrosive properties

Some information about irritation and corrosion caused by chemicals can be gained by checking the pH of the substance. Those with an extreme pH value (less than two or greater than 11.5) are generally presumed to be corrosive to the skin and/or severely irritant to the eyes (ECHA, 2008).

The existence of a SAR provides justification for the use of read-across methods. The presence or absence of structures and substructures within the substance that have corrosion or irritation potential can be used to predict these effects. Non-reactive chemicals do not normally exhibit corrosive or irritant effects, although irritant contact dermatitis can result from contact with substances that have defatting properties, such as organic solvents. An example of a simple SAR that is used as an alert for corrosivity is the hydroperoxide group of substances; this is because they are acidic and oxidising in nature. SAR have been incorporated into the BfR (Institute for Risk Assessment, Germany) rule-base and the Skin Irritation Corrosion Rules Estimation (SICRET) Tool. (Q)SAR and expert systems for skin irritation and corrosion have also been described. Mostly they have been developed from small datasets of specific compounds, but some larger and more diverse datasets do exist. Physicochemical properties such as acidity, basicity, hydrophobicity and molecular size are used for homologous substances, whereas common structural features are used for heterogeneous groups of chemicals (ECHA, 2008). Table 2.7 and Table 2.8 show the available (Q)SAR for skin and eye irritation or corrosion (ECHA, 2008).

It is important to remember that models have not necessarily been developed for regulatory purposes and so they should be assessed in terms of whether the predicted endpoint corresponds to the regulatory endpoint of interest (ECHA, 2008).

Tools and databases for prediction of sensitising potential

Skin sensitisation

The skin sensitisation potential of a chemical is related to its ability to react with proteins in the skin (the key molecular initiating event) and with subsequent recognition by the immune system. The initiating event is usually dependent on the electrophilic reactivity of the substance or its derivatives. There are various types of electrophile–nucleophile reactions in skin sensitisation; perhaps the most frequently encountered are Michael-type reactions, S_N2 reactions, S_NAr reactions, acylation reactions and Schiff-base formation. These chemical reaction mechanisms can serve as a means of describing the domain of applicability of a (Q)SAR. Because there are defined chemistries that are associated with the induction of skin sensitisation through the key molecular initiating event, this is likely to be one of the endpoints for which (Q)SAR are most useful. However, to date there are relatively few published in peer-reviewed literature; those that are described include local and global (Q)SAR and expert systems (ECHA, 2008).

The majority of local models for direct-acting electrophiles have been based on the relative alkylation index approach (RAI). This is based on the concept that the degree and magnitude of sensitisation depends on the degree of covalent binding to a carrier protein. It quantifies the relative degree of haptenation as a function of dose, chemical reactivity expressed as a relative rate constant for reactions with a model nucleophile, and the hydrophobicity (expressed as calculated log P values). This method has been used to assess a wide range of skin sensitising chemicals, for example sulfonate esters, sulfones, aldehydes and diketones. The method is mechanistically robust, but available models are still limited and require a reasonable knowledge of chemistry (ECHA, 2008).

Global statistical models have been developed by the application of statistical methods to biological data. These are perceived as able to predict outcomes for a wider range of chemicals, but they often lack a clear mechanistic reasoning and are difficult to interpret from a chemistry perspective. The domain of applicability of these models can be variable; some are well described, whereas others require judgement to determine the relevance of the model (ECHA, 2008).

Expert systems that have models for skin sensitisation are presented in Table 2.9 below.

Table 2.7 Overview of literature-based (Q)SAR models used for skin and eye irritation or corrosion

Reference	Type of model	Applicability domain	QMRF developed?	Notes
Barratt, 1995	Statistical model	Acids, bases, phenols & pK_a	No	Used pK_a to predict effect of acids and bases Current knowledge does not give clear recommendations about how to use pK_a information
Berner *et al.,* 1988, 1990	Mathematical model	pK_a related for acids	No	Provides evidence of pK_a as predictor of skin irritation for acids Current knowledge does not give clear recommendations about how to use pK_a information
Nangia *et al.,* 1996	Mathematical model	pK_a related for bases	No	Provides evidence of pK_a as predictor of skin irritation for bases Current knowledge does not give clear recommendations about how to use pK_a information
Barratt, 1996	Statistical model	Electrophiles	No	Discriminates between corrosives and non-corrosives Does not provide transparent algorithm, has low goodness-of-fit and poor predictivity, not recommended for regulatory use
Smith *et al.,* 2000	Statistical model	Esters	No	Discriminates between irritants and non-irritants Does not provide transparent algorithm; limited regulatory use
Barratt, 1996	Statistical model	Neutral organics	No	Discriminates between corrosives and non-corrosives Does not provide transparent algorithm, has low goodness-of-fit and poor predictivity, not recommended for regulatory use
Gerner & Spielmann 2005	Rule-based model	New chemicals database, organic chemicals	Yes	

Source: ECHA (2008)
QMRF: (Q)SAR model reporting format

The choice of model depends on the chemical of interest, the underlying dataset and the applicability domain of the model. For a well-characterised chemical domain, a local (Q)SAR will give a robust prediction for that particular domain; however, if the mechanism is not as well understood, one or more of the expert systems will provide a better estimate. Although these systems are not as transparent, they do provide supporting information to enable the user to evaluate the robustness of the prediction (ECHA, 2008).

The prediction of non-sensitisers is limited and model predictions should be interpreted carefully and considered alongside other information such as read-across, the potential for activation and skin permeability. For example, some substances are not themselves skin sensitisers but may be activated either metabolically or abiotically (e.g. by oxidation) to substances that can induce sensitisation. In evaluating the output of a model, it is therefore necessary to be aware if this has been taken into consideration or not. The use of physicochemical information and its position in the sequence that leads to prediction may also be important; Derek Nexus, for instance, regards any substance with a log $K_p < -5$ as an ‘impossible’ skin sensitiser, since it should not penetrate intact skin. However, it may be necessary to predict the skin sensitisation potential of such a substance if it is deliberately applied to damaged skin, such as may be the case with medicinal products.

Recently, attention has focused on an *in chemico* approach[9] to the prediction of skin sensitisation

---

[9] see www.inchemicotox.org/ [accessed March 2013]

Table 2.8 Overview of computerised models used for skin and eye irritation and corrosion

Name of model	Type of model	Applicability domain	QMRF developed?	Notes
TOPKAT	Mathematical model using connectivity descriptors	Organic chemicals	Yes (for skin irritation)	Predicts the probability that a substance is a weak/mild moderate/severe irritant Gives similar structures, the experimental result and whether the prediction is in the applicability domain of the model Algorithm not transparent – there is no external validation & no mechanic reasoning; cannot be used as a stand-alone method
Derek Nexus	Expert system using structural alerts	Organic chemicals and some metals	Yes	Rule-base for irritation that is regularly updated Has 25 structural alerts for prediction of skin irritancy/corrosion If Derek Nexus does not predict irritancy, this does not mean that there is no effect; could be that none of known alerts was present, or it was outside the applicability domain Algorithm is transparent but it not considered validated, so cannot be used as a stand-alone method Cannot be used to predict non-irritation/corrosion
MultiCASE	Mathematical model using fragments	Organic chemicals	No	Can be used to predict eye irritation, but relating scoring system to regulatory classifications is unclear Provides structural alert, information on internal validation and whether it is in applicability domain Does not have external validation or mechanistic reasoning, so cannot be used as a stand-alone method
HazardExpert	Organic chemicals using structural alerts	Organic chemicals	No	Includes membrane irritation among human health effects, but this endpoint not clearly defined Can be used as supplementary information in a WoE approach, but using it directly for assessment of skin or eye irritation not recommended
BfR rule-base	Rule-based model	New chemicals database, organic chemicals	Yes	Uses physicochemical properties and substructural molecular features to predict EU risk phrases for skin irritation and corrosion Algorithms and applicability domain are transparent, rulebase has been validated by JRC and RIVM and a mechanism of action can be deduced For chemicals within the applicability domain, rules can be used on their own to predict hazards, therefore can be used for classification; whether this is appropriate should be determined on case-by-case basis

Source: ECHA (2008)
Sources of these models are given in Appendix 3

Table 2.9 Computational tools to predict skin sensitising potential sensitisation

Software	Overview	Details
TOPKAT	Has two sets of models that calculate whether the chemical is a sensitiser or a strong sensitiser	The first set of models discriminates between non-sensitisers and sensitisers The second set of models examines the potency of the sensitiser (weak/moderate vs strong) TOPKAT only makes predictions for chemicals within the applicability domain of the model
MCASE (From the CASE family of models)	The MCASE models have been developed for skin sensitisation	Includes two sensitisation modules The CASE approach uses a probability assessment to determine if a structural fragment is associated with toxicity Estimates from one of these models are included in the Danish EPA (Q)SAR database (located on the JRC website)
Derek Nexus	The skin sensitisation knowledge base contains 70 alerts for skin sensitisation and photoallergenicity	Derek Nexus is based on toxicophores, literature references, comments and examples
TIMES-SS Tissue metabolism simulator for skin sensitisation	Evaluates reactivity of chemicals to predict skin sensitisation	A hybrid between knowledge-based & statistically based expert system Uses a skin metabolism simulator Contains 236 spontaneous and enzyme-controlled reactions Covalent interactions with skin proteins are described by 47 alerting groups

Sources of these tools are given in Appendix 3

potential. In this approach, chemical (protein) reactivity, mechanistic read-across (see Section 2.2) and physicochemical data are assessed. One method is the Predictive Quantitative Mechanistic Model, which can be used to predict sensitisation potency; one such model has been developed for Michael acceptors (Roberts, 2009).

Respiratory sensitisation

Attempts to model respiratory sensitisation have been hampered by the lack of a predictive test protocol for assessing chemical respiratory sensitisation. (Q)SAR models are available but these have largely been based on data for chemicals reported to cause respiratory hypersensitivity in humans. Respiratory sensitisation is often associated with defined structural alerts, in particular cyclic anhydride, isocyanate and diamine groups.

The MCASE group has developed three models for respiratory sensitisation. The Danish (Q)SAR Database has an in-house model for which estimates can be extracted from the online database[10]. Derek Nexus contains several alerts derived from a set of respiratory sensitisers (ECHA, 2008).

Tools and databases for mutagenicity prediction

There is a range of non-test methods that can help to provide information about the mutagenicity of a substance, including inspection of the chemical structure, read-across, expert systems, metabolic simulators, local and global (Q)SAR. If testing data exist for the substance, non-testing information can be used in a WoE approach, to gain a better understanding of the mechanisms or confirm results. If experimental data are not available, this information can help to design the tests needed (ECHA, 2008).

Certain structures or fragments in a chemical can be associated with mutagenicity, often through mechanisms of reaction with DNA. The following references have information about structures and fragments that are associated with mutagenicity, that is the super-mutagen model and the subsequent builds on this model (Ashby, 1988; Ashby, 1993; Munro *et al.,* 1996).

There are many local (Q)SAR available in the literature that predict data for genotoxic endpoints for chemicals with related structures. When the substance is within the applicability domain, these models may be the best tools to predict mutagenic and genotoxic endpoints. The predictivity of the models should be assessed on a case-by-case basis as the quality of reporting can vary. (Q)SAR that contain mechanistic information are preferred (ECHA, 2008).

---

[10] available at http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/DDB [accessed March 2013]

Most global (Q)SAR for mutagenicity are commercial, and so for many of these the data within the modelling dataset is not available to the user (ECHA, 2008). The most common genotoxicity endpoint for global models is the Ames test. Examples of systems that can predict this endpoint are Derek Nexus, MultiCASE, TOPKAT, OASIS and the Danish EPA (Q)SAR database (ECHA, 2008).

A number of databases and computational tools used for the prediction of mutagenicity are presented in Table 2.10.

The JRC conducted a review of publicly available and commercial models across all toxicology endpoints. They found that the most commonly used systems were Derek Nexus, MultiCASE, OECD Toolbox, TOPKAT

Table 2.10 Computational tools for the prediction of mutagenicity

Software	Overview	Details
The Danish EPA (Q)SAR Database	Contains predictions for over 166 000 chemicals and has a flexible system for searching parameters and chemical structure All models were derived from MultiCASE software	Contains assorted Ames models Can also model chromosomal aberrations, mouse lymphoma-TK, CHO/HPRT gene mutation assays, UDS (rat hepatocytes) information, *Drosophila* SLRL, mouse micronucleus, rodent dominant lethal, mouse SCE in bone marrow and mouse comet assay data
ESIS		Contains information on chemicals related to EINECS, ELINCS, NLP and biocide active substances
Enhanced NCI Database Browser Sponsored by the US National Cancer Institute	Contains predictions for over 250 000 chemicals for mutagenicity among other endpoints Modelling was done using PASS	Some of the endpoints contain mechanistic information, for example alkylating ability and microtubule formation inhibition Searchable by a wide range of parameters and structure combinations
OECD database on chemical risk assessment models	Contains information on models for mutagenicity and other endpoints	This was developed to help identify tools for use in research and development of chemical substances
CAESAR	Developed as an EU-funded project Statistically based model using 4225 compounds and the Kazius–Bursi mutagenicity database Simple to use	Provides data for defined endpoints Provides an estimate for the Ames test Provides a prompt when a chemical is outside the domain of applicability Provides estimates for six structural analogues within the database
Toxtree	Developed by the JRC as an EU project Rule-based approach using Benigni/Bossa and Tox Mic rule-bases Simple to use	Provides data for defined endpoints Provides an estimate of mutagenicity in *Salmonella* TA100 and micronucleus formation in rodents Toxtree applies set rules for evaluation but does not provide an alert when the chemical is not covered by the rule-base Chemical analogues have to be identified and entered separately
Derek Nexus	Contains 89 alerts for bacterial mutagenicity and 77 alerts for chromosome damage The chromosome damage alerts are based primarily on data from the *in vitro* chromosome aberration test, however additional assays have been considered when writing alerts	Derek Nexus 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

Source: ECHA (2008)
PASS: Prediction of Activity Spectra for Substances
Sources of these tools are given in Appendix 3

and Toxtree. The most highly developed models were those for genotoxicity. They also assessed the genotoxicity models using a diverse range of chemicals and found that Ames test data were highly reproducible. They also found that publicly available and commercial models have a very similar performance for genotoxicity assessment (JRC Report, 2011).

It has been reported that the accuracy of prediction of Ames mutagenicity by Derek Nexus is 65% (against a dataset of 400 chemicals) and the overall concordance between mutagenicity predictions and test data for TOPKAT is 73%. It has also been found that the predictions for external and independently chosen test 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 Nexus. These findings contrast with the usually good performances reported by the model developers, as assessed on small congeneric databases (JRC Report, 2011).

It has been recommended that when using computational models for regulatory purposes, predictions of genotoxicity should not be based on the use of any single model alone, but on a WoE approach including information, if possible, from all available sources (QSAR, read-across, *in vitro* test methods). In addition, the use of batteries of computational tools that combine high sensitivity models (to minimise false negatives) with high specificity models (thereby minimising false positives) is considered to improve the overall reliability of the prediction. 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 (JRC Report, 2011).

Tools and databases for prediction of carcinogenic potential

Prediction of carcinogenicity using non-testing data is currently very challenging because of the multitude of possible mechanisms that could occur. The most commonly modelled endpoint for carcinogenicity has been the rodent bioassay.

Certain structures or fragments in a chemical can be associated with carcinogenicity, often through mechanisms of reaction with DNA. Genotoxicity is therefore an important factor in carcinogenicity. The following references have information about structures and fragments that are associated with carcinogenicity (USA FDA, 1986; IARC, 2006).

An example of a model for a genotoxic carcinogen is described in Franke *et al.* (2001; ECHA, 2008).

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: to identify the electrophilic functional groups or substructures (i.e. to develop SAR models based on structural alerts), and to find molecular descriptors which can be quantitatively related to the activity of the chemicals (i.e. to develop QSAR; JRC Report, 2011).

Carcinogenicity that is not due to genotoxic effects is far more difficult to predict because of the vast array of different mechanisms that could be involved. A unifying scientific theory for the mode of action of non-genotoxic carcinogens is still missing. For this reason, (Q)SAR for non-genotoxic carcinogenicity are still in an early stage of development. Progress is being made in this area but the applicability of the models will be dependent on the mechanism involved and the chemical class (ECHA, 2008).

Global (Q)SAR exist which attempt to predict the carcinogenic hazard of diverse groups of substances. These models can be useful for screening, priority setting, deciding on testing strategies and hazard assessment in a WoE approach. Most of these models are commercial and include OncoLogic, Derek Nexus, MultiCASE and TOPKAT. The performance of this type of model is mixed and is dependent on how well the carcinogenicity is defined in those substances used to develop and test the model. Freely available models for carcinogenicity prediction are the Danish EPA (Q)SAR database and the enhanced US National Cancer Institute database.

A number of databases and computational tools used for prediction of carcinogenicity are presented in Table 2.11.

At present, (Q)SAR methods are more reliable for predicting genotoxic potential than carcinogenic potential. Carcinogenicity prediction represents a considerable challenge owing 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

Table 2.11 Computational tools for the prediction of carcinogenicity

Software	Overview	Details
The Danish EPA (Q)SAR Database	Contains predictions for over 166 000 chemicals and has a flexible system for searching parameters and chemical structure All models were derived from MultiCASE software	Contains eight MultiCASE FDA cancer models, rodent carcinogenic potency, hepatospecificity, oestrogenicity and aryl hydrocarbon receptor binding
ESIS	Database	Contains information on chemicals related to EINECS, ELINCS, NLP and biocide active substances databases
Enhanced NCI database browser Sponsored by the US National Cancer Institute	Contains data from more than 500 two-year cancer bioassays in two species	Also contains results relating to approx. 300 studies from shorter duration tests
Carcinogenic Potency Database (CPDB)	Searchable by chemical name, CAS number or author	Provides results of 6540 chronic animal cancer tests on 1547 chemicals
IARC monographs	Searchable by key word, chemical name or CAS number	Reviews of more than 900 chemicals of which 400 have been identified as known, probable or possible carcinogens
CAESAR	Developed as an EU-funded project Simple to use	Includes a regression model built on the analysis of 805 chemicals with rat TD_{50} values from the CPDB and a classification model A concordance of 68–74% for carcinogenicity has been reported
Derek Nexus	Contains 61 alerts for carcinogenicity	Note: Derek Nexus 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
OncoLogic	Knowledge-based system Uses a series of hierarchically ordered rules to describe and predict the carcinogenic potential of chemicals	Includes over 40 000 rules based on knowledge and generalisations derived from the examination of more than 10 000 chemicals belonging to approx. 50 chemical classes The user needs chemistry expertise

Source: JRC Report (2011)
Sources of these software tools are given in Appendix 3

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 (JRC Report, 2011).

Tools and databases for reproductive and developmental toxicity prediction

Some information on reproductive toxicity can be obtained from the physicochemical properties of the substance (ECHA, 2008).

The physicochemical properties of a substance can give information about whether it is likely to be absorbed by a particular exposure route and whether it can cross placental, blood–brain barriers or be secreted in milk. Physicochemical properties can also be used in a WoE approach (ECHA, 2008).

There are currently no formal criteria or structural alerts for reproductive toxicity and a large number of potential mechanisms are associated with it, 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.

At present, (Q)SAR models cannot adequately cover this endpoint, are not well validated for this use, and no firm recommendations can be made regarding their use in this area. In addition, there is a paucity of high-quality data on this endpoint suitable for model development. (Q)SAR can be used as part of a WoE approach alongside other data and they can be used as supporting evidence when assessing a substance by read-across or grouping, but a positive or negative (Q)SAR result is not sufficient as a stand-alone piece of evidence (ECHA, 2008).

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

Some of the databases and software tools that have been designed to predict reproductive toxicity are shown in Table 2.12 and Table 2.13 (Worth *et al.,* 2011).

Worth *et al.* (2011) found that the best combination of models based on the statistics was TOPKAT combined with PASS (for teratogenicity); this combination had a specificity of 32%, negative predictivity of 48% and a false negative rate of 14%. This was followed by combining Derek Nexus and PASS (for embryotoxicity), with a specificity of 56%, negative predictivity of 46% and a false negative rate of 27%. Combining models increased the negative predictivities, but these are still less than 50% and so would not be adequate for regulatory use.

A caveat that should be considered for all endpoints is that different regulatory bodies apply different criteria when evaluating raw data. Any conclusions should therefore be accompanied by a full description of the effects of the substance. It is also important to check the criteria used to distinguish between positive and negative outcomes, and it should be noted that (Q)SAR are often better suited for the identification of positives, with positive predictivities between 61–96% (Worth *et al.,* 2011).

Table 2.12 Databases containing information on reproductive toxicity

Database	Overview	Details	Other considerations
ToxRefDB	Toxicity test results for pesticides 387 chemicals for developmental toxicity & 316 chemicals for multigeneration reproductive toxicity	Potentially useful reference database for the development of new models and application of grouping and read-across	Worth *et al.,* 2011 considered ToxRefDB as a suitable public domain database to support grouping and read-across for developmental toxicity
TOXNET (Toxicology Data Network) & **DART** (Developmental & reproductive toxicology database)	Bibliographic database containing over 200 000 references to literature published since 1965 Developmental & reproductive toxicology	Users can search by subject term, title word, chemical name, chemical abstracts service registry number & author	
ICSAS Reprotox Database	Reproductive & developmental toxicology database Contains data records from animal studies from publicly available sources	Data from studies on reproductive toxicity in male & female animals, teratology, organ toxicity, non-specific effects to the fetus, behavioural toxicity in new-born pups Publicly available sources include: Shepard's Catalogue of Teratogenic Agents, TERIS, REPROTOX, RTECS, studies from EPA toxdata database & studies on drug labelling Contains data for 2173 chemicals, mostly pharmaceuticals	
NTPBSI Database (National Toxicology Program Bioassay Online Database)	Developmental toxicity dataset with data for 70 substances	Mixed dataset with different toxicological evaluations	
ILSI developmental toxicity database Currently under development*			

* April 2013
Sources of these databases are given in Appendix 3

Table 2.13 Software tools for predicting reproductive toxicity

Database	Overview	Details	Other considerations
CAESAR 2.0	Two classification models for developmental toxicity for 292 substances	Based on dataset from Arena *et al.* (2004)	Worth *et al.,* 2011 identified CAESAR as potentially useful; they noted that CAESAR is suitable for use by non-specialists
Derek Nexus 2.0	Contains alerts for developmental toxicity (3), teratogenicity (5), testicular toxicity (1) and oestrogenicity (4)	Some studies indicate that this might be useful for identification of developmental toxicants but not on its own (high positive predictivity 81–96%) Could be useful in a stepwise strategy in combination with grouping and read-across	Derek Nexus is facilitated by strong customer support from the developer
PALLAS HazardExpert v3.6.2.1	Structure-rule based prediction of organic compounds	Some studies indicate that this might be useful for identification of developmental toxicants but not on its own (high positive predictivity 81–96%) Could be useful in a stepwise strategy in combination with grouping and read-across	Worth *et al.,* 2011 identified HazardExpert as potentially useful
Leadscope Model Applier Version 1.3.3	Classification models for developmental toxicity in rodent fetus	Can model dysmorphogenesis (structural & visceral birth defects), developmental toxicity (fetal growth retardation & weight decrease) and fetal survival (fetal death, post implantation loss and preimplantation loss) The model for reproductive toxicity includes rodent male reproductive, rodent male sperm, female reproductive	
OSIRIS Property Explorer	Classification model based on RTECS database of >3500 substances	Predicts effects such as mutagenicity, tumourigenicity, irritating effects & reproductive effects	
OSIRIS Property Explorer	Classification models giving probability of adverse effects Demo available for free online and to download Commercial version has added functionality & predicted endpoints Has two models: PASS embryotoxicity and PASS teratogenicity	Embryotoxicity model predicts probability that a substance crosses placental membrane & causes a toxic effect, (e.g. fetal bradycardia, low birth rate) or death of an embryo Teratogenicity model predicts probability that a substance crosses placental membrane and causes abnormal development of a body system in the embryo Worth *et al.* (2011) found the PASS models to have the best performance of those tested in their study, with specificity of 59–62%, negative predictivity of 44–45% and false negative rate of 29–31% Some studies indicate that this might be useful for identification of developmental toxicants, but not on its own (high positive predictivity 81–96%) Could be useful in a stepwise strategy in combination with grouping and read-across	Worth *et al.,* 2011 identified both PASS models as being potentially useful
TOPKAT	Classification model for developmental toxicity of pesticides & industrial chemicals		Worth *et al.,* 2011 identified TOPKAT as potentially useful

Database	Overview	Details	Other considerations
MultiCASE (MC4PC)	Classification 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, developmental fetal weight decrease and fetal survival or death	
Endocrine Disruptor Knowledge Base (EDKB) database	Quantitative models to predict the binding affinity of compounds to the oestrogen and androgen nuclear receptor proteins		

Sources of theses software tools are given in Appendix 3

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. The International Life Science Institute (ILSI) and Leadscope have already started such an initiative[11]. In relation to databases, an important achievement has been the construction, from publicly available information sources, of the US FDA ICSAS Reprotox Database suitable for QSAR modelling (JRC Report, 2011).

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 oestrogen and androgen receptors (ER and AR). 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 as well as commercial models (JRC Report, 2011).

Tools and databases for prediction of neurotoxicity

Table 2.14 below shows software tools that are relevant for neurotoxicity prediction.

It was found that combining two models together (from Derek Nexus, PASS or HazardExpert) increases the negative predictivity to 48% and the false negative rate is 84–94%. Since the negative predictivity is less than 50% no individual model or two-model combination is adequate to predict the absence of neurotoxic potential. High positive predictivities mean that some tools (such as Derek Nexus and HazardExpert) might be useful for the identification of neurotoxicants as part of a stepwise strategy in combination with grouping or read-across (Worth *et al.,* 2011).

Worth *et al.* (2011) found that there was generally a lack of freely available software tools for neurotoxicity prediction. This will slow development of new (Q)SAR and limit read-across and grouping applications for this endpoint.

Table 2.14 Software tools for predicting neurotoxicity

Database	Overview	Details	Other considerations
Derek Nexus 2.0	Contains hundreds of alerts covering a wide range of toxicological endpoints Based on rules from mechanisms of action of a chemical class or empirical relationships from published data & experts	Has eight alerts for neurotoxicity: gamma diketone or precursor, acrylamide or glycidamide, nitroimidazole, carbon disulfide or precursor, pyrethroid, 1-methyl-1,2,3,6 tetrahydropyridine, lead or lead compounds & organophosphorus ester Derek Nexus performed the best for neurotoxicity predictions, with 80% specificity, 43% negative predictivity and 74% false negative rate Requires expertise for use, but has strong customer and developer support	Alerts not designed to identify absence of effects, so can only identify positives Applicability domains of expert systems not well developed. Worth *et al.* (2011) identified Derek Nexus as 'potentially suitable' for neurotoxicity prediction (Worth, 2011)

---

[11] more information available from: http://www.epa.gov/ncct/dsstox/CoordinatingPublicEfforts.html#ToxML [accessed March 2013]

Database	Overview	Details	Other considerations
PALLAS HazardExpert v3.6.2.1	Predicts toxicity of organic compounds based on toxic fragments Structure-rule-based system	Covers several endpoints including carcinogenicity, mutagenicity, teratogenicity, membrane irritation, immunotoxicity and neurotoxicity Requires specific expertise for use	Worth *et al.* (2011) identified HazardExpert as 'potentially suitable' for neurotoxicity prediction (Worth, 2011)
Leadscope Model Applier Version 1.3.3, neurotoxicity suite	Includes models from a predefined library of 27 000 hierarchically organised fragments that are typically found in small drug molecules & eight calculated molecular descriptors	Suite comprises three rodent new-born behaviour models Not directly relevant for this use as it is based on results from developmental neurotoxicity tests	
PASS v10.1	Predicts mutagenicity, carcinogenicity, teratogenicity, embryotoxicity, a range of mechanisms of action and pharmacological effects Free version is more restricted than commercial version in terms of functionality, endpoints & datasets	Version 10.1 of PASS includes a neurotoxicity model that predicts the probability of neurotoxicity	The neurotoxicity module and receptor-mediated activity modules not directly relevant for this use (Worth, 2011)
ADMET Predictor 5.0	Predicts probability of blood–brain barrier penetration as low or high	Not directly relevant, but could provide supporting information	Does not provide an apical endpoint
Accelrys ADME add-in	Predicts blood–brain barrier penetration from oral exposure in a quantitative linear regression model	Not directly relevant, but might provide supporting information	Does not predict an apical endpoint

Sources of these software tools are given in Appendix 3

2.2 Read-across and the analogue approach

2.2.1 The concept of grouping of chemicals

The term 'grouping' or 'chemical grouping' describes the general approach to assessing more than one chemical at the same time. It can include the formation of a chemical category or the identification of a chemical analogue for which read-across may be applied. Grouping can be employed for various purposes, for example in the development of a (Q)SAR model, or to predict certain properties of the grouped chemicals without the necessity to test them individually for all endpoints.

The term chemical grouping can be refined by the use of the more specific terms 'category approach' and 'analogue approach', which are two techniques used for grouping chemicals together.

A chemical category is a group of chemicals whose physicochemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity. The similarities may be based on the following.

- A common functional group (e.g. aldehyde, epoxide, ester, specific metal ion)
- Common precursors or breakdown products via physical or biological processes, which result in structurally similar chemicals (e.g. the metabolic pathway approach of examining related chemicals such as acid, ester or salt)

- An incremental and constant change across the category (e.g. a chain length category), often observed in physicochemical properties such as boiling point range
- Common constituents or chemical class, similar carbon range numbers

Within a category, a property can be estimated through read-across, trend analysis or (Q)SAR from those members of the category that have data to those that do not. Read-across is therefore a data gap-filling technique used to predict endpoint information for one chemical by using data from the same endpoint from another chemical which is considered to be similar in some way.

Whilst read-across can be used within the context of a category approach, it is also a useful tool for data gap-filling in cases where comparisons are based on a very limited number of chemicals. The reading-across of endpoint information for one chemical to predict the same endpoint for another chemical (i.e. a one-to-one read-across) is referred to as the analogue approach (ECHA, 2008).

2.2.2 Read-across

Introduction to read-across

The read-across technique is widely used and is often accepted as a valid approach for regulatory purposes. As explained above, read-across involves using endpoint information for one chemical to predict the same endpoint for another chemical considered to be similar in some important way that relates to the endpoint, usually on the basis of structural similarity, but possibly also on the mode of action, toxicokinetics or metabolism. Read-across can be used in the context of both the analogue approach and the category approach (OECD, 2007b). Read-across can be applied to characterise a range of properties, such as physicochemical properties, environmental fate, human health effects and ecotoxicity in a qualitative or quantitative manner (OECD, 2007b; ECHA, 2008).

Read-across can be performed in the following ways.

- One-to-one: a single analogue is used to estimate the endpoints for a single chemical.
- Many-to-one: two or more analogues are used to estimate the endpoints for a single chemical.
- One-to-many: a single analogue is used to estimate the endpoints for two or more chemicals.
- Many-to-many: two or more analogues are used to estimate the endpoints for two or more chemicals.

The guidance described here for read-across applies to both the analogue approach and the category approach. However, it should be noted that the category approach is generally considered to be more robust than the analogue approach because a greater amount of measured data is usually available in a category. As a consequence, the success of read-across depends strongly on the quality of the measured data and the appropriate selection of the characteristics and measures of similarity. In the read-across approach, the target chemical is the chemical for which an endpoint is being estimated and the source chemical is the chemical that is being used to make an estimate (OECD, 2007b; ECHA, 2008).

Read-across can be either qualitative or quantitative, as follows.

- Qualitative read-across: The presence or absence of a particular property or activity is inferred from the source chemical(s) to the target chemical(s). Qualitative read-across gives a yes/no answer, for example the substance is a skin sensitiser or not (OECD, 2007b; ECHA, 2008).
- Quantitative read-across: Properties from the source chemical(s) with known values are used to estimate the value of that property for the target chemical(s). Quantitative read-across yields a quantitative value for an endpoint, for example the potency of a carcinogen, an LD_{50} value or a LOAEC for hepatotoxicity (OECD, 2007b; ECHA, 2008).

Read-across is usually performed on the basis of similar chemical structures, properties and/or activities. These could be:

- a common functional group (e.g. aldehyde, epoxide, ester or metal ion);
- a common precursor and/or breakdown product resulting from physical or biological processes, such as a metabolic pathway;
- a common mechanism of action and similarities in biochemical reactivity.

Similarity plays a crucial role in grouping approaches. Although there is an intuitive understanding about the similarity of chemicals, attempts to formalise this understanding with universal practical validity are not, at present, advanced. The similarity of chemicals depends on a variety of factors. It may be accounted for by molecular topology (atom connectivity), chemistry (atom and bond types), the presence of functional or specifically acting groups, physicochemical properties, mechanism of interactions, endpoints, etc. (Cronin & Madden, 2010).

The absence of a unique measure of chemical similarity indicates that similarity should be considered as a context-dependent parameter. From a practical point of view, it is more useful if the grouping of similar chemicals is performed in a stepwise manner, applying different measures of similarity consecutively – topology, functionalities, properties, mechanism, etc. A consecutive approach to grouping of chemicals allows users to:

- start by grouping chemicals with less conservative conditions, thus allowing an initial collection of a large number of potential analogues;
- prune the initial set of analogues consecutively by applying different measures of similarity, starting with more general and ending up with more specific molecular features; and
- select a small number of very similar analogues in the context of the phenomenon studied (Cronin & Madden, 2010).

Traditionally, the assessment of chemical similarity has been performed on a case-by-case basis. Computational tools implementing various ranges of similarity metrics can facilitate identification of suitable analogues in a more efficient and consistent manner. The major challenge for computational tools that support grouping approaches centres around formalisation of expert knowledge, improved management and openness of toxicological data.

Read-across can be used to predict both the presence and absence of a property or effect. Read-across from a negative result is as valid as a positive result, provided the test design is adequate. It is particularly important to justify read-across of negative findings and this approach is more robust when a quantitative relationship can be established.

When deciding whether qualitative or quantitative read-across is necessary, it is first important to identify what type of data is required and the reason the endpoint is being estimated. If a specific value is required, then quantitative read-across will be necessary. If, however, the endpoint needs to be checked against a threshold or a classification, then qualitative read-across may be sufficient.

Qualitative read-across

The main application of qualitative read-across is in identifying and categorising hazards. In qualitative read-across, the presence or absence of a property is inferred from the established properties of one or more analogues. The application of qualitative read-across usually results in the allocation of the target chemical to the same hazard category as the source chemical. The arguments to support qualitative read-across are usually based on assessment of several factors and expert judgement (OECD, 2007b; ECHA, 2008).

For example, a common substructure that is normally responsible for a common property or effect could be affected by other parts of the chemical structure; this could be affected in either a qualitative or a quantitative manner. In addition to the problem of interactions between substructures, some differences in the chemical structure and/or in its physicochemical properties could completely alter the common effect altogether (OECD, 2007b; ECHA, 2008).

It should be noted that quantitative changes in potency between the source and the target chemical may warrant a different classification or banding; hence a quantitative change can also alter the qualitative assessment. Also, if the effect being measured is a direct effect in the source chemical(s) but a secondary, indirect effect in the target chemical(s), or vice versa, this can lead to a different classification of chemicals within the same category (OECD, 2007b; ECHA, 2008).

Quantitative read-across

In quantitative read-across, the known value of a property (e.g. the potency) for the source chemical is also used to estimate the unknown value of the same property for the target chemical.

There are four methods of applying quantitative read-across to estimate missing data points.

1 By using the same endpoint value as the source chemical

2 By using a (Q)SAR and the available experimental results from two or more source chemicals to estimate the endpoint for the target chemical

3 By processing the endpoint values from two or more source chemicals (e.g. averaging the endpoints or taking the most representative value)

4 By using the most conservative value of the analogues

If there are concerns that the target and source chemicals are different enough in terms of potency to affect the conclusions on hazard identification and characterisation, then it may be necessary to perform further testing to demonstrate these differences (OECD, 2007b; ECHA, 2008).

General considerations when performing read-across

There are several factors to consider when performing read-across.

- Whether the data for the source chemical are relevant and reliable. Particular consideration should be given to data that have not been produced to current OECD test methods.

- Whether the source and/or target chemical(s) are multifunctional and whether these functional groups can affect the reliability of the read-across.
- Whether impurities in the chemicals may influence the overall toxicity, and the effect these impurities may have on the robustness of the estimate. To consider this factor, the purity and impurity profiles of the target and source chemicals should be assessed. If all category members have the same sort of impurities then they may not have an influence on the read-across. However if one category member has a biologically active impurity, then read-across might not be appropriate.
- The physicochemical properties of the target and source chemicals should be compared, particularly their physical form, molecular weight, water solubility, particle size, particle structure, partition coefficient and vapour pressure. It should be noted that data on contaminants' physicochemical properties might not always be available.
- The toxicokinetics and any possible different metabolic pathways should be considered for the target and source chemicals.

Information from valid (Q)SAR can be used to inform on the need, extent and type of any additional testing (OECD, 2007b; ECHA, 2008).

Supporting information

When performing read-across, it is important to provide supporting information to strengthen the rationale for its use. This includes demonstrating that other properties relevant to the endpoint are also similar between the source and target chemicals. These additional properties could be identified through expert knowledge or be based on those molecular descriptors that have been found to be useful predictors of the endpoint in (Q)SAR models (OECD, 2007b; ECHA, 2008).

Useful supporting information could include the following.

- Whether the substance has any functional groups that could influence the behaviour of the chemical
- The similarities between the chemicals in terms of their physicochemical properties (e.g. molecular weight, log K_{ow} or water solubility)
- Similarities between the chemicals in terms of other toxicological and/or ecotoxicological data
- Similarities in the toxicokinetics of the substances, including metabolic pathways
- Information from valid (Q)SAR

Supporting information for human health endpoints

The type and amount of supporting information required for the read-across of human health endpoints will depend on the endpoint itself; current practice relies heavily on expert judgement. Physicochemical properties are generally useful, including those properties that determine bioavailability (e.g. molecular weight and partition coefficients (e.g. log K_{ow}), water solubility, pH, vapour pressure and viscosity). Particle size and structure are also relevant (OECD, 2007b; ECHA, 2008).

Reactivity with proteins or DNA can also provide useful supporting information. *In vitro* data can also be used as supporting information for some endpoints.

Stepwise approach for applying read-across within the analogue approach

The stepwise approach, described below, has been recommended by both the OECD and by the EU in relation to industrial chemicals under REACH. However, it should be noted that it is not the only approach (OECD, 2007b; ECHA, 2008).

1 Identification of potential analogues

There are several ways to identify analogues as source chemicals to use for read-across. Similar chemicals may be produced for comparable uses by the same company or industry sector, and are often used as potential analogues. No formal selection techniques are used in this scenario (OECD, 2007b; ECHA, 2008). A more formal search strategy may find additional analogues to compare with the target chemical, increasing the robustness of the read-across. As the number of chemicals included in a read-across increases, this approach gets closer to that used for categories. It should therefore be considered whether the chemical is best evaluated by the analogue or category approach. Factors that would influence this decision are whether the chemical is a member of a category that has already been evaluated, and the number of analogues identified. If a large number of analogues is identified, the category approach (Section 2.3) is recommended (OECD, 2007b; ECHA, 2008).

To select analogues for read-across, computational methods are available. These use a combination of expert knowledge and automated search tools for molecular similarity indices and electronic substructures (Table 2.15). Molecular similarity has predominantly been used by the pharmaceutical industry for virtual screening, estimation of absorption, distribution, metabolism, excretion, toxicity and prediction of physicochemical properties (OECD, 2007b; ECHA, 2008).

The fundamental basis for any read-across decision must be that the chemical structures are sufficiently

Table 2.15 Examples of Internet-based tools for analogue searching

Internet tool	Comments	Search options
AIM (Analog Identification Methodology)	Links to publicly available, experimental toxicity data for the target chemical and structural analogues Contains 31 031 records	Searchable by CAS, SMILES and (sub)structure
Ambit Developed by IdeaConsult Ltd	Databases and functional tools, includes a tool for defining applicability domain of (Q)SAR models Contains 463 426 records	Searchable by chemical name, CAS, SMILES and (sub)structure
CAS SciFinder	Commercially available and internet accessible portal to extensive collection of chemical and biochemical information from scientific literature and patents	Searchable by chemical name, (sub)structure, biological sequence and reaction, research topic, author and company
ChemFinder	Contains both publicly available and subscription scientific databases There is a diverse range of searchable parameters	Searchable by chemical name, synonyms, CAS, formula, chemical structure (exact match, substructure, similarity search), toxicological and physicochemical properties
ChemID Plus	Contains over 379 000 records	Searchable by chemical name and CAS number
Danish (Q)SAR Database Developed by the DK EPA	Contains 166 000 records	Searchable by chemical name, CAS, endpoint and (sub)structure
Hazardous Substances Database (HSDB)	More than 4800 peer-reviewed records	Searchable by chemical name, fragment name, CAS and subject terms
Leadscope	Commercially available databases and (Q)SAR functionalities	Searchable by chemical name, (sub)structure, toxic effect, study type and experimental conditions

Sources of these internet-based tools are given in Appendix 3

close that it can be reasonably expected that their effects would be similar. The more different the chemical structures, the lower the confidence in the prediction or read-across. Where biologically active functional groups are present, they should be in both chemicals, in the same orientation in a manner that would give similar biological activity. The purity of the two chemicals and subsequent impurities should be examined to ensure that the overall toxicity is the same (OECD, 2007b; ECHA, 2008).

2 Data gathering for the analogues

For the source analogues chosen, published and unpublished data should be gathered on standard physicochemical properties, environmental fate, ecotoxicological effects and toxicological effects.

If many analogues have been identified, some computational tools are available to help reduce the dataset to the closest analogues (OECD, 2007b; ECHA, 2008).

There are many web-based databases containing information or data on environmental and human health effects of numerous chemicals. One of the most extensive of these databases has been made available by the OECD in collaboration with ECHA (European Chemicals Agency) and other, global, organisations[12].

3 Evaluation of available data for adequacy

If the data have been obtained from a peer-reviewed source they can normally be used without further evaluation. However, if the data are not peer reviewed, they should be evaluated for adequacy.

4 Construct a matrix of data availability

A matrix of data availability should be constructed for the target endpoint and other relevant endpoints. It should include the target chemical and the analogue(s). If multiple analogues have been identified, they should be arranged in an order that reflects a trend or progression within the group. The cells of the matrix should indicate whether data are available or not (OECD, 2007b; ECHA, 2008).

5 Assess the adequacy of the analogue approach and fill the data gap

The decision on whether data from an analogue are suitable for read-across relies largely on expert

---

[12] http://www.echemportal.org/echemportal/index?pageID=0&request_locale=en [accessed March 2013]

judgement. Where possible, the read-across should be evaluated in terms of the mode of action and available data for the source and target chemicals. (Q)SAR can also be helpful for assessing the applicability of the read-across, predicting missing data or comparing experimental data with predictions.

If read-across is deemed to be suitable, the target chemical should be evaluated using data from the source chemical. If, however, read-across is not suitable, the following options are available.

- Identify alternative analogues
- Use the category approach
- Obtain the information directly by testing (OECD, 2007b; ECHA, 2008)

6 Document the analogue approach

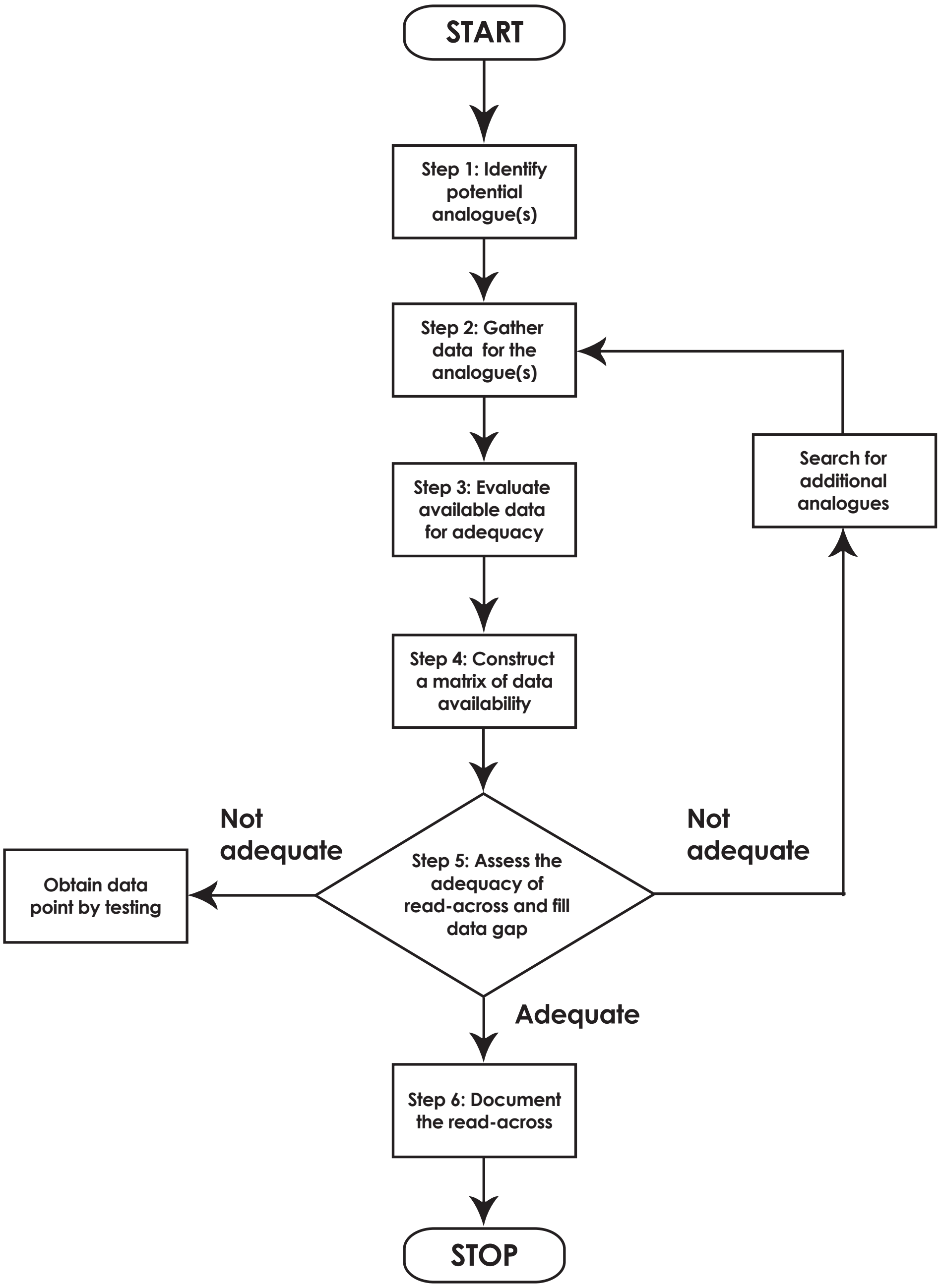
If read-across is appropriate, it should be documented in order to justify its use instead of testing. A suitable format is shown in Section 2.4. The justification for the read-across should include an explanation of the rationale, the assessment of the read-across and all supporting information, as shown in Figure 2.1 (OECD, 2007b; ECHA, 2008).

Reporting format for the analogue approach

The following reporting format for analogue read-across has been recommended by the OECD and at EU level in relation to industrial chemicals under REACH (OECD, 2007b; ECHA, 2008).

- **Hypothesis for the analogue approach:** Describe the molecular structure that a source chemical

Figure 2.1 Stepwise procedure to the analogue approach (ECHA, 2008)



must have in order to be suitable for read-across, identifying all functional groups. Provide a hypothesis on why the read-across can be performed; if there is a mechanistic reasoning, describe the foreseen mode of action for the source and target chemicals. Where relevant, describe the influence that the route of administration (oral, dermal or inhalation) would have. List the endpoints that are being predicted.

- **Source chemical(s):** Describe the source chemical(s) as fully as possible, providing CAS number(s), substance name(s) and chemical structure(s).
- **Purity/impurity:** Provide purity/impurity profiles for the target and source chemical(s), including their likely impact on the relevant endpoints being predicted.
- **Analogue approach justification:** Summarise how read-across is justified. The available experimental data should demonstrate that the anticipated toxicity is not affected by functional groups that are different in the source and target chemicals.
- **Data matrix**: Provide a matrix of data for endpoints versus target and source chemical(s). If experimental results are available, key study results should be included and should support the justification for read-across.
- **Conclusions:** For each endpoint and substance a conclusion should be reached. Uncertainties in the read-across should be discussed.

2.3 Guidance on grouping of chemicals

2.3.1 Chemical category approach

Introduction to the category approach

A chemical category is a group of chemicals that have structural similarity and, as a result, their physicochemical, human health, environmental toxicology and/or environmental fate properties are likely to be similar or follow a regular pattern. As the number of possible chemicals being grouped into a category increases, the potential for developing hypotheses for specific endpoints and making generalisations about the trends within the category will also increase and hence increase the robustness of the evaluation (OECD, 2007b; ECHA, 2008).

The principle of the category approach is that the properties of a series of chemicals with common structural features will have coherent trends in their physicochemical properties, toxicological effects or environmental fate properties. Trends are generally associated with a common underlying mode of action. As a result of this, it is possible to extend the use of measured data on tested chemicals within the category to similar untested chemicals (within the category) without the need for further testing. The category approach can also be used to help decide the nature and scope of any testing that might need to be carried out (OECD, 2007b; ECHA, 2008).

Chemicals within a category are assessed on the basis of the evaluation of the category as a whole. The category therefore provides an alternative to testing individual substances leading to a decrease in the use of animal testing (OECD, 2007b; ECHA, 2008).

Benefits of the category approach

There are a number of benefits to using the category approach.

- Missing data can be extrapolated from others in the category, reducing the amount of testing needed.
- The approach is based on a larger body of data than using a single chemical.
- Chemicals within a category can provide information about potential effects that might otherwise have been overlooked.
- It is helpful for the evaluation of chemicals that pose technical difficulties within standard test protocols.
- Category testing can be completed quicker than doing individual tests on several chemicals.
- Not every chemical needs to be tested for every endpoint. The overall dataset is used to estimate the missing data points.
- It allows for better consideration of the biological plausibility of grouping the chemicals within a category.
- The approach can be used to help identify and fill data gaps efficiently.
- Knowledge of expected biological effects of the chemicals in the category is helpful for deciding whether testing is needed and the nature and scope of any tests to be carried out. If confirmation is being sought that an individual category member does not have a particular property, this may be possible with a simple limit test, whereas when an individual category member is expected to have an effect, an *in vitro* test may be adequate for confirmation of the predicted effect.
- It allows for an evaluation of the biological basis for the effects seen in a group of chemicals within a category. When members of a certain chemical

category share a common mechanism of action, the confidence in the category is significantly greater than that associated with the use of the read-across approach to fill data gaps.

In a large category (more than 10 members), the presence or absence of hazards as well as the trend of an effect across the category can be identified more easily; this is a significant advantage over the read-across technique. For limited comparisons, it is harder to obtain the same level of confidence, particularly if the chemicals have multiple functional groups (OECD, 2007b; ECHA, 2008).

Explanation of concepts

Similarities within a chemical category may be based on the following.

- Common functional group(s) (e.g. aldehyde, epoxide, ester or specific metal ion)
- Common constituents or chemical classes (e.g. similar carbon range numbers)
- An incremental and constant change across the category (e.g. chain length) which is often related to an incremental change in physicochemical properties
- The likelihood of common precursors and/or metabolites which results in structurally similar chemicals

It is recognised that the formation of a chemical category is often dependent on which chemicals are manufactured by the companies sponsoring the category, or the regulatory context under which the evaluation is being made (OECD, 2007b; ECHA, 2008). Within a chemical category, data gaps may be filled by read-across (see Section 2.2), trend analysis (see below) and/or (Q)SAR. For a particular endpoint, category members are often related by a trend (increasing, decreasing or constant). Trend analysis can be carried out using a model based on the data for the members of the category. Data gaps can also be filled by an external (Q)SAR where the category under examination is a subcategory of the (Q)SAR model being used (i.e. the category falls within the applicability domain of the (Q)SAR; OECD, 2007b; ECHA, 2008).

Chemical categories can be described using a matrix consisting of the category members and a corresponding set of property and/or activity data (the category endpoints).

The choice of chemicals included in a category is often influenced by practical considerations. Ideally, the category should contain all potential chemical members when it is first developed; however, this would be very difficult to achieve.

Substances with several functional groups can belong to more than one category. If a category includes a high proportion of its potential members, conclusions drawn from the evaluation are likely to be more robust and the category more likely to be useful in reducing the amount of testing. If a substance is subsequently identified as a member of an existing category, it will be necessary to evaluate both the data for this substance in the light of the category evaluation and the category evaluation in the light of the data for the additional substance. If the initial category evaluation is sufficiently robust, the additional data are unlikely to alter significantly the initial evaluation.

In many cases, additional chemicals are identified which are at the boundaries of an existing category; these may need additional testing to confirm whether they belong to the category. The applicability domain of the category can also be used to help assess whether a substance is a member of an existing category. The applicability domain would specify the structural requirements and ranges of physicochemical, environmental fate, toxicological or ecotoxicological properties for which reliable estimations can be made (OECD, 2007b; ECHA, 2008).

A category can be divided into subcategories; this is helpful when an effect can be present for some but not all members of the category.

Subcategories are often endpoint specific and can arise for a number of reasons.

- Where the effect varies in intensity across the category such that some members may fall in one hazard class with other members in a different hazard class
- Where the effect may have a peak in activity or show a breakpoint in a trend
- Where trend analysis may apply to a subcategory but not the entire category

Subcategories improve the flexibility and practicality of the category approach, but they do not alter the scientific basis of the methodology (OECD, 2007b; ECHA, 2008).

Trend analysis and internal (Q)SAR models

For a particular endpoint, members of a category are often related by a trend across the category (e.g. increasing, decreasing or constant). The trend could be related to molecular mass, carbon chain length or to some other physicochemical property. Larger categories could have several different trends for a single endpoint, thereby defining subcategories. A breakpoint chemical is one that identifies a turning point in a trend, and sentinel chemicals are those that are at the opposite extremes of a trend (OECD, 2007b; ECHA, 2008).

A consistent trend is desirable in a chemical category; it indicates that the chemicals have a common mechanism of action and allows data gaps to be filled by scaling the measured values. Trends in experimental data are the basis for interpolation and extrapolation within the category. Interpolation is the estimation of a value using measured values from category members on both sides of the unknown data point. Extrapolation is the estimation of a value that is at or near the category boundary using measured values from internal category members. Interpolation is therefore generally preferred as it is usually a more reliable method. Extrapolation is necessary when there are data gaps for the boundary chemicals. It should be noted that extrapolation within a category is more robust than using read-across from an analogue (OECD, 2007b; ECHA, 2008).

When identifying trends in data there are several factors that should be considered, such as laboratory and experimental variations, different species or strains of animal used and different test protocols. Any deviations from the trend should be identified and explained where possible (OECD, 2007b; ECHA, 2008).

A quantitative trend between members will make the category approach more robust. An endpoint for which there is no toxic effect should be considered in terms of whether the tests failed to demonstrate the effect or whether it can be explained in terms of the mechanism of action of the substance (OECD, 2007b; ECHA, 2008).

A break in a trend within a category does not necessarily mean that the chemicals with different trends exhibit different toxicity pathways. Also, breaks should be distinguished from differences in hazard classification, which depend on an administrative cut-off for that property (OECD, 2007b; ECHA, 2008).

Data for a particular endpoint can be used to construct a (Q)SAR that describes the properties of the members of that category. Trends could also be used to construct a QAAR (quantitative activity–activity relationship; see below), based on the assumption that knowledge about the mechanism of action for one endpoint is applicable to a similar endpoint (OECD, 2007b; ECHA, 2008).

A chemical category can therefore be seen as a set of internal (Q)SAR (and possibly internal QAAR) for the different endpoints, with the advantage that all the underlying data are available to the assessor. Such models provide quantitative descriptions of the trends within a category and are derived directly from the experimental data for the category members. These models are based on only a relatively small dataset and usually work best for a homologous series of chemicals where the extrapolation is based on a simple parameter that can be linked to the physicochemical properties of the chemicals. If the category members are not a simple homologous series, it is essential that a parameter is established that can predict the trend across the category from which the missing values can be extrapolated (OECD, 2007b; ECHA, 2008).

Mechanistic basis of categories

A chemical category will often have a presumed common mechanism of action and the members will demonstrate this by the presence, absence or modulation of an effect (OECD, 2007b; ECHA, 2008).

In principle, a category is not endpoint specific, since structural changes across a category would be expected to change the whole spectrum of properties of the category members in a coherent and consistent manner. However, in practice it may not always be possible to identify the trends and changes for all the properties of interest. For some categories, the upper or lower ends of the series may show marked changes in their effects. A breakpoint within a category can indicate a change in the mode of action of the chemicals (OECD, 2007b; ECHA, 2008).

A common mechanism of action is an important factor when deciding which chemicals are relevant members of a category (OECD, 2007b; ECHA, 2008).

Robustness of a chemical category

There are several factors that contribute to the robustness of a category (OECD, 2007b; ECHA, 2008).

- The number of members in the category and the available data
- The quality of the experimental data for each of the endpoints
- The mechanistic basis that is thought to underpin the category
- The quality of the data estimated by external computational approaches.

The current guidance does not provide criteria to validate chemical categories; however, it does provide guidance about how to optimise the robustness of the category and how to document the justification for it (OECD, 2007b; ECHA, 2008).

Another important factor contributing to the robustness of a category is the definition of the applicability domain of the category. This is made up of the structural requirements and ranges of physicochemical, environmental fate, toxicological or ecotoxicological properties within which reliable estimations can be made for the category members.

Interdependence between categories and (Q)SAR

The concepts of chemical categories and (Q)SAR are strongly connected; in fact, the broad description of the concept for categories and the historical description of (Q)SAR are the same: *Chemical categories are formed and the measured data within them is used to estimate missing values for the untested members* (OECD, 2007b; ECHA, 2008). (Q)SAR often take the form of regression equations which are used to make predictions of activities or effects from a numerical measure of the chemical structure (OECD, 2007b; ECHA, 2008).

Similarly, a QAAR is a mathematical relationship between two biological endpoints or activities. QAAR are based on the assumption that knowledge about the mode of action for an endpoint is applicable to a similar endpoint because the main underlying processes of these related endpoints or activities are the same (e.g. partitioning, reactivity, enzyme inhibition). QAAR provide a means of performing trend analysis and filling data gaps (OECD, 2007b; ECHA, 2008).

The common scientific foundation between forming categories and QSAR/QAAR is that chemicals, once grouped together on the basis of common structural attributes, become chemical classes that exhibit consistent trends in their chemical properties and biological hazards. In addition, these trends in chemical activity are often related directly to trends in chemical structure expressed by (Q)SAR.

2.3.2 Stepwise procedure to develop categories

Chemical categories allow for hazard information to be obtained or estimated for all the individual chemicals in the category without the need to test each individual member. If there are sufficient experimental data to support the category evaluation, then each chemical within the category can be evaluated by read-across, (Q)SAR or trend analysis. If there are not sufficient experimental data to support the category then the following may be necessary (OECD, 2007b; ECHA, 2008).

- Perform limited and targeted testing
- Revise the category hypothesis and therefore the applicability domain in terms of its members and endpoints, or
- Abandon the category hypothesis; this should be a last resort

When using results from a category, it should be demonstrated that the category is robust by obtaining and documenting information about it. In order that this is done in a systematic and transparent way, it is recommended that the stepwise approach described in this section and shown in Figure 2.2 is used. Since there can be different starting points for the category formation (for example, from a single chemical, a small group of chemicals or a large group of chemicals) this method is flexible (OECD, 2007b; ECHA, 2008).

Step 0: Check whether the chemical is a member of an existing category

The first step is to check whether the query chemical is already a member of a category that has been evaluated. The following websites[13] have information on existing categories.

US EPA:
http://www.epa.gov/opptintr/newchems/pubs/chemcat.htm

OECD:
http://cs3-hq.oecd.org/scripts/hpv

United Nations:
http://www.chem.unep.ch/irptc/sids/OECDSIDS/sidspub.html

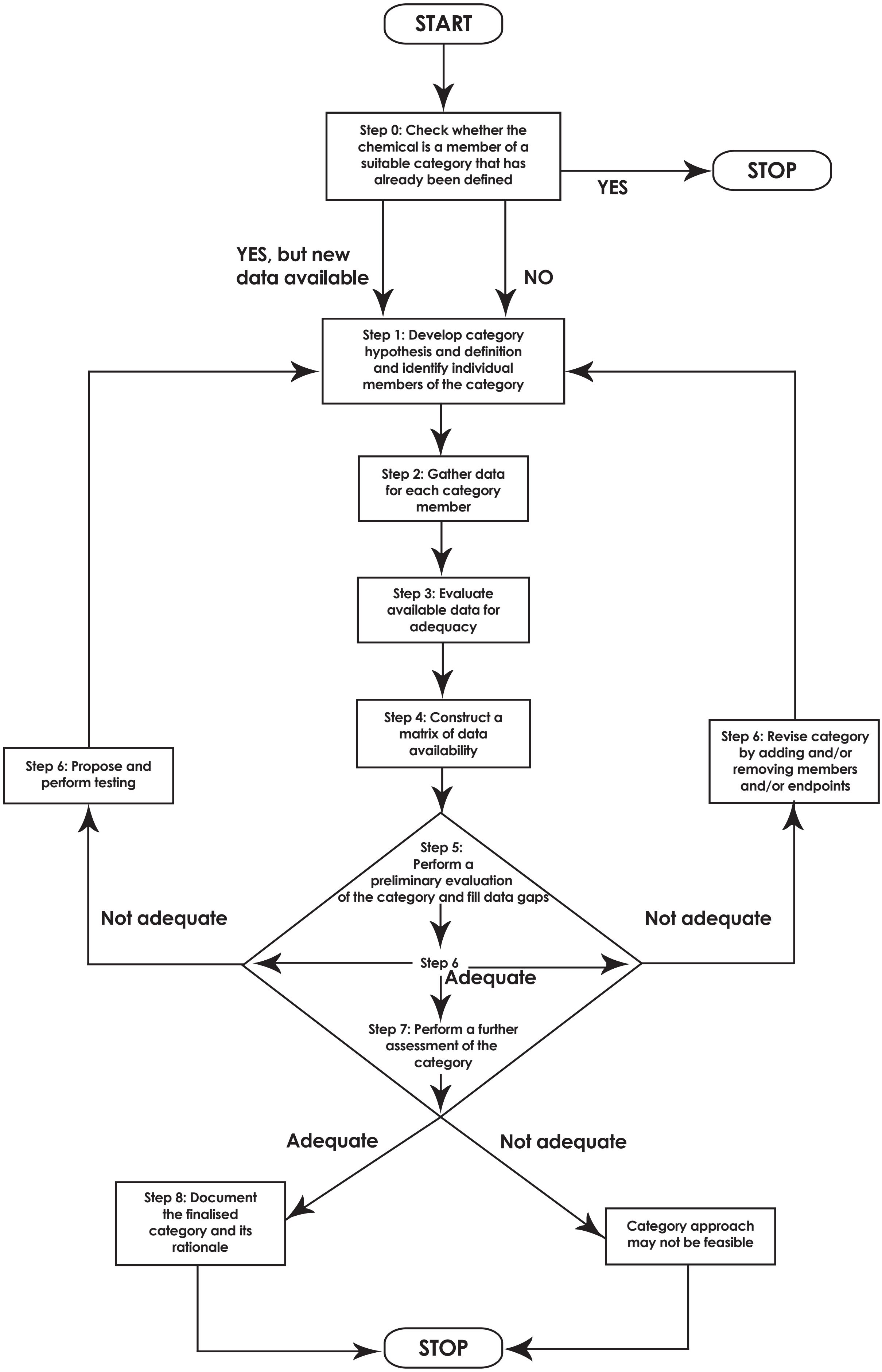
Some examples of industry sectors that apply grouping principles to evaluate health and environmental properties are petroleum substances, dyes and pigments, chlorinated paraffins, surfactants, hydrocarbon solvents, acrylate resins, petroleum additives, bitumen, flavours and fragrances (OECD, 2007b; ECHA, 2008).

Once a chemical is a member of a category that has been evaluated, the evaluation and results of the whole category can be used, taking account of the particular chemical's position within the category. New data can be used to verify the validity of the existing category and could also lead to a revision of the category, where necessary. In some cases, a relevant category may exist but the query chemical has not been specifically included in the category. In this case, it may be appropriate to extend the membership of the category to include the chemical of interest (OECD, 2007b; ECHA, 2008).

---

[13] Web addresses accessed March 2013

Figure 2.2 Stepwise procedure to category development



Step 1: Develop category definition and hypotheses and identify category members

A category definition should include a list of all the substances and endpoints covered. It could refer to chemical classes with a common functional group or chemicals with an incremental and constant change across the category. The chemical structure is usually the starting point, but the definition could also refer to a common mechanism of action or property. Some categories are defined by a metabolic pathway, where different members of the category are produced with each metabolic step. The category definition should include the molecular structure that is required to be in the category, with defined criteria such as carbon chain length and functionality (OECD, 2007b; ECHA, 2008).

Categories can be proposed for a specific endpoint or a selection of endpoints. All of the endpoints that can be expected to be relevant for the category should be included, because the wider the range of endpoints covered the more robust the results obtained from the category (OECD, 2007b; ECHA, 2008).

The hypothesis for the category should address the following (OECD, 2007b; ECHA, 2008).

- The similarities and trends in properties and/or activities that form the association between the members of the category
- The specific instances of read-across, trend analysis and computational methods that have been used
- The inclusion/exclusion rules that identify the ranges of values for which reliable estimations can be made for the given endpoint (also called applicability domain); these rules provide a way to extend the category to include other members at a future date

There are many approaches to selecting category members, from manual approaches to automated computer-based analogue searching methods.

Some categories have been formed using the SMILES notations. This approach is fast, simple and it can ensure that all possible members of a category are included. For new category proposals, computational methods can help develop the category hypothesis and define the category endpoints and members. The computational methods used depend on the starting point. If starting from a single chemical or a small group of chemicals, tools such as Derek Nexus, Leadscope or AIM would help build up the category, whereas tools such as TSAR and Cerius2 are helpful to identify substituents of a substructure (OECD, 2007b; ECHA, 2008).

Some examples of internet-based tools that can be used to find analogues are described in Table 2.15, Section 2.2.

It is important that all potential category members are described fully, including all the relevant CAS numbers (OECD, 2007b; ECHA, 2008).

The purity and impurity profiles of all potential category members should be collected as differences here could influence the toxicity. Category members should either have similar purity or, if different, the effects of the differences in purity should be known (OECD, 2007b; ECHA, 2008).

Step 2: Gather data for each category member

For each member of the category, both published and unpublished data should be gathered on the following.

- Physicochemical properties
- Environmental fate parameters
- Toxicity (human health)
- Ecotoxicity

There are many web-based databases containing information or data on environmental and human health effects of numerous chemicals. One of the most extensive of these databases has been made available by the OECD in collaboration with ECHA and other, global, organisations[14].

Step 3: Evaluate available data for adequacy

The available data should be evaluated for adequacy. The following factors should be taken into account when evaluating results for an individual substance within a category.

Different types of data can be available for the same endpoint. The scope of the estimated results cannot exceed the scope of the underlying data for the other members of the category. In addition, if the scope of the underlying data varies (e.g. a mix of results from screening tests and higher tier tests) for a particular endpoint, then it would be necessary to clarify the scope of the estimated results for the category members for which no experimental results are available. It may be appropriate to apply a WoE approach to all the data (OECD, 2007b; ECHA, 2008).

An effect that is defined by a cut-off point could yield different conclusions for individual substances within the same category. In this scenario the compounds should be carefully evaluated to decide whether there is a trend across the whole category or whether the

---

[14] http://www.echemportal.org/echemportal/index?pageID=0&request_locale=en [accessed March 2013]

data justify allocating substances to a series of subcategories (OECD, 2007b; ECHA, 2008).

If the data suggest possible breakpoints, they should be evaluated to ensure that this is genuine and not due to other factors, such as the testing being carried out in different laboratories, at different times or with different animal strains (OECD, 2007b; ECHA, 2008).

Datasets can contain apparent outlier(s). In this case, the results should be evaluated carefully to check whether it is a real difference in a mechanism of action across the category or whether the result should be questioned (OECD, 2007b; ECHA, 2008).

Step 4: Construct a matrix of data availability

A matrix of data availability that contains the category endpoints for each category member should be constructed. The matrix should be ordered to reflect any trends or progression within the category and it should indicate where data are available or missing. If possible, the cells should indicate the results of the key studies. For data-rich substances, the matrix could become very large and can be broken down into groups of endpoints (OECD, 2007b; ECHA, 2008). An example matrix is shown in Table 2.16.

Step 5: Perform a preliminary evaluation of the category and fill data gaps

A preliminary assessment of the category should be carried out to determine the following.

- That the rationale of the category is supported, i.e. the category does exhibit one or more of the trends postulated in the hypothesis.
- That the category is robust and contains sufficient, relevant and reliable information on the category members for the particular purpose. This should be checked for each endpoint, as there may be sufficient information for some endpoints but not for others.

The preliminary assessment is largely performed by expert judgement. The process is similar to the approach used to fill in data gaps. If both of the criteria above are satisfied for a particular endpoint, the data gaps can be filled and the category can be finalised and documented (OECD, 2007b; ECHA, 2008).

If both of the criteria are not satisfied, then the following options should be considered.

- If the data suggest that there is a pattern of effects for some of the chemicals in the group, then the category should be modified, for example, by dividing it into subcategories.
- If the category is reliable for one or more endpoints but adequate data do not exist, the category approach may still be proposed but may require additional testing (as described in Step 6 below). Animal welfare and financial issues would influence the choice of chemicals and endpoints for testing as well as the scientific basis for testing.
- If there are adequate data for an endpoint but no pattern, it could be that the category is not appropriate and should be abandoned.

Step 6: Perform and/or propose testing

If the preliminary assessment supports the rationale for the category but the data or information on the members are not sufficient, relevant or reliable, testing may be proposed or performed. When proposing testing, a number of factors should be considered, as follows.

- The proposed test should be influenced by the results of the preliminary evaluation of the category.
- If none of the members have any data for a particular endpoint, it may be appropriate to test a limited number of selected category members.
- Testing can be used to provide evidence of a particular effect that has been predicted by a trend in the category.

The test plan should summarise the existing data and how the proposed testing will be adequate. The matrix of data is a tool to present the available data, which displays the data for each endpoint for each category member. If toxicity is expected to change across the category (e.g. high to low toxicity), the members chosen for testing should bracket both ends of the toxicity range. If the category is large, testing should also be included for one or more members in the middle of the range. Subcategories or subsets of the category should be defined by the limits of the subcategory and accompanied by data. There are no rules as to the number of cells that must be filled within a data matrix; the acceptability is dependent on the members in the category, the endpoint and the confidence in the extrapolated data (OECD, 2007b; ECHA, 2008).

Test plans are intended to provide information about the properties of the group as a whole, not necessarily about any specific chemical. This means that key substances identified for testing may have little or no commercial importance and may even need to be synthesised specifically for this purpose. This may still be more economical, both in terms of costs and animal usage, than a conventional testing strategy (OECD, 2007b; ECHA, 2008).

Table 2.16 Example data matrix

Chemical name	Category member 1	Category member 2	Category member 3	Category member 4
CAS number	CAS #1	CAS #2	CAS #3	CAS #n
Physicochemical data				
Melting point				
Boiling point				
Density				
Vapour pressure				
Partition coefficient				
Water solubility				
Etc.				
Environmental fate and pathway				
Photodegredation				
Stability in water				
Transport and distribution				
Aerobic biodegredation				
Etc.				
Environmental toxicity				
Acute toxicity to fish				
Acute toxicity to aquatic invertebrates				
Toxicity to aquatic plants				
Etc.				
Mammalian toxicity				
Acute oral toxicity				
Acute dermal toxicity				
Acute inhalation toxicity				
Repeated dose toxicity				
Genetic toxicity *in vitro*				
- Gene mutation				
- Chromosomal aberration				
Genetic toxicity *in vivo*				
Reproductive toxicity				
- Fertility				
- Developmental toxicity				

Step 7: Perform a further assessment of the category

If new data are generated, the category should be revised and reassessed to determine whether the two criteria detailed in Step 5 are met.

If the results do support the category, the remaining data gaps can be filled and the category finalised and documented.

If the results do not support the category, the options are: a) to carry out further testing; b) to change the category (for example by dividing it into subcategories); or c) to abandon the category (OECD, 2007b; ECHA, 2008).

Step 8: Document the finalised category

The finalised category should be documented in a suitable reporting format (see Section 2.3.4). Categories can be revised in light of new data and/or experience, or on the basis of new chemicals being discovered or manufactured (OECD, 2007b; ECHA, 2008).

2.3.3 Guidance on specific types of category

Some examples of specific types of category (e.g. chain length, metabolic pathways, chemical reaction products and multi-constituent substances, isomers, complex substances, metals and metal compounds) are described in ECHA, 2008 (pp. 105–120).

2.3.4 Reporting format for the category approach

The following reporting format for the category approach has been recommended by the OECD and at the EU level in relation to industrial chemicals under REACH.

Category definition and its members

Category definition

1 Category hypothesis – Describe the molecular structure that a chemical must have to be included in the category. Provide a brief hypothesis for why the category was formed, the hypothetical relational features of the category, purported mechanisms and trends in properties that are

thought to collectively generate an association between the members. All functional groups of the category members need to be identified. If there is a mechanistic reasoning to the category, describe the foreseen mode of action for each category member and if relevant describe the influence of the mode of administration (oral, dermal, inhalation; OECD, 2007b; ECHA, 2008).

2. Applicability domain of the category – Describe the set of inclusion and/or exclusion rules that identify the ranges of values within which reliable estimations can be made for category members. Clearly indicate the borders of the category and for which chemicals the category does not hold (OECD, 2007b; ECHA, 2008).
3. List of endpoints covered – List the endpoints for which the category approach is applied. Also indicate if for some endpoints the category approach can only be applied to a subset of the members of the category (subcategories) (OECD, 2007b).

Category members

Describe all the category members as comprehensively as possible. Provide CAS numbers, names and chemical structures of all category members (OECD, 2007b; ECHA, 2008).

Purity/impurity

Provide purity/impurity profiles for each member of the category, including their likely impact on the category endpoints. Discuss the influence these impurities are thought to have on physicochemical parameters, fate and (eco)toxicology (OECD, 2007b; ECHA, 2008).

Category justification

Based on available experimental data (including appropriate physicochemical data and additional test results generated for the assessment of the category) summarise how these results verify that the category is robust. This should include an indication of the trend(s) for each endpoint. The data should also show that functional groups not common to all the (sub) category members do not affect the anticipated toxicity. The available experimental results in the data matrix should support the justification for the read-across (OECD, 2007b; ECHA, 2008).

Data matrix

Provide a matrix of data (category members vs endpoints). It should be constructed with the category members arranged in a suitable order (e.g. reflecting a trend or progression within the category). In each cell, the study result type should be indicated (e.g. experimental result, experimental study planned, read-across from supporting substance, trend analysis, (Q)SAR). If experimental results are available, the key study results should be shown in the data matrix (OECD, 2007b; ECHA, 2008).

Detailed discussion of how data gaps are filled for individual endpoints and individual category members and the rationale for the chosen method should also be provided (OECD, 2007b; ECHA, 2008).

2.3.5 Tools for category formation and read-across: The OECD (Q)SAR Toolbox

The OECD (Q)SAR Toolbox is computer software developed with the aim of making (Q)SAR approaches readily available and accessible and to improve their regulatory acceptance. The philosophy of the Toolbox is based on the 'chemical category' concept. The fundamental features of the Toolbox are as follows.

- Identification of relevant structural characteristics and potential mechanism or mode of action of a target chemical
- Identification of other chemicals that have the same structural characteristics and/or mechanism or mode of action (analogues)
- Use of existing experimental data to fill the data gap(s)

The identification of structural characteristics and potential mechanisms or modes of action is achieved with a set of 'profilers' in the module 'Profiling'. These profilers identify structural alerts involved in specific reactions or binding mechanisms relevant for different regulatory endpoints. Some of the profilers included in the Toolbox are 'Cramer classification', 'OASIS Acute Toxicity', 'protein binding', 'DNA binding', 'Benigni/Bossa rule-base', 'ER binding' and 'BfR rule-bases for skin and eye irritation/corrosion'.

The Toolbox includes six sequential modules.

1. Chemical input – provides several means of entering the chemical of interest or target chemical and of making sure the molecular structure assigned to it is the correct one.
2. Profiling – provides the means to profile the target chemical to identify relevant structural characteristics and potential mechanisms or modes of action. This leads to the identification of potential analogues of the target chemical.
3. Endpoints – provides a process of retrieving results for regulatory endpoints for the identified analogues which are stored in the Toolbox. This data gathering can be executed in a global fashion (i.e. collecting all data on all endpoints), or on a more narrowly defined basis (e.g. collecting data for a single or limited number of endpoints).

4 Category definition – provides several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule. This is the critical step in the workflow and several options are available in the Toolbox to assist the user in refining the category definition via sub categorisation.

5 Filling data gaps – provides three options for making an endpoint-specific prediction for the untested target chemical. These options, in increasing order of complexity, are: read-across, trend analysis and (Q)SAR models.

6 Report – the final module provides the user with a downloadable written audit trail of the functions the user performed using the Toolbox to arrive at the prediction.

2.3.6 Toxicologically meaningful categories

Category formation is a topic of paramount importance to predictive toxicology. The challenge in grouping chemicals for hazard assessment is placing them into groups that are meaningful for a particular toxicity endpoint. This challenge arises because of the discontinuity between chemical and toxicological spaces, which are often endpoint-related. Substances which are similar in structure can be dissimilar in terms of toxic action, including the ability to elicit a particular hazard endpoint, as well as potency within a particular hazard endpoint. Therefore, a means of segregating toxicological domains or forming a toxicologically meaningful category (TMC) is essential, as using TMC increases the confidence that the predictions made are accurate (Cronin & Madden, 2010).

Chemicals, historically, have been grouped together using a common structural feature exhibited by all the substances included in the group. However, inherent to forming a TMC is that the compounds in the category share a common mechanism of action or a common adverse outcome pathway (AOP) leading to coherent trends in their toxicological effects. An AOP is a set of chemical, biochemical, cellular and physiological responses that characterise the biological effects cascade resulting from a particular molecular initiating event (MIE).

As different hazard endpoints are determined by different AOP, it is unlikely that the members of a TMC will be the same for any two hazard endpoints. Hence, the membership of TMCs tends to be endpoint specific.

It should be noted that while it is relatively easy to form TMC for those hazard endpoints (e.g. irritation, genotoxicity) with simple AOP where plain chemical reactivity is the sole MIE, it is more difficult to form such categories for more elaborate hazard endpoints (e.g. repeated dose toxicity, developmental toxicity, carcinogenicity) with complex AOP where damage is the result of the propagation at the different biological levels of a multitude of multiple MIE.

Data to identify complex AOP and to measure the key events along the pathway are scattered and fragmentary. However, with the increasing development of advanced systems' biology, 'omic' technologies and other *in vitro* methods, there is the potential to generate such information.

An example of a preliminary attempt to form a category based on the concept of TMC comes from a recent research project undertaken by the nickel producers. In this project, different nickel compounds have been grouped together in relation to the reproductive toxicity endpoint and other systemic effects on the basis of bioavailability (bioaccessibility) data for the nickel ion in biological fluids relevant to different routes of exposures, such as sweat (for the dermal route), stomach fluid (for the oral route) and lung or interstitial fluid (for the inhalation route). The industry has interpreted the results as to show that the release of the nickel ion in relevant portal-of-entry biological fluids is a better predictor of the systemic availability and reproductive toxicity effects of nickel compounds than their water solubility. This is because water solubility is further removed from the molecular initiating event leading to the adverse outcome.

2.4 Stepwise approach for the use of non-testing data

2.4.1 Introduction and structured workflow

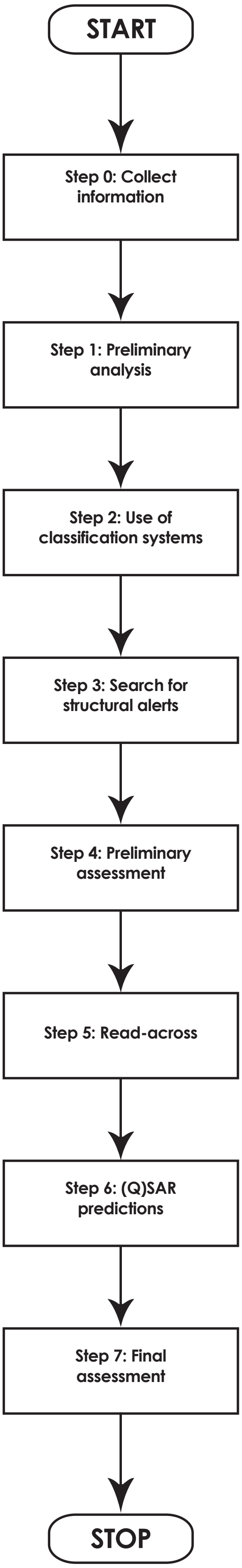
The following stepwise approach for the integrated use of non-testing approaches has been proposed at EU level in relation to industrial chemicals under REACH. The workflow described, and shown in Figure 2.3, is intended to be flexible so that it can be adapted to meet the specific needs of the user; however, it is recommended that all the steps are considered in order to increase confidence in the overall assessment (ECHA, 2008).

Step 0: Information collection

First the information that is required for the particular regulatory purpose should be considered (ECHA, 2008).

A representative structure of the chemical of interest should be selected. This is normally the two-dimensional structural formula of the chemical in the form of a SMILES code. The purity/impurity profile may also be useful (ECHA, 2008).

Figure 2.3 Flowchart for the use of non-testing approaches in the regulatory assessment of chemicals (ECHA 2008)



The structure of the chemical of interest should be verified; for example by checking that the structure agrees with the CAS number or formal chemical name. This can be done using a 'structure converter tool'; some online tools that can be employed at this step are listed below.

- ChemID can be used to check CAS number, chemical name and to identify the possible structure.
- Ambit can be used to convert CAS to SMILES.
- CAS SciFinder is a definitive source of CAS registry numbers, chemical names and structure information.
- The (Q)SAR Application Toolbox contains libraries that convert CAS numbers to SMILES.

The available information for the parent chemical should be collected. This should include physicochemical properties, toxicity data, experimental data and non-testing data. Some sources of data and information are listed below.

- The European Chemical Substances Information System (ESIS)
- The (Q)SAR application toolbox contains a database of available experimental data, a library of (Q)SAR models, chemical categories and a database of (Q)SAR predictions
- The (Q)SAR model database (QMDB) has robust summaries of (Q)SAR that can be searched by endpoint or chemical; it can provide information on whether the chemical is in the training or test set of some models
- The (Q)SAR prediction database (QPDB) can be used to generate predictions using models and to store these predictions
- Chemical category database contains an inventory of existing categories

A working matrix of data availability, endpoint by endpoint, should be built and the information gaps identified. From the working matrix, an endpoint can be selected for which a data gap has been identified, and processed through the remaining steps of the workflow (ECHA, 2008).

Step 1: Preliminary analysis of reactivity, uptake and fate

The preliminary analysis of reactivity, uptake and fate is based on existing information as well as inferences made by using physicochemical data.

First, information about the reactivity (biotic and abiotic) of the parent chemical should be collected;

this can be obtained from peer-reviewed literature and from tools and databases, including CAS SciFinder, MDL Reaction Database, TIMES, CATABOL, METEOR, META, MCASE and HYDROWIN as part of EPIWIN (for hydrolysis only). The stability of the parent chemical can be further estimated by analysing fragments and molecular orbital energy; at present this is performed by experienced chemists. The working matrix should be updated with the collated information by adding a row with information about metabolites and reaction products.

A preliminary analysis of uptake and fate is performed by taking into account the following considerations (ECHA, 2008).

- The effect of molecular weight, size, log K_{ow}, electric charge and stability on uptake (and toxicity)
- Whether ionisation can take place at the relevant pH and whether this affects uptake and fate
- The chemical reactivity (type of reactions) expected for the parent chemical
- The metabolites and reaction products that are generated

Finally, a suitable query chemical should be selected. The preliminary analysis should be used to determine which chemicals (parent or reaction product or metabolite) are suitable for modelling the endpoint of interest. The remaining steps are then applied to the query chemical(s) identified. This first step may also help assess the likelihood of exposure to the organism (tissue) or environmental compartment (ECHA, 2008).

Step 2: Use classification schemes for endpoint of interest

Information on the likely biological activity of the query chemical may be obtained using classification schemes for the endpoint of interest. For example, the classification scheme developed by Cramer *et al.* (1978) is useful for evaluating the likely systemic toxicity of a chemical. Toxtree implements the Cramer classification scheme.

This step may help to classify the mode of toxic action of the chemical. This information is useful in a later step when evaluating which (Q)SAR models should be applied (ECHA, 2008).

Step 3: Search for structural alerts for endpoint of interest

In this step, structural alerts for the endpoint of interest are searched. Software programs that are available to perform this analysis include Derek Nexus, MCASE and Leadscope. This step may help identify which hazards are likely to be present or absent (ECHA, 2008).

Step 4: Preliminary assessment of expected type of reactivity, uptake, toxicity and fate

In this step, which requires expert judgement, a preliminary assessment of the expected reactivity, uptake, toxicity and fate profile of the parent chemical is performed by using the outcomes of the previous steps applied to all the relevant query chemicals.

Step 5: Read-across

This step is aimed at filling data gaps for all the query chemicals using a read-across analogue approach. Read-across is based on the identification of similar chemicals, so the first step is to establish whether the query chemical belongs to an existing category. If the query chemical does not belong to an existing category and cannot reasonably be associated with one, an appropriate analogue should be searched (ECHA, 2008). Tools to help with analogue searching are listed below.

- Analog Identification Methodology (AIM)
- AMBIT (Ideaconsult Ltd)
- Danish (Q)SAR Database
- ChemFinder
- ChemID Plus Advanced
- Leadscope
- Superfragment (BioByte Inc)
- Toxmatch (ECB)

Once relevant analogues are identified, the experimental data available on these analogues should be collected and stored in the working matrix (ECHA, 2008).

Subsequently, read-across should be performed with the information from the relevant analogues. If, however, the read-across is unsuccessful in providing a reliable estimate, the search for other analogues should be expanded (ECHA, 2008).

Step 6: (Q)SAR

In this step, predictions of uptake, toxicity and fate are generated using (Q)SAR or expert systems.

First the user should retrieve the available predictions for the endpoint of interest from the (Q)SAR prediction database. If this database does not contain predictions for the query chemical, relevant models can be searched in the (Q)SAR model database. Information gathered in previous steps is useful to aid selection of a suitable model. Valid (Q)SAR should be assessed to check that the query chemical is within the applicability domain of the model. This assessment may be performed by using appropriate tools (e.g. Ambit Discovery; ECHA, 2008).

If relevant (Q)SAR have not been found using the databases above, other models should be searched. When relevant (Q)SAR are identified, predictions for the query chemical should be generated and added to the working matrix (ECHA, 2008).

Step 7: Overall assessment

Steps 1 to 6 are assessed for the chemical and the endpoint(s) of interest to generate an overall assessment. The toxicity of the parent chemical is assessed using the information obtained for all the query chemicals (metabolites, reaction products and analogues; ECHA, 2008).

2.5 Integrated testing strategies and the prediction of toxic hazard

Over the past two or three decades there has been mounting pressure to reduce reliance on laboratory animal testing for predicting hazards and assessing risks in relation to industrial chemicals and chemical products of various kinds including pharmaceuticals and agrochemicals, household and personal care products. Initially, this pressure was largely the result of ethical and animal welfare concerns, but the emphasis is now increasingly on scientific and economic considerations. It is increasingly recognised that animal tests are often not relevant for predicting hazard in humans because of species differences of various kinds. In addition, they are expensive and can a long time to perform (Cronin & Madden, 2010).

In response to this situation, a large amount of effort is being invested in the development of non-animal test procedures which make use of existing data and bioinformatics, novel *in chemico*, *in silico* and *in vitro* approaches and, where possible, ethical studies in humans. The result is an increasing complexity and variety of methods based on mechanisms of action at the molecular, cellular, tissue, organ and system levels, which are suited to answering highly specific questions rather than to providing information on toxic effects of all kinds (Cronin & Madden, 2010).

This means that the available tests will have to be used intelligently and selectively in combination as batteries and/or in tiered hierarchical schemes in what have come to be known as integrated testing strategies (ITS).

Before any ITS can be considered acceptable for use, especially in compliance with regulatory requirements, it must be established that it is relevant and reliable for a particular purpose (e.g. for classification and labelling, or risk assessment purposes).

ITS are hierarchical in nature and are used in a stepwise fashion, progressing from the evaluation of pre-existing data and a consideration of the physicochemical properties of a test item, via *in silico* modelling, to the use of *in vitro* methods followed – sometimes and where necessary – by *in vivo* tests in animals and in human volunteers. The stepwise approach usually represents a decision tree scheme, that is, at the conclusion of each step decisions can be made about whether or not further testing is advisable and, if so, which tests should or should not be performed (Cronin & Madden, 2010).

To secure the regulatory acceptance of ITS, there should be a rigorous evaluation, endpoint by endpoint, of which alternative systems could be considered worth the effort after objective comparison with their animal test rivals. This should be primarily the responsibility of companies producing the materials requiring testing, in collaboration with those capable of developing the alternative methods and the schemes. It is also important that regulatory bodies give preference to WoE approaches above very prescriptive and highly specific tests, and that testing requirements are adapted depending on the nature of the chemical, its projected use and the likely human exposure (Cronin & Madden, 2010).

ITS have been the focus of a recent European project developed under the Sixth Framework programme called OSIRIS[15]. The project aimed at optimising strategies for risk assessment of industrial chemicals through integration of non-test and test information. To this end, operational procedures that guide a transparent and scientifically sound evaluation of chemical substances in a risk-driven, context-specific and substance-tailored manner were developed, tested and disseminated. For the first time, ITS have been equipped with decision frameworks that include chemical and biological read-across, *in vitro* results, *in vivo* information on analogues, qualitative and quantitative SAR, thresholds of toxicological concern and exposure-based waiving.

The central software outcome of OSIRIS is the OSIRIS web tool. This tool guides in performing ITS on skin sensitisation, repeated dose toxicity, mutagenicity, carcinogenicity, bioconcentration factor and aquatic toxicity. The ITS are implemented as user interactive workflows that weigh different types of data (obtained from databases computed with models or input by the user). The ITS goal is to conclude whether there is sufficient information for classification and labelling and risk assessment purposes, and to suggest the appropriate test where not enough information is available for decision making.

---

[15] http://www.osiris.ufz.de [accessed March 2013]

3 Examples and Case Studies of Current Use of Predictive Approaches by UK Government Department and Agencies

The main objective of this section is to provide real examples of how predictive approaches are currently used by UK government departments and agencies for regulatory purposes.

It is hoped that these examples will help:

1. to identify issues requiring further work;
2. to illustrate the thought process behind the application of these techniques;
3. to increase the transparency of the decision-making process undertaken by UK government;
4. to identify the limits of these techniques;
5. to facilitate understanding of these approaches;
6. to enhance the regulatory acceptance of these tools;
7. to expand the opportunities for future application of predictive toxicology.

3.1 Case studies provided by HSE

3.1.1 A case study for the application of a read-across approach at HSE

Zineb (zinc ethylenebisdithiocarbamate) is an existing active substance for use in biocidal products. An evaluation of the risks to human health and the environment from the use of such zineb-containing products has been performed within the scope of Directive 98/8/EC (BPD – Biocidal Products Directive).

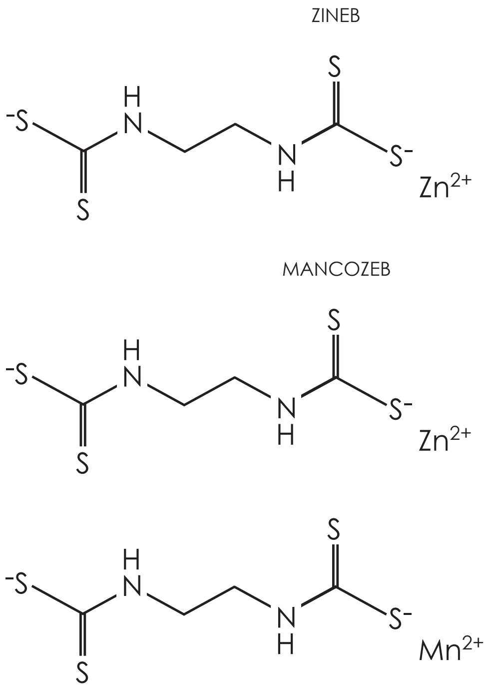
The toxicity dataset for zineb required under the BPD is not complete as, for some endpoints (toxicokinetics by the oral route, subchronic and chronic toxicity, mutagenicity, carcinogenicity, reproductive toxicity and neurotoxicity), studies are either not available or are not of the quality standard required by the BPD. For these endpoints, qualitative and quantitative read-across from data on mancozeb has been performed. The approach was proposed by the applicant.

Read-across from mancozeb is scientifically justified on the following basis.

a) Zineb and mancozeb have very similar molecular structures (Figure 3.1). They are both ethylenebisdithiocarbamate (EBDC) salts containing the metals zinc and/or manganese. In zineb the only metal ions present are Zn^{2+}, whereas mancozeb contains both Mn^{2+} and Zn^{2+} in the ratio 10:1.

b) The physicochemical properties associated with the dissociation of the molecules to release the EDBC^{2-}

Figure 3.1 Molecular structures of Zineb and Mancozeb



anions and the metal cations are very similar for zineb and mancozeb.

c) It is widely recognised that the effects of EDBC compounds are entirely due to the EDBC anion. The higher content of Zn^{2+} in zineb is not considered to be of toxicological relevance as zinc is an essential element of the body and is of relatively low toxicity in humans.

d) Zineb and mancozeb show comparable ADME profiles. Oral toxicokinetic data on mancozeb and public domain ADME studies on zineb show that both substances are rapidly and extensively absorbed by the oral route, are widely distributed around the body, are extensively metabolised to similar metabolites, with ethylene thiourea (ETU) being the major metabolite (~20% of the administered dose) for both substances, and are rapidly excreted, mainly via urine.

e) Where data are available for both substances, these show that zineb and mancozeb have comparable toxicological profiles. They are both of low acute toxicity, they are not skin or eye irritants and they both cause developmental toxicity in the presence of maternal toxicity. In repeated dose toxicity studies in rats and dogs, they both produce effects on the thyroid gland at similar dose levels. It is well documented that ETU, their common metabolite, is the entity responsible for the observed thyroid effects.

Conclusion: The read-across approach was accepted by all EU Member State regulatory authorities.

3.1.2 A case study for the application of a read-across approach at HSE where the proposition was rejected

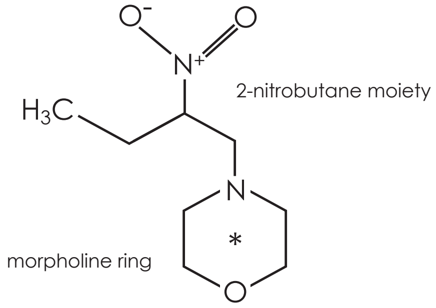
NBM (4-(2-nitrobutyl) morpholine) is an existing active substance for use in biocidal products. An evaluation of the risks to human health and the environment from the use of NBM-containing products is required within the scope of Directive 98/8/EC (BPD – Biocidal Products Directive). HSE, as the UK competent authority responsible for the implementation of the BPD, was the designated Rapporteur Member State in charge of the evaluation.

The toxicity dataset for NBM prescribed by the BPD was not complete as, for some endpoints (chronic toxicity and carcinogenicity), studies were not available. For these endpoints the industry duty holder had proposed that read-across from data (96-week drinking water study in mice and 104-week inhalation study in rats) on morpholine be performed, mainly on the basis that morpholine is the only major *in vivo* metabolite of NBM.

HSE assessed the applicant's request and came to the conclusion that the proposed read-across approach from morpholine had numerous shortcomings and was not sufficiently sound from a scientific point of view for the following reasons.

a) NBM contains the two moieties of 2-nitrobutane and morpholine (Figure 3.2).

Figure 3.2 Chemical structure of 4-(2-nitrobutyl) morpholine (NBM)



N-dealkylation of NBM (Dow Chemical Company BIOBAN P-1487) leads to the formation of morpholine and 2-nitrobutanal. At the same time, NBM could also be metabolised by reduction of the nitro group by the intestinal microflora to generate morpholine and 2-aminobutanal, as shown below (Figure 3.3).

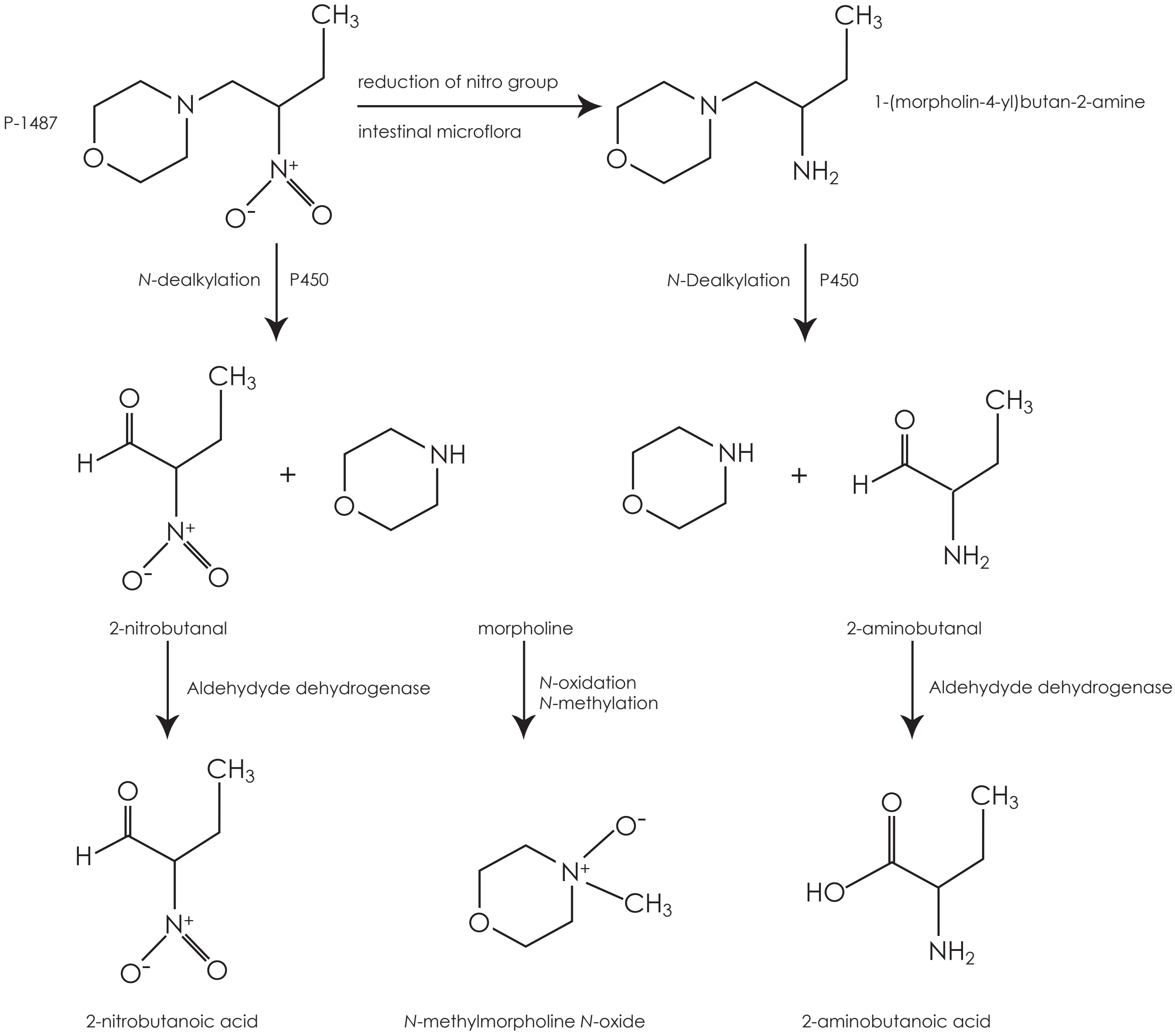
A well-conducted oral toxicokinetic study in rats with NMB (radiolabel located within the morpholine ring structure) showed that only up to a maximum of 30% of NBM could be estimated to be metabolised to morpholine and its derivative, *N*-methylmorpholine *N*-oxide. A significant amount of NBM (up to 70%) was predicted to be converted to 2-nitrobutanal and 2-aminobutanal and their carboxylic acid derivatives.

Overall, therefore, HSE considered that the evidence did not support the claim that morpholine was the only major *in vivo* metabolite of NBM. Performing a read-across from data on morpholine could not have taken into account the potential toxicological effects of the other NBM metabolites.

b) (Q)SAR model predictions generated by HSE for the two aldehyde derivatives (2-nitrobutanal and 2-aminobutanal) indicated a concern for mutagenicity and carcinogenicity which could not have been explored or addressed by the data on morpholine.

c) The two chronic/carcinogenicity studies available on morpholine were relatively old investigations of low reliability and quality.

Figure 3.3 Metabolic pathway for 4-(2-nitrobutyl) morpholine (NBM)



d) Where data were available for both NBM and morpholine, a reliable comparison of the toxicity profiles of NBM and morpholine was not possible owing to differences in species, study duration, dose levels, methodology, study quality and parameters investigated. Furthermore, the limited information available indicated not only potential quantitative (potency) differences, but also qualitative discrepancies (e.g. in some of the target organs of toxicity).

Conclusion: HSE rejected the proposed read-across approach and the NBM submission was eventually withdrawn by the applicant.

3.1.3 A case study for the application of a simple category approach at HSE

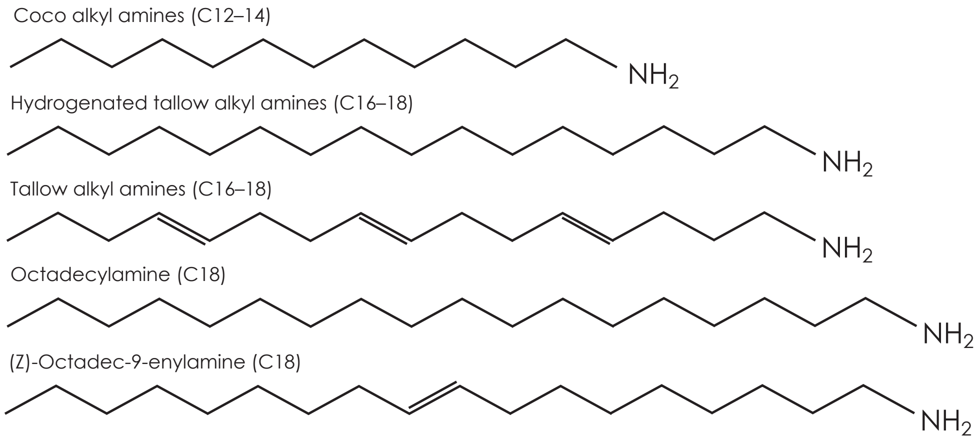
Coco alkyl amines, tallow alkyl amines, hydrogenated alkyl amines, octadecylamine and (Z)-octadec-9-enylamine are five primary alkyl amines included in the 4th priority list of the Existing Substances Regulation (EEC 793/93), an EU programme for identifying risks from industrial chemicals to human health and the environment.

The toxicity dataset for the majority of these alkyl amines is very limited. For some endpoints, data are available only on one of the five substances. To fill these data gaps, a category/grouping approach was used. The approach was proposed by industry.

The basis for applying such an approach is as follows.

a) The five alkyl amines under review share structural similarities, that is, the terminal amine group and the linear hydrocarbon chain (Figure 3.4). Each of the five alkyl amines is a mixture of primary straight chain alkyl amines with different degrees of saturation (no, 1, 2 or 3 double bonds).

Figure 3.4 Chemical structures of selected alkyl amines



b) There are two important trends in this group of alkyl amines: increasing chain length and increasing saturation (double bonds).

- Coco alkyl amines: C12–14, <10% unsaturated fraction
- Hydrogenated tallow alkyl amines: C16–18, <5% unsaturated fraction
- Tallow alkyl amines C16–18, >40% unsaturated fraction
- Octadecylamine C18, <5% unsaturated fraction
- (Z)-Octadec-9-enylamine C18, >70% unsaturated fraction

c) On the basis of these structural similarities and these trends, it can be predicted that as all members of the group have a terminal amine group, their pH is basic; hence they will all have irritant and/or corrosive properties. It can also be predicted that with the increasing degree of unsaturation, the reactivity will increase. In addition, the longer the alkyl chain attached to a double bond, the greater the reactivity. Based on these considerations, the reactivity in this group of alkyl amines is expected to increase in the following order: octadecylamine, hydrogenated tallow alkyl amines, coco alkyl amines, tallow alkyl amines, (Z)-octadec-9-enylamine.

d) For those endpoints where data were available for all members of the group, the predicted order of increasing reactivity was confirmed.

Conclusion: The category approach was accepted by all EU Member State regulatory authorities. For those endpoints where data were available only at the higher or more reactive end of the category, a straight read-across was performed as this represented the worst case. For those endpoints where data were available only for substances at the lower end or in the middle of the category, further testing of the higher-end members was requested.

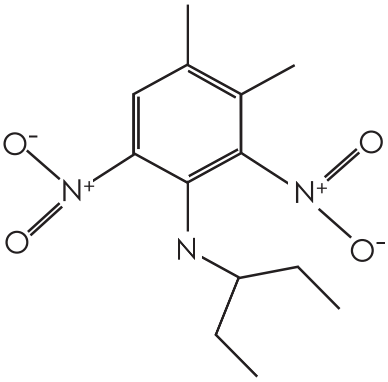
This category approach was also recently used by the European Chemical Agency’s Risk Assessment Committee when forming an opinion on a proposal for harmonising the classification and labelling of these substances. The trend in reactivity was found relevant for both irritant potential and the level of classification for acute toxicity, which is assigned on the basis of potency.

3.1.4 A case study for the application of a QSAR prediction at HSE

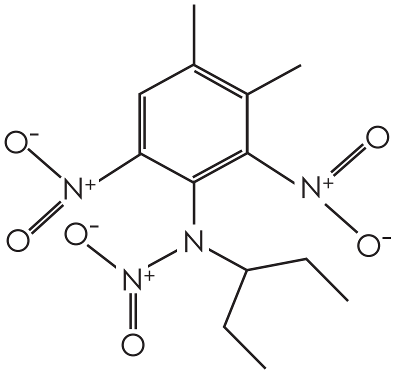
The toxicological significance of impurities in commercial sources of pesticide active substances and the toxicity of metabolites formed in crops or the environment needs to be assessed. One of the requirements is to perform a (Q)SAR evaluation of the metabolite or impurity. No specific guidance is given on how to perform the (Q)SAR, nor on how to interpret the output. HSE finds it useful if the (Q)SAR results for the metabolite or impurity are presented alongside those for the active substance. This permits the alerts to be set in context of the observed toxicity of the parent and effort focused on any alerts that are unique to the metabolite or impurity.

An example of this approach is presented below for a new manufacturing source of the pesticide pendimethalin, containing low levels of a nitro impurity (NB this impurity is in the public domain) not present in the original technical material, based on Derek Nexus. Chemical structures of the pesticide and impurity are given in Figure 3.5, with the output from Derek Nexus presented in Table 3.1.

Figure 3.5 Chemical structures of pendimethalin and *N*-nitro-pendimethalin



Pendimethalin



N-nitro-pendimethalin

Table 3.1 Alerts for pendimethalin and *N*-nitro-pendimethalin generated using Derek Nexus

Alert	Pendimethalin		*N*-nitro-pendimethalin	
	Present	Confidence	Present	Confidence
070 – Carcinogenicity (*N*-nitro or *N*-nitroso)	No	-	**Yes**	**Plausible**
105 – Carcinogenicity (aromatic nitro-)	Yes	Plausible	Yes	Plausible
544 – Hepatotoxicity (aromatic nitro-)	Yes	Plausible	Yes	Plausible
007 – Mutagenicity (*N*-nitro or *N*-nitroso)	No	-	**Yes**	**Plausible**
329 – Mutagenicity (aromatic nitro-)	Yes	Plausible	Yes	Plausible
007 – Chromosome damage (*N*-nitro or *N*-nitroso)	No	-	**Yes**	**Plausible**
329 – Chromosome damage (aromatic nitro-)	Yes	Plausible	Yes	Plausible
447 – Skin sensitisation (*N*-nitro or *N*-nitroso)	No	-	**Yes**	**Plausible**
566 – Teratogenicity (*N*-nitro or *N*-nitroso)	No	-	**Yes**	**Plausible**

In this instance, alerts 105, 544 and 329 can be discounted as they are triggered by structures common to both compounds.

The remaining alerts are related to the *N*-nitro grouping, but also triggered by *N*-nitroso compounds. Pendimethalin contains low levels of a number of *N*-nitroso compounds and it was considered acceptable to read-across from the *N*-nitroso compounds present in pendimethalin to *N*-nitro-pendimethalin. The material tested in studies covering the endpoints identified in the alerts and the current approved technical material contains higher overall levels of *N*-nitroso compounds than the total level of *N*-nitroso compounds plus *N*-nitro-pendimethalin in the new source.

Conclusion: It was concluded that the levels of *N*-nitro-pendimethalin present in the new source would be unlikely to have an impact on the toxicity profile of pendimethalin. The new source of pendimethalin was considered to be toxicologically equivalent to the existing source and authorised for use.

3.2 Case studies provided by HPA

3.2.1 Use of a surrogate marker approach by the HPA to assess the toxicity of polycyclic aromatic hydrocarbons in soil

Polycyclic aromatic hydrocarbons (PAH) are a group of organic compounds that contain two or more fused aromatic rings. Several PAH and mixtures of PAH have been shown to be genotoxic and to cause cancer in experimental animals. Former industrial 'brownfield' sites are important sources of exposure to PAH, for example former gasworks sites that are often contaminated with coal tar residues. Consequently it is important to assess the risk posed by PAH present in soil at such sites so that the risk to health can be reduced to a tolerable level.

Risk assessment challenges

It is difficult to assess mixtures of similar chemicals of PAH since the hazard, mechanism of toxicity and potency may vary among chemicals. One approach would be to determine the hazard posed by each individual PAH and to estimate the dose associated with a minimal risk of adverse health effects, thereby allowing the risk associated with exposure to each individual PAH to be assessed. Unfortunately the toxicity database and analytical methods available for these chemicals are insufficient to perform such a detailed risk assessment. Furthermore this approach would not take account of any possible combined effects of a mixture of PAH. The toxic equivalency approach is considered inappropriate because PAH cannot be considered to act by the same mechanism of action. Hence it is necessary to assess the risk posed by the mixture of PAH using other methods.

Surrogate marker approach

The surrogate marker approach estimates the toxicity of a mixture of PAH in an environmental matrix by using data from toxicity studies in which a PAH mixture of known composition was tested. Exposure to the surrogate marker (usually benzo[*a*]pyrene (BaP)) is assumed to represent exposure to all the PAH in the environmental matrix. Thus the level of toxicity ascribed to the surrogate represents (is proportional to) the toxicity of the PAH mixture.

This method is recommended by the HPA when providing advice on the risk assessment of PAH in soil to local authorities. It is based on the surrogate marker approach that is supported by European Food Safety Authority (ESFA) and Joint WHO/FAO Expert Committee on Food Additives (JECFA) for risk assessment of PAH in food. The critical toxicity study used was a two-year carcinogenicity study by Culp *et al.* (1998) in which mice were fed two coal tar mixtures containing several PAH.

Evaluation of the PAH profile in soil

Before recommending this approach, the HPA evaluated the analytical data from 52 contaminated sites (1848 individual soil samples) to assess the variability of the PAH profile (BaP and seven other genotoxic PAH) in soil. BaP was present in all sites that were reported to be contaminated with PAH. The absolute concentrations of the individual PAH were highly variable. However, the variability of the PAH profile in relation to level of BaP was low. The low variation means that the level of BaP is a good predictor of the levels of the other PAH that may be present in the soil. The relative profile was also similar to the relative profile of the eight PAH in the coal tar mixtures used in the Culp *et al.* (1998) study that is pivotal to the risk assessments by EFSA and JECFA.

The findings indicated that BaP is a suitable surrogate marker to represent the amount of the eight genotoxic PAH that are commonly measured in contaminated soil. There is some uncertainty in using this approach; for example, the unknown presence of a more potent PAH, such as dibenzo[*a,l*]pyrene, could affect the overall carcinogenicity of the PAH mixture.

Conclusion

The surrogate marker approach is used by the HPA and is recommended to the contaminated land community to assess the health risks of PAH in soil. The interim oral Health Criteria Value (minimal risk level) for PAH in soil, used and recommended by the HPA, is based on the $BMDL_{10}$ value ($0.1\ mg\ kg^{-1}\ bw\ day^{-1}$) proposed by JECFA and derived from the Culp *et al.* (1998) study. The uncertainty in this approach, such as the potential unknown presence of a more potent PAH such as dibenzo[*a,l*]pyrene, is acknowledged.

Further information on the surrogate maker approach is available on the HPA website[16].

3.2.2 A case study for the application of a read-across approach at FSA

Mixed halogenated dioxins

Dioxins, dibenzofurans and biphenyls (collectively referred to subsequently as dioxins) are chemical compounds formed as unwanted by-products in certain industrial processes and fires. They are persistent in the environment, and are known to cause a wide range of toxic effects in animals. Effects in humans have only been found at higher levels. Levels of dioxins in food and the environment have declined substantially since the 1980s.

A number of expert groups including the Scientific Committee on Food (SCF), WHO and the Committee on Toxicity (COT) have previously set a tolerable

---

[16] http://www.hpa.org.uk/Publications/ChemicalsPoisons/LandContamination/ContaminatedLandInformationSheets/1012ContaminatedLandinfosheetPAHs/ [accessed March 2013]

daily intake (TDI) for chlorinated dioxins (dioxins which contain chlorine but not bromine). The TDI is a level below which harmful effects are not expected to occur. The FSA monitors the presence of these compounds in food.

All dioxin-like compounds produce similar toxic effects, but some are more potent than others. The toxicity of each dioxin compound is expressed relative to that of a 'reference' compound, to give a toxic equivalency factor (TEF). The total dioxin activity in a food sample, expressed as toxic equivalent (TEQ), is determined by adding the results obtained by multiplying the concentration of each compound by its TEF.

The FSA funded research to develop methods to measure the levels of mixed halogenated dioxins (dioxins containing both chlorine and bromine) in food. Previous research had focused on dioxins containing either chlorine alone or bromine alone (chlorinated or brominated dioxins).

There are no agreed TEF for mixed halogenated dioxins. Applying the TEF for the equivalent chlorinated compounds made it possible to estimate the contribution of mixed halogenated dioxins to total dioxin activity. This approach had previously been used for brominated dioxins. If anything, this would be expected to overestimate that contribution, since the available evidence suggests that mixed halogenated (and brominated) compounds are less toxic than the equivalent chlorinated compounds.

The mixed halogenated compounds measured were selected on the basis of their expected levels in the environment and their toxicity in experimental animals or model systems, and with regard to practical considerations concerning their production. Measurements were made in similar foods to those tested in previous studies that have looked at levels of chlorinated and brominated dioxins.

Although only a small number of all the possible mixed halogenated dioxins were surveyed, those whose chemical structures were expected to result in the greatest toxicity were over-represented. The contribution of mixed halogenated dioxins to total dioxin activity in the foods tested was determined.

The major contribution to the total dioxin toxic activity in the foods measured came from chlorinated compounds. Brominated compounds made a much smaller contribution, and mixed halogenated compounds contributed even less.

Taking account of these relative contributions and dietary exposures to chlorinated dioxins, the COT concluded that the measured levels of mixed halogenated dioxins in food did not indicate a health concern (COT, 2010[17]).

---

[17] COT statement on occurrence of mixed halogenated dioxins and biphenyls in UK food. Available from http://cot.food.gov.uk/pdfs/cotstatementhalogenatedioxins201002.pdf [accessed March 2013]

4 Conclusions and Recommendations

4.1 Summary of current use by UK government

The primary objective of chemical safety regulation is to protect human health and the environment. This involves the identification of a chemical's hazards, associated risks as a consequence of the exposures that can or could arise, and the implementation of appropriate control measures. To facilitate this task, a comprehensive data package on the respective chemical is often required, which usually involves the conduct of toxicity testing, sometimes *in vitro* but frequently *in vivo*. However, societal, legislative and economic pressures are increasing the drive to reduce, refine and replace the use of animals in toxicity testing. There are also situations in which there is not the regulatory need to undertake toxicological testing, nor the regulatory leverage to compel such testing to be done. Substantial work has been undertaken by industry and academia to establish alternatives and/or supplementary additions to animal testing (including the development of the techniques described in Section 2), but regulatory acceptance is critical to ensure their wider use.

As part of this project we examined the use of predictive approaches (non-testing methods based on structure–activity relationship) in human health hazard assessment across UK government departments and agencies responsible for the safe use of pesticides, biocides, medicines, food and industrial chemicals and the assessment of contaminants. A workshop to explore in more depth the perception and use of predictive approaches across UK government regulatory bodies was held in January 2012. These techniques are not routinely utilised within these bodies, although there are certain situations in which they may be applied either by the regulators or by those submitting data. Read-across and grouping approaches are generally more commonly used than (Q)SAR. The scenarios in which these approaches are applied are discussed below, together with the requirements identified as being essential for their successful adoption.

Scenario (i) As part of a data package submitted by the duty holder

Several regulatory programmes require duty holders to complete and submit data packages on chemicals; the chemicals covered by such schemes include active substances used as pesticides, biocides, human and veterinary medicines; general (industrial) chemicals; food additives and food contact materials. In this scenario, the relevant legislation generally prescribes fairly rigid information requirements (with explicit statements of the test guidelines to be followed) that must be met for the data package to be in compliance; an exception to this is REACH, which encourages the use of alternatives to animal testing and includes 'rules for adaptation from the standard information requirements'.

In this scenario, a major barrier to the more widespread use of, in particular, (Q)SAR approaches is their lack of regulatory acceptance as stand-alone methods, as a consequence of the absence of consistently and demonstrably reliable predictions. Notwithstanding, in some cases the duty holder may be allowed to utilise *in silico* approaches to support existing data. However, most commonly read-across is used (e.g. non-clinical data from existing similar products for generic medicines).

The identified requirements of the regulator in this situation are for the submission to make clear how one was to:

- understand the prediction, in terms of
    - its relevance and suitability to the chemical and the endpoint under investigation
    - its reliability
    - clarity of justification and sufficiently detailed reporting

  - how it agrees or disagrees with other available information
- be confident that the substance is within the applicability domain of the method
- be able to accept, reject or request modification of a prediction with confidence.

It is noteworthy that in this scenario the regulator has the option to request further information from the duty holder to support the prediction, for example a justification that the model used is suitable for the test substance.

Scenario (ii) To support a non-standard submission made by industry

Predictive approaches may be used and/or promoted by the regulator to support a non-standard submission made by industry if a) the relevant legislation provides flexibility by allowing the use of predictive approaches to fulfil certain information requirements (as, for example, under the PPPR for groundwater metabolites, impurities and plant metabolites), or b) to provide support for and verification of information submitted (e.g. under NONS; Notification of New Substances). In this scenario, preference is given to computational tools rather than read-across and grouping approaches as they tend to be less resource intensive and provide answers more quickly. When the prediction is performed by the regulator, a (Q)SAR tool that is readily available within the department or agency is most commonly used.

A predictive approach to support a non-standard submission made by industry may be performed by the regulator or by industry at the regulator's request. When the prediction has been performed by industry, the requirements of the regulator would be the same as those listed under Scenario (i).

Where the prediction is performed by the regulator, the identified needs are to:

- have sufficient expertise to
  - select the most appropriate tool(s) for the chemical and the endpoint(s) under investigation
  - perform the analysis
  - evaluate the prediction in terms of its relevance and reliability
- have the appropriate tools available.

As in Scenario (i), the regulator has the option to engage in dialogue with the duty holder over the suitability and outcome of a prediction, and to request further justification of or information on its use.

Scenario (iii) There is no regulatory option to request further data from the duty holder

This scenario may include, for example, the assessment of environmental contaminants in drinking water, air or landfill sites, or, in some limited circumstances, classification and labelling of substances within the scope of the CLP Regulation. In this scenario there is no option in the legislation to request further test data (e.g., C&L) or it is very difficult to pinpoint one or more duty holders (e.g., contaminants). The prediction is performed by a regulatory authority. This may then require review, as is sometimes the case when classification and labelling proposals have been prepared by a regulatory body from another EU member state. Any tools that are available and with which the assessor is familiar may be used, such as (Q)SAR or a surrogate marker (see the HPA case study in Section 3.2).

The identified needs of the regulator in this situation are to:

- have sufficient expertise to
  - select the most appropriate tool(s) for the chemical and the endpoint(s) under investigation
  - perform the analysis
  - evaluate the prediction in terms of its relevance and reliability
- have the appropriate tools available.

Scenario (iv) A rapid answer is needed to a problem or question (with acceptance of its uncertainty)

This situation may arise in emergency situations such as incidents of food contamination. Important decisions are likely to be based on the outcome of a prediction, as there is no time to request or conduct testing; additional information to support or refute the prediction may not be available. Read-across has generally been applied in this situation.

The identified needs of the regulator in this situation are to have:

- immediate access to the tool(s)
- familiarity with and confidence in the use of the tool so that a prediction can be generated quickly
- expertise in the interpretation of the prediction
- confidence that the tool gives a robust prediction.

4.2 Recommended best practice

In Section 2, minimum criteria were proposed or established, to be taken into account when predictive approaches are used to perform and/or inform on risk assessment to ensure that the prediction is reliable and valid. The most critical ones are presented below.

- The chemical structure of the substance should be known, including the potential for isomers. In some models the ability to translate this into a 3-D structure improves predictivity, as 3-D descriptors are important in terms of receptor fit and binding.
- For mixtures, the composition of the mixture should be known with a degree of certainty, or at least the ability to exclude certain classes of highly toxic chemicals.
- The chemical structure should be within the range of structures that are covered by the approach being used (applicability domain).
- Any algorithms should be based on a broad range of reference chemicals (the training set) or, if not, the limitations should be specified.
- The model should use interpretable descriptors, be statistically sound and produce transparent and clear results.
- Where the output from running a substance through a model is that no alerts are generated, this should be checked to see if this is because the structure is known not to produce effects (reassuring) or that there is no alert because no data have ever been generated on a particular structure (not reassuring).
- Basic physicochemical properties such as volatility, solubility or partition coefficients and pH should be available. These can be relevant to the hazard (e.g. extreme pH) or potential for exposure (e.g. high volatility increases the need for information on inhalation exposures).
- The underlying data being used for extrapolation or interpolation should be of good quality. If the only available data are of poor quality they may still be used in a prediction, but the limitations should be acknowledged and if necessary taken into account in the final risk assessment. Some (Q)SAR programs provide either a quantitative or qualitative indication of the confidence of a prediction.
- Where possible, interpolation within a group of related chemicals is preferred to extrapolation from one chemical.
- Greater confidence is obtained where a combination of predictions from different methods (e.g. a statistical (Q)SAR and an expert system covering different chemical spaces) is used (consensus modelling).

In addition, when applying or evaluating data generated by predictive approaches, the following general principles should be taken into account.

- The principle of proportionality: the amount of information needed is dependent on the scale of the decision that will result from it.
- The principle of caution: the amount of information needed is dependent on the potential risk the substance might pose; the more severe the possible consequences, the more conservative the approach.

4.3 Overall concluding remarks

From the survey results and workshop discussions it was apparent that, across UK government regulatory bodies and agencies, there is scope and enthusiasm for a greater use of predictive approaches.

The decision of whether or not to use these methods as a stand-alone replacement for toxicity testing depends on two factors: the criticality of the outcome and the timeframe in which an answer is required. For example, regulations that govern active substances for use as pesticides, biocides and human and veterinary medicines, which ensure that the toxicology of these substances is thoroughly understood, have strict information requirements that must be met by the duty holder. In these cases, (Q)SAR, for example, does not have regulatory acceptance as a stand-alone method (whereas read-across does). However, in emergency situations in which the option to conduct new toxicity tests is not available, predictive approaches provide a valuable means to rapidly generate a predictive basis for assessment of the hazards, and therefore the possible risks involved.

In addition to being used as stand-alone methods, predictive approaches are perceived to be a useful tool to support regulatory decision making in a WoE analysis. Indeed, the use of WoE approaches and alternatives to animal tests is a central tenet of the REACH regulation, which states that new vertebrate testing should only be conducted as a last resort. The guidance on how industry can meet its information requirements under REACH includes an ‘integrated testing strategy’ (ITS) for each endpoint, which helps duty holders to decide whether sufficient information is available from existing data, predictive approaches and *in vitro* tests without the need to conduct new animal tests. The Council of Canadian Academies has recently published a report on the integrated testing of pesticides, which concluded that there will be an "evolution towards the use of integrated testing strategies in decision making, with the anticipation that this will better inform decisions … over the next two to ten years" (Expert Panel on the Integrated Testing of Pesticides, 2012)[18]. Weight-of-evidence approaches, and the use of predictive methods therein, will therefore become increasingly important over the coming years.

During the workshop discussions, all the government bodies highlighted the following three major barriers to the more extensive use of predictive approaches and in particular (Q)SAR models in regulatory decision making.

- A lack of confidence in their outputs given that accuracy and reliability of model predictions are still inadequate for regulatory use
- A lack of modelling expertise and adequate familiarity with more than one software package within the department or agency for appropriate interpretation and use of model predictions
- Limited availability of appropriate tools, owing to constraints imposed by cost and IT security issues (usually, any download of software must be approved and performed by the IT provider for the department or agency, even if the software is freely available on the internet)

These barriers apply especially to the use of (Q)SAR, which usually requires computational models, with which the regulatory bodies tend to have less practical experience. The findings were consistent with those of the EU ORCHESTRA survey (described in Section 1.3). There may also be a conflict between the preferred model that would best suit the chemical and the endpoint under investigation, and the software that is actually available. In this situation the regulator generally has to compromise and use the model(s) available to his institution, but be especially aware of potential limitations in the prediction(s).

The majority of the government bodies reported that it is uncommon for duty holders to submit information generated by predictive approaches to the regulators. Recently, however, a report written by the European Chemicals Agency on alternatives to animal testing in REACH registration dossiers submitted by industry indicated that 20–30% of dossiers, depending on the endpoint, had used read-across to fill data gaps. The UK REACH competent authority (situated within HSE) has reviewed a number of read-across arguments submitted by duty holders in support of their registration dossiers. Assessment of these data packages has identified that the use of read-across to fill data gaps in registration dossiers is frequently not accompanied by sufficient information to allow the regulator to make an assessment of the validity of the approach. For duty holders to take full advantage of such opportunities to use predictive approaches they need to be aware of what constitutes a robust justification for the use of these approaches and what information the regulator expects to be presented, for example, as outlined above in Section 4.2 'Recommended best practice'.

4.4 Recommendations

The following recommendations are made.

An 'Advisory Centre of Excellence' should be established

This would be available to all UK government regulatory bodies that use predictive approaches and would ensure that expert advice and appropriate tools were available to all departments and agencies, via 'consultants' in the Centre. It would also remove the duplication of effort that would occur where several people across different bodies try to achieve the level of expertise necessary to feel confident in the use and analysis of these methods.

Training on in silico *methods for regulators should be facilitated*

Regulators in the government bodies that use or have the potential to use predictive approaches should obtain training in these methods so that they have sufficient expertise to feel confident in their use and in the evaluation of data generated by them.

Best practice in the use of predictive approaches should be applied by duty holders

As far as possible, the OECD guidelines, as described in Section 2, should be applied when generating and assessing data from predictive approaches.

The regulatory community should issue advice to duty holders on the information expected in submissions that utilise predictive approaches

This would assist duty holders to submit good-quality justifications for the use of these approaches, so that they are more likely to be accepted by the regulator. The advice could include the following recommendations.

- To follow best practice, as described in guidance documents produced by ECHA and the OECD

---

[18] http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/Pesticides/Pesticides_Full_Report_EN.pdf

- That robust, valid justifications are required
- To use a weight-of-evidence (WoE) approach, including, where possible, the use of more than one model where (Q)SAR is used, rather than to rely on a single prediction (i.e., consensus modelling should be encouraged).

Integrated testing strategies and weight-of-evidence techniques should be promoted by the regulators for application by industry

Predictive approaches (and in particular *in silico* predictions) tend not to be accepted for regulatory purposes as stand-alone methods. However, through the use of ITS and WoE approaches, robust arguments can be built that negate the need for new animal testing.

The development of web-based tools (by model developers) which do not require local download of software should be encouraged by the regulators

This would bypass the cost and IT security issues experienced by many government departments and agencies.

The situation should be reviewed in a few years

As the use of predictive methods becomes more widespread, developments should be kept under review.

References

Arena, VC, Sussman, NB, Mazumdar, S *et al.* (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

Ashby, JPD (1993) The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. *Mutat Res,* 286(1), 3–74

Ashby, JTRW (1988) Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the US NCI/NTP. *Mutat Res,* 204(1), 17–115

Barratt, MD (1995) A quantitative structure–activity relationship for the eye irritation potential of neutral organic chemicals. *Toxicol Lett*, 80(1–3), 69–74

Barratt, MD (1996) Quantitative structure–activity relationships for skin irritation and corrosivity of neutral and electrophilic organic chemicals. *Toxicol* In Vitro*,* 10(3), 247–256

Bergstrom, CAWCM, Norinder, U, Luthman, K *et al.* (2004) Global and local computational models for aqueous solubility prediction of drug-like molecules. *J. Chem Inf. Comput Sci.,* 44(4), 1477–1488

Berner, BWDR, Guy, RH, Mazzenger, GC *et al.* (1988) The relationship of p*K*a and acute skin irritation in man. *Pharm Res***,** 5(10), 660–663

Berner, BWDR, Steffens, RJ, Mazzenga, GC *et al.* (1990) The relationship of p*K*a and skin irritation for a series of basic penetrants in man. *Fundam Appl Toxicol*, 15(4), 760–766

Cope, AC & Hancock EM (1939) Substituted Vinyl Barbituric Acids. II. (1-Methylpropenyl) Derivatives. *J Am Chem Soc*, 61(2), 353–354

Cramer, GM, Ford, RA *et al.* (1978) Estimation of toxic hazard – a decision tree approach. *Food Cosmet Toxicol*, 16(3), 255–276

Cronin, MT (1996) Quantitative Structure–Activity Relationship (QSAR) analysis of the acute sublethal neurotoxicity of solvents. *Toxicol In Vitro,* 10(2), 103–110

Cronin, MT & Madden, JC (2010) *In Silico* Toxicology – Principles and Applications. RSC Publishing. ISBN: 978-1-84973-004-4

Culp SJ, Gaylor DW, Sheldon WG, Goldstein LS & Beland FA (1998) A comparison of the tumors induced by coal tar and benzo[*a*]pyrene in a 2-year bioassay. *Carcinogenesis*, 19(1), 117–124

EC (2003) *Guidance document on the assessment of the relevance of metabolites in groundwater of substances regulated under council directive 91/414/EEC.* SANCO/221/2000-rev.10-final. [http://ec.europa.eu/food/plant/pesticides/approval_active_substances/docs/wrkdoc21_en.pdf (accessed June 2013)]

EC (2011) *Guidance document on the assessment of the equivalence of technical materials of substances regulated under regulation (EC) 1107/2009*. SANCO/10597/2003-rev.9. [http://ec.europa.eu/food/plant/protection/evaluation/guidance/wrkdoc23_en.pdf (accessed June 2013)]

ECHA (2008) *Guidance on information requirements and chemical safety assessment. Chapter R6: QSARs and grouping of chemicals*, European Chemicals Agency, Helsinki. [http://www.echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf (accessed June 2013)]

Expert Panel on the Integrated Testing of Pesticides (2012) *Integrating Emerging Technologies into Chemical Safety Assessment,* Council of Canadian Academies, Ottawa, Canada [http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/Pesticides/Pesticides_Full_Report_EN.pdf]

Franke, RGA, Giuliani, A & Benigni, R (2001) Prediction of rodent carcinogenicity of aromatic amines: a quantitative structure–activity relationships model. *Carcinogenesis*, 22(9), 1561–1571

Gerner, ILM & Spielmann, H (2005) Assessment of the eye irritating properties of chemicals by applying alternatives to the Draize rabbit eye test: the use of QSARs and *in vitro* tests for the classification of eye irritation. *Altern Lab Anim*, 33(3), 215–237

Hansch, C & Kurup A (2003) QSAR of chemical polarizability and nerve toxicity. 2. *J Chem Inf Comput Sci*, 43(5), 1647–1651

IARC (2006) *IARC Monographs on the evaluation of carcinogenic risk of chemicals to man.* Lyon, France, Preamble, IARC

JRC Report (2011) Worth A, Fuart-Gatnik M, Lapenna S *et al. The use of computational methods in the toxicological assessment of chemicals in food: Current status and future prospects*. Joint Research Centre, Ispra, Italy [http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/doc/EUR_24748_EN.pdf (accessed June 2013)]

Kimber, IHC, Westmoreland, C, Alepee, N *et al.* (2011) Computational chemistry, systems biology and toxicology. Harnessing the chemistry of life: revolutionizing toxicology. A commentary. *Journal of Applied Toxicology*, 31, 206–209

King, LA (1985) Ferguson's principle and the prediction of fatal drug levels in blood. *Hum Toxicol*, 4(3), 273–278

Munro, ICFRA, Kennepohl, E & Sprenger, JG (1996) Thresholds of toxicological concern based on structure–activity relationships. *Drug Metab Rev*, 28(1–2), 209–217

Nangia, AAPH, Berner, B & Maibach, HI (1996) High dissociation constants (p*K*a) of basic permeants are associated with *in vivo* skin irritation in man. *Contact dermatitis*, 34(4), 237–242

OECD (2005) Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Series on Testing and Assessment Number 34, Environment Directorate, Organisation For Economic Co-Operation And Development, Paris. [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2005)14&doclanguage=en (accessed June 2013)]

OECD (2007a) Guidance document on the validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] models, Series on Testing and Assessment No. 69 Environment Directorate, Health and Safety Division, Organisation for Economic Co-operation and Development, Paris [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en (accessed March 2013)

OECD (2007b) Guidance on grouping of chemicals, E. 80, Environment Directorate, Health and Safety Division, Organisation for Economic Co-operation and Development

Roberts, DWNA (2009) High throughput kinetic profiling approach for covalent binding to peptides: application to skin sensitisation potency of Michael acceptor electrophiles. *Chemical Research in Toxicology,* 22, 592–603

Smith, JSMOT, Sussman, NB, Luster, MI *et al.* (2000) A robust structure–activity relationship (SAR) model for esters that cause skin irritation in humans. *Toxicol Sci,* 55(1), 215–222

Tilaoui, L, Schilter, B, Tran, LA *et al.* (2007) Integrated computational methods for prediction of the lowest observable adverse effect level of food-borne molecules. *QSAR & Combinatorial Science*, 26, 102–108

TPHCWG (1997) Volume 4. Development of Fraction Specific Reference Doses (RfDs) and Reference Concentrations (RfCs) for Total Petroleum Hydrocarbons. Total Petroleum Hydrocarbon Criteria Working Group Series. Amherst: Amherst Scientific Publishers.

USA FDA (1986) Guideline for Threshold Assessment, Appendix I, Carcinogen Structure Guide

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

Worth, AFGM, Lapenna, S & Serafimova, R (2011) Applicability of QSAR analysis in the evaluation of developmental and neurotoxicity effects for the assessment of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Available from: http://www.efsa.europa.eu/de/supporting/doc/169e.pdf

Appendix

1 Tabulated results of questionnaires

Current use of predictive approaches in toxicological hazard assessment of chemicals*

UK Department/ agency	Programme/activity	Predictive approaches used and documented guidance available
HSE, CRD	BPD, active substance evaluation	Mainly read-across in waiving arguments (especially for chronic/cancer studies and reprotoxicity studies). Occasionally, model predictions (Derek Nexus), but in a WoE approach, not in isolation. No documented guidance has been produced by CRD or the EU – left to expert judgment
HSE, CRD	BPD, product authorisation	Mainly read-across from acute tox data and dermal absorption values on one formulation to another product. Also, use of the calculation method of the Preparation Directive is the ultimate application of predictive toxicology. No documented guidance has been produced by CRD or the EU – left to expert judgment
HSE, CRD	COPR, product approval	Mainly read-across from acute tox data and dermal absorption values on one formulation to another product. Also, use of the calculation method of the Preparation Directive is the ultimate application of predictive toxicology. No documented guidance has been produced by CRD – left to expert judgment
HSE, CRD	OECD SIDS	Read-across and category approaches used generally to fill data gaps. OECD guidance available
HSE, CRD	Pre-REACH: NONS, new chemicals	Mainly read-across to fill data gaps (sometimes for the complete dossier). Occasionally, model predictions (Derek Nexus & METEOR), but in a WoE approach, not in isolation. Documented guidance produced by CRD**
HSE, CRD	Pre-REACH, ESR, existing chemicals	Mainly read-across and categories approaches to fill data gaps. No documented guidance has been produced by CRD or the EU – left to expert judgment
HSE, CRD	Major hazards	Very occasionally read-across of acute toxicity data. No documented guidance has been produced by CRD – left to expert judgment
HSE, CRD	REACH	Use of predictive approaches a central tenet of REACH. Specific rules for adaptation of standard information requirements and general rules for use of existing data, WoE, (Q)SAR, *in vitro* methods, grouping and read-across, exposure-based waiving are laid out in the legal text. ITS is a key concept and requires that all available information be taken into account and a WoE approach used before the decision to conduct a new mammalian toxicity test is made. Detailed guidance on (Q)SAR, grouping/read-across, ITS is available on ECHA's website. The first registration deadline has only recently passed, so as yet no information on how commonly these methods have been used, but expect them to be fairly extensive. CRD has presented a 'thought-starter' paper to the Member State Committee (which makes decisions on testing proposals from registrants) that 90-day repeated dose study could be waived based on negative findings in 28-day and acute toxicity studies N.B. Defra has sponsored a project (Inchemicotox) to develop alternatives to animal testing for skin sensitisation (and acute fish toxicity) that included an ITS tool, databases, *in vitro* data, QSAR, *in chemico* analysis

UK Department/ agency	Programme/activity	Predictive approaches used and documented guidance available
HSE, CRD	C&L	Classification based on structural alerts (e.g. respiratory sensitisation for isocyanates, skin/eye irritation for organic peroxides) is explicitly allowed under CHIP & CLP. Also, prediction of corrosivity from phys–chem information leads to classification. Calculation method for mixtures under the Dangerous Preparations Directive and bridging principles/calculation under CLP allow for prediction. Read-across has very occasionally been used by CRD (malachite green to leucomalachite green for acute toxicity, skin sensitisation; benzidine to diaminobenzidine for mutation & carcinogenicity), and other member states have used grouping approach (nickel salts for skin sens & carc). There are some harmonised classifications for group entries (e.g. lead compounds, organic compounds of mercury, arsenic compounds). No documented guidance has been produced by CRD – left to expert judgement
HSE, CRD	PPP	Generally do not need to use predictive approaches as there is the option for requiring data to be submitted on nearly all aspects. Predictions often performed for animal welfare and resource/financial reasons. Take a precautionary approach if in doubt. In house QSAR, normally with Derek Nexus
HSE, CRD	PPP active substance, new or review evaluation	Predictive approaches are rarely used – extensive dataset normally submitted. Occasional read-across between related molecules (e.g. tumour mechanisms) and from racemic mixtures to resolved isomers – no guidance document
HSE, CRD	PPP active substance, source different to that used in original evaluation	EU guidance document stipulates submission of (Q)SAR on impurities but gives no guidance on performance or interpretation. Prediction of toxicity of impurities relative to original active substance as evaluated, e.g. levels present in rat metabolism; retention of known toxicophore; often request (Q)SAR on active substance as well as impurity to see if any alerts are common and whether they were seen in tests with the active substance. Even with guidance document can require much expert judgement. Some data e.g. acute or genotox often submitted
HSE, CRD	PPP plant metabolites	Some data e.g. acute or genotox submitted but often need to use read-across from data on active substance for overall risk assessment; (Q)SAR predictions; TTC; amount produced in rat metabolism studies with active substance. Guidance being reviewed at present
HSE, CRD	PPP groundwater metabolites	Use of structure considerations is mentioned in EU guidance document (no indication of how to use the results). Basic data requirements triggered by predicted levels in groundwater but risk assessment often relies on relating the likely toxicity of the metabolite to that seen with the active substance – e.g. rat metabolite; comparison of results in short-term studies
HSE, CRD	PPP/COPR product authorisation	Classification and labelling often based on properties of the components of the product – use directive 99/45 (Dangerous Preparations Directive) approach. Dermal absorption extrapolation from a formulation with data – based on similarity of formulations (composition, physical chemistry properties, irritancy) – guidance available but more being developed; a lot of expert judgement. Can use defaults e.g. 10% or 100% for dermal absorption
HSE, CRD	PPP route-to-route extrapolation	Repeat dose data are nearly always via oral route, non-consumer exposures are dermal or inhalation. Very simple predictions of route to route toxicity based on relative absorption (data on oral and often dermal, assume 100% for inhalation). Can require route specific data e.g. if acute studies via different routes indicate a significant concern. Guidance document available but under revision
HSE, CRD	PPP residues detected in food	For many substances EU/US/WHO evaluations are available and no need to use prediction. For some substance there is limited information available to CRD. Often need conclusion rapidly so no option to require more studies – use data on related molecules, TTC, (Q)SAR, determine MOE relative to any data available. Option to remove food from sale. No guidance document
FSA	Flavouring Group Evaluations	System is based largely on read-across from data on related substances using the approach described in the SCF opinion on Flavouring Group Evaluations (which is similar to but more rigorous in terms of genotoxicity than the JECFA approach) with modifications based on experiences gained as EFSA have undertaken the evaluations

UK Department/ agency	Programme/activity	Predictive approaches used and documented guidance available
FSA	Food additives	On the additives side there are examples of read-across when related compounds have been evaluated e.g. parabens and also when evaluating complex mixtures (especially those derived from botanicals) e.g. using representative rosemary extracts to cover the range of extracts used, increasingly in the re-evaluation of natural colours and gums. Steviol is another example. In all cases these are case by case and have developed from recognition that the SCF guidance for testing of food additives implied treating every extract as a single compound was not justifiable ethically or scientifically, the revised guidance includes this approach as a reduction strategy for animal testing
EFSA	Jelly minicups	In the case of jelly minicups physical (rather than toxicological) properties of various gums were read-across to identify the hazard and if this was likely to be a risk of choking
EFSA	Nutrient sources	The bulk of the comparisons of bioavailability for nutrients from nutrient sources have been based on dissolution data on the compound and comparison to dissolution and limited bioavailability data on one or more compounds
FSA	Food chemical incidents	For example there was read-across of Sudan I genotoxicity data to other Sudan dyes and often forced to use information on similar or related compounds when there are no data on the specific chemical identified
FSA	Food contact materials	Food contact material components especially plastics additives and printing inks where there may be data on one of a group of related products but not on every individual one. These are often trade name based (e.g. Irganox) where there are testing data on ‘medical grades’ but not necessarily on other ones
EFSA	Nanomaterials	The new EFSA nanomaterial guidance effectively uses a limited testing on the nanomaterial for comparison with and to validate read-across of the rest of the database from other non-nanomaterial forms
FSA	Environmental contaminants	The read-across of the chlorinated dioxins TEF to brominated and mixed halogenated dioxins for the risk assessment of the brominated and mixed dioxins was agreed by the Committee on Toxicity. This allowed the risks from the measured levels of brominated and mixed halogenated dioxins in a limited selection of foods to be evaluated. Whilst this approach is a logical extension of the TEF concept (which was recognised by WHO) based on the common mode of action, there are extremely few *in vivo* or *in vitro* data on the relative potency of brominated and mixed halogenated dioxins. These limited data suggest that chlorinated dioxins are likely to be more potent and using their TEF for the brominated and mixed halogenated dioxins would probably be precautionary Similarly there was an assumption of a common mode of action and read-across with brominated diphenyl congeners. All examples are case by case assessments accompanied by a description of the associated uncertainties and assumptions and there is neither a defined specific approach nor generic guidance
FSA	Process contaminants	As with environmental contaminants and incidents, read-across from similar chemicals has been used
FSA	ESR	Have previously agreed with read-across proposals in reviews prepared by HSE
FSA	Veterinary medicines	4-chloroaniline was noted to be an impurity within some veterinary medicinal products and was considered an *in vitro* genotoxin and carcinogen. *In vivo* genotoxicity was unclear such that it was possible that carcinogenicity could be non-genotoxic in line with an alternate theory. SAR and (Q)SAR assessments found alerts for positive *in vitro* and *in vivo* genotoxicity
FSA	Animal feed additives	Nicarbazin is a feed additive composed of two compounds, one of which (4,4’-dinitrocarbanilide, DNC) had mixed *in vitro* genotoxicity results and a single negative *in vivo* bone marrow micronucleus result. Predictive methods (Toxtree, Lazar and Osiris) found that DNC contained structural alerts for positive results within *in vitro* and *in vivo* assays and supported a line of requiring further *in vivo* data
FSA	Novel foods	Advice was needed at short notice on the toxicity of glycidol fatty acid esters (GFAE) found in fats and oils. Read-across from [genotoxic] glycidol and a manual consideration of whether structural alerts would be retained in the ester or not was made. This raised concerns over genotoxicity and some qualifying factors that might mitigate concern depending on what the actual molecular structures of poorly characterised GFAE were

UK Department/ agency	Programme/activity	Predictive approaches used and documented guidance available
FSA	Pesticides	Monocrotophos is a banned genotoxic pesticide and predictive methods (Toxtree, Lazar and Osiris) were used to consider whether a putative leaving group would retain the genotoxicity of the parent molecule or not
FSA	Food Chemicals Risk Assessment	Habitual exposures predicted from short-term food consumption survey data or if not available, portion size data for the food itself or an appropriate proxy. In the absence of survey data, exposure to chemicals in toddlers is predicted via the use of harmonised European databases (concise and comprehensive databases), or via extrapolation between age groups
HPA	COM/Mutagenicity	*In vitro* genotoxicity testing can be predictive of mutagenic potential. QSAR can be used for preliminary prediction of mutagenicity when no other data available, but additional data required prior to risk assessment. This is in line with draft revised COM strategy for testing document. COM discussed the use of SAR for the prediction of mutagenic potential of drinking water contaminants; however, the committee indicated that they required further information on the programs used before providing detailed comment
HPA	COC/Carcinogenicity	Use of results of mutagenicity tests to predict whether a chemical is genotoxic or non-genotoxic. This has implications for the risk assessment approach for carcinogenicity. Use of the Minimal Risk Approach to derive a pragmatic level for maximum exposure to a genotoxic carcinogen e.g. in advising on carcinogenic impurities in the pesticide 1-methylcyclopropene (http://www.iacoc.org.uk/publications/documents/guideline04.pdf Para 65ff) Use of TEF for dioxins to estimate carcinogenicity of a mixture of dioxins, when carcinogenicity data are lacking on most congeners (TEF apply in other categories also)
HPA	Contaminated land	Use of the Minimal Risk Level Approach for deriving tolerable exposure levels from genotoxic soil carcinogens from soil (http://archive.defra.gov.uk/environment/quality/land/contaminated/documents/legal-definition.pdf p. 13)
HPA	Drinking water	Not used QSAR and not aware of others in the HPA having used them for drinking water or contaminated land risk assessments. HPA has been faced with the problem of conducting risk assessments for chemicals where there is no available toxicity data e.g. drinking water product approvals and some private water supply test results. In this case it is stated that a risk assessment is not possible for such individual chemicals (where appropriate extremely small concentrations or declining concentrations can be noted). Could not find out the frequency with which such requests for a risk assessment cannot be given occur. For DWI product approval applications in some rare instances HPA has reported toxicity data for very similar compounds, noting that this could be useful information, but, clearly stating that no suitable data were available for the original compound and acknowledging the uncertainty in making comparisons (i.e. a read-across approach). **Example 1**: application for use of a particular product that might contaminate drinking water. A predictive approach of sorts was adopted. EFSA had already indicated that the substance is extensively metabolised to aniline. Aniline is genotoxic *in vivo* and also carcinogenic. Thus, it was considered prudent to regard the product as a potentially genotoxic carcinogen. We have also pointed out that certain structural characteristics can indicate potential mutagenicity and carcinogenicity such as nitrosamines and primary aromatic amines. **Example 2**: HPA was asked for advice on alkyl alkanes detected at $3\ \mu g\ l^{-1}$ in response to a customer taste/odour complaint. There was some uncertainty over the analysis. The approach recommended by the World Health Organization (WHO) for similar imprecise groups or hydrocarbon mixtures i.e. based on Total Petroleum Hydrocarbon Criteria Group (TPHCWG, 1997), was used. The TPHCWG (1997) derived a number of oral reference doses (RfD) for various hydrocarbon fractions that each contains alkyl alkanes (amongst other hydrocarbons). Thus, a conservative risk assessment can be undertaken by using the

UK Department/ agency	Programme/ activity	Predictive approaches used and documented guidance available
HPA	Drinking water	lowest RfD applicable to the various alkyl alkanes that could be present in drinking water, namely 0.1 mg kg^{-1} bw day^{-1} The RfD are derived from individual chemicals or hydrocarbon mixtures considered to be sufficiently representative of the toxicity of a hydrocarbon fraction (containing a large number of individual chemicals within certain carbon number range) Going forward, it may be useful to have access to and training for QSAR, particularly for potential mutagenicity. A COM view on suitable QSAR packages would be helpful (i.e. most appropriate QSAR packages) For a substantial number of chemicals (usually low water concentrations) no data are available. In such cases, software that uses pre-existing data on mode of action to predict other potential endpoints would not be of practical use (i.e. where there are no data at all, rather than just some data gaps). Generally, from the point of view of risk assessment of exposure to relatively low drinking water concentrations, it is the most sensitive endpoint that is crucial and additional information on other potential endpoints at higher doses may not be particularly informative
HPA	Chemicals in waste processes	Use of margin of exposure approach to risk assess concentrations of 1,2-dichloroethane around landfill sites. Use of EPAQS relative potency methodology to risk assess concentrations of PAH around landfill sites. Route-to-route extrapolation to derive reference concentration for systemic effects from styrene levels around landfill sites (Reference for all above: http://cot.food.gov.uk/pdfs/cotstatementlandfill201001.pdf)
HPA	Indoor air pollution	**Example**: HPA was asked when occupants could go back into building following a kerosene spill where there would be exposure to resultant hydrocarbon mixture We would recommend that the occupants could return to the house when the kerosene concentration has decreased to value within the range of 0.45 ppm to 1.5 ppm (3–10 mg m^{-3}), providing that the concentrations can be shown to be decreasing with time This is on the assumption that exposure is to a hydrocarbon fraction range of approximately C8–C16, which is commonly associated with kerosene and JP5/JP8 jet fuel. This would be in line with both the authoritative TPHCWG (1997) chronic (lifetime) reference concentration of 1 mg m^{-3} for JP8 jet fuel (C8–C16) multiplied by a factor of 10 for short-term exposure (JP5/JP8 jet fuel can be considered as sufficiently similar or representative of kerosene) [It may be useful to note that symptoms/discomfort, such as irritation (but not likely to cause serious harm) may be experienced at an acute exposure (up to 8 hours) at an AEGL-1 value of 44 ppm (290 mg m^{-3}) for JP-8. The symptoms would be transient and reversible upon cessation of exposure]
HPA	Emergency response	Not apparently used
HPA	Consumer products Veterinary medicines OECD HPV chemicals Biocides Pesticides	HPA does not lead on these areas, only provides comment, on request, on assessments made by other authorities/agencies
MHRA	Initial clinical trial dose setting considerations	Several predictive approaches used. Computer aided prediction of drug toxicity e.g. QSAR, Derek Nexus, MCASE, TOPKAT. Receptor occupancy of target receptors, based on *in vitro* studies. Use of PAD, which may be derived from appropriate pharmacodynamic–pharmacokinetic models (i.e. exposure–response relationships) in animal species and applied to determine a human PAD Also, use of NOAEL. The NOAEL is converted to HED and by scaling down by usually a factor of 10, the MRSD is derived. It is usually better to start with a dose justified as being safe, but not active, rather than a dose estimated as the PAD. The concept of PAD is close to that of the MABEL, however, MABEL is specific for humans and applies to a minimum level of biological activity whereas PAD implies

UK Department/ agency	Program/activity	Predictive approaches used and documented guidance available
MHRA	Initial clinical trial dose setting considerations	optimal activity. A safe starting dose for clinical trials in humans should be based on pharmacology and toxicology, adjusted for anticipated exposure in humans and for interspecies differences in potency *In vitro* studies in animal species/models in particular those designed to investigate the mechanism or mode of action of a drug
EA	REACH related risk assessments and generic human health risk assessments for environmental contaminants	OECD QSAR Toolbox Consult with HSE, CRD to use their prediction models

* See Glossary for abbreviations

** Hanway RH (2002) The use of toxicological read-across data in the notification of new chemicals. Industrial Chemicals Unit, Health and Safety Executive, Magdalen House, Stanley Precinct, Bootle, Merseyside, L20 3QZ, UK

Web addresses accessed March 2013

2 Goodness-of-fit and robustness of common algorithms used in (Q)SAR

Model	Introduction to model	Assessing goodness of fit	Avoiding over fitting	Notes
Multiple linear regression (MLR)	Traditional approach for deriving (Q)SAR models Relates a biological activity (dependent variable) to several molecular descriptors (independent variables)	The multiple correlation coefficient (R2) should be as close as possible to 1 Standard error(s) should be as small as possible Can also use mean squared error (MSE) R2 measures how well the model is able to mathematically reproduce the training set	Compare model error (standard error of estimate) with error in experimental data: the smaller the standard error, the higher the reliability of the prediction However a standard error smaller than the experimental data indicates over-fitted data	R2 value should be used with caution Exercise caution with models that appear to over fit the data Estimated data should not be more accurate than the experimental data
Partial least squares	An MLR approach Several activity variables can be modelled simultaneously	R2 should be used (the measure of explained variation)	Near perfect correlation can be achieved due to large number of variables, so high R2 not sufficient on its own Cross validation must be used and predictive ability calculated Find a balance between R2 and predictivity	Tolerates a certain amount of missing data Residual standard deviation (RSD) can be used to measure statistical fit. It should be similar to expected noise
Classification models (CM)	Used when the biological variability of some endpoints is too large to give a quantitative result. So chemicals are classed into one or more categories of toxic effect (e.g. yes or no)	Assessed using Cooper statistics, e.g. sensitivity, specificity, concordance, positive & negative predictivities, false positive & negative classification rates Variation in quality of underlying experimental data mean that it is not recommended to define inflexible criteria based on these statistics These statistics should not be used in isolation to judge a model	These models can over fit the data and there is particular danger of over fitting when one of the groups to be separated is small	Examples of linear CMs: multivariate discriminant analysis, logistic regression, classification trees Examples of non-linear CMs: neural networks and K nearest neighbour clustering The robustness can be tested by the total number of misclassifications estimated by the leave-one-out method The loss function is a measure of discrepancy between observed & predicted data Can increase robustness by combining predictions
Artificial neural networks	Learns from data in a manner emulating the human brain Two main types: supervised & unsupervised – these terms denote whether only descriptors or descriptors plus biological activities are involved in the learning process	The recall ability test is used to test goodness of fit It is also recommended that other tests are performed e.g. leave one out, leave many out, Y scrambling and assessment with independent samples Aim is to find best compromise between model simplicity and variability to minimise mean squared error R2 should be as close to 1 as possible and standard error should be small, but this is insufficient for validity on its own	The structure of a (Q)SAR should always be evaluated for over fitting, noise, sample specificity and unjustified complexity	Suitable for non-linear relationships and trends Useful for mathematical problems e.g. data exploration, pattern recognition

3 Computational tools

Tool	Available from*
Accelrys ADME add in (Accelrys Accord for Excel program)	Commercially available from Accelrys Inc. http://accelrys.com/products/discovery-studio/admet.html
ADME Boxes	ACD Labs http://www.acdlabs.com/products/pc_admet/adme/adme/
ADMET Discovery Studio	ACD LAccelrys Inc. http://accelrys.com/products/discovery-studio/admet.htmlabs
ADMET Predictor 5.0	Accelrys Inc. www.accelrys.com
Accelrys ADME add in (Accelrys Accord for Excel program)	Commercially available from Accelrys Inc. http://accelrys.com/products/discovery-studio/admet.html
AIM (analogue identification methodology)	Publicly available from the US EPA http://www.epa.gov/opptintr/sf/tools/aim.htm
AMBIT including AMBIT Disclosure & AMBIT Discovery	Freely available from http://ambit.sourceforge.net/downloads.html Also available from IdeaConsult Ltd
And ADMET Predictor 5.0	Commercially available from Simulations Plus, Inc. http://www.simulations-plus.com/
BfR rule-base	Free, available in house at BfR http://www.bfr.bund.de/en/home.html
CAESAR (Computer Assisted Evaluation of Industrial Chemical Substances According to Regulations)	Freely available from http://www.caesar-project.eu/
Carcinogenic Potency Database (CPDB)	Freely available from http://toxnet.nlm.nih.gov/cpdb/
CAS SciFinder	Commercially available from https://www.cas.org/products/scifinder
CASE (Computer Automated Structure Evaluation) including MultiCASE , META, MCASE CASETOX, TOXALERT	Commercially available from http://www.multicase.com
CATABOL	http://oasis-lmc.org/products/models/environmental-fate-and-ecotoxicity/catabol-301c.aspx
ChemFinder	Available both publicly & by subscription from http://www.chemfinder.com
ChemID / Plus / Advanced	Publicly available from the US National Library of Medicine (NLM) http://chem.sis.nlm.nih.gov/chemidplus
Danish (Q)SAR database	Free internet accessible version at http://qsar.food.dtu.dk/ or from the JRC Developed for JRC by the Danish EPA
DART (Developmental & Reproductive Toxicology Database)	Freely available from http://toxnet.nlm.nih.gov/
Derek Nexus	Commercially available from http://www.lhasalimited.org/
ECOSAR	http://www.epa.gov/oppt/newchems/tools/21ecosar.htm
Endocrine Disruptor Knowledge Base (EDKB) database	Freely available from the US FDA website http://www.fda.gov/scienceresearch/bioinformaticstools/endocrinedisruptorknowledgebase/default.htm
Enhanced NCI Database Browser	Freely available from http://cactus.nci.nih.gov/
EPIsuite	Freely available from the US EPA website at http://www.epa.gov/opptintr/exposure/pubs/episuite.htm
EPIWIN	Freely available at http://esc.syrres.com/esc/epi.htm
ESIS (European Chemical Substances Information System)	Freely available from http://esis.jrc.ec.europa.eu/
HazardExpert	Freely available from http://www.compudrug.com Developed by CompuDrug Ltd

Tool	Available from*
Hazardous Substances Database (HSDB)	Publicly available from the US National Library of Medicine (NLM) Toxicology Data Network (TOXNET) http://toxnet.nlm.nih.gov
HYDROWIN	Freely available at http://esc.syrres.com/esc/epi.htm
IARC monographs	Freely available from http://monographs.iarc.fr/index.php
ICSAS Reprotox Database	Freely available from http://www.fda.gov/
ILSI Developmental Toxicity database	Currently under development & not yet available Expected to be available from http://www.ilsi.org/ http://www.epa.gov/
JRC (Q)SAR Model Database	http://ihcp.jrc.ec.europa.eu/our_databases/jrc-qsar-inventory
Leadscope	Commercially available from http://www.leadscope.com/
MDL Reaction Database	Commercially available – more information at http://www.akosgmbh.de/pdf/PCD_Brochure.pdf
MDL–(Q)SAR	Freely available at http://mdl-qsar.software.informer.com/
MetabolExpert	www.compudrug.com
METEOR	Commercially available from http://www.lhasalimited.org
MolCode Toolboxes	Commercially available from http://www.compudrug.com/?q=node/26
NTPBSI Database (National Toxicology Program Bioassay Online Database)	Freely available, chemical list available from http://www.epa.gov/ Searchable online at http://www.epa.gov/ncct/dsstox/sdf_ntpbsi.html
OASIS	Commercially available at http://oasis-lmc.org/products/software/database-manager-and-centralized-database.aspx
OECD Chemical category database (eChemPortal)	Freely available from the OECD at http://www.echemportal.org/echemportal/index?pageID=0&request_locale=en
OECD database on chemical risk assessment models	Available from www.oecd.fr
OncoLogic	Freely available from the US EPA website at http://www.epa.gov/oppt/sf/pubs/oncologic.htm
OSIRIS Property Explorer	Free and commercially available versions from http://www.pharmaexpert.ru/PASSOnline/
PASS (Prediction of Activity Spectra for Substances)	Free and commercially available versions of PASS from http://www.pharmaexpert.ru/PASSOnline/
(Q)SAR Toolbox	http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm
REPROTOX®	Available at http://www.reprotox.org/Default.aspx
RTECS (Registry of Toxic Effects of Chemical Substances)	Available at http://www.cdc.gov/niosh/rtecs/
SICRET (Skin Irritation Corrosion Rules Estimation Tool)	Included and available from Toxtree at http://toxtree.sourceforge.net/skin.html
SPARC (SPARC Performs Automated Reasoning in Chemistry)	Available from http://www.archemcalc.com/sparc.html
TERIS Teratogen Information system	Available from http://depts.washington.edu/terisweb/teris/
TIMES (Tissue Metabolism Simulator)	Available from http://oasis-lmc.org Developed by Laboratory of Mathematical Chemistry (LMC)
TIMES-SS (Tissue Metabolism Simulator for Skin Sensitisation)	Available from http://oasis-lmc.org
TOPKAT	Commercially available from http://www.accelrys.com Developed by Accelrys Inc.

Tool	Available from*
Toxmatch	Available from the JRC website at http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/toxmatch
TOXNET (Toxicology Data Network)	Freely available from http://toxnet.nlm.nih.gov/
ToxRefDB (Toxicity Reference Database)	Freely available from http://www.epa.gov/NCCT/toxrefdb/
Toxtree	Freely available from http://sourceforge.net/projects/toxtree/ Developed for JRC by Ideaconsult Ltd

*Web addresses accessed March 2013

Abbreviations and Glossary

Abbreviations

ADME	Absorption, distribution, metabolism, excretion
ADMET	Absorption, distribution, metabolism, elimination, toxicity
AEGL	Acute exposure guideline level
AOP	Adverse outcome pathway
AR	Androgen receptor
BPD	Biocidal Products Directive
BfR	Institute for Risk Assessment, Germany
BPR	Biocidal Products Regulation
C&L	Classification and labelling
CEFIC	Conseil Européen des Fédérations de l'Industrie Chimique
CAS	Chemical Abstracts Service
CAS number	CAS Registry database number,.
CEBS	Chemical Effects in Biological Systems
CHIP	Chemicals (Hazard Information and Packaging for Supply) Regulations
CHO/HPRT	Chinese hamster ovary / hypoxanthine-guanine phosphoribosyl transferase
CLP	Regulation on classification, labelling and packaging of substances and mixtures
COC	Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment
COM	Committee on Mutagenicity of Chemicals in Food, Consumer products and the Environment
COPR	Control of Pesticides Regulations
CPDB	Carcinogenic Potency Database
DWI	Drinking Water Inspectorate
ECHA	European Chemicals Agency
EFSA	European Food Standards Agency
EINECS	European Inventory of Existing Commercial chemical Substances
ELINCS	European List of Notified Chemical Substances
EPA	Environmental Protection Agency
ER	Estrogen receptor
ESIS	European Chemical Substances Information System
ESR	Existing Substances Regulation
EU	European Union
FDA	Food and Drug Administration
GHS	Globally harmonised system

HED	Human equivalent dose
HPV	High production volume
IARC	International Agency for Research on Cancer
ICSAS	Informatics and Computational Safety Analysis Staff
ILSI	International Life Science Institute
ITS	Integrated testing strategy
JECFA	Join WHO/FAO Expert Committee on Food Additives
JRC	Joint Research Council
LOAEL	Lowest observed adverse effect level
MABEL	Minimal anticipated biological effect level
MIE	Molecular initiating event
MLR	Multiple linear regression
MRSD	Maximum recommended starting dose
MTD	Maximum tolerated dose
NLP	No-longer polymer
NOAEL	No observed adverse effect level
NONS	Notification of new substances
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
PAD	Pharmacologically active dose
PAH	Polycyclic aromatic hydrocarbons
PCA	Principal component analysis
PCR	Principal component regression
PPP	Plant Protection Products
PPPR	Plant Protection Products Regulation
QAAR	Quantitative activity–activity relationship
QMDB	(Q)SAR model database
QMRF	(Q)SAR model reporting format
QPDB	(Q)SAR prediction database
QPRF	(Q)SAR prediction reporting format
(Q)SAR	(Quantitative) structure–activity relationship
RAI	Relative alkylation index
REACH	Registration, evaluation, authorisation and restriction of chemicals
RfD	Reference dose
RIVM	National Institute of Public Health and the Environment, the Netherlands
RTECS	Registry of Toxic Effects of Chemical Substances
SAR	Structure–activity relationship
SIDS	Screening Information Datasets
SLRL	Sex-linked recessive lethal
SMILES	Simplified Molecular Input Line Entry System
TEF	Toxic equivalent factor
TK	Thymidine kinase
TMC	Toxicologically meaningful category
TPHCWG	Total Petroleum Hydrocarbon Criteria Group
TTC	Threshold of toxicological concern
UDS	Unscheduled DNA synthesis
WoE	Weight of evidence

Glossary

AEGL-1: the airborne concentration, expressed as ppm or mg m^{-3}, of a substance above which it is predicted that the general population, including susceptible individuals, could experience notable discomfort, irritation, or certain asymptomatic non-sensory effects. However, the effects are not disabling and are transient and reversible upon cessation of exposure.

Algorithm of a model: could be a mathematical model or a knowledge-based rule that forms the relationship between the chemical structure/substructure or other molecular descriptors of the substance and the predicted endpoint.

Analogue approach: the reading-across of endpoint information for one chemical to predict the same endpoint for another chemical (i.e., a one-to-one read-across).

Adverse outcome pathway (AOP): a set of chemical, biochemical, cellular and physiological responses which characterises the biological effects cascade resulting from a particular molecular initiating event.

Applicability domain (of a (Q)SAR model): the response and chemical structure space in which the model makes predictions with a given reliability.

Automated rule induction systems: a type of expert system that primarily use rules based on statistics, the models have been developed from statistical analysis of the data.

Breakpoint chemical: a chemical that identifies a turning point in a trend.

Chemical category: a group of chemicals whose physicochemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity.

Consensus modelling: combination of predictions from two or more expert systems and/or (Q)SAR.

Database: experimental data representing one or more effects of chemicals in biological systems.

Expert system: any formalised system (not necessarily computer-based) that enables a user to make rational predictions about the properties or activities of chemicals.

Extrapolation: the estimation of a value that is at or near the category boundary using measured values from internal category members.

Global (Q)SAR model: produces its prediction from a diverse dataset, covering a wide range of chemicals.

Grouping (or chemical grouping): the general approach to assessing more than one chemical at the same time, it can include the formation of a chemical category or the identification of a chemical analogue for which read-across may be applied.

Hybrid system: a type of expert system that combines both knowledge-based and statistically based rules.

Interpolation: the estimation of a value using measured values from category members on both sides of the unknown data point.

Knowledge based systems: a type of expert system that primarily uses rules derived from expert opinion.

Local (Q)SAR model: have a reduced applicability domain and produce their predictions from a narrower data subset, usually based on analogous chemicals.

Molecular initiating event (MIE): the initial chemical induced perturbation of some biological system.

Molecular descriptors: provide a means of representing molecular structures in a numerical form and are used to describe different features of a chemical. The number may be a theoretical attribute (e.g. relating to size or shape) or measurable property. The most simple and commonly used molecular descriptors are molecular weight, atomic composition indices and atomic count descriptors. Other commonly used descriptors are those representing structural fragments. More complex descriptors are the topological descriptors which tend to represent structural features of the molecule such as size, shape, symmetry, branching, cyclicity and bond multiplicity. Even more complex descriptors are the geometrical 3-D descriptors. These require knowledge about the relative positions of the atoms in 3-D space.

PAD: pharmacologically active dose – a dose that is estimated to have the intended pharmacological activity.

QAAR: a mathematical relationship between two biological endpoints or activities. QAAR are based on the assumption that knowledge about the mode of action for an endpoint is applicable to a similar endpoint because the main underlying processes of these related endpoints or activities are the same (e.g. partitioning, reactivity, enzyme inhibition).

(Q)SAR: a mathematical model that relates quantitative parameters from the chemical structure to a quantitative measure of a particular biological activity or property.

(Q)SAR model reporting format (QMRF): a reporting format for the use of (Q)SAR and includes a description of the algorithm, the model development and validation according to the OECD principles and evaluation studies performed with the model.

(Q)SAR prediction reporting format (QPRF): a reporting format for predictions from (Q)SAR, this should explain how the estimate was derived by applying a specific model to a specific substance. It should include information of the model prediction, the endpoint, the substance modelled, the relationship between the substance and the applicability domain, analogues of the substance and the relationship between the predicted endpoint and the regulatory endpoint.

Qualitative read-across: the presence or absence of a particular property or activity is inferred from the source chemical(s) to the target chemical(s). Qualitative read-across gives a yes/no answer (e.g. the substance is a skin sensitiser or not).

Quantitative read-across: properties from the source chemical(s) with known values are used to estimate the value of that property for the target chemical(s). Quantitative read-across yields a quantitative value for an endpoint (e.g. the potency of a carcinogen, an LD_{50} value or a LOAEC for hepatotoxicity).

Read-across: a data gap-filling technique used to predict endpoint information for one chemical by using data from the same endpoint from another chemical which is considered to be similar in some way. Read-across can be used in the context of both the analogue approach and the category approach.

RfD: the oral reference dose, defined as ‘an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily population exposure that is likely to be without an appreciable risk of deleterious, non-carcinogenic effects during a lifetime.

Rule-base: rules derived from experimental data that are used to analyse other chemicals.

SAR: a qualitative relationship between a particular molecular structure or substructure and the presence or absence of a biological activity, or the capacity to modulate the biological activity imparted by another substructure.

Sentinel chemicals: chemicals that are at the opposite extremes of a trend.

Similarity principle: theory that similar chemicals should have similar biological properties.

Subcategories: larger categories could have several different trends for a single endpoint, these are divided into smaller subcategories.

Test dataset: comprises a set of chemicals with their molecular descriptors and *estimated* endpoint values used to test the predictivity of the (Q)SAR model.

Training dataset: includes a set of chemical substances with their molecular descriptors and *measured* endpoint values used to develop a (Q)SAR model.

Members of the IGHRC Steering Committee

Dr Diane Benford
Food Standards Agency
Aviation House
125 Kingsway
London WC2B 6NH

Dr Ruth Bevan
IGHRC Secretariat
Institute of Environment and Health
Cranfield University
Cranfield
Bedfordshire MK43 0AL

Dr Ian Dewhurst
Chemicals Regulation Directorate
(CRD)
Health and Safety Executive
Mallard House
3 Peasholme Green
King's Pool
York YO1 7PX

Dr Steve Fairhurst
(Chairman)
Director of Chemicals Schemes
and Human Health
Chemicals Regulation Directorate
(CRD)
Redgrave Court (Bootle) and
Mallard House (York)

Mr Sam Fletcher
Safety Assessor
Veterinary Medicines Directorate
Woodham Lane
New Haw
Addlestone
Surrey KT15 3NB

Dr Paul Holley
Department of Health
523 Wellington House
135-155 Waterloo Road
London SE1 8UG

Home Office
Horseferry House
Dean Ryle Street
London SW1P 2AW

Ms Christine Knox
Consumer Product Safety
Department for Business, Innovation
& Skills (BIS)
Bay 424 1 Victoria Street
London SW1H 0ET

Professor Leonard Levy
IGHRC Secretariat
Institute of Environment and Health
Cranfield University
Cranfield
Bedfordshire MK43 0AL

Professor Joe Lunec
IGHRC Secretariat
Cranfield University
Cranfield
Bedfordshire MK43 0AL

Medical Research Council
Head Office
20 Park Crescent
London W1B 1AL

NERC
Environment & Human Health
Programme
Plymouth Marine Laboratory
Prospect Place
The Hoe
Plymouth PL1 3DH

Miss Laura Notton
Programme Manager, EBS
BBSRC
Polaris House
North Star Avenue
Swindon SN2 1UH

Dr Colin Powlesland
Air Quality and Health Policy
Manager
Environment Agency
Block One, Government Buildings,
Burghill Road
Westbury in Trym
Bristol BS10 6BF

Dr Mike Roberts
Defra
2A Nobel House
17 Smith Square
London SW1P 3JR

Dr Ovnair Sepai
Chemical Hazards and Poisons
Division
Health Protection Agency
Centre for Radiation
Chemical and Environmental
Hazards
Chilton
Didcot
Oxon OX11 0RQ

Mr Henry Stemplewski
Medicines and Healthcare Products
Regulatory Agency
Market Towers
1 Nine Elms Lane
London SW8 5NQ

Members of the IGHRC Executive Committee

Dr Ruth Bevan
IGHRC Secretariat
Institute of Environment and Health
Cranfield University
Cranfield
Bedfordshire MK43 0AL

Dr Susy Brescia
Chemicals Regulation Directorate (CRD)
Health and Safety Executive
2.3 Redgrave Court
Merton Road
Bootle
Merseyside L20 7HS

Environment Agency
Red Kite House
Howbery Business Park
Crowmarsh Gifford
Wallingford
Oxfordshire OX10 8BD

Dr David Gott
Chemicals Safety and Toxicology Division
Food Standards Agency
Aviation House
125 Kingsway
London WC2B 6NH

Dr Andy Hart
Risk Analysis Team
The Food and Environment Research Agency
Sand Hutton
York YO41 1LZ

Professor Leonard Levy
(Chairman)
Institute of Environment and Health
Cranfield University
Cranfield
Bedfordshire MK43 0AL

Professor Joe Lunec
IGHRC Secretariat
Cranfield University
Cranfield
Bedfordshire MK43 0AL

Miss Laura Notton
Programme Manager, EBS
BBSRC
Science and Technology Group
Polaris House
North Star Avenue
Swindon SN2 1UH

Dr Mike Roberts
Defra
2A Nobel House
17 Smith Square
London SW1P 3JR

Dr Stephen Robjohns
Health Protection Agency
Chemical Hazards and Poisons Division
Chilton
Didcot
Oxon OX11 0RQ

Dr Andy Smith
MRC Toxicology Unit
Hodgkin Building
University of Leicester
PO Box 138
Leicester LE1 9HN

The MHRA, NERC and VMD do not have representatives on the Executive Committee.

Publications

Risk Assessment and Toxicology Steering Committee

cr1 Developing New Approaches to Assessing Risk to Human Health from Chemicals

cr2 Risk Assessment Approaches used by UK Government for Evaluating Human Health Effects of Chemicals

cr3 Risk Assessment Strategies in Relation to Population Subgroups

cr4 Physiologically-Based Pharmacokinetic Modelling: A Potential Tool for Use in Risk Assessment

cr5 Exposure Assessment in the Evaluation of Risk to Human Health

cr6 From Risk Assessment to Risk Management: Dealing with Uncertainty

The Interdepartmental Group on Health Risks from Chemicals (IGHRC)

cr7 The Interdepartmental Group on Health Risks from Chemicals: First Report and Forward Plan to 2002

cr7A The Interdepartmental Group on Health Risks from Chemicals: Annexes to First Report and Forward Plan to 2002

cr8 Assessment of Chemical Carcinogens: Background to General Principles of a Weight of Evidence Approach

cr9 Uncertainty Factors: Their use in Human Health Risk Assessment by UK Government

cr10 Guidelines for Good Exposure Assessment Practice for Human Health Effects of Chemicals

cr11 The Interdepartmental Group on Health Risks from Chemicals: Final Report for Phase 1, 1999–2003 and Forward Plan to 2006

cr12 Guidelines on Route-to-Route Extrapolation of Toxicity Data when Assessing Health Risks of Chemicals

cr13 The Interdepartmental Group on Health Risks from Chemicals: Final Report for Phase 2 (2003–2007) and Forward Plan for Phase 3 (2007–2010)

cr14 Chemical Mixtures: A Framework for Assessing Risks to Human Health

cr15 Current Approaches to Exposure Modelling in UK Government Departments and Agencies

cr 16 Predictive Approaches to Chemical Hazard Identification and Characterisation: Current use by UK Government Departments and Agencies