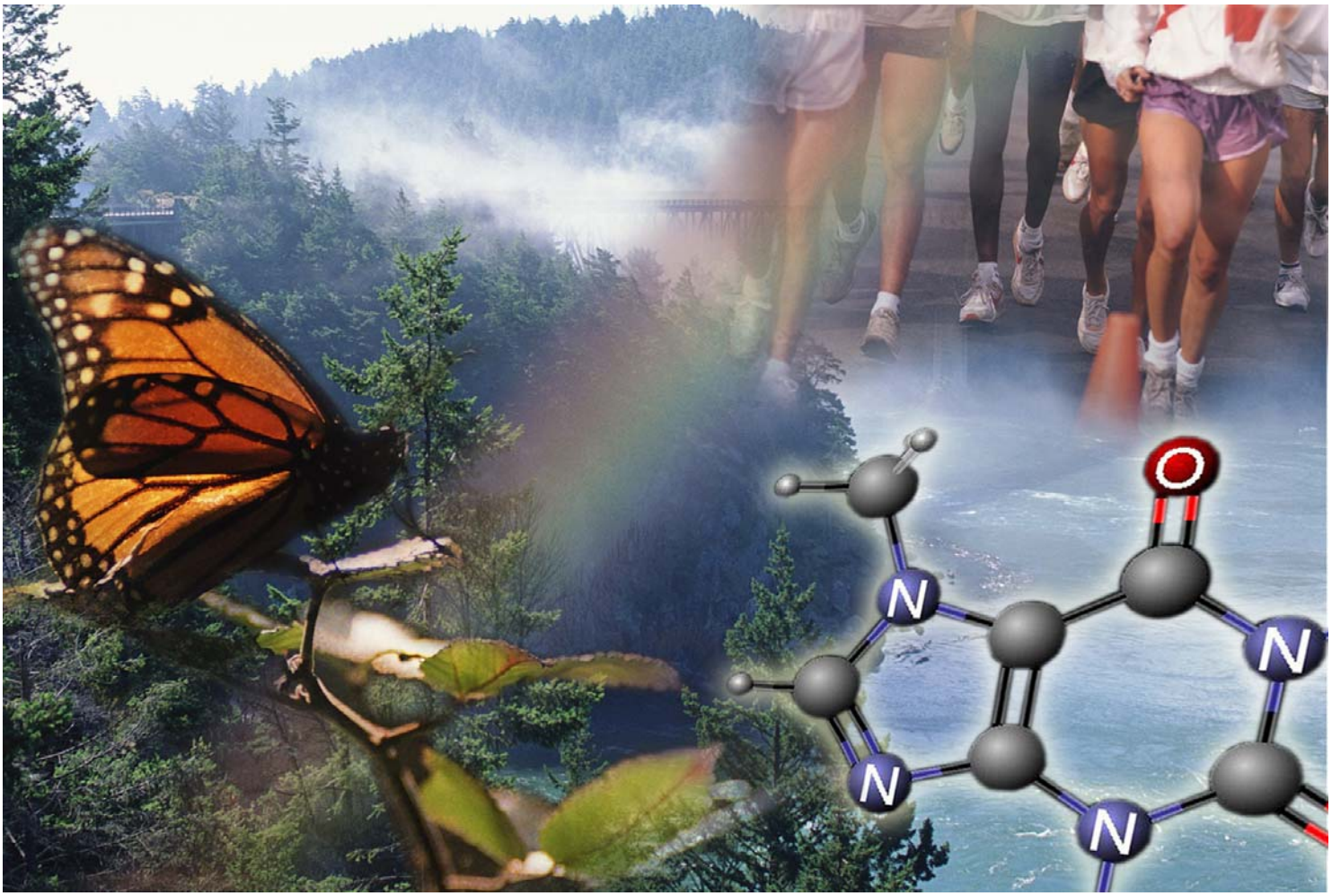
Guidance on information requirements and chemical safety assessment

Chapter R.6: QSARs and grouping of chemicals



May 2008

LEGAL NOTICE

This document contains guidance on REACH explaining the REACH obligations and how to fulfil them. However, users are reminded that the text of the REACH regulation is the only authentic legal reference and that the information in this document does not constitute legal advice. The European Chemicals Agency does not accept any liability with regard to the contents of this document.

PREFACE

This document describes the information requirements under REACH with regard to substance properties, exposure, use and risk management measures, and the chemical safety assessment. It is part of a series of guidance documents that are aimed to help all stakeholders with their preparation for fulfilling their obligations under the REACH regulation. These documents cover detailed guidance for a range of essential REACH processes as well as for some specific scientific and/or technical methods that industry or authorities need to make use of under REACH.

The guidance documents were drafted and discussed within the REACH Implementation Projects (RIPs) led by the European Commission services, involving stakeholders from Member States, industry and non-governmental organisations. These guidance documents can be obtained via the website of the European Chemicals Agency (http://echa.europa.eu/reach_en.asp). Further guidance documents will be published on this website when they are finalised or updated.

This document relates to the REACH Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006[1]

1 Corrigendum to Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC (OJ L 396, 30.12.2006); amended by Council Regulation (EC) No 1354/2007 of 15 November 2007 adapting Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) by reason of the accession of Bulgaria and Romania (OJ L 304, 22.11.2007, p. 1).

Convention for citing the REACH regulation

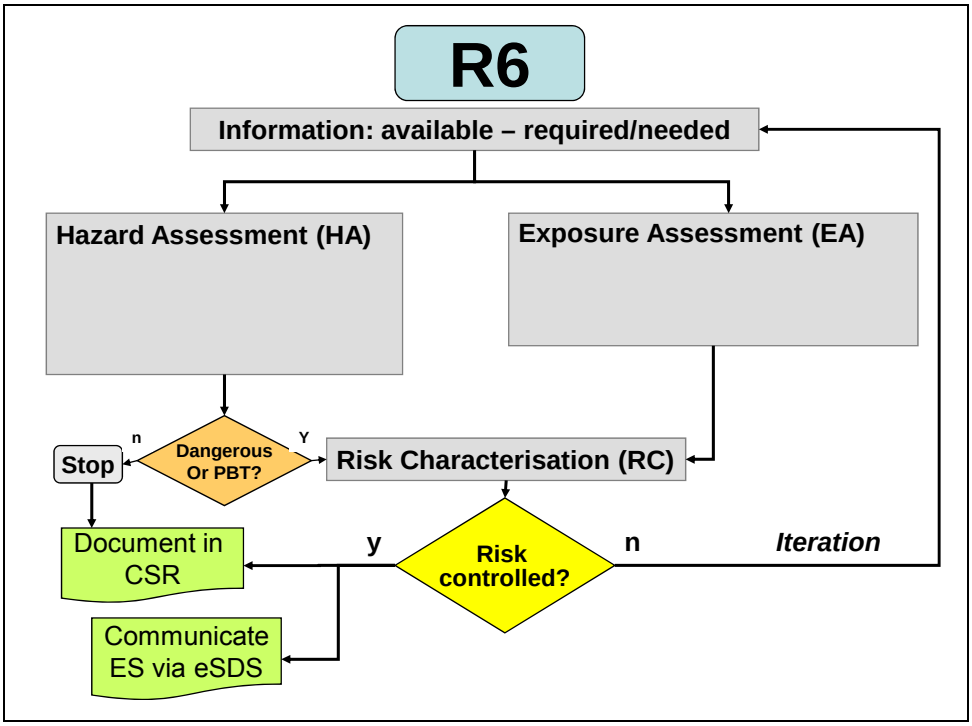
Where the REACH regulation is cited literally, this is indicated by text in italics between quotes.

Table of Terms and Abbreviations

See Chapter R.20

Pathfinder

The figure below indicates the location of chapter R.6 within the Guidance Document



CONTENTS

TABLES

FIGURES

R.6 GUIDANCE ON QSARS AND GROUPING OF SUBSTANCES

The previous sections provide advice on the interpretation and application of REACH and detail the overall process that should be followed in finding, assembling and evaluating all the relevant information that is required for the registration of a chemical under REACH. This chapter elaborates more detailed guidance on non-testing approaches such as QSAR and grouping that facilitate the evaluation of the intrinsic properties of chemicals. All of these approaches have a role in extending and extrapolating the existing information and improving the focus of new testing strategies and study design towards attaining the goal of protecting human health and the environment whilst minimising the need for additional vertebrate testing.

R.6.1 Guidance on QSARs

This document includes generic considerations on the use of (Q)SARs (and expert systems) only. Generic guidance on the grouping of substances (development of chemical categories and analogue read-across) is provided in Section R.6.2

The generic guidance in this report covers:

a) how to establish the validity of a (Q)SAR model

b) how to establish the adequacy of a (Q)SAR model result for regulatory purposes

c) how to document and justify the regulatory use of a (Q)SAR model

d) where to find information on (Q)SAR models

In relation to (d), this report describes the main expert systems that are currently available, and the major initiatives that are underway to provide the IT tools for implementing non-testing methods under REACH.

Guidance on the use of specific (Q)SARs (or expert systems) and grouping approaches within the context of endpoint-specific Integrated Testing Strategies (ITS) is *not* covered. This specific guidance is provided in the individual endpoint specific sections.

Since the field of computational toxicology (including (Q)SARs) is rapidly developing, and experience in the regulatory use of computational approaches (including their reporting) is increasing, this guidance document should be considered as a step in a continuously evolving process.

R.6.1.1 Explanation of the (Q)SAR concept

Non-testing data can be generated by three main approaches: a) grouping approaches, which include read-across and chemical category formation; (quantitative) structure-activity relationships ((Q)SARs); and c) expert systems. The development and application of all kinds of non-testing methods is based on the *similarity principle*, i.e. hypothesis that similar compounds should have similar biological activities.

SARs and QSARs, collectively referred to as (Q)SARs, are theoretical models that can be used to predict in a qualitative or quantitative manner the physico-chemical, biological (e.g. toxicological) and environmental fate properties of compounds from a knowledge of their chemical structure. The two terms can be defined as follows:

A **SAR** is a qualitative relationships that relates a (sub)structure to the presence or absence of a property or activity of interest. The substructure may consist of adjacently bonded atoms, or an arrangement of non-bonded atoms that are collectively associated with the property or activity.

A **QSAR** is a mathematical model (often a statistical correlation) relating one or more quantitative parameters derived from chemical structure to a quantitative measure of a property or activity (e.g. a (eco)toxicological endpoint). QSARs are quantitative models yielding a continuous or categorical result.

The term (Q)SAR is not used in a consistent way: in some cases, the term *quantitative* is used to refer to the nature of the endpoint, whereas in others it refers to the nature of the parameters and model. The latter usage is recommended, as reflected in the definitions above. In other words, the term *quantitative* in QSAR refers to the nature of the parameter(s) used to make the prediction. The presence of a quantitative parameter enables the development of a quantitative model. Such a model can be used to predict a qualitative or quantitative endpoint.

The parameters used in a QSAR model are also called (molecular) descriptors.

The most common techniques for developing QSARs are regression analysis, neural nets and classification methods. Examples of regression analysis include ordinary least squares, multiple least squares and partial least squares. Examples of classification methods are discriminant analysis, classification trees and distance based methods of similarity analysis. Expert systems are a diverse group of models consisting of combinations of SARs, QSARs and databases (see Section R.6.1.6 for examples).

R.6.1.2 The REACH framework for using (Q)SARs and grouping approaches

The obligation to carry out vertebrate testing only as a last resort, and to consider all other options before performing (or requiring) testing is laid down in REACH Article 25 (1). This includes the need to gather all existing information on physico-chemical, toxicological and ecotoxicological properties of a substance, including information generated by (Q)SARs an chemical grouping methods.

REACH Article 13 (1) lays down the basic rules for generating information, whether by testing, (Q)SARs or other means.

REACH Annex XI foresees the use of (Q)SARs and grouping methods when *testing does not appear necessary* because the same level of information can be obtained by means other than (vertebrate) testing. Regarding the use of (Q)SARs, Annex XI contains the following wording:

Results obtained from valid qualitative or quantitative structure-activity relationship models ((Q)SARs) may indicate the presence or absence of a certain dangerous property. Results of (Q)SARs may be used instead of testing when the following conditions are met:

- *results are derived from a (Q)SAR model whose scientific validity has been established,*
- *the substance falls within the applicability domain of the (Q)SAR model,*
- *results are adequate for the purpose of classification and labelling and/or risk assessment, and,*
- *adequate and reliable documentation of the applied method is provided.*

This wording emphasises the principle that information generated by (Q)SARs may be used instead of experimental data, provided a number of conditions are met.

In the ideal situation, (Q)SAR results can be used *on their own* for regulatory purposes if they are considered relevant, reliable and adequate for the purpose, and if they are documented in an appropriate[2] manner. In practice, there may be uncertainty in one or more of these aspects, but this does not preclude the use of the (Q)SAR estimate in the context of a *Weight of Evidence* approach, in which additional information compensates for uncertainties resulting from a lack of information on the (Q)SAR. These concepts of relevance, reliability and adequacy, as they relate to (Q)SARs, are discussed in more detail in Section R.6.1.3. Guidance on the provision of appropriate documentation is given in Section R.6.1.6.

In principle, (Q)SARs can be applied in a number of ways, namely to:

a. provide information for use in priority setting procedures;
b. guide the experimental design of an experimental test or testing strategy;
c. improve the evaluation of existing test data;
d. provide mechanistic information (which could be used, for example, to support the grouping of chemicals into categories);
e. fill a data gap needed for hazard and risk assessment.
f. fill a data gap needed for classification and labelling;
g. fill a data gap needed for PBT or vPvB assessment

The first four applications (a-d) are more general regulatory applications of QSARs, whereas the last three applications (e-g) are more REACH-specific.

In some situations, (Q)SARs could be used to replace test data, whereas in other situations, the models would be used to provide supplementary information to experimental data. In practice, it is foreseen that (Q)SAR information will most often be used to supplement experimental test data within chemical categories and endpoint-specific Integrated Testing Strategies (ITS). However, it is expected that (Q)SARs will be used increasingly for the direct replacement of test data, as relevant and reliable models become increasingly available, and as experience in their use becomes more widespread.

A stepwise approach to the use of non-testing data, integrating (Q)SAR and grouping approaches, is proposed in Section R.6.1.7.

R.6.1.3 The validity, applicability and acceptance of (Q)SARs

As mentioned previously, a number of conditions need to be met in order for (Q)SAR results to provide an *acceptable* alternative to experimental data. The aim of this chapter is to explain some basic concepts concerning the validity, applicability and acceptability of (Q)SAR models.

There is widespread agreement that models should be *scientifically valid* or *validated* if they are to be used in the regulatory assessment of chemicals. In the EU, the concept of *scientifically valid model* is incorporated into the legal text of the REACH regulation, as described previously. Since the concept of validation is incorporated into legal texts and regulatory guidelines, it is important to clearly define what it means, and to describe what the validation process might entail.

For the purposes of REACH, an assessment of (Q)SAR model validity should be performed by reference to the internationally agreed OECD principles for the validation of (Q)SARs. These were adopted by the OECD Member Countries and the Commission in November 2004 (see below). The

[2] In this document, the term "appropriate" documentation interprets what is meant by "adequate and reliable" documentation in Annex XI

validation exercise itself may be carried out by any person or organisation, but it will be the industry registrant (i.e. manufacturer or importer) of the chemical who needs to argue the case for using the (Q)SAR data in the context of the Registration process. This is consistent with a key principle of REACH that the responsibility for demonstrating the safe use of chemicals lies with industry. The need to demonstrate the validity of (Q)SARs does not necessarily imply that the models will have been validated by means of a formal validation process[3], such as the process that has been applied to some *in vitro* tests. The justification for using the (Q)SAR information should be based on the use of the QSAR Reporting Formats described in Section R.6.1.6.

The principles for (Q)SAR validation identify the types of information that are considered useful for the assessment of (Q)SARs for regulatory purposes. The principles constitute the basis of a conceptual framework, but they do not in themselves provide criteria for the regulatory acceptance of (Q)SARs. Fixed criteria will be difficult, if not impossible, to define in a pragmatic way, given the highly context-dependent framework in which non-testing data will be used. Instead, experience and common understanding should be gained by a learning-by-doing approach, and by documenting the learnings (see Section R.6.1.5.).

Under REACH, there will be no formal adoption process for (Q)SARs. The information generated on the characteristics of a (Q)SAR model, and reported to the authorities with the registration dossier (using the reporting formats described in Section R.6.1.6.) will be used as the basis for deciding whether the information on the substance, taken as a whole, is adequate for the regulatory purpose. This process will therefore involve an initial acceptance of the data (including non-testing data) by the industry registrant and the subsequent evaluation, on a case-by-case basis, by the authorities. Information on (Q)SAR models, including peer-reviewed documentation, is likely to be available from various sources, including the JRC QSAR Model Database at - http://qsardb.jrc.it

R.6.1.3.1 OECD principles for (Q)SAR validation

The first step towards a harmonised definition of (Q)SAR model validation, in the context of chemical hazard and risk assessment, was made during an international workshop on the "Regulatory Acceptance of QSARs for Human Health and Environment Endpoints", organised by the International Council of Chemical Associations (ICCA) and the European Chemical Industry Council (CEFIC), held in Setubal, Portugal, on 4-6 March, 2002 (Jaworska *et al*, 2003; Eriksson *et al*, 2003, Cronin *et al*, 2003). During this workshop, a set of six principles were proposed for assessing the validity of (Q)SARs.

Subsequently, an Expert Group established by the OECD carried out an extensive assessment of the six principles (referred to as the *Setubal principles*) by applying them to a range of different (Q)SARs, including literature-based models and models in expert systems (OECD, 2004). On the basis of this assessment, the OECD Expert Group on (Q)SARs reworded the six principles and combined two of the principles into a single principle, to produce a set of five principles. In November 2004, in the context of the 37th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology, this set of five principles was adopted at a policy level by the OECD Member Countries and the European Commission.

The OECD Principles for (Q)SAR validation state that in order:

[3] A "formal" validation process refers to a process managed under the auspices of a "formal" or officially recognised validation body or group.

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

1. *a defined endpoint;*
2. *an unambiguous algorithm;*
3. *a defined domain of applicability;*
4. *appropriate measures of goodness-of-fit, robustness and predictivity;*
5. *a mechanistic interpretation, if possible."*

According to Principle 1, a (Q)SAR model should be associated with a *defined endpoint*, where endpoint refers to any physico-chemical property, biological effect (human health or ecological) environmental fate parameter that can be measured and therefore modelled. The intent of this principle is to ensure transparency in the endpoint being predicted by a given model, since a given endpoint could be determined by different experimental protocols and under different experimental conditions.

According to Principle 2, a (Q)SAR model should be expressed in the form of an unambiguous algorithm. The intent of this principle is to ensure transparency in the description of the model algorithm.

According to Principle 3, a (Q)SAR model should be associated with a *defined domain of applicability*. The need to define an applicability domain expresses the fact that (Q)SARs are reductionist models which are inevitably associated with limitations in terms of the types of chemical structures, physico-chemical properties and mechanisms of action for which the models can generate reliable predictions.

According to Principle 4, a (Q)SAR model should be associated with *appropriate measures of goodness-of–fit, robustness and predictivity*. This principle expresses the need to provide two types of information: a) the internal performance of a model (as represented by goodness-of-fit and robustness), determined by using a training set; and b) the predictivity of a model, determined by using an appropriate test set.

According to Principle 5, a (Q)SAR should be associated with a *mechanistic interpretation*, wherever such an interpretation can be made. Clearly, it is not always possible to provide a mechanistic interpretation of a given (Q)SAR, which is why a majority of the OECD Expert Group preferred to add the wording *if possible* to this principle. The intent of this principle is therefore to ensure that there is an assessment of the mechanistic associations between the descriptors used in a model and the endpoint being predicted, and that any association is documented. Where a mechanistic interpretation is possible, it can add strength to the confidence in the model already established on the basis of Principles 1-4.

A preliminary guidance document was produced by the European Chemicals Bureau (ECB) to provide practical guidance on the interpretation of these OECD principles (Worth *et al*, 2005). Following some minor revisions, the document was broadly accepted by the OECD ad hoc QSAR

group[4] at its meeting of June 2006. The document (OECD, 2007) was adopted by the Joint Meeting in December 2006 and is publicly available.

R.6.1.3.2 Validity of (Q)SAR model

According to the OECD Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (OECD, 2007), the term *validation* is defined as follows:

"…the process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose"

In the context of (Q)SARs, this definition is rather abstract and difficult to interpret in relation to the OECD validation principles. Thus, for the practical validation of (Q)SAR models intended for use in the regulatory assessment of chemicals, the following operational definition has been proposed (Worth *et al*, 2005 and 2006): *"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."*

In this definition, the *performance* of a model refers to its goodness-of-fit, robustness and predictive ability, whereas *purpose* refers to the scientific purpose of the (Q)SAR, as expressed by the defined endpoint and applicability domain. The first part of the definition (*performance*) refers to *statistical validation*, whereas the second part (mechanistic interpretation) refers to the physical or chemical interpretation of the descriptors (where possible) and to the establishment of a hypothesis linking the descriptors with the endpoint.

This definition captures all of the five validation principles, which collectively reflect the validity (reliability and relevance) of the model.[5] The relevance part of validity can be regarded as the mechanistic relevance of the model descriptors to the endpoint predicted. This can be regarded as the *scientific relevance* of the model, which does not necessarily imply regulatory relevance. The *regulatory relevance* of a (Q)SAR expresses the usefulness of the predicted endpoint in relation to the information needed for the regulatory purpose. A (Q)SAR can be valid without being relevant for a given regulatory purpose. In other words, the scientific purpose of a (Q)SAR need not have an association with a possible regulatory application. In fact, many such (Q)SARs can be found in the scientific literature, because in many cases, the models were not developed with specific regulatory needs in mind.

R.6.1.3.3 Reliability of (Q)SAR prediction

A valid (Q)SAR will be associated with at least one defined applicability domain in which the model makes estimations with a defined level of accuracy (reliability). When applied to chemicals within its applicability domain, the model is considered to give *reliable results*. There is no unique measure of model reliability, and no criteria for (Q)SAR reliability are offered in this document. Model reliability should be regarded as a relative concept, depending on the context in which the model is applied. In other words, a greater or lesser degree of reliability may be sufficient for a

[4] The OECD ad hoc Group on QSARs, established in 2006, is the successor the OECD Expert Group on (Q)SARs. The membership of the group was extended in order to directly involve the regulatory end-users of (Q)SARs, as well as (Q)SAR specialists.

[5] When referring to models, the term reliability is often used synonymously with validity (i.e. the relevance aspect is implicit). However, when referring to individual predictions, this can be misleading, because a QSAR estimate might be generated by a valid model, and yet still considered unreliable for the specific purpose.

given regulatory application[6]. This implies that the applicability domain can be defined to suit the regulatory context.

If a model is applied to a chemical outside its applicability domain, it is possible that the estimated result may be not sufficient reliable for the purpose. It is therefore important to determine the applicability of the model to the chemical of interest.

R.6.1.3.4 Adequacy of (Q)SAR prediction

The OECD principles for (Q)SAR validation focus on the scientific validity (relevance and reliability) of a model. The REACH text emphasises the need to demonstrate the adequacy of the (Q)SAR result (i.e. the adequacy of the estimate generated by the (Q)SAR model), which involves additional considerations.

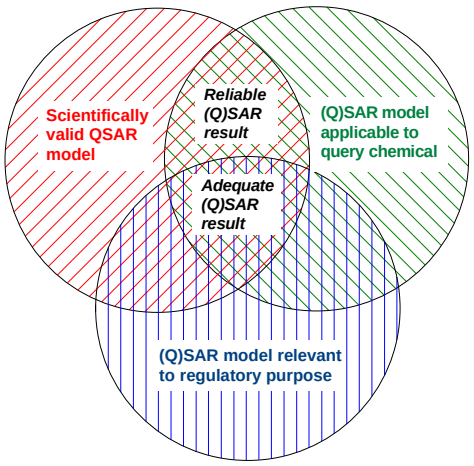
In summary, in order for a (Q)SAR result to be adequate for a given regulatory purpose, the following conditions must be fulfilled:

1. the estimate should be generated by a valid (relevant and reliable) model
2. the model should be applicable to the chemical of interest with the necessary level of reliability
3. the model endpoint should be relevant for the regulatory purpose

These conditions are illustrated in Figure R.6-1. When applying these conditions in the context of a chemical assessment, it is also necessary to consider the *completeness* of the overall information (see Section R.6.1.5.).

If a registrant intends to use (Q)SAR data instead of experimental data, the adequacy of the (Q)SAR results should be documented by using the appropriate QSAR Reporting Formats (Section R.6.1.6) These reporting formats are intended to help the registrant to provide *adequate and reliable documentation of the applied method*, as required by REACH Annex IX.

Figure R.6-1: Interrelated concepts of (Q)SAR validity, reliability, applicability, adequacy, regulatory relevance



[6] This is also referred to as the "fitness-for-purpose" concept.

The circles refer to (Q)SAR models whereas the intersections refer to (Q)SAR results with certain features. In order for a (Q)SAR result to be reliable for a given chemical, it should be generated by a scientifically valid (Q)SAR that is also applicable to the chemical of interest. This (Q)SAR estimate may or may not be adequate (fit for purpose), depending on whether the endpoint predicted is relevant to the particular regulatory purpose, and whether the estimate is sufficiently reliable for that purpose.

R.6.1.4 Regulatory use of QSARs – current experience

A useful starting point for developing guidance on the acceptability of (Q)SAR data under REACH is to understand the accepted practices under current EU legislation. Of course, this is only a starting point because REACH is based on a different paradigm for using non-testing data.

Examples of the use of (Q)SARs under different regulatory programmes, including EU programmes, is provided in the TAPIR report (ECB, 2005). A recent OECD report (OECD, 2006a) also documents case studies in OECD Member Countries. In addition, the ECB has been compiling summaries of situations in which non-testing methods have been considered within the main regulatory groups (TC NES, TC C&L and PBT WG). These surveys, which are currently being reviewed by the different working groups, show that (Q)SARs (and especially grouping approaches) have been used quite widely in the EU regulatory programmes. However, little documentation is available that captures the reasoning why a particular non-testing approach was eventually accepted or not.

This chapter provides a short summary of how (Q)SARs have been used under current EU legislation.

R.6.1.4.1 Use of (Q)SARs for risk assessment

Within the Existing Substances Regulation (ESR), data needs for risk assessment purposes have generally been fulfilled by performing tests. Measured values are always preferred to estimates. Nevertheless, there has been regulatory acceptance of non-testing methods in some circumstances.

For basic physico-chemical properties, (Q)SAR predictions have not been routinely used, since the ESR requires provision of a base set of physico-chemical data. In cases where predictions have been used, this has been to supplement experimental data. However, in a few cases, the estimates were used instead of experimental data; examples include the vapour pressure of V6 (2,3-Bis(chloromethyl)trimethylene bis(bis(2-chloroethyl) phosphate), due to practical difficulties in performing a test, and the vapour pressure of the trichloroethylene degradation product dichloroacetic acid, since no existing data were available. Another exception concerns explosive and oxidising properties, for which the absence of certain structural alerts has often been used to justify the omission of a test. The same arguments have been used occasionally to justify the omission of surface tension and some of the flammability tests; examples include V6 and TCPP (Tris(2-chloro-1-methylethyl) phosphate)).

QSAR estimates have been used routinely for predicting key environmental fate parameters of organic substances, partly because the experimental determination of these parameters can be difficult and/or expensive, and partly because the information is not normally required in the regulatory submissions. For example, the AOPWIN program (Syracuse Research Corporation (SRC), NY, USA) has been used to derive atmospheric degradation rate constants, and logKow has been used as a predictor of the organic carbon-water partitioning coefficient (Koc). For a few

chemicals (e.g. trichloroethylene, nonylphenol), QSAR-generated BCF values have been used instead of a range of measured values.

Degradation rates in sediment and soil are, however, assumed by default to be reduced if the substance is highly sorptive (since it is less available), and this is governed by the K_p value (derived from K_{ow} or K_{oc}). This is a type of SAR which is directly implemented in the TGD and in EUSES. Furthermore, QSAR model predictions have been used in sensitivity analysis concerning the properties of selected constituents in multi-constituent substances, such as SCCP & MCCP in relation to degradability and other environmental fate related properties.

For ecotoxicological endpoints, several QSARs are recommended in the TGD (EC, 2003). These QSARs have occasionally been used instead of test data, generally when it has not been technically possible to provide such data (e.g. 1,3-butadiene). More often, the QSAR estimates have been used to supplement experimental data on the acute or chronic toxicity to algae, fish and Daphnia. When data have been available for two but not all three species, QSAR estimates has been used to provide arguments about mode of action and the relative sensitivities in ecotoxicity tests, thereby justifying the use of lower assessment factors for PNEC derivation and avoiding the need for one or more chronic tests (e.g. styrene, trichloroethylene, naphthalene).

For human health effects, non-testing methods have rarely been used, and where they have been used, it is generally in the form of grouping rather than QSAR. In other cases, the QSAR result is used as supplementary information to experimental data (e.g. a QSAR analysis of the oestrogen receptor-binding ability of 1,2,4-trichlorobenzene).

In summary, in the context of risk assessment, (Q)SAR and read-across approaches have been used to:

- provide data when testing is not technically possible. Examples include vapour pressure of V6 (2,3-Bis(chloromethyl)trimethylene bis(bis(2-chloroethyl) phosphate, aquatic toxicity of 1,3-butadiene);
- provide data when they are not available for a non-prioritised substance (e.g. degradation products and components of complex substances);
- help assess the reliability of measured data, occasionally supporting the choice of an experimental value from a range of values);
- estimate properties for a range of components in multi-component substances. For such substances (e.g. chlorinated paraffins), single experimental values of basic properties (vapour pressure, K_{ow} and K_{oc}) were chosen for use in the risk assessment. However, estimated values for a range of components were additionally used as input to a sensitivity analysis;
- provide, either alone or in combination with experimental data, information on environmental effects for classification and labelling purposes (e.g. when Annex I of Directive 67/548/EEC was revised to include the environmental classification);
- estimate environmental fate data, especially partitioning behaviour and abiotic degradation (e.g. atmospheric oxidation and hydrolysis);
- argue against the need for certain tests due to lack of reactive substructures (e.g. explosivity, ozone depleting effects and hydrolysis);
- justify the need to request unusual tests (e.g. plant toxicity via atmospheric exposure);
- provide arguments about mode of action and relative sensitivities in ecotoxicity tests, to justify the use of lower assessment factors for PNEC derivation and avoid the need for chronic tests;
- provide supporting information on modes of uptake (e.g. in sediment tests) or toxicokinetics (e.g. dermal absorption).

R.6.1.4.2 Use of (Q)SARs for classification and labelling

Current EU classifications in Annex I of Directive 67/548/EEC are produced according a consensus process in which the EU Member State authorities agree on the classification. However, the classification criteria in Annex VI of Directive 67/548/EEC are also implemented by the manufacturer and/or importer to provisionally classify and label chemicals under Article 6 of Directive 67/548/EEC, and a number of industry sectors have published guidance for the *self-classification* of chemicals within their responsibility. Self-classification by industry is important because a lack of test data on the individual chemicals may imply *no classification* because the specific classification criteria largely refer to test data. Lack of hazard classification may however in such cases be misleading, because it is not necessarily due to the harmless nature of the chemical.

Use of (Q)SARs for self-classification

To support the self-classification process, the Danish EPA published an *advisory list for self-classification of dangerous substances*. The list of suggested hazard classifications was derived by using predictions from (Q)SAR models obtained or developed by the Danish EPA for the following endpoints: acute oral toxicity, skin sensitisation, mutagenicity, carcinogenicity and danger to the aquatic environment. The QSAR models were used to make predictions for the approximately 47,000 discrete organic substances in the EINECS. This Danish Advisory List contains 20,624 chemical substances with suggested classifications for one or more of the dangerous properties, and is searchable via the internet (http://glwww.mst.dk/homepage). The Danish (Q)SAR database (described above) is also accessible via http://ecbqsar.jrc.it.

Use of (Q)SARs in EU classification according to Directive 67/548/EEC

The *EU Labelling Guide* (Annex VI) contains criteria that are based largely on the interpretation of experimental test results. Nevertheless, Section 1.6.1 of the Annex recognises that *validated* QSARs can be used for the classification and labelling of substances with the following wording:

"For substances the data required for classification and labelling may be obtained: …The results of validated structure-activity relationships and expert judgement may also be taken into account where appropriate."

The use of QSAR in Annex VI can be illustrated by the use of predicted log K_{ow} values in the classification of long term aquatic hazard (bioaccumulation). When valid test data on the preferred predictor of bioaccumulation (fish BCF) are not available, the BCF value can be calculated by using a QSAR or by using a decision rule based on the (experimental or calculated) log K_{ow} value, provided that the QSAR is considered valid for the chemical in question. Classifications based on log K_{ow} values are more conservative than those based on experimental BCF data (i.e. application of log K_{ow} -based trigger results in the classification of more chemicals).

QSARs were used when the EU List of Dangerous Substance (Annex I) was updated in the early '90s to include classification for environmental hazards. The ECB generated QSAR estimates of the aquatic toxicity and (lack of ready) biodegradation for each Annex I entry. In some cases, where experimental data were lacking, the QSAR estimates were used directly as the basis of classification (Hansen *et al*, 1999; Loonen *et al*, 1997). In other cases, the QSAR estimates were used alongside experimental data.

The use of SARs in Annex VI can be illustrated by the assumption that an isocyanate is likely to be a respiratory sensitizer, unless there is evidence to the contrary. Similarly, organic peroxides are assumed to be skin irritants, unless evidence suggests otherwise. In addition, read-across from

structural analogues that are known sensitizers or carcinogens can be used as supporting evidence for classifications regarding sensitisation or carcinogenicity.

R.6.1.4.3 Use of (Q)SARs for PBT (vPvB) assessment

The assessment of PBT (Persistence Bioaccumulation and Toxicity) and vPvB (very Persistent and very Bioaccumulative) potential (referred to hereafter as PBT assessment) is treated separately, because in the EU, the identification of such potential is not part of the classification and labelling process.

PBT assessment has been carried out in accordance with the strategy and criteria proposed in the TGD, and in the framework of the European Commission's *interim strategy for the management of PBT and vPvB substances* (EC, 2001). The work was carried out by the PBT working group, which is a subgroup of the TC NES.

In general, QSARs have been used in combination with experimental data, but have also been used on their own for the selection of PBT candidates where experimental data did not exist or was considered unreliable, and alongside experimental data to confirm or negate their PBT status. An initial screening exercise, based on the use of both experimental and QSAR data for persistency, bioaccumulation and toxicity (aquatic and mammalian), led to the selection of 125 candidate PBTs with tonnages in the range ≥1000 metric tonnes. This screening exercise was based on input from the UK, the Nordic Countries, Denmark, Germany and the ECB. Most of this input concerned substances in the tonnage ≥1000 t/y and included use of both test data and QSAR predictions. The Danish input concerned the tonnage bands ≥10 t/y and employed only QSAR predictions.

The TGD criteria for identifying PBT candidates on the basis of QSAR estimates alone are similar, but not identical, to those referring to the use of screening test data. The guiding criteria for the selection of candidate PBTs based on QSAR predictions and the relationship between these criteria and those referring to screening test data are given in RIP 3.2 (PBT) – section R.11.1.2

The subsequent assessment of the candidate PBTs, using both existing experimental data and QSAR predictions in a *Weight of Evidence* approach, has led to many chemicals being deselected from the list, whereas others have been confirmed as PBTs, or targeted for further assessment.

For persistence, the EPIWIN models available within the EPI Suite (SRC, NY, USA) have been used, in addition to a MultiCASE model developed by the Danish EPA. In addition, the BIOHCWIN model, recently developed for predicting the degradability of hydrocarbons, and CATABOL, have been employed in a few cases.

For bioaccumulation, the BCFWIN model has been used, in addition to the TGD BCF model and in a few cases assisted by prediction of metabolism by the MCASE programme, METABOL. Furthermore, for deselecting PBT candidates from further consideration due to high bioaccumulation potential, use of a newly proposed criteria based on molecular size have been accepted. These criteria are based on empirical but scarce evidence for the lack of high bioconcentration in fish when the length or diameter of the substance is above certain indicative cut off values. To make this assessment of steric hindrance of uptake (which implies lack of high bioconcentration), it is necessary to calculate the length and diameter of the candidate substance in 3D, taking into account to the various conformers of the candidate molecule.

For toxicity, QSARs for short-term aquatic toxicity to algae, fish and *Daphnia* have been used, generally when test data were available for one or more of the three organisms, but lacking for the remaining ones. QSARs for chronic mammalian toxicity, reproductive toxicity and mutagenicity have been proposed, but have not been decisive for T assignment. Read-across has been used on a

case-by-case basis and grouping approaches have also been used (e.g. diarylide pigments, in which different functional groups attached to a common substructure are thought to account for differences in bioconcentration). In addition to single substances, QSARs (and experimental data) have been used to evaluate whether possible constituents of multi-component mixtures fulfil the PBT criteria.

R.6.1.5 Regulatory use of QSARs – a framework for REACH

R.6.1.5.1 Steps in assessing adequacy of (Q)SAR results

The determination of whether a (Q)SAR result may be used to replace a test result can be broken down into three main steps:

1. an evaluation of the scientific validity (relevance and reliability) of the model
2. an assessment of the applicability of the model to the chemical of interest and the reliability of the individual model prediction
3. an assessment of the adequacy of the information for making the regulatory decision, including an assessment of *completeness*, i.e. whether the information is sufficient to make the regulatory decision, and if not, what additional (experimental) information is needed.

To be used as a full replacement of an experimental test, all three conditions need to be fulfilled. In cases where some information elements are missing, (Q)SAR results may still be used in the context of a *Weight of Evidence* approach (see Section R.6.1.5.

R.6.1.5.2 Evaluation of the model validity

When using (Q)SARs, it should be remembered that (Q)SARs are models and are therefore inevitably associated with a degree of uncertainty. This uncertainty is caused predominantly by two different reasons: a) the inherent variability of the input data; and b) the uncertainty resulting from the fact that a model can only be a partial representation of reality (in other words does not model all possible mechanisms and types of chemicals). Despite this uncertainty, it should also be remembered that a (Q)SAR is not only a model, but is associated with an underlying dataset. As a representation of this dataset, the model averages the uncertainty over all chemicals. Thus, it is possible for an individual model estimate to be more accurate than an individual measurement.

The validity of a model should be evaluated in accordance with the OECD validation principles (OECD, 2004; Worth et al, 2005; OECD, 2007). These principles provide a systematic framework for describing and evaluating the characteristics of a (Q)SAR model.

Evaluation of a model in terms of a defined endpoint

One of the factors that influences the reliability of a (Q)SAR is the nature of the experimental test data used in the training set. Therefore, information about the underlying experimental data significantly increases the transparency of the model. In the development of a (Q)SAR model, the ideal but rare scenario is to use data generated by a single well-defined testing protocol with well controlled exposure conditions. If the training set of data is derived from a single laboratory, this is likely to maximise the statistical performance of the (Q)SAR model, since interlaboratory

differences in test data are excluded. If the data are collected from a single laboratory, this is likely to maximise the statistical performance of the (Q)SAR model, since interlaboratory differences in test data are excluded. If the data are collected from multiple laboratories, this is likely to reduce the model performance, due to interlaboratory variability.[7] However, it can be argued that a model with variation in the training set data caused by variations in the testing method employed and by using data from more laboratories more realistically reflects the real-world situation of empirical data. Test data for an endpoint are typically derived by using test results from multiple laboratories by use of similar, but not identical, testing methods. In other words, small variations in testing procedures and interlaboratory variability are implicitly built into the model. In general, a higher performance can often be obtained for (Q)SAR models having a more precisely defined biological endpoint and based on test data having less variance. In this guidance, no preference is expressed for using single or multiple laboratory data or for accepting variations in testing methods concerning the same endpoint in model development. The important point is to adequately document the nature and sources of the data, so that the user can make an informed evaluation.

Evaluation of a model in terms of an unambiguous algorithm

In order to establish the validity of a model rigorously, both the (Q)SAR method and its underlying data should be transparent and available. This means that documentation should be provided on the algorithm, the compounds used during the parameterisation of the model, and the correct application of the model. For example, it is necessary to know whether each parameter (descriptor) should be measured (and if so, according to which experimental protocol) or calculated (and if so, according to which algorithm / program). If calculated descriptors are used, additional information may be needed to provide guidance on the correct application of the model; for example, the ionisation and configuration states of the molecule.

For some freely available and most commercial (Q)SAR tools, full transparency is rarely, if ever, achieved. In other words, a complete set of information according to OECD principles is unlikely to be available. This should not necessarily preclude the use of such models, since it should be possible to benchmark the predictivity of the model on compounds that are similar to the chemical under investigation. For some commercially developed expert systems, such as Derek (Greene *et al*, 1999) and TOPKAT (Enslein, 1988), whilst the training sets and to an extent the algorithms are latent in the software program, both systems do provide some information to assist in benchmarking. Derek provides representative example chemicals and explanations of the mechanistic basis for the SAR used. TOPKAT flags whether a chemical of interest is in its training database and hence enables a search and retrieval of *similar chemicals* within the database with associated test data. Other commercial systems (e.g. MCase) have similar functionalities.

Evaluation of a model in terms of a defined applicability domain

An important issue in model validation is the definition of its applicability domain (Netzeva *et al*, 2005). (Q)SAR models are based on empirical knowledge about specific chemicals and therefore they are associated with limitations in terms of chemical structures, physico-chemical properties and the mechanisms of action for which the models can reliably be used. A thorough analysis of ways to formulate applicability domains for (Q)SAR models is given in (Netzeva *et al*, 2005; Jaworska *et al*, 2005; Nikolova and Jaworska, 2003; Dimitrov *et al*, 2005). It is emphasised that there is no single and absolute applicability domain for a given model. In general, a trade-off exists between breadth of applicability and predictivity. Therefore, it is important to carefully define the applicability domain and document the approach used in defining the domain. The applicability

[7] Similar considerations apply to the use of alternative test methods

domain of a model should be taken into account when estimating prediction accuracy. This has been illustrated in relation to the prediction of ADMET properties (Tetko *et al*, 2006). The development of statistical and mathematical methods for defining applicability domains is an active field of current research.

Evaluation of a model in terms of its statistical characteristics

As a result of the inherent uncertainty of a model, the statistical validation of a (Q)SAR is an important part of its overall development. The statistical characteristics of a model can be evaluated in terms of its goodness-of-fit, the robustness and predictive ability.

For regression models, the goodness-of-fit is based on the multiple correlation coefficient R^2, which should be as close as possible to one, and on the standard error of the estimate *s*, which should be small as possible. R^2 measures how well the model is able to mathematically reproduce the training set but on its own is an insufficient measure of model validity. R^2 can generally be increased by adding additional predictor variables to the model, even if the added variables do not contribute to reduce the unexplained variance of the dependent variable. Thus, the R^2 value should be used with caution. It is not recommended to define inflexible criteria for judging QSAR models on the basis of R^2 values, because the greater the underlying experimental error in the endpoint, the lower the R^2 is expected to be. Another statistic used to characterise the uncertainty of QSAR models is the mean squared error (MSE), which is calculated from the measured and predicted values of the endpoint. This error can be compared with the underlying error in the experimental data.

Caution should be exercised with models that appear to overfit the data. One way of checking this is to compare the model error (the standard error of estimate) with the error inherent in the experimental data. The standard error of estimate measures the dispersion of the observed values about the regression line. The smaller the value of the standard error of the estimate, the higher the reliability of the prediction. However, it is not recommended to have the standard error smaller than the experimental error of the biological data, because this is an indication of an overfitted model. The basic principle is that estimated data should not be more accurate than the experimental data upon which they are based.[8]

For classification models, the goodness-of-fit is often expressed by the so-called Cooper statistics: sensitivity, specificity, concordance, positive and negative *predictivities*, and false positive and negative classification rates. It is not recommended to define inflexible criteria for judging classification-based QSARs on the basis of these statistics, since there are variations in the quality of the underlying experimental data. Furthermore, these statistics should not be used in isolation to judge a model, because they capture different aspects of the overall model performance. For example, a classification model may have a low sensitivity (i.e. correctly identifies a small percentage of known positive chemicals), but it may also have a high positive predictivity (i.e. if the model makes a positive prediction, it is almost certain to be right). Such a model would be useful in a tiered testing approach, on the assumption that some, but not all, positive chemicals could be reliably identified by the model. In other words, a model should not be dismissed just because one of the Cooper statistics is low. The classification models can also overfit the data and there is a particular danger of overfitting when the size of one of the groups to be separated is small.

In the case of SARs, because of their qualitative nature, validation may necessitate the use of specific approaches, e.g. the application of similarity analysis to datasets containing experimental

[8] This should be assessed on the basis of the model dataset as a whole. In contrast, the model prediction for an individual chemical may be more accurate than its experimental value.

data for the predicted endpoint. While it is possible, in principle, to validate individual SARs, the information obtained is necessarily limited. It is therefore important to consider how the SARs are used in practise. For example, individual structural alerts from an expert system are used in combination to make a prediction. In such a case, the alerts should be validated by an integrated approach, assessing all the rules at the same time and taking account of any hierarchies in their use. When SARs are applied in order to alert for potential hazard (or enhanced hazard), a warning should be given that a lack of alert does not always mean lack of hazard since a hazardous chemical functionality might not be known as such. A hazard might appear when functional groups, which are otherwise not recognise as alerting, appear close to each other in the molecule, or are positioned in a way that triggers specific (receptor) interaction. The alert, if recognised in a molecule, can be modulated in both directions of activating and deactivating, and for this reason the evaluation on a case-by-case basis is recommended.

The most rigorous form of statistical validation is *external validation*, i.e. to use an external set of substances (i.e. substances that have not been used for establishment of the model) if these compounds are shown to be representative of the class of substance to be predicted. However, if external validation is not possible, *internal validation* techniques such as cross-validation and permutation testing can be used (Eriksson *et al*, 2001) to provide indications of model robustness and predictive power. A cross-validated regression coefficient (usually referred to as Q^2) is computed by dividing the dataset into a number of subsets and then developing a series of models from some but not all of these subsets. The subset that is left out of model development is used to assess the statistical performance of the model. This is repeated for each "parallel" model developed, and the computed Q^2 reflects the average performance of all models. From a scientific perspective, a $Q^2 > 0.5$ is generally regarded as good and a $Q^2 > 0.9$ as excellent, but these guidelines strongly depend on the specific case (Eriksson *et al*, 2003). Cross-validation can be carried out in various ways, using approaches such as leave-one-out (LOO), leave-many-out (LMO), randomisation, stratified randomisation and bootstrapping; further information is provided elsewhere (Eriksson *et al*, 2001; Efron and Gong, 1983; Gramatica, 2004). The diversity of internal validation approaches available to the model developer underlines the need for transparency in documenting the approach chosen.

An external validation can be performed when it is possible to find new compounds belonging to the same chemical domain in a statistically significant number (Gramatica, 2004). Often, external validation is carried out by the model developer as part of the model development process. In this case, the external validation is achieved by rationally splitting the available input data set into a training set (for model development and assessing goodness-of-fit) and a validation or test set (for assessing predictivity). The model developed using only training set chemicals is then applied to the validation set to verify the predictive ability of the model. An external Q^2 can then be computed by using the measured and predicted values of the validation set. This approach can provide a reliable indication of model predictivity, but only if the splitting is performed by partitioning the compounds in a well-defined and rational way, since the external test set should be representative of the model applicability domain. Strictly, when performing external validation, the test set should not be used for the development of the QSAR model, which means that the overall number of data available is reduced. This may be critical when the overall number of test data is limited. Thus, careful attention should be paid to the selection of the training set and test set compounds. In practise, data used for external validation are often used to improve the model validated. In such cases, it is important to remember that the validation statistics do not refer to the final model developed (and possibly used). The fact that multiple choices can be made in external validation underlines the importance of transparent documenting the approach chosen. The relative merits of cross-validation and external validation have been debated in the academic literature. For example, a comparison of validation procedures as well as approaches to the division of experimental datasets

into training and test sets can be found in (Gramatica, 2004; Golbraikh *et al*, 2002; Golbraikh and Tropsha, 2002; Kraker *et al*, 2006).

Evaluation of a model in terms of its mechanistic interpretation

When establishing the validity of a (Q)SAR model, the reliability can be associated with the statistical characteristics of the model whereas the relevance can be associated with the mechanistic interpretation of the model. A mechanistic interpretation refers to the assignment of physical/chemical/biological meaning to the descriptors used in the model and an explanation of the relationship between the descriptors (predictors) and the predicted endpoint.

An understanding of the mechanistic basis of a (Q)SAR increases the confidence in the model based on the other validation principles, and in some cases is an integral aspect of the applicability domain assessment. When there is a choice between multiple models for an endpoint, the identification of the mode or mechanism of action[9] may be a necessary prerequisite for selecting the appropriate model and for avoiding models that might give less reliable predictions (Schultz *et al*, 2006).

Evaluation of models based on novel (Q)SAR approaches

In principle, the QSAR validation principles can be applied to model developed by using more recently developed approaches, such as neural network modelling. In recent years, Kohonen neural networks and counter propagation neural networks have become an important tool in QSAR modelling. The validation of such models is performed in terms of recall ability test, which assesses how well the model recognises the training objects, cross-validation procedure (LOO and LMO) as well as external validation. The application of OECD principles to a neural network model has been illustrated by Vracko *et al* (2006). This case study demonstrates that a QSAR model derived using counter propagation neural network satisfies most of the OECD validation principles.

Another QSAR modelling approach is 3D QSAR, and in particular CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Indices Analysis). These are useful for providing insights into the mechanisms of molecular action, such as ligand-protein interactions. The CoMFA and CoMSIA QSAR models are derived by the PLS (Partial Least Squares) method. The performance of these models are generally expressed in terms the cross-validated Q^2, the optimal number of components (N_{opt}) and the cross-validated standard error of prediction (SEP_{cv}). The models can also be validated by applying external validation.

Consensus modelling approaches (or *battery* approaches) make *consensus* predictions on the basis of results generated by multiple QSAR models. The main assumption in consensus modelling is that multiple models will effectively describe more aspects of relationship between chemical structure and the endpoint of interest than a single model (Golbraikh *et al*, 2003). The development and application of these approaches is a field of active research. At present, it is not possible to give firm guidance on how to use such approaches in the regulatory context.

R.6.1.5.3 Assessment of the reliability of the individual model prediction

Assessment of model validity is a necessary but not sufficient step in assessing the acceptability of a QSAR result. Assuming that the model is considered valid, the second and crucial step is to evaluate the reliability of prediction for a specific compound. The question being asked is "Is this

[9] The mechanism of toxic action can be defined as what happens at the molecular/biochemical level , while a mode of toxic action can be defined as what happens at the cellular/physiological level.

QSAR appropriate for the compound of interest?" This is not a trivial question, but it can be broken down into the following questions:

1. is the chemical of interest within the scope of the model, according to the defined applicability domain of the model?
2. is the defined applicability domain suitable for the regulatory purpose?
3. how well does the model predict chemicals that are *similar* to the chemical of interest?
4. is the model estimate reasonable, taking into account other information?

When addressing question 1, it is important to bear in mind that the applicability domain of a model can be defined in one or more of the following ways:

- descriptor domain (do the descriptor values of the chemical fall within defined ranges?)
- structural fragment domain: does the chemical contain fragments that are not represented in the model training set?
- mechanistic domain: does the chemical of interest act according to the same mode or mechanism of action (e.g. a ligand-receptor interaction assumed to be responsible for the biological effect observed) as other chemicals for which the model is applicable?
- metabolic domain: does the chemical of interest undergo transformation or metabolism, and how does this affect reliance on the prediction for the parent compound?

Clearly, the more explicit the definition of the model domain, the easier it will be to answer these questions. In practice, not all of this information will be available.

Question 2 arises because most currently available models were not tailor-made for current regulatory needs and inevitably incorporate biases which may or may not be useful, depending on the context of prediction. A model can be biased toward certain types of chemicals (e.g. a model optimised to calculate values for those training substances that most closely matched measured ones), or toward a certain type of prediction (e.g. a model optimised to correctly identify positives at the expense of correctly identifying negatives). Such biases do not affect the validity of the model, but they do affect its applicability for specific purposes. Information on these biases can therefore help the user determine whether or how the model is suitable. For example, many QSARs for predicting biodegradation are biased towards predictions of non-ready biodegradability. The predictions generated by such models may be used in a conservative manner to predict non-ready biodegradability, but predictions of biodegradability might not be reliable. Another example relates to (Q)SARs developed for specific chemical classes. For some classes, models have been developed but there is no regulatory need to predict chemicals from such classes. For other classes, there is a regulatory, but models are lacking. Therefore, in the ideal situation, there will be a good match between the (Q)SAR applicability domains and the regulatory inventory of interest.[10]

Question 3 provides a simple way of checking whether a model is appropriate by checking its predictive capability for one or more analogous compounds that are similar to the one of interest and for which measured values exist. These analogues may be selected from the training set of the model (if this is available) and/or additional datasets. Addressing question 3 is effectively using a

[10] In the case of REACH, this is the EU Inventory of Chemicals.

read-across argument to support the reliability of the (Q)SAR prediction. When using analogue data it is the important to consider factors that might affect the quality of the measured endpoint (e.g. molecular weight, absorption, water solubility, volatility, and ionic dissociation). Guidance on judging the adequacy of read-across is covered in Section R.6.2.3.1.

When addressing question 3, the choice of similarity metric is important. If similarity is assessed by using the same descriptors that are included in the QSAR, the argument becomes tautologous, because structures identified as *similar* on this basis are likely to have similar predictions. Thus, the choice of such analogues should be justified case-by-case, using specific arguments in relation to the endpoint in question.

A more generic check, expressed by question 4, is whether the predicted value *seems reasonable*. This inevitably implies an expert judgement, which should be clearly rationalised. One approach could be done to cross-referencing the calculated value(s) for the substance of interest (and possibly also its analogues) to measured ones for related endpoints. For example, a calculated boiling point of >300°C should correlate with a low measured vapour pressure.

In general, it is recommended that the reliability of a QSAR estimate is assessed by using a *Weight of Evidence* approach, based on the above-mentioned considerations. These considerations do not necessarily need to be applied in any fixed order, but a stepwise approach might be useful. For example, a stepwise approach for determining the reliability of (Q)SAR model predictions for skin sensitisation, incorporating many of these considerations, has recently been proposed (Dimitrov *et al*, 2005).

In commercial QSAR tools, the domains are characterised to a greater or lesser extent. For example, in Derek, it is relatively straightforward to judge whether the compound has a particular structural feature which fires the alert, although a chemistry judgement to decide whether the alert is wholly relevant for the compound of interest is required. Within TOPKAT, it is possible to obtain an assessment of whether the compound of interest falls within the applicability domain of the model (both with respect to the fragment and descriptor space). TOPKAT also flags whether the chemical of interest is in its own database and retrieves similar compounds with associated test data. The model estimate, the assessment of applicability domain and other similar analogues builds up a package of supporting information to enable a user to assess the reliability of a given prediction result.

R.6.1.5.4 Assessment of adequacy

The third and last step of the evaluation considers the regulatory requirements and the extent to which non-testing data adequately fulfils these requirements, either alone or in combination with other information (including test data). Even though computer-based estimation tools are becoming increasingly available, these tools are intended to facilitate the process of (Q)SAR acceptance and cannot substitute the need for expert judgement and dialogue between industry and authorities. The use (Q)SAR predictions *in an automatic way*, without considering validation results, regulatory purpose and use of WoE judgements is not recommended. Having said that, on the basis of current experience, it is difficult to give detailed guidance on how to use (Q)SAR estimates for regulatory purposes. Indeed, it is debatable to what extent it will be possible to codify accepted practise in terms of rules-of-thumb, although some attempts must be made along these lines.

The approach proposed is that experience in the regulatory use of non-testing data should be obtained by following a learning-by-doing approach, with the learnings being documented as

examples for reference purposes.[11] In this way the possibilities for enhanced use of non-testing methods in general under REACH will be optimised whilst avoiding long bureaucratic and formal adoption schemes.

Only limited guidance on the acceptance of (Q)SARs can be given at this moment. However, two important principles can be outlined, as already explained in the TAPIR report (ECB, 2005):

- the **principle of proportionality** expresses the relationship between the amount of information needed and the severity of the decision. For example, more data would be needed to ban a substance compared with the data needed for classification of the substance as a skin irritant. Another example is that a higher accuracy or confidence in any data point, including a (Q)SAR prediction, is generally needed when the value is close to a regulatory decision point (e.g. a classification cut off or a risk quotient close to 1);
- the **principle of caution**[12] (or conservativeness) expresses the relationship between the amount of information needed and the (likely) consequence(s) of the decision based on that information being wrong. For example, if there is higher uncertainty in the data and/or the more severe the consequence of being wrong, the more conservative the approach in extrapolating data to *safe* exposure levels (i.e. margin of safety or higher assessment factors are used).

As a consequence of these two principles (which also apply to test data), the relationship between scientific validity and regulatory acceptability is not a constant, but varies according to the decision being made.

The TAPIR report (ECB, 2005) also argues that non-test data should be used in the same way, and according to the same principles and criteria, irrespective of whether the information is required according to the tonnage-dependent requirements of REACH or not. This could be called the **principle of consistency**.

In summary, further work and discussion is necessary to build a common understanding on the acceptability of individual (Q)SARs for specific regulatory purposes. This guidance, which according to REACH Annex XI will be documented in the form of examples, should take into account the principles of proportionality, precaution and consistency.

R.6.1.5.5 The acceptance of (Q)SAR data under REACH

The process of (Q)SAR acceptance under REACH will involve initial acceptance by industry and subsequent evaluation by the authorities, on a case-by-case basis. It is not foreseen that there will be a formal adoption process, in the same way that test methods are currently adopted in the EU and OECD. In other words, it is not foreseen that there will be an official, legally binding list of (Q)SAR methods. With reference to the acceptance criteria in REACH Annex XI, it is stated that "*the Agency in collaboration with the Commission, Member States and interested parties shall*

[11] The need for such documentation is expressed in Annex XI, where it states that the "Agency in collaboration with the Commission, Member States and interested parties shall develop and provide guidance in assessing which (Q)SARs will meet these conditions and provide examples."

[12] This should not be confused with the Precautionary Principle.

develop and provide guidance in assessing which (Q)SARs will meet these conditions and provide examples". Standardised reporting formats for QSAR models and their predictions are provided in Section R.6.1.6.

R.6.1.6 QSAR Reporting Formats

R.6.1.6.1 The need for appropriate documentation on (Q)SARs

According to Annex XI of the REACH regulation, one of the conditions for using (Q)SARs instead of test data is that *adequate and reliable documentation of the applied method is provided*. Here the term *appropriate* is used to interpret what is intended in REACH Annex XI by *adequate and reliable* documentation. At present, an extensive summary of appropriately documented (Q)SARs is not available. Therefore, the ECB in consultation with the EU QSAR Working Group, took the initiative to start building a database of evaluated (Q)SARs, which should help to identify (Q)SAR models suitable for the regulatory purposes of REACH. This database (the JRC QSAR Model Database) will be made freely available from the website (http://qsardb.jrc.it).

In the wider international context, the content of the JRC QSAR Model Database could also be used in the (Q)SAR Application Toolbox, a project currently being led by the OECD. The (Q)SAR Application Toolbox is intended to be a set of tools supporting the use of QSAR models in different regulatory frameworks by providing estimates for commonly used endpoints together with guidance on the interpretation of estimated data.

The requirement for appropriate documentation of (Q)SARs has led to discussions on what information is required for (Q)SARs and how this information should be structured. Different types of (Q)SAR Reporting Formats (QRFs) are being developed to provide a standard framework for summarising and structuring key information about (Q)SAR models and their predictions. The reporting formats are not meant to limit the use of (Q)SAR approaches or impose what methods should be used – they are simply meant to provide sufficient and up-to-date information so that informed choices can be made regarding the use of (Q)SARs and so that the same information is available to Industry registrants, the MS authorities, and the European Chemicals Agency.

R.6.1.6.2 Different types of QSAR Reporting Formats (QRFs)

Three different reporting formats have been proposed to capture the different types (or levels) of information.[13] The description of a particular (Q)SAR model (i.e. description of the algorithm, of its development and validation based on the OECD principles) will be stored in the **(Q)SAR Model Reporting Format (QMRF)**. This should involve an input from the developer(s) and/or proponent of the model, as well as information from any evaluation studies performed with the model. The **(Q)SAR Prediction Reporting Format (QPRF)** will explain how an estimate has been derived by applying a specific model or method to a specific substance. This should include information on the model prediction(s), including the endpoint, a precise identification of the substance modelled, the relationship between the modelled substance and the defined applicability domain, and the identities of close analogues. Another important piece of information is the relationship between the predicted endpoint and the regulatory endpoint of interest: in cases where the predicted endpoint is not the endpoint of regulatory interest, the relevance of the former to the latter should be described.

[13] The development of these formats started in the context of the QSAR Experience Project, coordinated by RIVM (NL), which was subsequently subsumed into the activities of the QSAR Working Group.

In the overall assessment of a given chemical, it will often be necessary to integrate the QSAR estimates with other sources of information (e.g. *in vitro* and *in vivo* test data). This data integration should be based on *Weight of Evidence* considerations, which are perhaps better thought of as *totality of evidence* considerations, because it is not necessarily the case that weights will be attached to individual pieces of information. The QSAR Working Group has discussed the idea that this level of integration should be documented in a specific reporting format (called a Totality of Evidence Reporting Format (TERF) or *Weight of Evidence* Reporting Format (WERF)).

Collectively, these three levels of reporting formats would provide a comprehensive description of the use of the (Q)SAR and other approaches applied during the classification and safety assessment of a given substance for a specific endpoint, and for justifying any further testing considered necessary to obtain adequate and complete information.

The QRFs should be regarded as a communication tools to enable an efficient and transparent exchange of (Q)SAR information between Industry and MS authorities. Ideally, these reports would be attached to the registration dossier.

The structure of the formats needs further discussion. However, they should be designed to ensure transparency, consistency, and acceptability:

> **Transparency**: Information on the (estimation) methods, predictions and reasoning should be clearly reported and explained to facilitate interpretation of conclusions. Ideally, all of this information should be in the open domain.
>
> **Consistency**: Information related to different approaches should be reported in a common format to enable a comparison of different models used and predictions made.
>
> **Acceptability**: The reports should include all relevant information required to evaluate the adequacy and completeness of the (Q)SAR information for a given substance and endpoint. It should also be auditable, i.e. the rationale is clearly linked back to a regulatory decision.

The general form of the QMRF and QPRF, as developed to date, are described below. In addition, read-across and category formats have been developed for grouping methods (see Section R 6.2.6). The contents of an eventual TERF (or WERF) will depend on the progress made in understanding how to integrate testing and non-testing data and further discussion on this subject will be needed.

Under REACH, reporting formats could be submitted to the Agency as attached files in an IUCLID dossier. In some cases, it may be sufficient for the registrant to make reference to a pre-existing reporting format (accessible, for example, via the JRC QSAR Model Database at http://qsardb.jrc.it).

R.6.1.6.3 The (Q)SAR Model Reporting Format (QMRF)

The QMRF provides the framework for compiling robust summaries of (Q)SAR models and their corresponding validation studies. The structure of this format has been designed to include the essential documentation that can be used to evaluate the concordance of the (Q)SAR model with the OECD principles.

The QMRF contains information on the source, type, development, validation, and possible applications of the model. The set of information that are provided in the QMRF should be used to facilitate regulatory considerations of (Q)SARs, and for this purpose, the structure of the QMRF is devised to reflect as much as possible the OECD principles for the validation, for regulatory purposes, of (Q)SAR models. In the QMRF each of the OECD principles is associated with a set of

fields; the different sections forming the QMRF are listed below in Section R.6.1.9 with brief explanations.

Section 3.2. of the QMRF (see Section R.6.1.9.1)requires an endpoint to be selected from a predefined list. Section R.6.1.9.2. lists the endpoints grouped according to four types of effect: physico-chemical, environmental fate, ecotoxicological, human health.

Information about the identity of the chemicals contained in both training and test sets can also be included in the QMRF (where possible): a) Chemical Name (IUPAC); b) Chemical Name (Not IUPAC); c) CAS Number; d) SMILES (Simplified Molecular Input Line Entry System); e) InChI (IUPAC International Chemical Identifier); f) Mol file; g) Structural formula; e) Values for the dependent variable; f) Values for the descriptors.

In commercial models, the training and test sets (or parts thereof) are proprietary and hidden from the end-user. The issue of how to provide sufficient transparency within commercial models yet maintain the confidentially of proprietary information is needs to be discussed and resolved.

The QMRF is evolving, with input from the EU QSAR Working Group, the OECD ad hoc group, and other interested parties/persons.[14] Some of the specific issues being discussed include:

- How to ensure transparency and completeness of the report for models where certain information is confidential (e.g. algorithm, training set).
- How to implement a flexible QMRF which is capable of accommodating all sorts of different (Q)SAR models.
- The level of resolution required within the QMRF to fully evaluate the concordance of the model with the OECD principles for (Q)SAR validation.

As mentioned above, the ECB has started building a freely-accessible inventory of evaluated (Q)SARs (the JRC QSAR Model Database), which should help to identify valid (Q)SAR models for regulatory purposes. For this reason, ECB is implementing an application that will manage the creation, storage and download of QMRFs. A web-based interface will allow for the retrieval of QMRFs in a suitable readable format and for the submission of a QMRF in, for example, excel format.

R.6.1.6.4 The (Q)SAR Prediction Reporting Format (QPRF)

The framework to describe the evaluation of a specific substance by a specific model will be provided by the QPRF which makes reference to the QMRF. In the QPRF, the prediction outcome is presented with some reasoning. The reliability of the prediction should also be assessed and provided.

A scheme for ranking the reliability of the predictions, analogous to the Klimisch (1997)might be misleading for non-testing data and appears therefore not useful. This is because non-testing data is generally used in combination with other information in a *Weight of Evidence* approach. Thus, the level of confidence in an individual estimate is highly context-dependent, and is based not only on the validity and performance of the model but also on the availability and quality of other data.

[14] The ECB launched a beta test of the QMRF in June 2006. Based on the outcome of the beta test in December 2006, an updated version (January 2007) is given in Section R.6.1.0.

The structure of the QPRF is evolving. A preliminary proposal is given in Section R.6.1.10

R.6.1.6.5 Conclusions

This chapter has presented work in progress on the development of reporting formats intended to ensure the appropriate documentation of non-testing methods under REACH. The QMRF has already been extensively discussed and agreed at the OECD level, although small adaptations have more been proposed by the QSAR Working Group. The QPRF is at an earlier stage of development, whereas the need for a TERF (WERF) is still questionable. To some extent, the QPRF is REACH-specific, since judgements are included on the adequacy and completeness of (Q)SAR estimates for the regulatory goals of REACH. In addition to these reporting formats, equivalent reporting formats for read-across and category approaches are available (see Section R.6.2.6.). The formats will need to evolve as further experience is gained, rather than being *fixed* in their current forms.

R.6.1.7 Stepwise approach for the use of non-testing data

R.6.1.7.1 Meeting regulatory requirements with computational tools

In Section R.6.1.8. , the most commonly used (Q)SAR tools are reviewed. It is anticipated that some, but not all, of the existing tools will be useful for addressing the requirements of REACH. Some tools will be useful, but not widely available, due to their proprietary nature. Other tools are currently under development, or will need to be developed in the near future.

Due to the limited availability of freely-accessible (Q)SAR software, there is a need to develop a range of transparent and open-source tools, which should eventually be available to all stakeholders in the REACH process (especially industry, authorities and the Agency). The essential functionalities needed for implementing REACH should ideally be available in the form of a Decision Support System (DSS) in which different needs (functionalities) are addressed by different (but mutually compatible) components tools. The different components of such a DSS should enable the user to generate non-testing information within the context of a structured workflow, and to obtain guidance on the applicability of the information for the regulatory goals of REACH.

The need for a DSS is not new, and was extensively discussed in an ECB workshop in May 2005 and in an ECB contractor's report (Gini, 2005). These discussions have led to the development, at the OECD level, of a prototype DSS called the QSAR Application Toolbox.[15]

This chapter presents current thinking by ECB on how different commercially and publicly available tools, including those described in Section R.6.1.8, could be integrated into a DSS that enables the generation of non-testing data for REACH. The intent of this chapter is to illustrate how a diverse range of different tools can be used in the context of a single workflow. The development and evaluation of this workflow represents a work in progress.

R.6.1.7.2 Structured workflow for the generation and use of non-testing data

The workflow proposed for the generation and use of non-testing data comprises a sequence of operations exploiting the functionalities of a wide array of Information Technology (IT) tools and databases. Some of these tools are already available, whereas others need to be developed. The

[15] The QSAR Application Toolbox is intended to be broadly applicable in the international context. It will nevertheless take into account as far as possible the specific needs of national/regional legislations, including REACH.

description of the workflow in this chapter tries to identify useful tools that could be used in association with different steps of the process, but due the large number of available applications, only some of them are mentioned.

The proposed stepwise approach is intended to help the registrant meet the general requirements for using non-test methods laid down in REACH Annex XI (e.g. a QSAR prediction for a substance should fall within the applicability domain of the (Q)SAR model, and appropriate documentation of the applied method should be provided).

The workflow is summarised in Figure R.6-2 and includes the following steps:

Step 0: Information collection

Step 1: Preliminary analysis

Step 2: Use of classification schemes

Step 3: Search for structural alerts

Step 4: Preliminary assessment

Step 5: Read-across

Step 6: (Q)SAR predictions

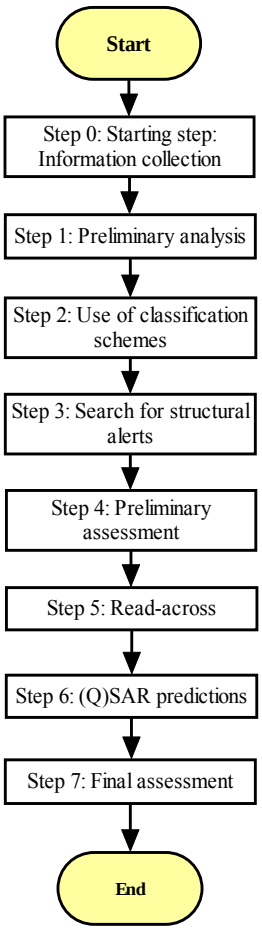
Step 7: Final assessment

The details of the various steps of the workflow are explained below in separate sections. As the user proceeds through the workflow, a *Working Matrix* is built. The *Working Matrix* stores all the information collected during the workflow. Different rows store information for different compounds, and different columns refer to specific types of information (e.g. a physico-chemical properties).

It is emphasised that the workflow is intended to be flexible, so that it can be adapted to meet the specific and context-dependent needs of the user. For example, it might be more efficient, depending on the substance, endpoint of interest and regulatory purpose, to omit certain steps or perform them in a different order. However, even if some of the steps do not provide useful information for certain chemicals and endpoints, it is recommended to consider all of the steps because it will increase the confidence in the overall assessment.

The guidance below is based on the assumption that each chemical is a subject of potential transformation (either biotic or abiotic), independently of whether it actually transforms under a defined set of conditions. The term *parent compound* is introduced to distinguish between the main chemical of interest (the *parent*) and its potential products.

Figure R.6-2: Flowchart for the use of non-testing approaches in the regulatory assessment of chemicals



In the starting step, information on experimentally determined and estimated properties is collected. Step 1 involves a preliminary analysis of the reactivity, uptake and fate profile expected for both the substance of interest and its (chemical or metabolic) transformation products. Step 2 solicits further information on likely biological activity of the compound using classification schemes (where available) for the endpoint of interest. Step 3 involves an investigation for the presence of structural alerts within the chemical(s) of interest. Step 4 involves a preliminary assessment of the expected uptake, toxicity and fate profile. Step 5 explores the use of grouping approaches, whereas Step 6 uses (Q)SARs. Finally an overall assessment is carried out in Step 7. Depending on the particular substance, endpoint of interest and regulatory purpose, certain steps may be omitted, or performed in a different order.

R.6.1.7.3 Step 0 - Information collection

Assess information requirements under REACH

The workflow begins by considering the information requirements under REACH, which are largely tonnage-dependent and specified in Annexes VII-X.

Select a representative structure for the assessment

The composition of the substance (main chemical component, other components, impurities) should be clearly defined, and a specific compound is selected for the study. This operation is necessary

because predictions from (Q)SAR methods and category/read-across approaches are generated by feeding them with a single well-defined structure (generally the two-dimensional structural formula in the form of a SMILES code). The purity/impurity profile might be useful at a later stage to explain discrepancies between experimental and non-testing data. In the case of multi-constituent substances (mixtures), it may be necessary to model two or more structures, if a single representative structure is not considered sufficient.

For multi-constituent substances, a similar workflow may be relevant for individual components. The selection of relevant components will depend on the particular substance, endpoint of interest and regulatory issue.

Verify structure of parent compound

If the parent compound is known by CAS or EC number or by name, it is essential to derive its structure (e.g. in the form of the SMILES code) to be used in the prediction generation process. This can be achieved using a *Structure Converter* tool. If the structure is known, it is important to verify that the structural information agrees with the CAS number or with the name. Some online tools that can be employed at this step are:

- ChemID (National Library of Medicine), which can be used to check the CAS number, the chemical name, and to identify the corresponding possible structure
- Ambit (IdeaConsult Ltd), which can be used to convert CAS to SMILES
- CAS SciFinder (commercial), which is a definitive source of CAS registry numbers matched with chemical name and structure information.
- The (Q)SAR Application Toolbox will contain libraries which convert CAS to SMILES.

Collect available information for parent compound

Available chemical information (including physico-chemical properties and toxicity data) about the parent compound can be retrieved from the ESIS, the European chemical Substances Information System, accessible from the ECB website.

The (Q)SAR Application Toolbox will contain a resident database with available experimental data (e.g. the aquatic toxicity data from the AQUIRE database) and will the Toolbox will allow the user to add missing experimental results to the resident database.

In addition, the use of non-testing data will benefit from the implementation of the following databases, which could be queried through ESIS:

(Q)SAR Model Database (QMDB): this database will be an inventory of robust summaries of (Q)SARs that can be searched, for example, by endpoint or by chemical. The search by chemical could provide information on whether the chemical in question is present among the training and test sets of some models. The QMDB will provide information on evaluated models documented in the form of (Q)SAR Model Reporting Formats (QMRFs);

(Q)SAR Prediction Database (QPDB): the models that are documented in the QMDB can be used to generate predictions for various chemicals. These predictions will be stored in the (Q)SAR Prediction Database, so that each prediction is associated with a robust summary of the model used to generate it. For individual predictions, the QMDB will provide links to the appropriate (Q)SAR Prediction Reporting Formats (QPRFs);

Chemical Categories Database: an inventory of existing categories will be useful to apply category/read-across approaches. This database should include all the information necessary to adequately document the use of a specific category for generating predictions.

The (Q)SAR Application Toolbox will contain a library of (Q)SAR models, Chemical Categories as well as a database of (Q)SAR predictions. Efforts are currently on-going to ensure that the information available through ESIS will also be available via the (Q)SAR Application Toolbox.

Search external databases

External databases can be searched to obtain additional relevant information on the physico-chemical, toxicological, ecotoxicological properties of the parent compound. A list of useful external databases is provided in Section R.6.1.8.

A tool capable of interfacing different on-line databases and that allows for the retrieval of the entire set of available information for the compound of interest from all databases in a single run would be a very useful application. Attempts will be made to implement such a tool within the (Q)SAR Application Toolbox.

Build working matrix and identify information gaps

All pieces of information collected in the previous phases are stored in the Working Matrix (WM), which is used as a growing summary of the workflow process. It is then possible to identify information gaps by comparing the REACH information requirements and the collated information.

If necessary the search for existing information is refined taking into consideration specific information gaps.

An endpoint for which non-testing data is needed and which can be generated by means of (Q)SAR methods and category/read-across approaches is then selected, and one or more of Steps 1-7 are followed to obtain the non-testing data along with guidance on how to interpret the data for regulatory purpose. In addition, (Q)SAR data that is not specifically referred to in the Information Requirements, but which may nevertheless contribute to the regulatory assessment, can be obtained by following Steps 1-7.

R.6.1.7.4 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 physico-chemical data.

Collect information on the reactivity of the parent compound

At this stage, information on the reactivity/stability of the parent compounds is collected/generated. Available information on biotic and abiotic reactions involving the parent compound can be retrieved from the peer-reviewed literature and from available tools and databases, including the following resources:

- CAS SciFinder (commercial)
- MDL Reaction database (commercial)
- TIMES (commercial) developed by LMC, University of Bourgas, Bulgaria
- Catabol (commercial) developed by LMC, University of Bourgas, Bulgaria
- KEGG

- METEOR, Lhasa (commercial)
- META, MCASE (commercial)
- HYDROWIN, as part of EPIWIN (for hydrolysis only)

Not many freeware software applications are available for analysing the metabolic fate of chemicals. The development of a freeware tool that can generate a list of plausible metabolites would be very useful and is being planned by ECB. The (Q)SAR Application Toolbox will contain maps of estimated metabolic pathways for a large number of chemicals.

The stability/reactivity of the parent compound may be further estimated by analysing fragments and molecular orbital energy (like HOMO, LUMO). At present this kind of analysis is performed by experienced chemists but a tool capable of making simple descriptor-based predictions of reactivity would be highly desirable.

On the basis of the collated information, the Working Matrix is updated. Additional rows are added with information on metabolites and reaction products.

Preliminary analysis of uptake and fate

A preliminary assessment of expected reactivity, uptake and fate is performed on the basis of the information for the abiotic and biotic reactions involving the parent compound. The following considerations should be taken into account:

- how molecular weight, size, log Kow, electric charge and stability/reactivity parameters affect uptake and consequently toxicity
- whether ionisation can take place at the relevant pH (role of pKa) and whether this affects uptake, toxicity and fate
- what chemical reactivity (what type(s) of reactions) is expected for the parent compound
- which metabolites and reaction products (i.e. hydrolysis products) are generated

Select suitable query compound(s)

The preliminary analysis of uptake and fate is used to determine which compound(s) (parent compound and/or reaction products and/or metabolites) are suitable for modelling the endpoint of interest. Having identified the suitable query compounds, Steps 2-6 are applied for each compound.

R.6.1.7.5 Step 2 - Use classification schemes for endpoint of interest

Further information on the likely biological activity of the compound may be obtained using classification schemes (where available) for the endpoint of interest. For example, classification schemes by Verhaar *et al* (1995) and Russom *et al* (1997) can be used when assessing the mode of action for acute fish toxicity. The classification scheme developed by Cramer *et al* (1978) is useful for evaluating the likely systemic toxicity of a compound. The Verhaar and Cramer classification scheme have been automated in Toxtree, a freeware application developed by Ideaconsult Ltd., (Sofia, Bulgaria) and accessible from the ECB website.

R.6.1.7.6 Step 3 - Search for structural alerts for endpoint of interest

In this step, structural alerts (where available) for the endpoint of interest are searched. Both 2D and 3D structural alerts can be used, although 2D SAR approaches are likely to be more readily used

since they are more intuitive and more easily automated. Several commercial software programs are available for performing this kind of analysis:

- Derek, Lhasa, (commercial)
- MCASE (commercial)
- Leadscope (commercial)

The use of non-testing data would be facilitated by the implementation of freeware tools encoding diverse SAR models available in the literature (e.g. the BfR rules for eye irritation/corrosion and skin irritation/corrosion). The (Q)SAR Application Toolbox will contain a number of structural alerts for a number of endpoints and will allow the user to add their own structural alerts.

R.6.1.7.7 Step 4 – Preliminary assessment of expected type of reactivity, uptake, toxicity and fate

In this step, which requires expert judgement, a preliminary assessment is made of the expected reactivity/uptake/toxicity/fate profile of the parent compound is performed, using the outcomes of Steps 1-3 applied to all relevant query compounds.

The preliminary analysis in Step 1 (physico-chemical properties, metabolites, reaction products) may help to assess the likelihood of exposure to the organism (or tissue) or environmental compartment of interest.

The application of Step 2 may help to classify the mode of toxic action of the compound. This information is useful in a later step when evaluating which (Q)SAR models should be applied. Step 2 also helps to make Threshold of No Concern estimations, i.e. to predict exposure levels below which there would be no appreciable risk to human health or environmental species.

The application of Step 3 (SARs) may help to identify which hazards are likely to be present or absent.

This evaluation step should also help to define the hazard and risk assessment strategy that is be further supported by applying the subsequent steps.

The outcome of Step 4 is also used to update the Working Matrix for future reference.

R.6.1.7.8 Step 5 – Read-across

Select a suitable query compound

This step is aimed at filling data gaps for all the query compounds using a read-across (or analogue) approach, where the endpoint information for one or more *source chemicals*, is used to make a prediction of the endpoint for the *target chemical*. Read-across is based on the identification of similar compounds.

Step 5a. Determine whether the query compound belongs to an existing category

A straightforward way to find analogues of the query compound is to browse existing categories where the compound may be listed as a member. Chemical categories are groups of chemicals whose physico-chemical and toxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity. These structural similarities may create a predictable pattern in any or all of the following parameters: physico-chemical properties, environmental fate and environmental effects, and human health effects.

It is also possible to apply expert knowledge to link the compound in question to an existing category even though the compound is not explicitly listed as a member.

The availability of a database of existing categories would be useful for this phase. The (Q)SAR Application Toolbox will inform the user whether a query compound either already has been assessed as part of a category or whether it can be associated with an existing category e.g. a category assessed within the OECD HPV Chemicals Programme or a category defined within the new chemical notification scheme and the HPVC challenge programme of the US-EPA. If the compound of interest does not belong to, or cannot reasonably be associated with, any existing category, a similarity assessment is performed (Step 5b).

Step 5b. Similarity assessment

If it is not possible to associate the compound to any existing category, similar compounds can be identified by performing a similarity assessment procedure (pair-wise similarity or similarity to a group). In fact, it is always helpful to perform a search for analogues (even if the chemical can be associated with existing category) since new and valuable information could be obtained. This step may lead to the identification of multiple analogues which might form the basis of a new category. Tools to identify analogues are:

- Analog Identification Methodology (AIM) (under development by US-EPA)
- AMBIT (Ideaconsult Ltd)
- Danish (Q)SAR Database
- ChemFinder
- ChemID Plus Advanced
- Leadscope (commercial)
- Superfragment (under development by BioByte Inc.)
- Toxmatch (under development by ECB)

In one type of grouping (which may be called *descriptor-based grouping*), the structural similarities of the analogues can be explored by means of statistical approaches such as Principal Component analysis (PCA) or pattern recognition approaches (e.g. Kohonen neural maps). Firstly a wide array of descriptors is generated (constitutional, topological, and geometrical descriptors, molecular connectivity indices, physico-chemical properties) for all the analogues; secondly a suitable plot (e.g. PCA plot) is obtained to visualise similarities, trends and possible outliers.

A second type of grouping (which may be called *endpoint-based grouping*) makes use of different QSAR predictions that can be generated for all the analogues and endpoints of interest. This information can be employed to predict trends as well as breakpoints in trends, and therefore possible subcategories. As far as possible, the predictions and trends established by QSAR methods should be verified by comparison with experimental data.

The (Q)SAR Application Toolbox will contain a number of *descriptor-based grouping* methods which are in the public domain. It will also allow the user to perform *endpoint-based grouping*, both with the experimental results from the resident database as well as estimated results from (Q)SAR models which reside in the (Q)SAR library. More guidance for the application of these two types of grouping is needed.

The similarity assessment procedure may lead to the development of a new category whose details can be conveniently stored in database for future use. The reasoning behind the formation of the

category is an important piece of information which should also be included in the appropriate reporting format and stored in the database.

Collect information for analogues and update working matrix

Experimental data for relevant analogues are collected as necessary and stored in the working matrix to be used in the subsequent read-across approach. Toxicological information on the analogues can be obtained from the available in-house databases, such as the (Q)SAR Prediction Database, and from querying external databases.

Perform read-across and update working matrix

Endpoint information for the query compound can be obtained using the corresponding information for relevant analogues. The Working Matrix is updated with the results from the read-across and the suitability of the analogues is documented in the appropriate reporting format.

If the read-across approach was not successful in providing relevant and reliable estimates for the query compound, it might be useful to expand the search for analogues. The analogue searching could be expanded by using the same query compound or by selecting an additional query compounds.

R.6.1.7.9 Step 6 – (Q)SAR

In this step, predictions for toxicity/fate/uptake are generated by using (Q)SAR models or expert systems that incorporate such models.

Retrieve available estimates for endpoint of interest

Estimates for the query compound can be directly retrieved from the (Q)SAR Prediction Database, along with the appropriate reporting formats - (Q)SAR Prediction Reporting Formats and the (Q)SAR Model Reporting Formats, respectively. It is important that predictions from valid and applicable models are selected.

Consult the (Q)SAR model inventory to identify relevant (Q)SARs

If the (Q)SAR Prediction Database does not include predictions for the query compound, relevant (Q)SAR models can be searched in the (Q)SAR Model Database. The information gathered in the previous steps (e.g. information from the classification schemes) may be useful for selecting a suitable model, which is crucial for assessing the reliability of the QSAR result. If valid (Q)SAR model(s) can be found, it is important to verify whether the query compound falls within the applicability domain of the model(s). This assessment may be performed by using appropriate tools (e.g. AmbitDiscovery).

Consult other sources to identify relevant (Q)SARs

If relevant (Q)SAR models for the query compound cannot be obtained from the in-house facilities, namely the (Q)SAR Model Database and the (Q)SAR Prediction Database (or the (Q)SAR Application Toolbox), other models can be searched in the literature, external databases and tools or by consulting experts. If the query compound falls within the applicability domain of the model, meaning that the model is likely to generate a reliable prediction for the compound, the model details should be documented in the (Q)SAR Model Reporting Format.

How much effort should be devoted for finding relevant models for the query compound will depend on the importance of the query compound itself in the final assessment of the fate/toxicity/uptake of the parent compound.

Relevant (Q)SAR models are used to generate predictions, which are used to ,update the Working Matrix, and their application to the query compound is documented by means the (Q)SAR Prediction Reporting Formats. If necessary, additional (Q)SAR Prediction Reporting Formats are compiled and added to the QMDB.

If relevant and reliable predictions cannot be generated for the query compound, it might be useful to perform an additional search for relevant (Q)SARs. When more than one relevant and reliable prediction is available (i.e. more that one adequate model has been found), a judgment of the relative reliability of the different predictions might be necessary if the predictions do not lead to the same conclusions.

R.6.1.7.10 Step 7 – Overall assessment

In the final step, expert judgement is used to reach an overall assessment of the outcome of Steps 1-6 for the chemical and endpoint(s) of interest. The toxicity of the parent compound is assessed using the information obtained for all the query compounds (metabolites, reaction products, analogues).

The overall assessment should make use of all the available information (testing and non-testing data). Decision analysis tools, based on decision theory, might be useful to evaluate multiple options and to help the user reach the best decision.

The overall assessment should also take into account information on the validity of the different models applied within the workflow.

There is still relatively little experience with this type of data integration, and further research into the application of decision analysis methods is required before detailed guidance can be provided.

R.6.1.8 Computational tools for applying (Q)SARs

A wide variety of publicly available and commercial computational tools have been developed, or are under development, that are suitable for the development and application of (Q)SARs. Such tools include methods for a range of (Q)SAR-related tasks, including data management and data mining, descriptor generation, molecular similarity analysis, analogue searching, and hazard assessment.

Among these tools, QSAR-based expert systems enable predictions of chemical toxicity to be obtained directly from chemical structure. All are built upon some experimental toxicity data with rules derived from the data (Dearden, 2003; Dearden *et al*, 1997). The rules are based on expert judgment (e.g. SARs describing reactive chemistry) and/or statistical induction (e.g. QSARs). Examples of QSAR rule-based systems include TOPKAT and MCASE. Knowledge-based systems include Derek, OncoLogic® and HazardExpert, whereas other systems, such as TIMES and ECOSAR, are hybrids.

This chapter provides an overview of some of the better known computational tools for (Q)SAR analysis. As with any (Q)SAR model, if it is intended to use a given tool as a standalone replacement for experimental data, the underlying model should be characterised according to the OECD validation principles (Section R 6.1.3.) and documented by using the appropriate reporting formats (Section R.6.1.6). QSAR reporting formats for selected models are available from JRC QSAR Model Database at http://qsardb.jrc.it.

R.6.1.8.1 Molecular descriptors

Molecular descriptors play a fundamental role in computational chemistry. They are used to describe different features of chemicals, to compare different chemicals structures or different conformations of the same chemical, and relate structure to activity (i.e. develop QSARs). In QSARs, molecular descriptors are used as the independent variables that are used to predict a dependent variable (e.g. an endpoint of regulatory interest). If relevant descriptors for an endpoint are identified, these can also be used to support the adequacy of a read-across for that endpoint (Section R.6.2.2.1 and section R.6.2.3.1).

A molecular descriptor has been defined as (Todeschini and Consonni, 2000): "the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment to measure a molecular attribute".

In this definition, *useful* can be taken to mean that the number provides insight into the *interpretation* of the molecular properties *and/or* is able to take part in a QSAR model for the *prediction* of some other property.

Stated more simply, a molecular descriptor provides 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 a measurable property.

A number of molecular descriptors have been proposed in recent years which have been derived from different theories and approaches to predict the physico-chemical and biological properties of molecules. The information content of a molecular descriptor depends on the molecular representation used and on the defined algorithm for its calculation. The following classification is often used:

- 0D (zero dimensional)
- 1D (mono dimensional)
- 2D (two dimensional)
- 3D (three dimensional)

Zero-dimensional (0D) descriptors are the most simple and commonly used descriptors, reflecting the molecular composition of a compound and derived by counting atom-types or bonds. Examples of these descriptors are molecular weight, atomic composition indices and atomic count descriptors. They are easily and rapidly computed, and as a consequence are often used in database screening. However, they are prone to high degeneracy (i.e. equal values for different molecules), and they are not able to differentiate isomers or chirality.

One-dimensional (1D) descriptors are simple descriptors derived by counting structural fragments in the molecule. They are also used in database searching although a limitation is that they provide *local* information, i.e. do not take into account possible interactions between structural fragments.

Two-dimensional (2D) descriptors comprise a wide variety of descriptors computed by many different methods. They are derived from algorithms applied to a topological representation of the molecule (molecular graph) and are therefore sometimes called topological descriptors. They are obtained by applying algebraic operators to matrices representing molecular graphs and their values are independent of vertex numbering or labelling. Examples of these descriptors include the Balaban index, Zagreb index, Gutman Molecular Topological Index, Wiener W index, Kier symmetry index, Randic shape index, 2D Petitjean shape index. An advantage of these descriptors

is that they can be rapidly derived from SMILES representations, since they do not need optimised structures. They are sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode information concerning atom type and bond multiplicity, and they can differentiate isomers. A possible disadvantage is that the interpretation of some of these descriptors is not always apparent, but many correlate strongly with molecular properties such as volume or surface area.

Geometrical or three-dimensional (3D) descriptors comprise a more complex class of molecular descriptors. These are derived from geometrical representations, i.e. involving knowledge about the relative positions of the atoms in 3D space, i.e. the atomic (x,y,z) coordinates of the atoms. Geometrical descriptors provide more information and discriminatory power. Examples include surface area parameters, 3D-Wiener index, 3D-Balaban index, average geometric distance degree, gravitational indices, WHIM descriptors, GETAWAY descriptors. Despite their high information content, there are some drawbacks in using geometrical descriptors. They require geometry-optimised structures (using, for example, MDL mol files, Hyperchem files, SDF, Sybyl Mol2 files, MacroModel files) and therefore a transparently described means of deriving them. Furthermore, for flexible molecules, several molecular conformations may be available, which results in new information that can be exploited but with added complexity. Most geometrical descriptors need alignment rules in order to achieve molecule comparability. For these reasons, these descriptors do not lend themselves to rapid database screening. Their main utility lies in searching for relationships between molecular structures and receptor-mediated biological activities.

A number of commercial software programs have been developed for the calculation of the molecular descriptors. Some of them are provided in Table R.6-1, in alphabetic order.

Table R.6-1 Commonly used software packages used for the calculation of molecular descriptors

Software	Description
Accord for Excel Accelrys Inc., San Diego, CA, USA	A tool which uses Accord Chemistry Engine to handle chemical structures and incorporates a number of add-ins to perform chemical calculations based on the structure of a compound in a record. http://www.accelrys.com/products/accord
ADAPT Prof. P.C. Jurs, PennState University, University Park, PA 16802, USA	A QSAR toolkit with descriptor generation (topological, geometrical, electronic, and physico-chemical descriptors), variable selection, regression and artificial neural network modelling. http://zeus.chem.psu.edu
CODESSA Semichem Inc. – 7204 Mullen, Shawnee, KS 66216, USA	Calculation of several topological, geometrical, constitutional, thermodynamic, electrostatic, and quantum-chemical descriptors, including tools for regression modelling and variable selection. http://www.semichem.com
DRAGON Talete srl, via Pisani 13, 20124 Milano, Italy	Calculation of several sets of molecular descriptors from molecular geometries (topological, geometrical, WHIM, 3D-MoRSE, molecular profiles). http://www.talete.mi.it
GRIN/GRID Molecular Discovery Ltd. – West Way House, Elms Parade, Oxford OX2 9LL, UK	Calculates the GRID empirical force field at grid points.
HYBOT-PLUS Prof. O. Raevsky – Russian Academy of Science, IPAC.	Calculation of hydrogen bond and free energy factors. http://www.ipac.ac.ru/qsar/index.htm
MOLCONN-Z Prof. L.H. Hall – 2 Davis Street, Quincy, MA 02170, USA	Successor to MOLCONN-X, MOLCONN-Z calculates the most known topological descriptors, including electrotopological and orthogonalised indices. Last release: 3.0. http://www.eslc.vabiotech.com/molconn/manuals/310s/preface1.html
OASIS Laboratory of Mathematical Chemistry. Prof. O. Mekenyan – Bourgas University, 8010 Bourgas, Bulgaria	Calculation of steric, electronic, and hydrophobic descriptors. http://www.oasis-lmc.org
POLLY Prof. S. Basak - University of Minnesota, 5013 Miller Trunk Highway, Duluth, MN 55811, USA	Calculation of topological connectivity indices.
SYBYL/QSAR Tripos Inc. – 1699 South Hanley Rd., St.Louis, MO 63144-2913, USA	SYBYL module for the calculation of EVA descriptors, CoMFA and CoMSIA fields, also including several QSAR tools. http://www.tripos.com
TSAR Accelrys Inc., San Diego, CA, USA (formerly Oxford Molecular Ltd, UK)	Statistical and database functions with molecular and substituent property calculations. Within TSAR 3D package. http://www.accelrys.com

The following chapters provide a brief overview of various computational tools/databases that are either publicly or commercially available. Guidance on the use and interpretation of individual tools for the regulatory assessment of specific endpoints is outside the scope of this document.

R.6.1.8.2 Computational tools developed by CEFIC

Ambit

Ambit (http://ambit.acad.bg) is freely available software for data management and QSAR applications, including searchable databases and tools for grouping and applicability domain assessment. The AMBIT database stores chemical structures, their identifiers such as CAS, INChI numbers, attributes such as molecular descriptors, experimental data together with test descriptions, and literature references. The database can also store QSAR models. In addition, the software can generate a suite of 2D and 3D molecular descriptors. Search options include searching by name, CAS number, SMILES, substructure and structure-based similarity, and by chemical identifier (experimental property, molecular descriptor) ranges. AMBIT Discovery performs chemical grouping and assesses the applicability domain of a QSAR by offering a variety of methods: statistical methods that rely on descriptor space; approaches based on mechanistic understanding such as the Verhaar and Cramer classification schemes; and several approaches based on structural similarity. Ambit was developed jointly by Ideaconsult Ltd (Sofia, Bulgaria) and Procter & Gamble (Dr Joanna Jaworska) with funding from the CEFIC LRI project.

R.6.1.8.3 Computational tools developed/implemented by the ECB

Danish QSAR database

The Danish Environmental Protection Agency (EPA) constructed a database of (Q)SAR predictions made by some 70 models for about 166,000 organic chemicals for a wide range of different endpoints. An internet-accessible version of downsized version of this database (as of 1.1.2004) is available at http://ecbqsar.jrc.it. It contains around 60 model predictions for each chemical including a simple yes/no statement on all MultiCase predictions in relation to whether it is inside or outside the applicability domain of the model. Different types of searching are possible including structure (substructure/exact match) searching, ID (CAS number, name) searching and parameter (endpoint) searching. The (Q)SAR models encompass endpoints for physico-chemical properties, fate, eco-toxicity, absorption, metabolism and toxicity.

Toxtree

Toxtree, developed by Ideaconsult Ltd under contract to ECB, is a freely available application (http://ecb.jrc.it/qsar/qsar-tools) which is able to estimate different types of toxic hazard by applying structural rules. Toxtree includes options for applying the Verhaar scheme (Verhaar *et al*, 1995) and the Cramer decision tree (Cramer *et al*, 1978).

The Cramer classification scheme (tree) is probably the best known approach for structuring chemicals in order to make an estimation of the Threshold of Toxicological Concern (Cramer *et al*, 1978). The tree relies primarily on chemical structures and estimates of total human intake to establish priorities for testing. The procedure uses recognised pathways for metabolic deactivation and activation, toxicity data and the presence of a substance as a component of traditional foods or as an endogenous metabolite. Substances are classified into one of three classes:

Class 1 contains substances of simple chemical structure with known metabolic pathways and innocuous end products which suggest a low order of oral toxicity;

Class 2 contains substances that are intermediate. They possess structures that are less innocuous than those in Class 1 but they do not contain structural features that are suggestive of toxicity like those in Class 3;

Class 3 contains substances with structures that permit no strong initial impression of safety and may even suggest a significant toxicity.

The Verhaar scheme for acute toxicity to aquatic organisms (mainly fish) is a widely used scheme for assigning the mode of action of chemicals into four groups: non-polar narcotics, polar narcotics, reactive chemicals and specifically-acting chemicals (Verhaar *et al*, 1995).

Toxmatch

Chemical similarity is often perceived as a structural similarity but there are a number of methods of characterising chemicals in terms of their physico-chemical, topological, geometrical, and surface properties and these numerical representations lend themselves to comparisons using so-called similarity indices. Thus, the similarity indices based on the presence of substructures represent a special case of similarity indices. The ECB has commissioned the development of a software tool, Toxmatch, that will encode different types of similarity measures which can be used to facilitate the development of generic and endpoint-specific categories. The tool will include a functionality to facilitate read-across, as well as to compare chemicals of interest with existing categories. A prototype is being developed using several endpoints, including skin sensitisation, skin irritation, aquatic toxicity and bioaccumulation.

JRC QSAR Model Database

The JRC QSAR Model Database, which is currently under development will be a searchable tool for linking chemicals of interest to a collection of robust summaries of (Q)SAR models. The summaries are being compiled by using a standard (Q)SAR Model Reporting Format (QMRF). A database with a web-based interface will be implemented to allow on-line access to the JRC QSAR Model Database at http://qsardb.jrc.it Different search options will be possible, such as by chemical (CAS or EC number, structure), endpoint, descriptors, and model author.

DART

DART (Decision Analysis by Ranking Techniques), developed by Talete srl (Milan, Italy) under ECB contract, is a user-friendly tool to support the priority setting of chemicals for risk assessment. Different kinds of ranking (sorting) methods, roughly divided into total order (or even-scoring) and partial order ranking methods (Hasse diagram technique), are implemented in DART. The ranking methods can be applied to experimental and/or estimated data.

These methods can be used to rank chemicals on the basis of more than one criterion. In the case of total order ranking methods the different criteria values are combined into a global ranking index and chemicals are ordered sequentially according to the numerical value of the ranking index. This requires the transformation of each criterion (variable) independently by using an arbitrary function that transforms the actual value of each chemical into a value between 0 and 1. A total of 19 different kinds of functions are implemented in DART to allow the user to explicitly define the *best* condition for each criterion used in the ranking process. A weighting scheme is also implemented. All the algorithms can be applied in the presence of missing data. The tool allows the user to derive basic statistics and provides a user-friendly graphical user interface for visualising the results of the analyses. More advanced features include the possibility to calculate several ranking indices for degeneracy, stability, and discrimination power.

R.6.1.8.4 Computational tools in development by the OECD

QSAR Application Toolbox

The OECD QSAR Application Toolbox, for which a pilot version is currently under development, will be an application linking a number of existing tools as well as a library of existing (Q)SAR models which will allow a user to:

- Make estimations for single chemicals, and receive the results of all the (Q)SAR estimates for all the models covering the appropriate domain, for the relevant endpoints that the user wishes to estimate.
- Receive summary information on the validation results of the model according to the OECD validation principles so that the user can decide for which regulatory purpose the estimate can be used. The (Q)SAR models would be incorporated into the toolbox as they come forward from member countries with the information on their validation according to the OECD Principles.
- Receive a list of analogues, together with their (Q)SAR estimates.
- Receive estimates for metabolite activation/detoxification information.

The Toolbox will link a number of public domain tools which are described in this chapter and make them available to the user according to a flexible workflow which is currently being discussed at OECD level. The user will be able to use the Toolbox to implement the stepwise approaches described in Section R.6.2.3. and section R.6.2.4.

R.6.1.8.5 Computational tools developed by the US-EPA

AIM

The Analog Identification Methodology (AIM) has been developed by the U.S. Environmental Protection Agency to facilitate read-across and chemical grouping by identifying chemical analogues that have existing test data publicly available. AIM is a web-based, computerised tool that identifies chemical analogues based on structure. The tool also provides the user with pointers or links to publicly available experimental data on the closely related chemical(s).

AIM identifies chemical analogues from a default database that currently contains 31,031 compounds that have some type of toxicity data publicly available. AIM employs a fragment-based search method to identify analogous compounds using a set of 645 pre-defined fragments and correction factors, and a "three-pass" searching strategy to locate structures through defined rules and allowable substitution patterns for different types of structural features. AIM can be searched on the basis of structure, SMILES or CAS number, though it cannot be searched by chemical name.

The tool provides a simple means of identifying analogues that have some kind of toxicity data available, but it does not categorise or rank the analogues returned. This approach leaves it to individual users need to determine when a specific analogue is suitable for a specific assessment, as the determination of what structure is 'appropriate' can vary depending on the endpoint assessed.

The available test data is accessed in the form of hyperlink pointers. The data is not structured in any way and cannot be downloaded into Excel or other tools for analyses. Some hyperlinks point to a general webpage, e.g. IUCLID homepage or RTECS homepage, so the user may need the appropriate licenses to be able to extract available information. Other links take the user directly to

the data source. Thus, the pointer informs that there is a record for the chemical, but does not always indicate the specific type of data available.

AIM allows users to rapidly categorise multiple chemicals, focus available resources, facilitate read-across, and streamline assessment exercises.

OncoLogic®

The Cancer Expert System or OncoLogic® is an expert system that assesses the potential of chemicals to cause cancer. OncoLogic® was developed under a cooperative agreement between the EPA's Office of Pollution Prevention and Toxics (OPPT) and LogiChem, Inc. It predicts the potential carcinogenicity of chemicals by applying the rules of SAR analysis and incorporating what is known about the mechanisms of action and human epidemiological studies. OncoLogic® has the ability to reveal its line of reasoning just as human experts can. After supplying the appropriate information about the structure of the compound, an assessment of the potential carcinogenicity and the scientific line of reasoning used to arrive at the assessment outcome are produced. This information provides a detailed justification of a chemical cancer causing potential. The Cancer Expert System is comprised of four subsystems that evaluate fibres, metals, polymers, and organic chemicals of diverse chemical structures. The OncoLogic® Cancer Expert System was previously distributed exclusively by LogiChem, Inc. The US-EPA has recently purchased the right to the system and is currently updating the system for free distribution to the public (available by contacting Dr Yin-tak Woo; email: woo.yintak@epa.gov).

ECOSAR

ECOSAR uses a number of class-specific log K_{ow}-based QSARs in order to predict the toxicity of chemicals to aquatic organisms (fish, daphnids, green algae). The QSARs are developed for chemical classes based on measured test data that have been submitted by industry to the US Environmental Protection Agency (US-EPA). ECOSAR produces warnings in several occasions (e.g. when the water solubility is very low, or when the prediction is outside the range of log K_{ow} in the training set). The software is freely available from the US-EPA (downloadable from http://www.epa.gov/oppt/newchems/tools/21ecosar.htm).

EPI Suite

The EPI (estimation program interface) Suite program integrates a number of estimation models for the prediction of environmental and physical/ chemical properties in one convenient interface. EPI Suite is freely available from the US-EPA website (http://www.epa.gov/oppt/exposure/pubs/episuite.htm). These models include K_{ow}Win (for estimating log K_{ow}), AopWin (for predicting gas-phase reaction rates), HenryWin (for Henry's Law constant), MPBPVP (for predicting melting point, boiling point, and vapour pressure), WsKow (for estimating water solubility and log K_{ow}), Hydro (for estimating hydrolysis rate constants for specific organic classes), DermWin (for estimating the dermal permeability coefficient (K_p)), ECOSAR (described above) and BCFWin (for estimating the bioconcentration factor). EPI Suite also estimates a chemical's rate of volatilisation from a model river and lake to the atmosphere as well as its expected fate in a sewage treatment plant and level III fugacity model.

Commercially available tools

A wide range of commercially available software tools are available, of which a few are described below. Some of the available tools have been evaluated by an ECETOC Task Force (ECETOC, 2003).

Leadscope®

Leadscope® is a data management and data mining tool developed and commercialised by Leadscope Inc. (http://www.leadscope.com). It is possible to import additional datasets and perform comparisons with existing databases on the basis of the 27,000 chemical fingerprints. A number of statistical algorithms are also embedded to enable functionalities such as clustering of chemicals and data, extraction of structural rules, development of QSAR models as well as development of chemical categories. Leadscope possesses a unique chemical hierarchy containing over 27,000 chemical fingerprints which represent functional groups, chemical groupings and pharmacophores. The software can be purchased with a toxicity database and/or known drugs database. The toxicity database contains integrated information on over 160,000 chemical structures from multiple sources including the FDA PAFA Database, the US National Toxicology Program (NTP), RTECS®, and the DSSTox Carcinogenicity Potency Database (CPDB). The database covers a range of endpoints including acute and multiple dose studies, such as subchronic liver, carcinogenicity, genetic toxicity, reproductive and irritation. The database can be searched by structure (such as substructure or similarity), type of study, toxic effect, species, sex, dosage, duration and route of exposure. Results can be viewed and exported in convenient formats, such as Excel files.

Derek

Derek is a knowledge-based expert system created with knowledge of structure-toxicity relationships and an emphasis on the need to understand mechanisms of action and metabolism (EC, 2001; Sanderson and Earnshaw, 1991; Judson, 2002). It is marketed and developed by Lhasa Ltd, a not-for-profit company and educational charity (http://www.lhasalimited.org).

Within Derek, there are over 504 alerts covering a wide range of toxicological endpoints. An alert consists of a toxicophore (a substructure known or thought to be responsible for the toxicity) and is associated with literature references, comments and examples. The Derek knowledge base covers a broad range of toxicological endpoints, but its main strengths lie in the areas of mutagenicity, carcinogenicity and skin sensitisation. All the rules in Derek are based either on hypotheses relating to mechanisms of action of a chemical class or on observed empirical relationships. Information used in the development of rules includes published data and suggestions from toxicological experts in industry, regulatory bodies and academia. The toxicity predictions are the result of two processes. The program first checks whether any alerts in the knowledge base match toxicophores in the query structure. The reasoning engine then assesses the likelihood of a structure being toxic. There are nine levels of confidence: certain, probable, plausible, equivocal, doubted, improbably, impossible, open, contradicted. The reasoning model considers the following information: a) the toxicological endpoint; b) the alerts that match toxicophores in the query structure; c) the physico-chemical property values calculated for the query structure; and d) the presence of an exact match between the query structure and a supporting example within the knowledge base.

A further application of Derek is its integration with the Meteor system to enable predictions of toxicity for both parent and metabolites.

HazardExpert

HazardExpert is a module of Pallas software developed by CompuDrug Limited (http://www.compudrug.com). Along with toxicity predictions it can also consider the bioavailability and bioaccumulation of the compounds by calculation of logP and pKa. The default knowledge base of the system is based on the report of US-EPA (Brink and Walker, 1987) and the scientific information collected by CompuDrug Limited. The rule-based system of the program has open architecture, allowing the user to understand, expand, modify or optimise the data on which the toxicity estimation relies. It covers the following endpoints: oncogenicity, mutagenicity,

teratogenicity, membrane irritation, sensitisation, immunotoxicity, neurotoxicity. A further application of the program is prediction the toxicity of the parent compound and its metabolites by link with MetabolExpert system (another module of Pallas software).

TOPKAT

TOPKAT is a statistical system developed by Accelrys, Inc (http://www.accelrys.com) consisting of a suite of QSAR models for a range of different endpoints. There are currently 16 modules 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, logKow, rabbit skin irritancy, rat chronic LOAEL, rat inhalation toxicity LC_{50}, rat Maximum Tolerated Dose (MTD), rat oral LD_{50}, skin sensitisation, and *Weight of Evidence* rodent carcinogenicity (Cronin *et al*, 2003).

TOPKAT models are typically based on the analysis of large datasets of toxicological information derived from the literature. The molecular descriptors used include structural (e.g. molecular bulk, shape, symmetry), topological and electrotopological indices. The QSARs are developed by regression analysis for continuous endpoints and by discriminant analysis for categorical data (Enslein, 1988). TOPKAT also works in batch mode. It estimates the confidence in the prediction by applying the patented Optimal Predictive Space (OPS) validation method. The OPS is unique multivariate descriptor space in which the model is applicable. When a query is within the OPS for a given model, the probability of the prediction to be accurate is as good as the cross-validated statistical performance of the model.

The CASE family of methods

The CASE methodology and all its variants have been developed by Klopman and Rosenkranz (http://www.multicase.com). There are a multitude of models for a variety of endpoints and hardware platforms. There are many forms of the CASE models, and the software is variously called CASE, MULTICASE, MCASE, CASETOX and TOXALERT, depending on the endpoint and the hardware platform.

The CASE approach uses a fragment based technology (Klopman and Chakravarti, 2003). It is based on a hierarchical statistical analysis of a database composed of a number of chemicals associated with their toxicity data. The program discovers substructures that appear mostly in active molecules thus being with high probability responsible for the observed activity. At the beginning it identifies the statistically most significant substructure within the training set. This fragment, labelled the top biophore, is seen responsible for the activity of the largest possible number of active molecules. The active molecules containing this biophore are then removed from the database, and the remaining ones are submitted to a new analysis for identification of the next biophore. The procedure is repeated until either the activity of all the molecules in the training set has been accounted for or no additional statistically significant substructure can be found. Then for each set of molecules containing a specific biophore, the program identifies additional parameters called modulators, which can be used to derive QSAR within the reduced set of congeneric molecules. The modulators consist of certain substructures or physico-chemical parameters, such as HOMO/LUMO energies, logP, water solubility, location of hydrogen donors/acceptors, lipophilic centres with respect to biophore, etc, that significantly enhance or diminish the activity attributable to the biophore. QSARs are then performed with these modulators. The knowledge that the program gains during the training process can be then used to predict the biological activity of new chemicals not included in the training set (ECETOC, 2003). The program covers a range of endpoints, including carcinogenicity, mutagenicity, teratogenicity, irritation, developmental toxicity, acute toxicity, biodegradation.

TIMES

The Tissue MEtabolism Simulator (TIMES) integrates on the same platform metabolic simulators (see Section R.6.1.8.6.) and QSAR models for predicting toxicity of selected metabolites. The TIMES platform has been used to predict skin sensitisation, mutagenicity, chromosomal aberration and ER/AR binding affinities of chemicals, while accounting for metabolic activation (http://www.multicase.com). Recently, it has incorporated models to predict the toxicity to aquatic species (OASIS/TIMES). OASIS/TIMES uses an approach for modelling acute toxicity for two types of toxicochemical domains: reversible (non-covalent) acting chemicals and irreversible covalent bioreactive chemicals.

TerraQSAR™

TerraQSAR™ (http://www.terrabase-inc.com) is a collection of computation programs for the prediction of biological effects and physico-chemical properties of organic compounds. The available models developed using a probabilistic neural network (PNN) methodology include: DM 24 hr EC_{50} for *Daphnia magna*, E2-RBA estrogen receptor binding affinity (RBA), FHM 96-h LC_{50} for fathead minnow (*Pimephales promelas*), log P octanol/water partition coefficient, OMAR mouse and rat oral LD_{50}, RMIV rat and mouse *intravenous* LD_{50} as well as SKIN a skin irritation potential model.

R.6.1.8.6 Tools and databases to help in the assessment of metabolism

A variety of databases and software tools have been developed to help in the assessment of metabolism. Some of these are highlighted in the following paragraphs. For more detailed information, literature reviews are available (Payne, 2004). Guidance on the use and interpretation of these tools is outside the scope of this document.

COMPACT

The computer-optimised molecular parametric analysis of chemical toxicity (COMPACT) system was developed at the University of Surrey (UK) by Lewis and co-workers (Lewis, 2001 and 2003). COMPACT has modules that assess the ability of xenobiotics to form enzyme substrates complexes and undergo metabolic activation by the CYP1A and CYP2E subfamilies of cytochrome P450s. The system is used mainly in-house by the group at Surrey University, and is not commercially or publicly available.

META

The META system is a commercially available tool developed by Klopman and Tu (1999) at Case Western Reserve University (OH, USA). It is an expert system capable of predicting the sites of potential enzymatic attack and the nature of the chemicals formed by such metabolic transformations. The program uses dictionaries of biotransformation operators which are created by experts in the field of xenobiotic metabolism to represent known metabolic pathways. A query structure is entered and the program applies biotransformation operators according to the functional groups detected. After each biotransformation a stability check is performed on the reaction product by using quantum mechanical calculations to detect unstable atom arrangements. The program then evaluates the stable metabolites formed and attempts to transform them further until water soluble metabolites that are deemed to be excretable are formed.

MetabolExpert

MetabolExpert is a commercially available software product composed of a database, a knowledge base and several prediction tools (Darvas, 1987). The basic biotransformation database contains 179 biotransformations, developed as *if-then* rules derived from the literature by experts.

Meteor

Meteor is a commercially available tool that uses a knowledge-base of structure-metabolism rules to predict the metabolic fate of a query chemical structure. The system is developed and marketed by Lhasa Ltd (Leeds, UK) and evolved from the Derek system for toxicity prediction (Greene *et al*, 1999). Meteor's biotransformation rules are generic reaction descriptors rather than simple entries in a reaction database. To limit over prediction, Meteor has an integrated reasoning engine based on a system of non-numerical argumentation, which uses a repository of higher level reasoning rules. The reasoning model allows the system to evaluate the likelihood of biotransformation taking place and to make comparisons between potentially competing biotransformations. The user can choose to analyse queries at a number of available search levels. At the *high likelihood* level, only the more likely biotransformations are requested for display. The system is also supplied with a knowledge base editor so that users can add their own (proprietary) rules. The metabolic tree can be searched and metabolites of specific molecular mass and or molecular formula highlighted. The generated tree is also structure-searchable. Individual biotransformations can be viewed with generalised graphical descriptions of their scope. It is possible to generate sequences automatically and to generate metabolites from an individually chosen biotransformation. It is possible to search for either phase I or phase II biotransformations only. Additionally, Meteor is provided with a link to ClogP to identify biotransformations that are not likely to occur, due to very low lipophilicity.

CATABOL

CATABOL applies a mechanistic approach for quantitative assessment of biodegradability of chemicals by simulating their biodegradation pathways and predicting physico-chemical and toxic endpoints of stable degradants across biodegradation pathways of the chemicals. The core of CATABOL is the biodegradability simulator including a library of hierarchically ordered individual transformations (abiotic and enzymatic reactions). The catabolic transformations are derived from set of most plausible metabolic pathways predicted by experts for each chemical from the training set. The transformation probabilities are adjusted to best reproduce documented degradation pathways for over 500 chemicals. The current developments of CATABOL are oriented to predicting the extent of biodegradation at different time frames.

TIMES

The Tissue MEtabolism simulator (TIMES) is a heuristic algorithm that aims to produce plausible biotransformation pathways from a query molecule by using rules developed from a comprehensive library of biotransformations and abiotic reactions (Meykenyan *et al*, 2004). The transformation probabilities can be calibrated to specific reference conditions, and the generation of metabolites by TIMES can be limited to the most likely ones or can be extended to include less likely ones and allow predicting the quantity of generated metabolites with consideration of water solubility, log K_{ow} and other physico-chemical properties.

MDL Metabolite

MDL Metabolite (http://mdli.com) is a commercial database containing a browsing interface. The database is the uses information from multiple studies to assemble structural metabolic database entries for particular parent compounds. The focus is on xenobiotic compounds and

biotransformations of medicinal drugs. Experimental data is abstracted from *in vitro* and *in vivo* studies. In addition to structural information, the database contains enzyme information, species information, physiological activity, parent compound toxicity, bioavailability, analytical methodology, route of administration, excretion routes, quantitative and qualitative yield, CAS number of parent compound and references to the original literature.

The Accelrys Biotransformation database

This database, commercially available as a CD ROM from Accelrys Inc, comprises biotransformations of chemical entities, including pharmaceuticals, agrochemicals, food additives and environmental and industrial chemicals. The database is indexed with original citations, test systems and a variety of keywords for generic searching and is fully cross referenced to a series of books (Hawkins, 1996).

KEGG

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) is a freely available bioinformatics resource being developed by Kyoto University and the University of Tokyo (http://www.genome.jp/kegg). The KEGG project was initiated in May 1995, with a view to providing a tool that helps to understand the basic principles and practical applications of the relationships between genomic information and higher order functional information.

KEGG consists of: a) the PATHWAY database providing information on molecular interaction networks such as pathways and complexes; b) the GENES database providing information about genes and proteins generated by genome sequencing projects; c) the LIGAND database providing information about chemical compounds and metabolic pathway information; d) limited amounts of experimental gene expression data in the EXPRESSION and BRITE databases; and e) the SSDB database, containing information about amino acid sequence similarities among all protein-coding genes in the complete genomes.

University of Minnesota Biocatalysis/Biodegradation Database

The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://umbbd.ahc.umn.edu) contains compound, enzyme, reaction and pathway information for microbial catabolism of primarily anthropogenic materials. It has been available on the web for over 10 years, and has grown from 4 to almost 150 pathways. It currently contains information on over 900 compounds, over 600 enzymes, over 1000 reactions and about 350 microorganism entries. Along with pathway data, Biochemical Periodic Tables (http://umbbd.ahc.umn.edu/periodic) and a Biodegradation Pathway Prediction System (PPS) (http://umbbd.ahc.umn.edu/predict) are also available.

R.6.1.9 The QSAR Model Reporting Format (QMRF)

R.6.1.9.1 QMRF – version 1.2[16]

Please try to fill in the fields of the QMRF for the model of interest. If the field is not pertinent with the model you are describing, or if you cannot provide the requested information, please answer "no information available". **The set of information that you provide will be used to facilitate regulatory considerations of (Q)SARs.** For this purpose, the structure of the QMRF is devised to

[16] Version of January 2007. For more information consult the following website: http://ecb.jrc.it/qsar/qsar-tools

reflect as much as possible the OECD principles for the validation, for regulatory purposes, of (Q)SAR models. You are invited to consult the OECD "Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship Models" that can aid you in filling in a number of fields of the QMRF (OECD, 2007).

1.	**QSAR identifier**
1.1.	**QSAR identifier (title):** *Provide a short and indicative title for the model including relevant keyword. Some possible keywords are: endpoint modelled (as specified in field 3.2, recommended), name of the model, name of the modeller, and name of the software coding the model. Examples: "BIOWIN 1 for Biodegradation"; "TOPKAT Skin Irritation Acyclics (Acids, Amines, Esters) MOD v SEV Model".*
1.2.	**Other related models**: *If appropriate, identify any model that is related to the model described in the present QMRF. Example: TOPKAT Skin Irritation Acyclics (Acids, Amines, Esters) NEG/MLD v MOD/SEV Model" is related to the model mentioned in 1.1: "TOPKAT Skin Irritation Acyclics (Acids, Amines, Esters) MOD v SEV Model".*
1.3.	**Software coding the model***: If appropriate, specify the name and the version of the software that implements the model. Examples: "BIOWIN v. 4.2 (EPI Suite)"; "TOPKAT v. 6.2".*
2.	**General Information**
2.1.	**Date of QMRF**: *Report the date of QMRF drafting (month/year). Example: "5 November 2006".*
2.2.	**QMRF author(s) and contact details**: *Indicate the name and the contact details of the author(s) of the QMRF (first version of the QMRF).*
2.3.	**Date of QMRF update(s):** *Indicate the date (day/month/year) of any update of the QMRF. The QMRF can be updated for a number of reasons such as additions of new information (e.g. addition of new validation studies in section 7) and corrections of information.*
2.4.	**QMRF update(s):** *Indicate the name and the contact details of the author(s) of the updates QMRF (see field 2.3) and list which sections and fields have been modified.*
2.5.	**Model developer(s) and contact details:** *Indicate the name of model developer(s)/author(s), and the corresponding contact details; possibly report the contact details of the corresponding author.*
2.6.	**Date of model development and/or publication:** *Report the year of release/publication of the model described in the current QMRF.*
2.7.	**Reference(s) to main scientific papers and/or software package:** *List the main bibliographic references (if any) to original paper(s) explaining the model development and/or software implementation. Any other reference such as references to original experimental data and related models can be reported in field 9.2 "Bibliography"*
2.8	**Availability of information about the model:** *Indicate whether the model is proprietary or non-proprietary and specify (if possible) what kind of information about the model cannot be disclosed or are not available (e.g., training and external validation sets, source code, and algorithm). Example: "The model is non-proprietary but the training and test sets are not available"; "The model is proprietary and the algorithm and the data sets are confidential"*
2.9	**Availability of another QMRF for exactly the same model:** *Indicate if you are aware or suspect that another QMRF is available for the current model you are describing. If possible, identify this other QMRF.*
3.	**Defining the endpoint – OECD Principle 1** *PRINCIPLE 1: "A DEFINED ENDPOINT". ENDPOINT refers to any physico-chemical, biological, or environmental effect that can be measured and therefore modelled. The intent of PRINCIPLE 1 (a (Q)SAR should be associated with a defined endpoint) is to ensure clarity in the endpoint being predicted by a given model, since a given endpoint could be determined by different experimental protocols and under different experimental conditions. It is therefore important to identify the experimental system that is being modelled by the (Q)SAR.*
3.1	**Species:** *Indicate the species for the endpoint being modelled.*
3.2.	**Endpoint**: *Choose the endpoint (physicochemical, biological, or environmental effect) from the pre-defined classification. If the pre-defined classification does not include the endpoint of interest, select "Other" and report the endpoint in the subsequent field 3.3.*
3.3.	**Comment on the endpoint**: *Include in this field any other information to define the endpoint being modelled. Specify the endpoint further if relevant, e.g. according to test organism such as species, strain, sex, age or life stage; according to test duration and protocol; according to the detailed nature of endpoint etc. You can also define here the endpoint of interest in case this is not listed in the pre-defined classification (see field 3.2) or you can add information about a second endpoint modelled by the same model. Example: Nitrate radical degradation rate constant: kNO3.*

3.4.	**Endpoint units**: *Specify the units of the endpoint measured.*
3.5.	**Dependent variable***: Specify the relationship between the dependent variable being modelled and the endpoint measured since the two quantities may be different. Example: For modelling purposes all rate constants (i.e. Nitrate radical degradation rate constant KNO3) were transformed to logarithmic units and multiplied by -1 to obtain positive values. The dependent variable is: -log (kNO3).*
3.6.	**Experimental protocol**: *Make any useful reference to a specific experimental protocol (or protocols) followed in the collection of the experimental datasets.*
3.7.	**Endpoint data quality**: *Provide available information about the test data selection and evaluation and include a description of the data quality used to develop the model. This includes provision of information about the variability of the test data, i.e. repeatability (variability over time) and reproducibility (variability between laboratories) and sources of error (confounding factors which may influence testing results).*
4.	**Defining the algorithm – OECD Principle 2** *PRINCIPLE 2: "AN UNAMBIGUOUS ALGORITHM". The (Q)SAR estimate of an endpoint is the result of applying an ALGORITHM to a set of structural parameters which describe the chemical structure. The intent of PRINCIPLE 2 (a (Q)SAR should be associated with an unambiguous algorithm) is to ensure transparency in the model algorithm that generates predictions of an endpoint from information on chemical structure and/or physico-chemical properties. In this context, algorithm refers to any mathematical equation, decision rule or output from a formalised modelling approach.*
4.1.	**Type of model**: *Specify the type of model (e.g., SAR, QSAR, Expert System, Neural Network, etc.).*
4.2.	**Explicit algorithm**: *Report the algorithm (only the algorithm) for generating predictions from the descriptors; more text information about the algorithm can be reported in the following fields of this section or as supporting information (see field 9.3). If the algorithm is too long and complicated and thus cannot be reported here, include in this field a reference to a paper or a document where the algorithm is described in detail. This material can be attached as supporting information.*
4.3.	**Descriptors in the model**: *Identify the number and the name or identifier of the descriptors included in the model. In this context, descriptors refers to e.g. physicochemical parameters, structural fragments etc*
4.4.	**Descriptor selection**: *Indicate the number and the type (name) of descriptors /decision rules initially screened, and explain the method used to select the descriptors and develop the model from them.*
4.5.	**Algorithm and descriptor generation**: *Explain the approach used to derive the algorithm and the method (approach) used to generate each descriptor.*
4.6.	**Software name and version for descriptor generation**: *Specify the name and the version of the software used to generate the descriptors. If relevant, report the specific settings chosen in the software to generate a descriptor.*
4.7.	**Descriptors/Chemicals ratio**: *Report the following ratio: number of descriptors to number of chemicals (chemicals from the training set), if applicable. If not, explain why.*
5.	**Defining the applicability domain – OECD Principle 3** *PRINCIPLE 3: "A DEFINED DOMAIN OF APPLICABILITY". APPLICABILITY DOMAIN refers to the response and chemical structure space in which the model makes predictions with a given reliability. Ideally the applicability domain should express the structural, physicochemical and response space of the model. The CHEMICAL STRUCTURE (x variable) space can be expressed by information on physicochemical properties and/or structural fragments. The RESPONSE (y variable) can be any physicochemical, biological or environmental effect that is being predicted. According to PRINCIPLE 3 a (Q)SAR should be associated with a defined domain of applicability. Section 5 can be repeated (e.g., 5.a, 5.b, 5.c, etc) as many time as necessary if more than one method has been used to assess the applicability domain.*
5.1.	**Description of the applicability domain of the model**: *Describe the response and chemical structure and/or descriptor space in which the model makes predictions with a given reliability. Discuss if relevant whether: a) fixed or probabilistic boundaries define the applicability domain; b) structural features, a descriptor or a response space defines the applicability domain; c) in the case of SAR, there exists a description of the limits on its applicability (inclusion and/or exclusion rules regarding the chemical classes to which the substructure is applicable); d) in the case of SAR, there exist rules describing the modularity effects of the substructure's molecular environment; e) in the case of QSAR, there exist inclusion and/or exclusion rules that define the descriptor variable ranges for which the QSAR is applicable; f) in the case of QSAR, there exist inclusion and/or exclusion rules that define the response variable ranges for which the QSAR is applicable; g) there*

	exists a (graphical) expression of how the descriptor values of the chemicals in the training set are distributed in relation to the endpoint values predicted by the model.
5.2.	**Method used to assess the applicability domain**: *Describe the method used to assess the applicability domain of the model.*
5.3.	**Software name and version for applicability domain assessment**: *Specify the name and the version of the software used to apply the applicability domain method, where applicable. If relevant, report the specific settings chosen in the software to apply the method.*
5.4.	**Limits of applicability**: *Describe for example the inclusion and/or exclusion rules (fixed or probabilistic boundaries, structural features, response space) that define the applicability domain.*
6.	**Defining goodness-of-fit and robustness – OECD Principle 4** *PRINCIPLE 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY". PRINCIPLE 4 expresses the need to perform validation to establish the performance of the model. GOODNESS-OF-FIT and ROBUSTNESS refer to the internal model performance.*
6.1.	**Availability of the training set**: *Indicate whether the training set is somehow available (e.g., published in a paper, embedded in the software implementing the model, stored in a database) and appended to the current QMRF as supporting information (field 9.3). If it is not available, explain why. Example: "It is available and attached" "It is available but not attached"; "It is not available because the data set is proprietary"; "The data set could not be retrieved".*
6.2.	**Available information for the training set**: *Indicate whether the following information for the training set is reported as supporting information (see field 9.3): a) Chemical names (common names and/or IUPAC names); b) CAS numbers; c) SMILES; d) InChI codes; e) MOL files; f) Structural formula; g) Any other structural information.*
6.3.	**Data for each descriptor variable for the training set**: *Indicate whether the descriptor values of the training set are available and are attached as supporting information (see field 9.3).*
6.4.	**Data for the dependent variable (response) for the training set**: *Indicate whether dependent variable values of the training set are available and attached as supporting information (see field 9.3).*
6.5.	**Other information about the training set**: *Indicate any other relevant information about the training set (e.g. number and type of compounds in the training set (e.g. for models predicting positive and negative results the number of positives and the number of negatives in the training set)).*
6.6.	**Pre-processing of data before modelling**: *Indicate whether raw data have been processed before modelling (e.g. averaging of replicate values); if yes, report whether both raw data and processed data are given.*
6.7.	**Statistics for goodness-of-fit**: *Report here goodness-of-fit statistics (r^2, r^2 adjusted, standard error, sensitivity, specificity, false negatives, false positives, predictive values etc).*
6.8.	**Robustness – Statistics obtained by leave-one-out cross-validation**: *Report here the corresponding statistics.*
6.9.	**Robustness – Statistics obtained by leave-many-out cross-validation**: *Report here the corresponding statistics, the strategy for splitting the dataset (e.g. random), the percentage of left out compounds and the number of cross-validations.*
6.10.	**Robustness – Statistics obtained by Y-scrambling**: *Report here the corresponding statistics and the number of iterations.*
6.11.	**Robustness – Statistics obtained by bootstrap**: *Report here the corresponding statistics and the number of iterations.*
6.12.	**Robustness – Statistics obtained by other methods**: *Report here the corresponding statistics.*
7.	**Defining predictivity – OECD Principle 4** *PRINCIPLE 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY". PRINCIPLE 4 expresses the need to perform validation to establish the performance of the model. PREDICTIVITY refers to the external model validation. Section 7 can be repeated (e.g., 7.a, 7.b, 7.c, etc) as many time as necessary if more validation studies needs to be reported in the QMRF.*
7.1.	**Available information for the external validation set**: *Indicate whether an external validation set is available and appended to the current QMRF as supporting information (field 9.3). If it is not available,*

	explain why.
7.2.	**Data for each descriptor variable for the external validation set**: *Indicate whether the following information for the external validation set is reported as supporting information (see field 9.3): a) Chemical names (common names and/or IUPAC names); b) CAS numbers; c) SMILES; d) InChI codes; e) MOL files; f) Structural formula; g) Any other structural information.*
7.3.	**Data for the dependent variable for the external validation set**: *Indicate whether descriptor values of the external validation set are somehow available and attached as supporting information (see field 9.3).*
7.4.	**Other information about the external validation set**: **Data for the dependent variable for the external validation set:** *Indicate whether dependent variable values of the external validation set are somehow available and attached as supporting information (see field 9.3).*
7.5.	**Other information about the external validation set:** *Indicate any other relevant information about the validation set. Example: "External validation set with 56 compounds appended".*
7.6.	**Experimental design of test set:** *Indicate any experimental design for getting the test set (e.g. by randomly setting aside chemicals before modelling, by literature search after modelling, by prospective experimental testing after modelling, etc.).*
7.7.	**Predictivity – Statistics obtained by external validation:** *Report here the corresponding statistics. In the case of classification models, include false positive and negative rates.*
7.8.	**Predictivity – Assessment of the external validation set:** *Discuss whether the external validation set is sufficiently large and representative of the applicability domain.* *Describe for example the descriptor and response range or space for the validation test set as compared with that for the training set. Here the descriptor values of the chemicals predicted by the model (training set) should be compared with the descriptor value range of the test set. In addition the distribution of the response values of the chemicals in the training set should be compared to the distribution of the response values of the test set.*
7.9.	**Comments on the external validation of the model**: *Add any other useful comments about the external validation procedure.*
8.	**Providing a mechanistic interpretation – OECD Principle 5** *PRINCIPLE 5: "A MECHANISTIC INTERPRETATION, IF POSSIBLE". According to PRINCIPLE 5, a (Q)SAR should be associated with a mechanistic interpretation, if possible.*
8.1.	**Mechanistic basis of the model**: *Provide information on the mechanistic basis of the model (if possible). In the case of SAR, you may want to describe (if possible) the molecular features that underlie the properties of the molecules containing the substructure (e.g. a description of how sub-structural features could act as nucleophiles or electrophiles, or form part or all of a receptor-binding region). In the case of QSAR, you may give (if possible) a physicochemical interpretation of the descriptors used (consistent with a known mechanism of biological action). If it is not possible to provide a mechanistic interpretation, try to explain why.*
8.2.	**A priori or a posteriori mechanistic interpretation**: *Indicate whether the mechanistic basis of the model was determined a priori (i.e. before modelling, by ensuring that the initial set of training structures and/or descriptors were selected to fit pre-defined mechanism of action) or a posteriori (i.e. after modelling, by interpretation of the final set of training structures and or descriptors).*
8.3.	**Other information about the mechanistic interpretation**: *Report any other useful information about the (purported) mechanistic interpretation described in the previous fields (8.1 and 8.2) such as any reference supporting the mechanistic basis.*
9.	**Miscellaneous information**
9.1	**Comments:** *Add here other relevant and useful comments (e.g. other related models, known applications of the model) that may facilitate regulatory considerations on the model described. Include if relevant experience obtained by use of model prediction for various types of regulatory decisions (incl. references as appropriate).*
9.2	**Bibliography**: *Report useful references other than those directly associated with the model development (references describing the model development are reported in field 2.7).*

9.3	**Supporting information**: *Indicate whether supporting information is attached (e.g. external documents) to this QMRF and specify its content and possibly its utility.*[17]
10.	**Summary for the JRC QSAR Model Database** *The summary section is specific for the JRC Database. If the model is submitted to JRC for inclusion in the database of QSAR models, then this summary is compiled by JRC after QMRF submission. The QMRF author does not have to fill in any of the fields of the summary section.*
10.1	**QMRF number:** *A unique number (numeric identifier) is assigned to any QMRF that is published in the JRC QSAR Model Database. The number encodes the following information: model described in the QMRF (as derived from field 4.2), software implementing the model (as derived from field 1.3), version of the QMRF for the same model and the same software (as derived from the information included in field 2.4) and author of the QMRF (as derived from field 2.2). The number is unique for any QMRF uploaded and stored in the JRC Database.*
10.2	**Publication date**: *The date (day/month/year) of publication in the JRC QSAR Model Database is reported here.*
10.3	**Keywords:** *Any relevant keywords associated with the present QMRF are reported here.*
10.4	**Comments:** *Any comments that are relevant for the publication of the QMRF in the JRC Database (e.g., comments about updates and about supporting information) are reported here.*

R.6.1.9.2 Endpoint Classification

The predefined endpoint classification included in the QMRF is:

1.	**Physico-chemical effects**
1.1.	Melting point
1.2.	Boiling Point
1.3.	Water solubility
1.4.	Vapour pressure
1.5.	Surface tension
1.6.	Octanol-water partition coefficient (K_{ow})
1.7.	Octanol-water distribution coefficient (D)
1.8.	Octanol-air partition coefficient (K_{oa})
1.9.	Air- water partition coefficient (Henry's law constant, H)
1.10.	Dissociation constant
2.	**Environmental fate parameters**
2.1.	Persistence: Abiotic degradation in water a. Hydrolysis, b. Oxidation, c. Others
2.2.	Persistence: Abiotic degradation in air (Phototransformation) a. Direct photolysis b. Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)
2.3.	Persistence: Biodegradation a. Ready/not ready biodegradability

[17] Supporting information may include the training and test sets submitted in defined file formats.

	b. Biodegradation time frame (primary, ultimate degradation)
2.4.	Bioconcentration a. BCF fish b. BCF other organisms
2.5.	Bioaccumulation a. BAF fish b. BAF other organisms
2.6	Organic carbon-sorption partition coefficient (organic carbon; Koc)
2.7.	Adsorption/Desorption in soil
2.8.	Adsorption/Desorption in sediment
2.9.	Vegetation-water partition coefficient
2.10.	Vegetation-air partition coefficient
2.11.	Vegetation-soil partition coefficient
3.	**Ecotoxic effects**
3.1.	Acute toxicity to Daphnia (immobilisation)
3.2.	Acute toxicity to algae (inhibition of the exponential growth rate)
3.3.	Acute toxicity to fish (lethality)
3.4.	Long-term toxicity to Daphnia (lethality, inhibition of reproduction)
3.5.	Long-term toxicity to fish (egg/sac fry, growth inhibition of juveniles, early life stage, full life cycle)
3.6.	Microbial inhibition (activated sludge respiration inhibition, inhibition of nitrification, other)
3.7.	Toxicity to soil microorganisms (inhibition of C-mineralisation, inhibition of N-mineralisation, other)
3.8.	Toxicity to earthworms (survival, growth, reproduction)
3.9.	Toxicity to plants (leaves, seed germination, root elongation)
3.10.	Toxicity to soil invertebrates (survival, growth, reproduction)
3.11.	Toxicity to sediment organisms (survival, growth, reproduction)
3.12.	Toxicity to birds a. Short term toxicity (feeding, gavage, other) b. Long-term toxicity (survival, growth, reproduction)
4.	**Human health effects**
4.1.	Acute inhalation toxicity
4.2.	Acute oral toxicity
4.3.	Acute dermal toxicity
4.4.	Skin irritation/corrosion
4.5.	Acute photoirritation
4.6.	Skin sensitisation
4.7.	Respiratory sensitisation
4.8.	Photosensitisation
4.9.	Eye irritation/corrosion
4.10.	Mutagenicity
4.11.	Photomutagenicity
4.12.	Carcinogenicity

4.13.	Photocarcinogenicity
4.14.	Repeated dose toxicity
4.15.	In vitro reproductive toxicity (e.g. embryotoxic effects in cell culture such as embryo stem cells)
4.16.	In vivo pre-natal-developmental toxicity
4.17.	In vivo pre-, peri-, post natal development and/or fertility (1 or 2 gen. study or enhanced 1 gen. study)
4.18.	Endocrine activity a. Receptor-binding (specify receptor) b. Receptor binding and gene expression (specify receptor) c. Other (e.g. inhibition of specific enzymes involved in hormone synthesis or regulation, specify enzyme(s) and hormone)
5.	**Toxicokinetics**
5.1.	Skin penetration
5.2.	Ocular membrane penetration
5.3.	Gastrointestinal absorption
5.4.	Blood-brain barrier penetration
5.5.	Placental barrier penetration
5.6.	Blood-testis barrier penetration
5.7.	Blood-lung barrier penetration
5.8.	Metabolism (including metabolic clearance)
5.9.	Protein-binding
5.10	DNA-binding
6.	**Other**

R.6.1.10 The QSAR Prediction Reporting Format (QPRF)

R.6.1.10.1 QPRF – version 1.1[18]

Please fill in the fields of the QPRF with information about the prediction and the substance for which the prediction is made. The information that you provide will be used to facilitate considerations on the adequacy of the prediction (model result) in relation to a defined regulatory purpose.

The adequacy of a prediction depends on the following conditions: a) **the (Q)SAR model is scientifically valid**: the scientific validity is established according to the OECD principles for (Q)SAR validation; b) **the (Q)SAR model is applicable to the query chemical**: a (Q)SAR is applicable if the query chemical falls within the defined applicability domain of the model; c) **the (Q)SAR result is reliable**: a valid (Q)SAR that is applied to a chemical falling within its applicability domain provides a reliable result; d) **the (Q)SAR model is relevant for the**

[18] Version of May 2008. For more information consult the following website: http://ecb.jrc.it/qsar/qsar-tools

regulatory purpose: the predicted endpoint can be used directly or following an extrapolation, possibly in combination with other information, for a particular regulatory purpose.

A (Q)SAR prediction (model result) may be considered adequate if it is reliable and relevant, and depending on the totality of information available in a weight-of-evidence assessment (see Section 4 of the QPRF).

1.	**Substance** *This section is aimed at defining the substance for which the (Q)SAR prediction is made.*
1.1.	**CAS number**: *Report the CAS number*.
1.2	**EC number:** *Report the EC number.*
1.3.	**Chemical name**: Report the chemical name (IUPAC and CAS names)
1.4.	**Structural formula**: Report the structural formula.
1.5.	**Structure codes**: *Report available structural information for the substance, including the structure code used to run the model. If you used a SMILES or InChI code, report the code in the corresponding field below. If you have used any another format (e.g. mol file), please include the corresponding structural representation as supporting information.* **a. SMILES**: *Report the SMILES of the substance (indicate if this is the one used for the model prediction).* **b. InChI:** *Report the InChI code of the substance (indicate if this is the one used for the model prediction).* **c. Other structural representation**: *Indicate if another structural representation was used to generate the prediction. Indicate whether this information is included as supporting information. Example: "mol file used and included in the supporting information".* **d. Stereochemical features:** *Indicate whether the substance is a stereo-isomer and consequently may have properties that depend on the orientation of its atoms in space. Identify the stereochemical features that may affect the reliability of predictions for the substance, e.g. cis-trans isomerism, chiral centres. Are these features encoded in the structural representations mentioned above?*
2.	**General information** *General information about the compilation of the current QPRF is provided in this section.*
2.1.	**Date of QPRF**: *Report the date of compilation of the QPRF. Example: January 2007.*
2.2.	**QPRF author and contact details**: *Report the contact details of the author of the QPRF.*
3.	**Prediction** *The information provided in this section will help to facilitate considerations on the scientific validity of the model (as defined in the OECD Principles for the validation of (Q)SAR models) and the reliability of the prediction. Detailed information on the model are stored in the corresponding QMRF which is devised to reflect as much as possible the OECD principles. Remember that the QMRF and the QPRF are complementary, and a QPRF should always be associated with a defined QMRF.*
3.1.	**Endpoint (OECD Principle 1)** **a. Endpoint**: *Define the endpoint for which the model provides predictions (this information should correspond to the information provided in the QMRF under fields 3.2 and 3.3). Example: "Nitrate radical degradation rate constant KNO_3".* **b. Dependent variable:** *Report the dependent variable for which the model provides predictions including any transformations introduced for modelling purposes (note that this information should correspond to the information provided in the QMRF under field 3.5). Example: "-log (KNO_3)".*
3.2.	**Algorithm (OECD Principle 2)** **a. Model or submodel name:** *Identify the model used to make the prediction and possibly report its name as stored in the corresponding QMRF; in the QMRF the model name is*

	reported in the field QSAR identifier. Examples: "BIOWIN for Biodegradation"; TOPKAT Skin Irritation Model". If applicable identify the specific submodel or algorithm applicable to the specific chemical Examples: "BIOWIN 1"; TOPKAT Skin Irritation Acyclics (Acids, Amines, Esters) MOD v SEV Model"; "ECOSAR esters model". **b. Model version:** *Identify, where relevant, the version number and/or date of the model and submodel.* **c. Reference to QMRF:** *Provide relevant information about the QMRF that stores information about the model used to make the prediction. Possible useful pieces of information are: availability, source, reference number (if any) of the QMRF. Examples: "The corresponding QMRF named 'BIOWIN 1 for Biodegradation' has been downloaded from the JRC QSAR Model Database"; "The corresponding QMRF named 'TOPKAT Skin Irritation Acyclics (Acids, Amines, Esters) MOD v SEV Model' has been newly compiled".* **d. Predicted value (model result):** *Report the predicted value (including units) obtained from the application of the model to the query chemical. For an expert system such as Derek for Windows, report the alert triggered together with the reasoning. Example: " aromatic amine - mutagenicity, plausible".* **e. Predicted value (comments):** *If the result is qualitative (e.g. yes/no) or semi-quantitative (e.g. low/medium/high), explain the cut-off values that were used as the basis for classification. In reporting the predicted value, pay attention to the transformations (e.g. if the prediction is made in log units, apply anti-logarithm function).* **f. Input for prediction:** *Specify what kind of input was used to generate the prediction (SMILES, mol file, graphical interface etc). Please provide the structure code used to generate the prediction (unless already provided in section 1.5).* **g. Descriptor values:** *Where appropriate, report the values (experimental or calculated data) for numerical descriptors and indicate which values were used for making the prediction.*
3.3.	**Applicability domain (OECD principle 3)** a. **Domains:** *Discuss whether the query chemical falls in the applicability domain of the model as defined in the corresponding QMRF (section 5 of QMRF, Defining the applicability domain – OECD Principle 3). If additional software/methods were used to assess the applicability domain then they should also be documented in this section. Include a discussion about:* **i.** *descriptor domain* **ii.** *structural fragment domain (e.g., discuss whether the chemical contains fragments that are not represented in the model training set)* **iii.** *mechanism domain (discuss whether the chemical is known or considered to act according to the mechanism of action associated with the used model)* **iv.** *metabolic domain, if relevant* b. **Structural analogues:** *List the structural analogues that are present in the training or test sets, or accessible from other sources (in this case you should explain how the structural analogue was retrieved[19]) and why they are considered analogues). For each analogue, report the CAS number, the structural formula, the SMILES code, and the source (e.g., training set, test set or other source). For an expert system (like Derek for Windows or TOPKAT), the example compounds or structurally related analogues with their experimental data should be provided.* c. **Considerations on structural analogues:** *Discuss how predicted and experimental data for analogues support the prediction of the chemical under consideration.*
3.4.	**The uncertainty of the prediction (OECD principle 4)** *If possible, comment on the uncertainty of the prediction for this chemical, taking into account relevant information (e.g. variability of the experimental results).*

[19] Various software tools (e.g. the OECD (Q)SAR Toolbox) could be used to support the search for analogues.

3.5.	**The chemical and biological mechanisms according to the model underpinning the predicted result (OECD principle 5).** *Discuss the mechanistic interpretation of the model prediction for this specific chemical. For an expert system based on structural alerts (e.g. Derek for Windows, Oncologic™) the rationale for the structural alert fired should be provided.*
4.	**Adequacy (optional)**[20] *The information provided in this section might be useful, depending on the reporting needs and formats of the regulatory framework of interest. This information aims to facilitate considerations about the adequacy of the (Q)SAR prediction (result). A (Q)SAR prediction may or may not be considered adequate ("fit-for-purpose"), depending on whether the prediction is sufficiently reliable and relevant in relation to the particular regulatory purpose. The adequacy of the prediction also depends on the availability of other information, and is determined in a weight-of-evidence assessment.*
4.1.	**Regulatory purpose**: *Explain the regulatory purpose for which the prediction described in Section 3 is being used.*
4.2.	**Approach for regulatory interpretation of the model result**: *Describe how the predicted result is going to be interpreted in light of the specific regulatory purpose (e.g. by applying an algorithm or regulatory criteria). This may involve the need to convert the units of the dependent variable (e.g. from log molar units to mg/l). It may also involve the application of another algorithm, an assessment factor, or regulatory criteria, and the use or consideration of additional information in a weight-of-evidence assessment.*
4.3.	**Outcome**: *Report the interpretation of the model result in relation to the defined regulatory purpose.*
4.4.	**Conclusion**: *Provide an assessment of whether the final result is considered adequate for a regulatory conclusion, or whether additional information is required (and, if so, what this additional information should be).*

R.6.1.11 References for the guidance on (Q)SARs

Brink RH, Walker JD. (1987). EPA TSCA ITC Interim Report. Dynamic Corporation, Rockwille, 6/4/1987.

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

Cronin MTD, Walker JD, Jaworska JS, Comber MHI, Watts CD, Worth AP (2003). Use of quantitative structure-activity relationships in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. *Environmental Health Perspectives* **111**, 1376-1390.

Cronin MTD, Jaworska JS, Walker JD, Comber MHI, Watts CD, Worth AP (2003). Use of quantitative structure-activity relationships in international decision-making frameworks to predict health effects of chemical substances. *Environmental Health Perspectives* **111**, 1391-1401.

Darvas F (1987). MetabolExpert, an expert system for predicting metabolism of substances. In *QSAR in Environmental Toxicology*. Kaiser K, ed. Riedel, Dordrecht, The Netherlands. Pp 71-81.

Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PM, Payne MP, Richard AM, Tichy M, Worth AP, Yourick JJ (1997). The development and validation of expert systems for predicting toxicity. The report and recommendations of an ECVAM/ECB workshop (ECVAM workshop 24). *ATLA* **25**, 223-252.

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

Dimitrov S, Dimitrova G, Pavlov T, Dimitorva N, Patlewicz G, Niemala J, Mekenyan O (2005)**.** A stepwise approach for defining the applicability domain of SAR and QSAR models. *Journal of Chemical Information and Computer Sci*ence **45**, 839-849.

[20] In any case, adequate and reliable documentation is required (section 1.3 of REACH Annex XI). The format provides guidance on which elements are needed.

EC (2001). An interim strategy for management of PBT and vPvB substances. Joint Meeting of the Competent Authorities for the Implementation of Council Directive 67/548/EEC (New Substances) and Council Regulation (EEC) 793/93 (Existing Substances). ENV/D/432028/01. NOTIO/36/2001. DG Environment, Brussels.

EC (2003). Technical guidance documents on risk assessment in support of Commission Directive 93/67/EEC on risk assessment for new notified substances, Commission Regulation (EC) No. 1488/94 on risk assessment for existing substances, and Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market. EUR 20418 EN. European Commission, Joint Research Centre, Ispra, Italy.

ECB (2005). Scoping study on the development of a technical guidance document on information requirements on intrinsic properties of substances. Report prepared by CEFIC, DK-EPA, Environmental Agency of Wales and England, ECETOC, INERIS, KemI and TNO. European Chemicals Bureau, Joint Research Centre, European Commission, Ispra, Italy. Accesible from: http://ecb.jrc.it

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

Efron B, Gong G (1983). A leisurely look at the bootstrap, the jackknife, and crossvalidation. *The American statistician* **37**, 36-48.

Enslein K (1988). An overview of structure-activity relationships as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation, and acute oral toxicity. *Toxicological Industrial Health* **4**, 479-498.

Eriksson L, Johansson E, Kettaneh-Wold N (2001). Multi and megavariate data analysis-Principled and applications. Umea, Sweden: Umetrics AB.

Eriksson L, Jaworska JS, Worth AP, Cronin MTD, McDowell RM, Gramatica P (2003). Methods for reliability, uncertainty assessment, and applicability evaluations of classification and regression based QSARs. *Environmental Health Perspectives* **111**, 1361-1375.

Gini G. (2005). Scoping Study for the Development of an Internet Based Decision Support System For (Quantitative) Structure Activity Relationships. Final report of JRC contract 22518-2004-11 F11ED ISP IT. European Chemicals Bureau, Joint Research Centre, European Commission, Ispra, Italy. http://ecb.jrc.it/DOCUMENTS-/QSAR/Final_report_DSS_scoping.pdf

Golbraikh A, Tropsha A. (2002). Beware of q2! *Journal of Molecular Graphics & Modelling* **20**, 269-276.

Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A (2003). Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer Aided Molecular Design* **17**, 241-253.

Gramatica P (2004). Evaluation of different statistical approaches for the validation of quantitative structure-activity relationships. Final report for JRC Contract ECVA-CCR.496576-Z. European Chemicals Bureau, Joint Research Centre, European Commission, Ispra, Italy. http://ecb.jrc.it/DOCUMENTS/QSAR/Report_on_QSAR_validation-_methods.pdf

Greene N, Judson P, Langowski J, Marchant CA (1999). Knowledge based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. *SAR & QSAR in Environmental Research* **10**, 299-313.

Hansen B., Hart J. & Loonen H. (1999). QSARs and SARs in EU Hazard Identification and Risk Assessment. Proceedings of the 8th. International Workshop on Quantitative Structure-Activity Relationships (QSARs) in the Environmental Sciences. SETAC, 1998.

Hawkins DR. (1996). Biotransformations. Volumes 1-7, Royal Society of Chemistry, Cambridge, UK.

Jaworska JS, Comber M, Auer C, van Leeuwen CJ (2003). Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. *Environmental Health Perspectives* **111**, 1358-1360.

Jaworska J., Nikolova-Jeliazkova N & Aldenberg T (2005). Review of methods for QSAR applicability domain estimation by the training set. *ATLA* **33**, 445-459.

Judson PN (2002). Prediction of toxicity using DEREK for Windows. *Journal of Toxicological* Sciences **27**, 278.

Klimisch HJ, Andreae M, Tillmann U (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicology and Pharmacology* **25**, 1-5

Klopman G, Tu MH (1999). META - a program for the prediction of products of mammalian metabolism of xenobiotics. In *Drug Metabolism, Databases in High Throughput Testing during Drug Design and Development*. Erhardt PW, ed. Blackwell, Oxford, UK. pp. 271-276.

Klopman G, Chakravarti SK. (2003). Screening of high production volume chemicals for estrogen receptor binding activity (II) by the MultiCASE expert system, *Chemosphere* **51**, 461-468.

Kraker JJ, Hawkins DM, Basak SC, Natarajan R, Mills D. (2006). Quantitative Structure–Activity Relationship (QSAR) modeling of juvenile hormone activity: Comparison of validation procedures. *Chemometrics and Intelligent Laboratory Systems*, in press.

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

Lewis DVF (2003). Quantitative structure-activity relationships (QSARs) within the cytochrome P450 system: QSARs describing substrate binding, inhibition and induction of P450s. *Inflammopharmacology* **11**, 43-73.

Loonen H, Hansen B, Hart J, Karcher W (1997). Application of QSAR in hazard and risk assessment. Presentation at Seventh Annual Meeting of SETAC-Europe. Amsterdam, The Netherlands, April 6 - 10, 1997.

Mekenyan OG, Dimitrov S, Schmieder P, Veith G (2004). A systematic approach to simulating metabolism in computational toxicology. I. The TIMES heuristic modelling framework. *Current Pharmaceutical Design* **10**, 1273-1293.

Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, van de Sandt JJM, Tong W, Veith G, Yang C (2005). Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. The Report and Recommendations of ECVAM Workshop 52. *ATLA* **33** 155-173.

Nikolova N, Jaworska J (2003). Approaches to measure chemical similarity - a review. *QSAR & Combinatorial Science* **22**, 1006-1026.

OECD (2004). The Report from the Expert Group on (Quantitative) Structure Activity Relationship ([Q]SARs) on the Principles for the Validation of (Q)SARs. OECD Series on Testing and Assessment No. 49. ENV/JM/MONO(2004)24. Organisation for Economic Cooperation and Development, Paris, France. 206 http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html

OECD (2006b). Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals. Organisation of Economic Cooperation and Development. Paris, France. http://www.oecd.org/document/-30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html

OECD (2007). Guidance Document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Models. OECD Series on Testing and Assessment No. 69. ENV/JM/MONO(2007)2. Organisation for Economic Cooperation and Development, Paris, France. 154 pp. http://www.oecd.org/document/-30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html

Payne MP. (2004). Computer-based methods for the prediction of chemical metabolism and biotransformation within biological organisms. In *Predicting Chemical Toxicity and Fate*. Cronin MTD, Livingstone DJ, eds. CRC Press, Boca Raton, FL, USA. pp. 205-227.

Russom CL, Bradbury SP, Broderius SJ, Hammermeister DE, Drummond RA (1997). Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). *Environmental Toxicology and Chemistry* **16**, 948–967.

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

Schultz TW, Yarbrough JW, Koss SK (2006). Identification of reactive toxicants: structure-activity relationships for amides. *Cell Biology and Toxicology* **22**, 339-349.

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

Todeschini R, Consonni V (2000). *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim.

Verhaar HJM, Mulder W, Hermens JLM (1995). QSARs for ecotoxicity. In: Overview of structure-activity relationships for environmental endpoints, Part 1: General outline and procedure. Hermens, J.L.M. (Ed), Report prepared within the framework of the project "QSAR for Prediction of Fate and Effects of Chemicals in the Environment", an international project of the Environmental; Technologies RTD Programme (DGXII/D-1) of the European Commission under contract number EV5V-CT92-0211.

Vracko M, Bandelj V, Barbieri P, Benfenati E, Chaudhry Q, Cronin M, Devillers J, Gallegos A, Gini G, Gramatica P, Helma C, Mazzatorta P, Neagu D, Netzeva T, Pavan M, Patlewicz G, Randic M, Tsakovska I, Worth A (2006). Validation of counter propagation neural network models for predictive toxicology according to the OECD principles: a case study. *SAR & QSAR in Environmental Research* **17**, 265-284.

Worth AP, Bassan A, Gallegos A, Netzeva TI, Patlewicz G, Pavan M, Tsakovska I, Vracko M (2005). The Characterisation of (Quantitative) Structure-Activity Relationships: Preliminary Guidance. JRC report EUR 21866 EN. European Chemicals Bureau, Joint Research Centre, European Commission, Ispra, Italy. http://ecb.jrc.it/DOCUMENTS/QSAR/QSAR_characterisation_EUR_21866_EN.pdf

Worth AP, Bassan A, de Bruijn J, Gallegos Saliner A, Netzeva TI, Patlewicz G, Pavan M, Tsakovska I, Eisenreich S (2006). The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods. *SAR and QSAR in Environmental Research* **18**, 111-125.

R.6.2 Guidance on the Grouping of Chemicals

Under REACH, Annex XI opens the possibility for substances to be assessed by the use of grouping approaches. Annex XI requires the Agency, after consulting with relevant stakeholders and other interested parties, to issue guidance on technically and scientifically justified methodology for the grouping of substances sufficiently in advance of the first registration deadline for phase-in substances. This chapter provides the first draft of the guidance prepared in order to fulfil this requirement of Annex XI.

The guidance first explains what a category is and relevant concepts that will enable the document to be better read (Section R.6.2.1). The mechanistic basis for categories is explained and the advantages derived from using a category described. Section R.6.2.1 also describes the close relationship that exists between (Q)SAR and categories, both in terms of the concepts and in the use of (Q)SAR for data evaluation and gap-filling. Section R.6.2.2. describes the main approaches that are used for data gap filling: read-across, trend analysis and QSARs. While Sections R.6.2.1 and section R.6.2.2. provide explanations on the scientific and methodological background of the analogue and category approaches, respectively, Section R.6.2.3 and section R 6.2.6 focus more on practical aspects for forming and documenting analogue and chemical category approaches. Separate Section R.6.2.3 and section R.6.2.4 were elaborated to provide guidance on stepwise procedures for analogue read-across and chemical categories, so that the guidance document can be used in a *modular fashion*, making it possible to use parts of the guidance only. Therefore a number of repetitions of texts were also necessary. Section R.2.5 elaborates on some specific issues that need to be addressed with specific types of categories. Finally, in Section R.6.2.6, a Category Reporting Format is proposed as a tool for documenting chemical categories.

R.6.2.1 Explanation of the chemical category approach

Under REACH, testing requirements for individual substances are based on the specific information requirements shown in Annexes VI-X. As an alternative approach, Annex XI opens the possibility of evaluating chemicals not on a one-by-one basis, but by grouping chemicals in categories.

In this guidance, the terms *category approach* and *analogue approach* are used to describe techniques for grouping chemicals, whilst the term *read-across* is reserved for a technique of filling data gaps in either approach. A *chemical category* is a group of chemicals whose physico-chemical 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 (or other similarity characteristic). In principle, more members are generally present in a chemical category, enabling the detection of trends across endpoints. 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. The term *analogue approach* is used when the grouping is based on a very limited number of chemicals, where trends in properties are not apparent.

Categories of chemicals are selected based on the hypothesis that the properties of a series of chemicals with common structural features will show coherent trends in their physico-chemical properties, and more importantly, in their toxicological (human health/ecotoxicity) effects or environmental fate properties. Common behaviour or consistent trends are generally associated with a common underlying mechanism of action, or where a mechanism of action exhibits intensity changes in a consistent manner across the different members of a category.

The use of a category approach will mean that it is possible to identify properties which are common to at least some members of the category. The approach also provides a basis on which to identify possible trends in properties across the category. As a result, it is possible to extend the use of measured data to similar untested chemicals, and reliable estimates that are adequate for classification and labelling and/or risk assessment can be made without further testing. In addition, knowledge of the expected effects of the category together with information on use and exposure will help in deciding not only whether additional testing is needed, but also the nature and scope of any testing that needs to be carried out.

The assessment of chemicals by using a category approach differs from the approach of assessing them on an individual basis, since the effects of the individual chemicals within a category are assessed on the basis of the evaluation of the category as a whole, rather than based on measured data for any one particular substance alone. For a category member that lacks data for an endpoint, the data gap can be filled in a number of ways, including by read-across from one or more other category members. In some circumstances, it may only be necessary to use data from one category member using read-across principles to adequately characterise the member lacking data. The category approach is important since it provides an alternative to testing individual substances and as a result should lead to a decrease in the use of animal testing.

R.6.2.1.1 Benefits of the chemical category approach

Assessment of a large number of chemicals as a category can be more efficient and accurate than assessment of single compounds for a number of reasons:

- data from one or more chemicals can be interpolated or extrapolated to other chemicals, reducing the need to test for every endpoint for every chemical;
- since existing data can be applied to additional chemicals without the need for additional testing, the use of animal testing is reduced;
- the category evaluation is based on a greater body of data than on data on a single compound;
- the identification of compounds as members of a category provides an insight into the potential effects of the compounds that might otherwise be overlooked;

- the use of a category approach may also provide significant advantages in the evaluation of compounds that are often considered as difficult, in the sense that they can present technical difficulties when carrying out standard test protocols (examples are given in Hart, 2007; Comber and Simpson, 2007);
- the approach provides a valuable tool in cases where animal models do not always reliably predict effects on humans (examples are given in Hart, 2007),
- in most cases, category testing can be completed earlier than individual tests for each chemical that requires notification, submission or inclusion,
- in order to gain future efficiencies, category proposals may be expanded via the inclusions of chemicals that may be addressed under various global programs,
- in the category approach, not every chemical needs to be tested for every endpoint. Rather, the overall data for that category must prove adequate to support a hazard assessment. The overall data set must allow the estimation of the hazard for the missing data points,
- a category approach allows for better consideration of the biological plausibility of grouping the chemicals within a category.

Use of a category approach can also provide significant efficiencies and benefits when identifying data gaps and filling data needs that are ultimately deemed necessary. A category test plan is designed to provide information to characterise the group as a whole rather than to fill every data point for every chemical in the category. This reflects an approach that is more efficient from a testing perspective than test plans for obtaining data on individual chemicals of commercial interest. Knowledge of the expected biological effects of the category will be helpful in deciding not only whether testing is needed, but also the nature and scope of the test to be carried out. Where confirmation is sought that an individual category member does not have a particular property (e.g. acute oral toxicity), a simple limit test might be adequate to provide the necessary confirmation. Where an individual category member is expected to have an effect (e.g. skin irritation or corrosion), a simple *in vitro* test might provide adequate confirmation of the predicted effect.

Another benefit of using a category approach is that this approach allows for an evaluation of the biological basis for the effects seen in a group of chemicals within a category. When it is known that members of a chemical category share a presumed common mechanism of action, the confidence in the category is significantly greater than that associated with the use of a read-across approach to fill data gaps. This confidence increases with increasing numbers of chemicals included in the category. For a large category[21], both the presence and absence of certain hazards, as well as the trend of an effect across a category, can be identified. This provides a basis on which the properties of individual members of the category can be identified with the necessary confidence. For more limited comparisons, particularly with chemicals containing multiple functional groups, it is harder to obtain the same level of confidence. A category approach can provide significant advantages compared to the read-across techniques for filling data gaps, in that it is possible to analyse trends in properties. Read-across techniques between chemical analogues have been extensively used (e.g. within the OECD HPV Chemicals Programme, the EU Existing Chemicals Programme or for Classification and Labelling in the EU), often on an *ad hoc* basis and it is foreseen that they will continue to be used extensively. Nevertheless, an important consideration in

[21] Based on the current experience at OECD, any category with more than 10 members is a large category.

preparing this Guidance is to encourage the replacement of these *ad hoc* approaches by a more wide-ranging approach that can provide a greater degree of confidence in the result.

Guidance on the analogue approach is provided in Section R.6.2.3, and guidance on category formation is provided in Section R.6.2.4

R.6.2.1.2 Explanation of relevant concepts

The term *grouping* or *chemical grouping* describes the general approach to assessing more than one chemical at the same time. It can include formation of a chemical category or identification of a chemical analogue for which read-across may be applied. In this document, the more specific terms *chemical category* and *analogue approach* are used.

A *chemical category* is a group of chemicals whose physico-chemical 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:

- common functional group(s) (e.g. aldehyde, epoxide, ester, specific metal ion);
- common constituents or chemical classes, similar carbon range numbers. This is frequently the case with complex substances often known as substances of Unknown or Variable composition, Complex reaction products or Biological material (UVCB substances);
- an incremental and constant change across the category (e.g. a chain-length category), often observed in physico-chemical properties, e.g. boiling point range;
- the likelihood of common precursors and/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/salt).

Categories can be developed systematically on the basis of structure (or other similar characteristic) alone. It is recognised that in many cases the formation of a chemical category is also dependant on which chemicals are manufactured by the consortium of companies sponsoring the category and/or the regulatory context under which the evaluation is being made. While these considerations can legitimately influence the formation of a category, they are independent of the scientific analysis of a category.

Within a chemical category, data gaps may be filled by read-across, trend analysis and QSARs. *Read-across* is a technique used to 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 (on the basis of structural similarity and similar properties and/or activities). For a given category endpoint, the category members are often related by a trend (e.g. increasing, decreasing or constant) in an effect, and a *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 QSAR* model, where the category under examination is a subcategory of the wider QSAR. Further details are given in Section R.6.2.2

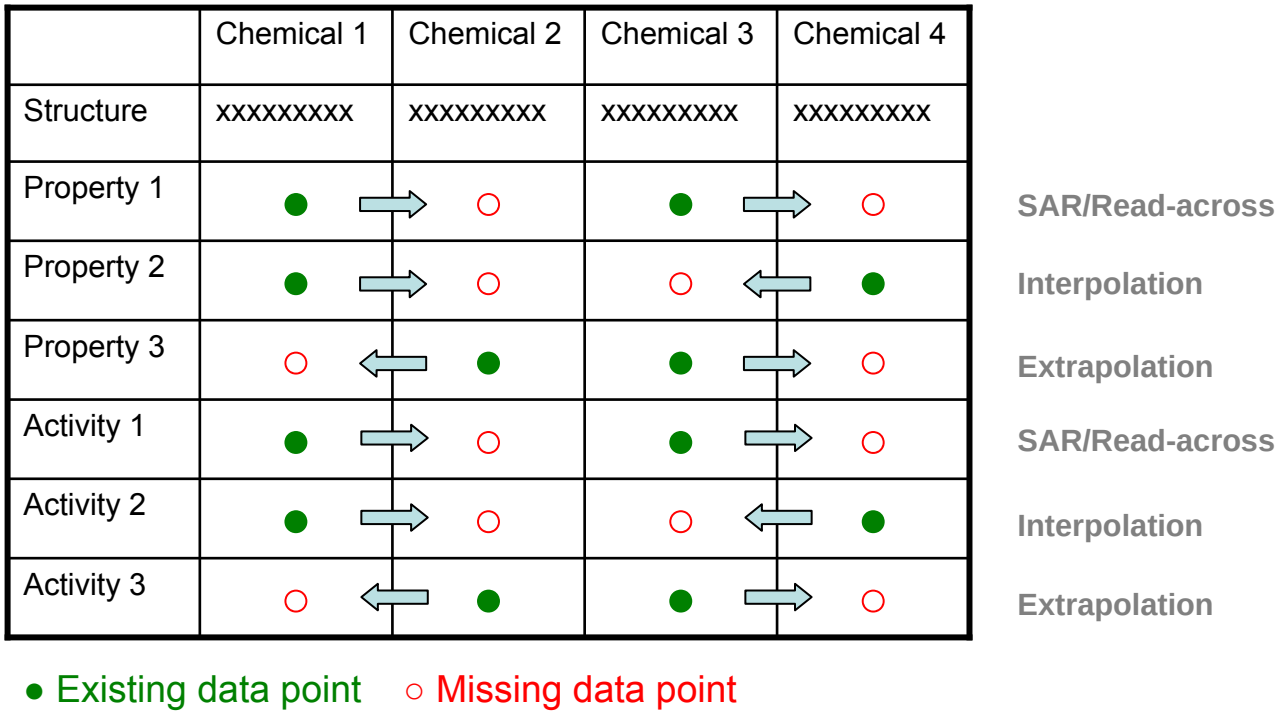
While read-across is a technique for data gap filling within the context of a category approach, it is also a useful tool for data gap filling in cases where comparisons are based on a very limited number of chemicals. The simplest example of the *category approach* is a comparison between two chemicals. This form of evaluation is often called a read-across approach, and this is the term used

in Annex XI of REACH. This approach has been used extensively in the evaluation of chemicals under a number of different evaluation programmes, and, although the approach has been used on a largely *ad hoc* basis, there are a number of examples on which guidance can be based. Whilst sharing many characteristics in common with a category approach, the evaluation of a very limited number of chemicals does present a number of differences compared to the evaluation of larger, systematically derived categories, for which there is more limited experience. In order to avoid confusion, evaluations of a very limited number of chemicals using largely read-across to fill data gaps is described in this guidance as the *analogue approach*. The term *read-across* is therefore limited to the technique for filling data gaps described in Section R.6.2.2

In the *analogue approach* endpoint information for one chemical is used to predict the same endpoint for another chemical, which is considered to be *similar* in some way (usually on the basis of structural similarity and similar properties and/or activities). General guidance on how to use the analogue approach is provided in Section R.6.2.3

A chemical category can be described by a matrix consisting of the *category members* and by a corresponding set of properties and/or effects data (the *category endpoints*), (see Figure R.6-3). General guidance on how to build categories is provided in Section R.6.2.4, whereas specific guidance for different types of categories is given in Section R.6.2.5

Figure R.6-3: Graphical representation of a chemical category and some approaches for filling data gaps



As illustrated in Figure R.6-3 , data gap filling can be done using read-across from one tested chemical to an untested chemical. In general, interpolation is preferred to extrapolation between category members; this is discussed in more detail in Section R.6.2.2.2. Other approaches which include trend analysis, (Q)SARs/Expert systems are also covered in Section R.6.2.2 More specific guidance on the application of these data-filling techniques in the analogue approach is given in Section R.6.2.3, and in a broader category approach in Section R.6.2.4. Examples of the data matrices used to report the use of this approach is shown in Section R.6.2.6

Category membership

In an ideal situation, a category would include all potential members of the category when first developed. This ideal situation will be difficult to achieve in practice. For example, even when a category includes all the single compounds that can be included, it may not necessarily include the additional commercial products that are complex substances containing a mixture of compounds which are also included in the category.

Practical considerations will often influence the choice of chemicals included in the category. Since categories have often been developed in the context of a High Production Volume Chemicals programme, the selection of the chemicals that are included in a particular chemical category has frequently been guided by the fact that the chemicals in the category are produced in high volumes and likely to be dependant on which chemicals are manufactured by the consortium of companies sponsoring the category.

However, it should be noted that the category may also contain substances that are not produced in high volumes, or indeed, substances that are not necessarily commercially available, as well as other substances put on the market by companies not involved in the category evaluation. Substances included in the category that are not formally evaluated have previously been described as *surrogate* substances. This term is not used in the guidance as these substances may subsequently be assessed, e.g. if their production volume changes.

There are significant potential advantages associated with the evaluation of a category which contains a high proportion of its potential members. The conclusions drawn from the evaluation are likely to be more robust, since the category evaluation is less likely to be affected by the subsequent addition of other substances, and the potential advantages of limiting animal and other testing are also likely to be greater.

As chemical categories submitted to authorities for review often do not contain all potential members of a category, due to the practical considerations outlined above, they are evaluated based on the data available for the chemicals submitted. If subsequently chemicals are assessed which fit within the definition and rationale of the category, the category might have to be re-evaluated based on the available data for those additional chemicals.

A substance can potentially belong to more than one category. For example, a multifunctional compound can belong to a category based on function A as well as to the category based on function B. The properties of the compound will be influenced by the presence of both functional groups.

Assessment of categories and individual compounds in a category

The successful use of a category approach should lead to the identification and characterisation (qualitative or quantitative) of the hazards for all the members of the category, irrespective of their production volume or whether or not they are produced by the companies carrying out the category evaluation.

Under REACH, however, as substances are registered on a substance-by-substance basis, a category evaluation does not necessarily result in all the individual substances included in the category evaluation being registered to the Agency, although the data from these substances will be included in the category report in support of the registration.

If a substance is assessed and 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 is unlikely to alter the conclusions of the initial evaluation significantly. Since subsequent assessments of additional members of a category are possible at any time, there is an incentive to ensure that as many potential members of a category are included in the initial evaluation. This would ensure that the evaluation is sufficiently robust in order to minimise the potential revisions as a result of additional data at a later date.

Experience has shown that in many cases additional chemicals are identified which fall on either the lower or upper end of an existing category. In those cases additional testing might be necessary to confirm that the chemicals belong to the category. In these cases, best professional judgement and *Weight of Evidence* (see Section R.6.2.2.4) are used together in making recommendations/decisions about whether to test or not.

When assessing whether a substance could be a member of an existing category (but it is not already listed as such), the concept of *applicability domain* may be useful. The *applicability domain* (AD) of a (sub)category would identify the structural requirements and ranges of physico-chemical, environmental fate, toxicological or ecotoxicological properties within which reliable estimations can be made for the (sub)category members. For example, there may be a trend of increasing acute aquatic toxicity with increasing chain length from C_2 up to a carbon chain length of C_{12}, after which no aquatic toxicity is seen because the water solubility has decreased with increasing chain length. Thus the applicability domain for aquatic toxicity would be C_2 to C_{12}.

Subcategories

In some cases, an effect can be present for some but not all members of the category. An example is the glycol ethers, where the lower members of the category show reproductive toxicity whilst higher members do not. In other cases, the category may show a consistent trend where the resulting potencies lead to different classifications. Examples include the lower aliphatic ethers, where aquatic toxicity is insufficient to lead to classification for aquatic toxicity with the lower members of the category, but does lead to classification for this effect with higher members (Hart and Veith, 2007).

In these cases it can be helpful to divide the category into subcategories. Examples which have been encountered within the OECD HPV program (http://cs3-hq.oecd.org/scripts/hpv) include the case of mono-, di-, tri-, tetra-, and penta- ethylene glycols, when a subcategory was denoted by a cut off of chain length of 6-8 to account for the change in physical form from liquid to solid and a decrease in uptake. A slightly different approach was used in the case of Oxo alcohols C_9 to C_{13} where clear trends in properties were seen with increasing chain length (Caley *et al*, 2007). For environmental hazards, two category members exhibited higher ecotoxicity than the other five members and thus formed a subcategory in the assessment. For the long chain alcohols ($C_{6\text{-}22}$ primary aliphatic alcohols), decreasing water solubility and increasing lipophilicity is observed with increasing chain length, leading to a cut off for acute aquatic toxicity effects at C_{13} to C_{14} and around C_{15} for chronic effects. At C>18, biodegradability is reduced. Three distinctive subcategories can be identified using the GHS classification criteria for aquatic toxicity based on the trends in toxicity and biodegradability.

Subcategories may arise for a number of reasons and are often endpoint specific:

- an effect which varies in intensity across the category, such that some members of the category meet the criteria for one hazard classification for the particular endpoint, whereas other members of the category meet the criteria for another. These subcategory definitions can be qualitative (i.e. they have degrees of hazard potential or different regulatory

classifications) or quantitative (the numerical values of the endpoint include values on either side of a breakpoint).

- an effect where there is a peak in activity or a breakpoint in a trend can also lead to the formation of subcategories.
- it is possible that a trend analysis may apply to a subcategory but not to the whole category.

The concept of subcategories has been introduced to improve the practicality and flexibility of the category approach and it does not alter the scientific basis of the category approach.

R.6.2.1.3 The mechanistic basis of chemical categories

A category of chemicals will often show the presence, absence or modulation of a particular effect for all members of the category, based on the presumption of a common mechanism of action. This can be expected to apply to many different categories of chemicals for many aliphatic hydrocarbons, aliphatic amines, nitriles, aldehydes, alcohols, and ethers (Jäckh, 2007). Additional examples can be found from the OECD HPV Chemicals Programme (http://cs3-hq.oecd.org/scripts/hpv).

If the data for a category includes one or more exceptions to the effects expected from a common mechanism of action, a review of the toxicological data for the category should be able to explain the difference in behaviour. Excluding the exception(s) from the category would decrease the information content of the category and hence its robustness. The presence of such *outlying* effects underlines the importance of developing an understanding of the (toxic) mechanisms of action within categories.

A category may be justified on more than one basis, for example both a chain length and metabolic pathway category (Caley *et al*, 2007). Multiple justifications could increase confidence in the category. This increased confidence is largely a result of the more detailed evidence that the common mechanisms of toxic action have been properly identified.

In principle, a category is not endpoint specific, since the structural changes across the category would be expected to produce changes that would affect the whole spectrum of properties of the individual members in a coherent and consistent manner. The changes in properties across a category, for each parameter, would be the result of related rather than purely arbitrary differences. However, it is recognised that in practice it may be possible to identify the trends and changes for some but not all of the properties of potential interest, and hence it may not be possible to use a category approach to identify all relevant effects.

One example is the use of a metabolic pathway approach where the category approach will be able to address the common toxicological mechanism for endpoints related to systemic effects, whereas it may not predict the local effects (on skin and other membranes) due to the parent compound [see for example the category of monoethylene glycol ethers and their acetates or diethylene glycol ethers and their acetates (http://cs3-hq.oecd.org/scripts/hpv) (Caley *et al*, 2007)].

For some series of compounds, the lower or upper end of the series may show marked changes in effects. At the lower end of the series, the methyl analogue may have exceptional properties. Examples are the differences shown in acute toxicity between methyl alcohol and ethyl alcohol, and for carcinogenicity between butter yellow and its ethyl homologue or between methylcarbamate and ethylcarbamate. This may be the result of specific differences in metabolism, such as the differences in carcinogenicity between benzene and toluene, due to the possibility of metabolism of the methyl group with carboxylate formation (Jäckh, 2007).

The presence of a breakpoint can indicate a change in the mode of action or the effect of a consistent tendency across a category. In a homologous series of organic compounds, there is often a breakpoint e.g. the loss of aquatic toxicity as carbon chain length increases and solubility decreases.

The importance of a common mechanism of action is also a factor in deciding what chemicals would not be expected to be relevant members of a category. Variations in chemical structure can affect both toxicokinetics (uptake and bioavailability) and toxicodynamics (e.g. interactions with receptors and enzymes). For example, the introduction of a carboxylate or sulfate function often decreases bioavailability and toxicity to mammals, whilst halogen substituents tend to increase lipophilicity and increase toxicological activity (see example in Worth *et al*, 2007). Thiols and esters are not considered as relevant analogues for evaluation of ether activity (see example in Hart and Veith, 2007).

R.6.2.1.4 Application of the chemical category approach

In cases where the approach to chemical hazard and risk assessment is based on the evaluation of substances on an individual basis (e.g. the approach taken for the notification of new substances) testing requirements are primarily based on the production volume of the chemical. This approach is consistent with the fact that the legal obligations are placed on individual producers, and as a result, producers are legitimately concerned to provide information on their own product, but do not necessarily have any interest in acquiring data on related substances in which they have no commercial interest.

As stated in Section R.6.2.1.2, since categories have often been developed in the context of a High Production Volume Chemicals (HPVC) programme, the selection of a particular chemical category has normally been guided by the presence of a number of chemicals in the category that are produced in high volumes. However, it should be noted that a category may also contain other substances that are not HPV chemicals (or indeed, are not necessarily commercially available). These chemicals are still members of the category, and may prove to be relevant candidates for further testing in order to evaluate the properties of the category as a whole.

The formation of a category has in many cases also been dependant on which chemicals are manufactured by the consortium of companies sponsoring the category. Different industry sectors may cooperate on category assessments. This guidance recognises that it is a challenge for Industry to include all relevant members based on the basic properties excluding use pattern/exposure. There may be different needs for hazard information for different members of a consortium depending on uses and thereby the outcome of the risk assessments for the individual members of the chemical category. It is therefore important to develop incentives or articulate benefits for industry taking this approach, as it would be desirable for the consortium to check with other producers/manufacturers for appropriate support and information.

R.6.2.1.5 Robustness of a chemical category

The chemical category approach can be very beneficial when information from other category members help to fill data gaps for untested chemicals. However, the approach may not always be straight forward, especially when a category has many members, when the trend analysis does not show an obvious ‘trend’, and/or when different kinds of information (e.g., computational data as well as experimental data) lead to different results are available within a category. The experience from the OECD HPVC program, where industry has had a possibility to discuss their category approach with a Sponsor Member Country, has shown that this collaboration is very helpful. The

possibility for registrants to consult 'relevant regulatory bodies' regarding their application of the chemical category approach remains. For substances that are part of the OECD HPVC program, the OECD will continue to support collaboration between Industry and OECD member countries. Robustness of a chemical category

A number of factors contribute to the robustness of a category. Useful considerations might include:

a) membership of the category characterised by the number of members in a category and the available data.

b) the density and distribution of the category (both in terms of the chemicals represented and the data available).

c) the quality of the underlying experimental data for each of the endpoints covered.

d) the presumed mechanistic basis underpinning the category for a particular endpoint.

e) the quality of the data estimated by the external computational approaches.

The current document does not provide criteria for validation of chemical categories. Instead the document provides guidance on how to optimise the robustness of chemical categories and how to document the justification for each category.

R.6.2.1.6 The interdependence between categories and QSARs

The chemical category and QSAR concepts are strongly connected. The concept of forming chemical categories and then using measured data on a few category members to estimate the missing values for the untested members is a common sense application of QSAR. The reason this concept is so compatible with QSAR is that this broad description of the categories concept and the historical description of QSAR are one and the same (see Figure R.6-4).

A *Quantitative Structure-Activity Relationship* (QSAR) is a quantitative (mathematical) relationship between a numerical measure of chemical structure, and/or a physico-chemical property, and an effectpactivity (Figure R.6-4). QSARs often take the form of regression equations, and can make predictions of effects/activities that are either on a continuous scale or on a categorical scale. Thus, in the term QSAR, the qualifier *quantitative* refers to the nature of the relationship, not the nature of the endpoint being predicted. An example of a QSAR is the prediction of acute toxicity to an invertebrate species (*Tetrahymena pyriformis*) by means of a regression equation with the partitioning behaviour (log K_{ow} value) of the chemical as a descriptor (Schultz *et al*, 2002).

Similarly, a *Quantitative Activity-Activity Relationship* (QAAR) is a mathematical relationship, but between two biological endpoints (Figure R.6-4), which can be in the same or different species. QAARs are based on the assumption that knowledge about the mechanism or mode of action, obtained for one endpoint, is applicable to the *same* endpoint in a different species, or to a similar endpoint in the same species, since the main underlying processes are the same (e.g. partitioning, reactivity, enzyme inhibition). QAARs provide a means of performing trend analysis and filling data gaps[22].

[22] The experience with QAAR is currently limited and therefore this approach has not been routinely used. The concept is presented in this document for completeness sake. Further experience in the application of this concept will lead to revisions of this document.

Figure R.6-4: Graphical representation of a QSAR/QAAR

A QSAR can make extrapolations from chemical structure and/or physico-chemical properties to other properties or activities. A QAAR makes an extrapolation from one activity to another related activity.

	Chemical 1	Chemical 2	Chemical 3	Chemical 4
Structure	xxxxxxxxx	xxxxxxxxx	xxxxxxxxx	xxxxxxxxx
Property 1	●	●	●	●
Property 2	○	○	○	○
Activity 1	●	○	○	○
Activity 2	●	●	●	●
Activity 3	○	○	●	○

● Existing data point ○ Missing data point

The common scientific foundation between forming categories and QSARs/QAARs is that chemicals, once grouped together on a basis of common structural attributes, become chemical classes which 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 QSARs.

In many cases, QSARs are quantitative models of key mechanistic processes which result in the measured activity of the chemicals. The importance of this mechanistic understanding is two fold. First, the structure-activity relationships provide useful models for hypothesis testing which increases the reliability and causality of the QSAR model. Secondly, the mechanistic understanding can be described as a series of structural requirements which define the mechanism boundaries on reliable domain of application of QSAR model.

The categories concept creates a practical and powerful approach for describing these structural requirements of toxicity mechanisms. Chemicals can be grouped together initially using expert judgement which is reflected by the chemicals included. Further discussion may question the similarity of some chemicals based on measured data, evidence of anomalous behaviour or other information about the chemical attributes which suggest some chemicals may fit more than one category. The careful use of expert judgement to define the boundaries of a chemical category is crucial to the reliable application of QSAR models or other methods to estimate values for untested chemicals. A formal definition of which chemicals should be included in a category and which chemicals should be excluded can lead to much more reliable estimates of missing values than the use of QSAR models with poorly defined domains. The expert judgement should be described in a transparent manner in order to be evaluated by others.

A QSAR estimate is the result of an assumption and a prediction about the chemical. The assumption is that of the predominant interaction mechanisms of the chemical, and thus leads to selection of a QSAR model. The prediction is the quantitative estimation of the intensity, or potency, of the chemical structure within the specific mechanisms of interaction. Both the assumption of mechanism and the prediction bear heavily on the reliability of that overall QSAR estimation.

However, the errors created in selecting the proper QSAR model for a specific chemical are greater than those related to the potency estimate of the QSAR model. For example, in ecotoxicity studies, some phenols are polar narcotics, some are uncouplers, and others are electrophilic. QSAR models for each mechanism have comparable uncertainty, but the potency of the latter mechanism can be orders of magnitude greater than polar narcotics. The use of a category approach can thus help to ensure that the QSAR estimates are based on mechanistically valid models by aiding correct selection of the model.

Further information on the use of (internal) QSARs to express trends in categories, and on the use of (external) QSARs to provide additional support for trends, is given in Section R.6.2.2. and section R.6.2.2.3, respectively.

Within a chemical category, the primary difference between hazard identification and classification and labelling is that the classification and labelling is performed in the context of risk management thresholds established by the regulator. It is possible that the risk management threshold is defined simply as a positive test result in a hazard identification test guideline and the majority of a category would be expected to be classified similarly. However, if the risk management threshold is a specific value along a large range of possible potency values for a specific hazard endpoint, it is reasonable to expect some member to be above and below that threshold and still belong to the chemical category. For classification and labelling, the QSAR models may be designed to either provide a potency estimate or to estimate the likelihood that the potency would be above or below the risk management threshold.

Estimation methods work best for homologous series of chemicals where the metric for extrapolating from one chemical to another is a simple molecular weight, number of carbon atoms or a similar parameter which can be linked to physico-chemical properties of the chemicals. However, when the members of the category are not a simple homologous series, it is essential that some parameter which predicts the trend across the members be established in order to extrapolate the measured values to the missing values. For example, the vapour pressure is mechanistically related to the acute inhalational toxicity (LC_{50}) of ethers because it is a surrogate for the thermodynamic activity of the chemical in the blood and tissues (Hart and Veith, 2007); but it is not directly related to carbon number or molecular weight because the degree of branching is significantly different among the category members. An estimate using carbon number would not produce defensible extrapolations within this category. In contrast, vapour pressure is a more reliable parameter to extrapolate the results from measured values to missing values.

In addition to the concern over which parameter to use in the estimation, it is necessary to make an assumption about the proportionality factor so that the structural differences between a measured and unmeasured chemical can be proportioned into a difference in toxicity. For example, the acute inhalational toxicity (LC_{50}) of ethers does not increase with vapour pressure with a proportionality of 1.0, but rather with a proportionality of 0.7 (see example taken from Hart and Veith, 2007). The advantage of a more rigorous use of QSAR models within categories is that one can base the estimate in the large context of a mechanistic model where the parameter for extrapolation and the proportionality factor(s) are easily justified and explained in transparent terms.

R.6.2.2 Approaches to data gap filling in chemical categories

The absence of relevant, reliable and sufficient experimental data for a chemical, results in one or more data gaps which need to be filled in order to finalise the hazard and/or risk assessment. This chapter explains the following non-testing techniques for filling data gaps:

- read-across
- trend analysis and use of computational methods based on internal models
- use of computational methods based on external models

In principle, these techniques can be used to indicate either the presence or the absence of an effect. In certain cases, the application of these techniques to assess a particular chemical may benefit from the generation of test data for one or more other chemicals in the category. In other words, the generation of additional experimental data by strategic testing may be useful

In this document, the term *model* refers to any formalised method for estimating the properties of chemicals, and typically refers to a QSAR, QAAR or expert system. These models are only useful for data gap filling when they are based on data of sufficiently high quality. This is particularly important when applying a model to the interpretation of boundary substances.

The use of these three techniques is described in more detail below. It should however be recognised that whilst these three techniques are described separately in the following section, there are many elements that are common to all three approaches. All three techniques can be used with varying degrees of applicability in the context of both the analogue approach and the wider category approach. Experience from current practice shows that the first of these three techniques, the use of qualitative or quantitative read-across is already widely used and is often accepted as a valid approach for regulatory purposes. Whilst computational approaches based on SARs, QSARs, QAARs or expert systems can also provide a basis for filling data gaps, experience shows that additional supporting evidence is often required for acceptance of these estimates.

R.6.2.2.1 Read-across

In the *read-across technique*, endpoint information for one chemical is used to predict the same endpoint for another chemical, which is considered to be *similar* in some way (usually on the basis of structural similarity). In principle, read-across can be applied to characterise physico-chemical properties, environmental fate, human health effects and ecotoxicity. For any of these endpoints, read-across may be performed in a qualitative or quantitative manner. In practice, read-across for basic physico-chemical properties is not generally recommended, since reliable data should normally be available or easily obtainable, does not involve the use of animals and provides key information for the assessment of a chemical. However, there may occasionally be practical problems, especially for UVCBs, when the use of these techniques will be required.

Within a group of chemicals, read-across can be performed in the following ways to fill data gaps:

- one-to-one (one analogue used to make an estimation for a single chemical)
- many-to-one (two or more analogues used to make an estimation for a single chemical)
- one-to-many (one analogue used to make estimations for two or more chemicals)
- many-to-many (two or more analogues used to make estimations for two or more chemicals)

The transition between comparisons using an analogue approach involving more than two chemicals and a more comprehensive category approach described in the following chapter is of

course arbitrary. The guidance on read-across given below applies both to the analogue approach described in Section R.6.2.3 as well as to the categories approach described in Section R.6.2.4

It should be recognised that the robustness of a category approach would be expected to be considerably greater than that of an analogue approach, since the basis for evaluating any individual chemical in the category is greater, and there is usually more measured data available in such a wider approach. The following sections contain guidance particularly with respect to supporting information that is more relevant for the use of an analogue approach, as a category approach will in itself provide additional support for the robustness of the estimates.

A chemical being used to make an estimate can be referred to as a *source chemical*, whereas a chemical for which an endpoint is being estimated can be referred to as a *target chemical*.

Read-across can be qualitative or quantitative. In *qualitative read-across*, the presence (or absence) of a property/activity for the target chemical is inferred from the presence (or absence) of the same property/activity for one or more source chemicals. Qualitative read-across gives a 'yes/no' answer. In *quantitative read-across*, the known value(s) of a property for one or more source chemicals is used to estimate the unknown value of the same property for the target chemical. Quantitative read-across is used to obtain a quantitative value for an endpoint, such as a dose-response relationship.

Most often, structural similarity and similar properties and/or activities between chemicals is used as a basis for read-across. Thus, endpoint information is read-across from a structural analogue. A structural analogue is a source chemical whose physico-chemical and toxicological properties are likely to be similar to the target chemical as a result of structural similarity. The similarity may be based on the following:

- a common functional group (e.g., aldehyde, epoxide, ester, metal ion). An example is the ethylene glycols category assessed in the OECD HPV Chemicals Programme (http://cs3-hq.oecd.org/scripts/hpv),
- a common precursor and/or breakdown product, that results via physical or biological processes (metabolic pathway similarity). This is used to examine related chemicals, such as acid/ester/salt. Examples are certain azo dyes based on carcinogenic components such as benzidine or other carcinogenic aromatic amines, where the carcinogenic aromatic amine is formed by the metabolism of the dye.

Analogies between chemicals can also be drawn on the basis of common mechanisms of action and similarities in chemical (or biochemical) reactivity.

In principle, it is possible to predict the presence or absence of a property/effect by applying the read-across approach. Read-across from a negative result is regarded as equally valid and convincing as a positive result provided the test design, concentrations tested etc. have been chosen adequately. For example, if all tested chemicals of a category are shown not to be mutagenic and if there is scientific justification that the untested chemical rightly belongs in the category, it is justified to assume that the untested chemicals are also not mutagenic. However, if the mutagenicity test system that has been used is inappropriate to demonstrate the genotoxicity of the group of chemicals, then a conclusion that the category would not be mutagenic would not be valid. There is extensive experience of read-across of negative findings or absence of effect in the EU risk assessment and classification and labelling work and the OECD HPV Programme. For example, in the ESR risk assessment of medium-chain chlorinated paraffins, data from the short-chain chlorinated paraffins was used as supporting evidence for lack of genotoxicity, low acute dermal

toxicity and absence of skin sensitisation potential. It is particularly important to adequately justify read-across of negative findings. The read-across approach is most robust when a quantitative trend between the analogues can be established.

A stepwise approach for performing read-across on a limited number of chemicals (analogue approach) is given in Section R.6.2.3. The use of this approach for filling data gaps in a larger category approach is shown in Section R.6.2.4.

a) Choice of qualitative or quantitative read-across

Before deciding on the type of read-across approach which is necessary, it is important to determine why the data gap is being filled and what type of data is required. Is a specific value required or does the endpoint need to be checked against a threshold or hazard banding/cut off (for example a classification banding)? Read-across has been used for a range of different reasons to date. For example:

- To fill a data gap for a specific endpoint - both threshold and non-threshold values[23]
- To reduce an assessment factor used to derive a PNEC[24]
- To flag a concern for further testing[25]
- To read-across classification and labelling[26]

In deciding on whether to use quantitative or qualitative read-across, the nature of the property should also be considered. It may be expressed on a numerical or categorical scale. In most cases, a specific value is required for risk assessment, such as a NOEC or NOAEL, environmental half-life or partition coefficient. A numerical value obtained by quantitative read-across would normally be needed. For conducting a hazard assessment, PBT assessment or assigning classification and labelling, one generally needs to know whether that substance fits the particular hazard criteria. Identification of the hazard by qualitative read-across may be adequate.

Under REACH, the result of read-across should be adequate for classification and labelling, risk assessment or PBT (vPvB) assessment, which implies the need for both qualitative and quantitative read-across, depending on the particular situation.

An issue that may arise when read-across is carried out in the context of a category is that the experimental results for different category members may be available for different test methods or species relating to the same general endpoint. For example, in the case of reproductive toxicity, only screening studies may be available for some category members, whereas two-generation studies may be available for other members. As the estimated results from the category approach have to be useful for risk assessment and classification, the uncertainty associated with the underlying results has to be ascertained. It is clear that the scope of the estimated results for a member of a category

[23] For example, the ESR risk assessment of short chain chlorinated paraffins CAS 85535-84-8 where the NOAEL for effects via lactation was read-across from medium chain chlorinated paraffins.

[24] For example, the ESR risk assessment of medium chain chlorinated paraffins CAS 85535-85-9 where aquatic toxicity data from short chain chlorinated paraffins was used to show invertebrates are most sensitive and thus reduce the assessment factor from 50 to 10 to derive the PNECaquatic.

[25] For example, the ESR risk assessment of p-t-butylphenol CAS 98-54-4 where data from p-t-pentylphenol were used to request further testing on endocrine disruption in fish (Tsakovska and Worth, 2007).

[26] For example, the common EU classifications for skin irritation and sensitisation agreed for sulphate, dichloride, nitrate and carbonate salts of nickel (Hart, 2007).

cannot exceed the scope of the underlying data for the other members of the category, e.g. if for genotoxicity, only *in vitro* results are available for some members of the category (source chemicals), only conclusions on *in vitro* genotoxicity can be reached for the members of the category for which experimental results are lacking (target chemical). If the scope of the underlying experimental results for an endpoint vary (e.g. a mix of results from screening tests and higher tier tests), it is necessary to clarify the scope of the estimated results for the category members for which no experimental results are available. It may be possible to apply a *Weight of Evidence* approach to all the data, which could lead to the same hazard identification for all the members of the category, irrespective of the data available for the individual compounds.

b) Qualitative read-across

In qualitative read-across, the presence or absence of a property is inferred from the established properties of one or more analogues. The main application of qualitative read-across is in hazard identification, and usually results in the allocation of the target chemical(s) to the same hazard category as the source chemical(s).

The arguments to support the read-across are normally based on expert (eco)toxicological judgement. Several factors can be considered in making this judgement. The assumption that a common substructure is responsible for the common property or effect could be affected by interactions between the substructure and other parts of the chemical structure. Another substructure could alter the property/effect in a qualitative manner (in which case the assumption may be false) or a quantitative manner (i.e. change the degree to which the substance exhibits the property). One example could be changes in the degree of branching of a carbon chain which can affect biodegradability and toxicity. In addition to interactions between substructures, differences in one or more whole-molecule properties could alter the assumption of commonality (e.g. differences in aqueous solubility could affect the read-across of a classification for aquatic toxicity). These factors are assessed by a process of expert judgement. An example is the read-across of carcinogenicity for musk ketone, which was evaluated by the SCHER (2006).

If a regulatory classification is used to express the property or effect, a quantitative change in the potency of the chemical could be sufficient to warrant a different classification, depending on the classification threshold. If a difference in the potency between source and target chemicals is suspected, for example based on trends in the available data, a quantitative read-across approach rather than a qualitative approach would usually be required. This is particularly important where the target chemical is suspected to have a more stringent classification than the source chemical. A different classification can be considered where the classification criteria are based on the strength of the available evidence rather than a quantitative cut off. In addition, differences between a direct and an indirect effect can lead to a different classification of the target chemical than the source chemical. An example is the classification of benzidine azodyes as category 2 carcinogens whilst benzidine itself is classified as a Category 1 carcinogen.

c) Quantitative read-across

In addition to identifying a particular property for a target chemical, in quantitative read-across the known value of a property for the source chemical(s) is also used to estimate the unknown value of the same property for the target chemical.

When applying quantitative read-across, there are four general ways of estimating the missing data point:

- by using the endpoint value of a source chemical, e.g. the closest analogue in a

(sub)category[27]

- by using an internal QSAR to scale the available experimental results from two or more source chemicals to the target chemical[28]
- by processing the endpoint values from two or more source chemicals (e.g. by averaging, by taking the most representative value)
- by taking the most conservative value of the closest analogues or the most conservative value in the (sub)category[29]

Quantitative read-across can also be utilised for complex substances/UVCBs, typically by applying data from physico-chemically similar substances (e.g. substances with similar boiling ranges, carbon ranges, composition) or by applying data from key/major constituents. However, this must be done carefully, may be more applicable for indication of ranges and requires an understanding of the key structures that may drive the behaviour of UVCBs. This is further discussed in Chapter R.6.2.5.5.

In risk assessment, a dose descriptor is used as a quantitative basis for deriving a Predicted No Effect Concentration (PNEC) or Derived No Effect Level (DNEL), depending on the endpoint. To account for various sources of uncertainty in the derivation of the PNEC or DNEL, an assessment factor is applied to a numerical value of the dose descriptor (see Chapter R.8 and R.10 for guidance on deriving PNECs and DNELs).

When conducting a risk assessment, a NOAEL, NOEC or other effect concentration such as EC_{10} may be read-across in order to derive a DNEL or PNEC for the target chemical, provided that this is justified. Read-across of the PNEC or DNEL itself from the source to target chemical is not recommended since the range of available data for a chemical must be considered when deriving the DNEL or PNEC. The size of the assessment factor used to derive a PNEC or DNEL depends on the confidence with which it can be derived from the available data. Generally, lower assessment factors can be used with larger more relevant datasets.

When deriving a DNEL or PNEC based on an endpoint which has been read-across, it is important to ensure that the read-across is sound and that the target chemical is unlikely to be more potent than the source chemical. In cases where there are multiple source chemicals, and consequently a range of possible values for read-across, the use of the most conservative (lowest) value may be sufficient to account for the uncertainty in the read-across. In particular, the read-across is likely to be conservative when the target chemical has a lower bioavailability than the source chemical. If there is any uncertainty in the read-across, and thus the DNEL or PNEC derived from it, it may be necessary to conduct testing for that endpoint.

In the ESR risk assessment of medium chain chlorinated paraffins CAS 85535-85-9, aquatic toxicity data from short chain chlorinated paraffins was used to show that invertebrates are most

[27] For example, the OECD HPV Gluconates category, where aquatic toxicity data for Sodium D-gluconate were read-across to the calcium and potassium salts, D-Gluconic acid and Glucono-delta-lactone (Caley *et al*, 2007).

[28] For example, OECD HPV C_{6-22} Aliphatic Alcohols category where internal QSARs were developed to predict aquatic toxicity based on Kow and thus derive aquatic toxicities for the target chemicals (http://cs3-hq.oecd.org/scripts/hpv/).

[29] For example, the ESR risk assessment of Zinc distearate used aquatic toxicity data from the more soluble zinc salts (chloride, sulphate) to derive the PNECaquatic for Zinc distearate (Tsakovska and Worth, 2007).

sensitive and thus reduce the assessment factor from 50 to 10 to derive the $PNEC_{aquatic}$ despite the fact that no chronic fish test was available for medium chain chlorinated paraffins.

There is no experience to date with the use of DNELs for human health risk assessment so further guidance should be developed on the use of read-across data in DNEL derivation once experience is gained with its use.

In cases where there are concerns that the relative potency of the different chemicals may be sufficiently large to affect the conclusions of either hazard identification (in cases where the criteria contain a quantitative cut off) or risk assessment (based on an estimated PNEC/DNEL), additional testing specifically designed to demonstrate differences in potency across a category can be considered.

d) Choice of endpoints for the application of read-across

In principle, read-across can be applied for any property or endpoint, irrespective of whether it is a physico-chemical property, environmental fate parameter, human health effect, or ecotoxicological effect.

In practice, read-across is not encouraged for basic physico-chemical properties (e.g. water solubility, log K_{ow}) since these properties provide key information for the assessment of a chemical in particular for the assessment of the environmental properties, and experimental data or valid QSAR predictions should normally be available (or should be reasonably obtainable).

e) General considerations when performing read-across

Irrespective of the type of read-across, it is important to consider a number of factors (Hanway and Evans, 2000):

- Whether the data point of the source chemical is relevant and reliable for the purpose of the read-across. If read-across data have not been produced using the most current OECD test methods, particularly careful consideration of the quality and suitability of a method is important.
- Whether the source and/or target chemical is a multi-functional compound and whether the additional functionality may therefore affect the reliability of the read-across.
- The purity and impurity profiles of the target and source chemicals need to be assessed. There is a need to identify those impurities which might influence the overall toxicity of the source chemicals and to discuss the consequences these impurities will have for the robustness of the chemical category and hence for the read-across. If all category members have the same sort of impurities, then they may not have any relevant influence on the read-across. If there is a very biologically active impurity (e.g. CMR substances) in one category member, but not the other members, then the results from that category member might not be appropriate for read-across.
- Comparison of the physico-chemical properties of the target and source chemicals, particularly the physical form, molecular weight, water solubility, particle size and structure30, partition coefficient and vapour pressure, provides useful information as to their similarity.
- The likely toxicokinetics of the substances, including the possibility of different metabolic pathways coming into play, needs to be considered where possible.

30 There is debate ongoing on the regulatory application (classification and derivation of dose-descriptors).

- Information from valid (Q)SARs may be used where possible to inform decisions on the need, extent and type of additional testing.

In the case of UVCBs (Section R.6.2.5.5), it should be considered whether the differences between the UVCBs in a specific group would actually give rise to different effects, bearing in mind the internal consistency of the basic structural families and assumption of similarity of action or reaction.

f) Supporting information

It is important to provide supporting information to strengthen the rationale for the read-across. Thus, in addition to the property/endpoint being read-across, it is also useful to show that additional properties, relevant to the endpoint, are also (qualitatively or quantitatively) similar between the source and target chemicals. Such properties could be known or suspected determinants of the endpoint, or they could be limiting factors.

Relevant molecular properties of the source chemical should be of comparable value to those of the target chemical. The selection of relevant molecular properties depends on the endpoint for which the read-across is being performed. The identification of these properties could be based on expert knowledge, or could be based on the use of properties (molecular descriptors) that have been found to be useful predictors of the endpoint in QSAR models.

In the case of single substances, irrespective of the endpoint being read-across, useful considerations might include:

- the presence or absence of additional functional groups or substituents that could influence the behaviour of a chemical.
- similarity in physico-chemical profiles (e.g. molecular weight, log Kow, water solubility).
- similarity in other toxicological and/or ecotoxicological data.
- the likely toxicokinetics of the substances, including the possibility of different metabolic pathways coming into play, needs to be considered where possible.
- information from valid (Q)SARs may be used where possible to inform decisions on the need, extent and type of additional testing.

In cases where there are convincing arguments for a read-across approach, the need to generate new data with tests on vertebrates should require a strong and convincing argument, whether to remove an unwanted classification or confirm a non-classification. In such cases, if test data demonstrate the measured value differed considerably from the estimated, the read-across and the resultant category, if applicable would have to be carefully reconsidered. A *Weight of Evidence* analysis (Section R.6.2.2.4) may be useful for determining whether the read-across or the test data was suspect.

g) Supporting information for environmental endpoints

What constitutes appropriate supporting information will depend on the environmental endpoint being read-across. However, basic physico-chemical properties that determine environmental distribution and fate (e.g. molecular weight, partition coefficients such as log K_{ow}, water solubility) will generally be useful. Particle size and structure[30] may also be relevant.

For example, in the case of aquatic toxicity, similar log K_{ow} and aqueous solubility values between the source and target chemicals could be used to support the read-across, because log K_{ow} is known

to be a determinant of the toxicity in aquatic organisms when the effect is mediated by mechanisms of narcosis. If the chemical is known or expected to act by a non-narcotic mode of action, additional properties might provide useful supporting information. For example, experience with new chemicals in the EU suggests that tests such as acute toxicity to Daphnia can provide additional confidence that read-across of other data is possible, i.e. if toxicity differences are found between the source and target chemical then further testing for other endpoints may be appropriate (Hanway and Evans, 2000). The acute Daphnia toxicity test raises few animal welfare issues while providing good confirmation of the comparability of aquatic toxicity.

Furthermore, in the case of read-across of aquatic toxicity endpoints, results (fish, invertebrates and algae) for source and target chemicals should be compared. For example if a read-across to acute toxicity to fish is based on a presumed mode of action, and if this mode of action is applicable to invertebrates and algae, the available results for invertebrates and algae for the source and target chemicals should confirm the applicability of the read-across.

h) Supporting information for human health endpoints

What constitutes appropriate supporting information will depend on the human health endpoint being read-across. However, physico-chemical properties that determine biokinetics and bioavailability (e.g. molecular weight, partition coefficients such as log K_{ow}, water solubility, pH, vapour pressure, viscosity) will generally be useful. Particle size and structure[30] may also be relevant.

In general, current practice relies heavily on expert judgement. The type and amount of supporting evidence needed may vary with the endpoint concerned.

In the case of musk ketone, the target chemical, read-across for carcinogenicity can be based on the data for musk xylene, the source chemical (SCHER, 2006). Important considerations for the read-across were:

- musk ketone (the target chemical) has similar physico-chemical properties as musk xylene (the source chemical)
- there are structural similarities between the two chemicals
- both chemicals have been tested for mutagenicity; neither chemical is genotoxic
- both nitro musks are inducers of cytochrome P450 2B1

However, musk xylene effects on the liver cytochrome P450 activities are different from those of musk ketone. While both musk xylene and musk ketone induce CYP 2B gene expression, the induced cytochrome P450 2B protein is present in an inactivated form after musk xylene administration resulting in a much lower CYP 2B1 associated catalytic activity. Due to its chemical structure, musk ketone cannot be reduced to an enzyme inhibiting p-amino metabolite and therefore induces, but does not inactivate CYP 2B enzymes in mice. Hence, high levels of active cytochrome P450 2B are present after administration of musk ketone.

- The mode of action of musk xylene in both mice and rats seems to be identical, while some species differences in the pattern of cytochrome P450 induction by musk ketone are observed
- The role of enzyme induction in the development of liver tumours by musk xylene in mice and in the toxicity of repeated administration of musk ketone is not well defined.
- There are similarities of the effects of both musk xylene and musk ketone to effects of phenobarbital, which also induces liver tumours in rodents by a non-genotoxic mode of

action and is also an inducer of cytochrome P450 2B.

- Assuming that the induction of cytochrome P450 2B is a relevant mode of action for liver tumours induction by musk xylene, read-across based on enzyme induction and structural and physico-chemical properties may be sufficient as a basis for read-across since musk ketone is also an inducer of this enzyme. More detailed information on the mechanisms of enzyme induction by musk ketone is not available.

For some endpoints, such as skin sensitisation or mutagenicity, chemical reactivity might provide useful supporting information. For skin sensitisation, one of the necessary hurdles a chemical has to undergo is to form a stable association with a skin protein. This is thought to be a covalent association where the chemical behaves as an electrophile and the protein as a nucleophile. A similar analogy is relevant for mutagenicity but where DNA represents the nucleophile. An experimental system that quantifies the electrophilic reactivity would be useful to support a read-across for skin sensitisation, (Aptula *et al*, 2006) or mutagenicity (Benigni *et al*, 2005).

In vitro data might also provide useful supporting information. For example, if acute mammalian toxicity is being read-across, it might be appropriate to refer to similarity of *in vitro* cytotoxicities of the source and target chemicals, if it is known (or suspected) that cytotoxic effects underlie the acute systemic effect. Relationships between *in vitro* cytotoxic effects and acute systemic toxicity has been investigated by a number of workers (e.g. Clemedson *et al*, 2002).

R.6.2.2.2 Trend analysis and computational methods based on internal models

For a given category endpoint, the category members are often related by a trend (e.g. increasing, decreasing or constant). The trend could be related to molecular mass, carbon chain length, or to some other physico-chemical property. For larger categories, it is possible that several different relationships can be established for a single endpoint, thereby defining subcategories. A chemical that identifies a turning point in a trend is called a *breakpoint chemical* (see also Section R.6.2.1.2). Category members falling at the opposite extremes of a trend and within which interpolations are considered reliable are called *sentinel (boundary) chemicals*.

A demonstration of consistent trends in the behaviour of a group of chemicals is one of the desirable attributes of a chemical category and one of the indicators that a common mechanism for all chemicals is involved. When some chemicals in a category have measured values and a consistent trend is observed, missing values can be estimated by simple scaling from the measured values to fill in the data gaps.

The observation of a trend (increasing, decreasing or constant) in the experimental data for a given endpoint across chemicals can be used as the basis for interpolation and possibly also extrapolation (see Figure R.6-3). Interpolation is the estimation of a value for a member using measured values from other members on *both sides* of that member within the defined category spectrum, whereas extrapolation refers to the estimation of a value for a member that is near or at the category boundary using measured values from internal category members. Interpolation between measured analogues may give a more reliable result depending on the reliability of the measured data. Interpolation can be performed when the series of values is monotonic (all increasing or decreasing) or when non-monotonic (e.g. parabolic). In such circumstances the extent to which the available data describe the trend will determine the level of confidence in the prediction.

In general, interpolation between category members is preferred to extrapolation. However, it may be the case that whilst data is available for several members of a category, there can be data gaps for the boundary chemical. In this case extrapolation will be necessary. It should be noted that extrapolation based on a clearly established trend will be substantially more robust than the use of

read-across from analogues to fill a data gap. The robustness of any extrapolations used to fill data gaps will be closely related to the general evaluation of the whole category.

When establishing trends in data, laboratory and experimental variations should be considered. Similar species/strains, endpoints and test protocols should be compared. Deviations from a trend should be clearly identified and possible reasons for the deviations laid out in the category analysis.

In principle, it is possible to predict the presence or absence of a property/effect by applying trend analysis The category approach is most robust when a quantitative trend between the category members can be established. A lack of observed toxic effects for a chemical substance in a study of a specific endpoint (especially if no dose-relationship can be established because no effects are observed at some of the doses tested) requires further consideration and in such circumstances, the data need to be carefully evaluated. It is important to distinguish between cases where the lack of response can be explained on the basis of the mechanistic understanding for that endpoint, or whether the tests have failed to demonstrate the absence of an effect for the category as a whole.

The larger the category, the more likely that there may be breaks in trends which may affect the reliability of interpolation or extrapolation. The observation of a *break* in a trend among some members of a category is a warning sign, but is not necessarily an indication that the chemicals with different trends exhibit different toxicity pathways. Bioassay measurements frequently are only comparable over a narrow range of chemical properties with the result that different pharmacodynamic factors are controlling the bioassay results for different chemicals. The bilinear or multilinear nature of trends in measured data, if observed, can be used to confine the methods for scaling intensity of the endpoint to specific members of the category.

The observation of a trend *break* should not be confused with differences in the hazard classification of the members of a category. When the cut off dividing different classification bands is between the extreme values of the trend, then the members of the category will be classified differently. If all members of the category have properties above or below the administrative cut off agreed for that property, the trend analysis may be useful for judging the adequacy of forming the category but apparent breaks in the trends would not lead to differences in the classification.

There is little current experience in the use of the type of formal trend analysis shown here. However, there is good reason to believe that arguments based on this approach would be acceptable to estimate missing data, and that this technique provides a basis for a robust estimate.

The data for a particular endpoint can be used to construct a QSAR that describes the properties of the members of the category. A Quantitative Structure-Activity Relationship (QSAR) is a quantitative (mathematical) relationship between a numerical measure of chemical structure, or a physico-chemical property, and an effect/activity. QSARs often take the form of regression equations, and can make predictions of effects/activities that are either on a continuous scale or on a categorical scale. Thus, in the term QSAR, the qualifier *quantitative* refers to the nature of the relationship, not the nature of the endpoint being predicted.

An example of a QSAR is the prediction of acute toxicity to an invertebrate species (*Tetrahymena pyriformis*) by means of a regression equation with the partitioning behaviour (log K_{ow} value) of the chemical as a descriptor (Schultz *et al*, 2002).

A trend might also be expressed as a quantitative activity-activity relationship (QAAR). A Quantitative Activity-Activity Relationship (QAAR) is a mathematical relationship between two biological endpoints, which can be in the same or different species. QAARs are based on the assumption that knowledge about the mechanism or mode of action, obtained for one endpoint, is

applicable to the *same* endpoint in a different species, or to a similar endpoint in the same species, since the main underlying processes are the same (e.g. partitioning, reactivity, enzyme inhibition).

Thus, a chemical category can be seen as a set of *internal* QSARs (and possibly also internal QAARs) for the different endpoints, with the advantage that all the underlying data are transparently available to the assessor. Such models provide quantitative descriptions of the trends within a category and are referred to as *internal* QSARs (or QAARs) because they are derived directly from the experimental data for the category members. These models are also likely to be *local* models in the sense that they are based on a relatively small data set. Such an internal local model was for example developed for acute aquatic toxicity for the category of long-chain alcohols ($C_{6\text{-}22}$ primary aliphatic alcohols) assessed within the OECD HPV Chemicals Programme (http://cs3-hq.oecd.org/scripts/hpv).

Such methods work best for homologous series of chemicals where the metric for extrapolating from one chemical to another is a simple molecular weight, number of carbon atoms or a similar parameter which can be linked to physico-chemical properties of the chemicals. However, when the members of the category are not a simple homologous series, it is essential that some parameter which predicts the trend across the members be established in order to extrapolate the measured values to the missing values. For example, the vapour pressure is mechanistically related to the acute inhalational toxicity (LC_{50}) of ethers (Hart, 2007) because it is a surrogate for the thermodynamic activity of the chemical in the blood and tissues; but it is not directly related to carbon number or molecular weight because the degree of branching is significantly different among the category members. An approach using carbon number would not produce defensible extrapolations within this category. In contrast, vapour pressure is a more reliable parameter to extrapolate the results from measured values to missing values.

R.6.2.2.3 Computational methods based on external models

In this guidance, the term *external model* is used in distinction to the *internal model* described in the section above and can refer to any model (QSAR, QAAR or expert system) that was not developed as part of the category formation process. If such models are used to fill data gaps in a category, they should be based on experimental data that are obtained from a wider range of chemicals than those used in the category. Such external models are also known as "global models" since the data on which they are based comes from a relatively large number of chemicals in comparison with those in the category. In this sense, the category under evaluation is a subcategory of this wider QSAR. More guidance exists on the availability and use of specific (Q)SARs for individual endpoints in other chapters, e.g. R.6.1.8.

USE OF COMPUTATIONAL METHODS FOR SUPPLEMENTING EXPERIMENTAL DATA

The predictions made by an external model may be used to provide additional support for the trend (even though reliance is placed on the experimental data rather than the model estimates). To be applicable the prediction should be considered as reliable and the comparison between the predicted value and the experimental value available for other members of the category or the analogue should be taken into account. For example, a parabolic QSAR could be used to characterise the trend in bioconcentration factor (BCF) values across a series of chemicals of increasing molecular weight.

In other cases, model predictions may be used to identify and rationalise category members that deviate from a trend. For example, a QSAR or expert system might indicate that certain chemicals

in a series have anomalous behaviour due to metabolism, although this would need to be confirmed by consideration of the biological plausibility of the differences.

If multiple experimental data are available for a single substance, the result of a computational model can be helpful in choosing a valid data point.

The result of one or more computational models can be used to increase the confidence in an experimental measurement for a single substance. For example, within the ESR, estimated results obtained with two QSAR models for biodegradation were used to support an experimental observation of ready biodegradability for acrylaldehyde (Tsakovska and Worth, 2007).

R.6.2.2.4 Weight of Evidence considerations

Since the data used in a hazard or risk assessment should be relevant, reliable and sufficient for the regulatory purpose, it is necessary to base the assessment on the totality of available information, i.e. to apply *Weight of Evidence* (WoE) considerations. The WoE assessment can be based on experimental data as well as estimated data (obtained by applying one or more non-testing approaches). In most cases, estimated data might be used to supplement and increase confidence in the available experimental data, whereas in some others, such data might be used instead of experimental data.

Guidance on how to apply WoE consideration in hazard and risk assessments is provided in Chapter R.6.2.3, and in the OECD Manual for Investigation of HPV Chemicals (OECD, 2007b).

R.6.2.3 Guidance on a stepwise procedure to perform the analogue approach

The guidance in this chapter primarily provides guidance on how to estimate missing data from a single or limited number of compounds using the analogue approach.

This guidance is primarily based on the widespread current experience in the application of the analogue approach using non-formalised approaches. However, the guidance also provides indications of where computer-based methods can be included to facilitate the process. A stepwise approach to analogue evaluation is proposed, in which the use of formalised computational approaches can be integrated.

In the EU, there is considerable experience in the application of read-across using the analogue approach in the classification and labelling group (See Appendix 9 in the Tapir final report: ECB, 2005; Comber and Simpson, 2007; Gallegos Saliner *et al*, 2007; Hart, 2007; Hart and Veith, 2007; Schoeters and Verougstraete, 2007). More recently additional experience has been gained in the risk assessment of Existing Chemicals (ESR programme; (Tsakovska and Worth, 2007), and in the Notification of New Substances (NONS programme; Hanway and Evans, 2000).

There is also considerable experience on the use of analogue approaches in the OECD HPV programme and by the US-EPA (See Appendix 9 in the Tapir final report: ECB, 2005). Within the OECD HPV Chemicals Programme, read-across has been extensively performed since 1998. Examples of initial hazard assessments that rely on data from analogues, and which have been published, include: isobutanol (CAS No 78-83-1), p-chlorotoluene (CAS No 106-43-4), and methyltriacetoxysilane (CAS No 4253-34-3). These initial assessments are available from UNEP Chemicals (http://www.chem.unep.ch/irptc/sids/OECDSIDS/sidspub.html).

Much of this experience has taken place in the context of consultation in either the EU Technical Committees or at the OECD, and reflects a consensus on the use of expert judgement between experts from the Member States.

The current practice in the EU is often based on an empirical identification of an appropriate analogue. The choice of analogue is normally fairly straightforward, as any potential analogue has to be data-rich in order to form a basis for comparison. In many cases the choice is governed by the availability of data on an analogue manufactured by the same producer or an analogue where data is available from detailed regulatory evaluations (OECD HPV programme or the EU ESR programme) or from the open literature. For example, under the EU ESR programme, data for ETBE was estimated by comparison with the data collected for MTBE and TAME (Tsakovska and Worth, 2007).

It is foreseen that read-across using non-formalised methods within the analogue approach will continue to be the more frequently used method for filling data gaps over the next few years. Based on a learning-by-doing approach, the experience gained in application of this approach will lead to further improvements of this guidance in the future.

In the case of single substances, or complex substances where there are dominating constituents, read-across by non-formalised approaches generally involves the identification of a chemical substructure that is common to the target chemical and its analogue (or their respective breakdown products) and the assumption that:

a) in the case of qualitative read-across, the presence (or absence) of a property/activity for the chemical of interest (target chemical) can be inferred from the presence (or absence) of the same property/activity for the analogue (source chemical).
b) in the case of quantitative read-across, the known value of a property for the analogue (source chemical) can be used to estimate the unknown value of the same property for the chemical on interest (target chemical). In the case of a toxicological effect (human health or ecotoxicological), this assumption implies that the potency of an effect shared by the two chemicals is similar or follow a regular pattern.

In the case of complex substances, the basis for comparison is likely to be different. For example, complex substances derived from certain process streams may share common structures.

With limited information it can be difficult to judge the degree of uncertainty associated with the assumption of commonality for a particular read-across. To provide the most robust read-across possible, other relevant properties should be compared between the source and target chemicals.

R.6.2.3.1 Stepwise procedure for applying read-across within the analogue approach

The following stepwise approach is recommended, but should be regarded as flexible and not the only possible approach. This is presented in Figure R.6-5

Step 1: Identification of potential analogues

There are a number of different possible ways of identifying potential analogues as source chemicals with data with which the target chemical can be compared.

In many cases, the choice of a source chemical is straightforward. Similar chemicals produced for similar uses by the same company (or sector group of companies) are often used as potential analogues. In this case, no formal selection techniques are used.

However, a more formal search strategy may indicate additional potential analogues for comparison, and hence, increase the robustness of the read-across. It should be noted that with increasing numbers of chemicals included in a read-across, the closer this approach is to the approach used for categories described in the next chapter. One starting point would therefore be to consider whether the chemical is best evaluated by an analogue approach, or whether a wider

category approach should be used. One factor that would affect the choice is whether the chemical is a member of a category that has already been evaluated. Another factor would be the number of analogues identified: if a significant number of analogues are identified, then a wider category approach would be justified, as outlined in the next chapter.

Information on categories that have been evaluated by the US-EPA is available from http://www.epa.gov/opptintr/newchems/pubs/chemcat.htm

Information on categories that have been evaluated within the OECD HPV Chemicals Programme is available from http://cs3-hq.oecd.org/scripts/hpv/

There is no single information source on categories evaluated within the EU. However, information can be found in the Tapir final report (ECB, 2005), Gallegos Saliner *et al* (2007), and Tsakvoska and Worth (2007).

A number of industry sectors have applied the principles of *grouping* for use in evaluation of health and environmental hazard properties. Examples, including rationales for grouping, include petroleum substances (Concawe, 2001), dyes and pigments (ETAD, 2001), chlorinated paraffins (CPIA, undated), surfactants (CESIO, 2000, 2003) hydrocarbon solvents (HSPA, 2002), acrylate resins (UV/EB Acrylate Resins, 2003), petroleum additives (ATC, 2000a, b) and bitumen (Eurobitume, 2002) (see Appendix 9 of the Tapir final report: ECB, 2005).

Categorisation approaches have been applied to flavours and fragrances (Salvito, 2007) under JECFA, USHPV, Environment and Health Canada DSL Program, SPORT, and the safety assessment of fragrance ingredients under RIFM.

Computational methods for analogue selections are expert knowledge in combination with electronic substructure searching and automatic tools using molecular similarity indexes (e.g. the Tanimoto similarity index). The pharmaceutical industry, which are the predominant users of the concept of molecular similarity, are employing similarity methods in a wide range of applications e.g. virtual screening, estimation of absorption, distribution, metabolism, excretion and toxicity (ADME/Tox) and prediction of physico-chemical properties (solubility, portioning, etc.). Whilst these techniques have not been widely used in this context, the use of such techniques should be considered when searching for relevant source chemicals for comparison.

A non-exhaustive list of possible analogue-searching tools is given in Table R. 6-2

The identification strategy is an exploratory process, and is not intended to be an element of the read-across rationale. If a large number of analogues are identified, the use of the categories approach described in the next chapter is recommended. It should also be noted that the use of a category approach reduces the demands on extensive data for any individual source chemical, as this approach draws on the cumulative data available for all the individual chemicals in the category.

The structural similarity and the purity and impurity profiles of the substance and the structural analogue need to be assessed. The fundamental basis for any read-across decision must be that the chemical structures of the analogues are sufficiently close for there to be a reasonable expectation of similar effects. The more divergent the structures, the lower will be confidence in making such a prediction. In general, where biologically active functional groups are present, they should be present in both structures and be in the same structural orientation so that any biological activity would be unaffected.

The extent to which differences in the purity or impurities are likely to influence the overall toxicity (Hanway, 2000), needs to be addressed and, where technically possible, excluded.

Step 2: Data gathering for the analogues

For the source analogues chosen, published and unpublished data should be gathered on standard physico-chemical properties, environmental fate and transport properties, ecotoxicological and toxicological effects. Standard physico-chemical properties include physical state, molecular weight, log K_{ow} and other partition coefficients (e.g. the Henry's Law coefficient, soil organic-carbon partition coefficient), aqueous solubility, particle size and structure[30], vapour pressure, melting point and boiling point. Since these physico-chemical properties provide basic information on environmental distribution, fate and bioavailability, they can often provide supporting information for the read-across. The data gathering should include all existing relevant data, including both experimental data and data generated by non-testing methods.

If a large number of analogues are identified, it is recommended to consider forming a larger chemical category (see Section R.6.2.4). If this is not feasible, e.g. for practical reasons, computational tools such as (Q)SARs can help to reduce the dataset to a subset of the closest analogues, e.g. homologues for which properties similar to the target chemical are estimated (see Section R.6.2.2.1).

Data is already available on many high volume chemicals that have been thoroughly assessed. Information on substances assessed by the OECD is available from the OECD (http://cs3-hq.oecd.org/scripts/hpv) and the United Nations:

(http://www.chem.unep.ch/irptc/sids/OECDSIDS/sidspub.html).

Information on chemicals assessed in the EU can be found on the ECB website (http://ecb.jrc.it).

Information on the environmental and human health effects of chemicals can be found from a large number of internet-accessible databases. A list of such databases, including internet links, has been compiled by the European Chemicals Bureau

(http://ecb.jrc.it/QSAR/information_sources/information_databases.php).

Step 3: Evaluation of available data for adequacy

Where data is available from relevant peer-reviewed sources such as the OECD HPV Chemicals programme, EU risk assessment programme or other comparable sources, the data can normally be used without further evaluation.

In other cases, the available experimental data should be evaluated for adequacy according to Chapter R.4 or by using the OECD Guidance for Determining the Quality of Data for the SIDS Dossier (see Section 3.1 of the OECD Manual for Investigation of HPV Chemicals).

If read-across data have not been produced using the most current test methods, required under REACH, particularly careful consideration of the quality and suitability of a method is important (Hanway and Evans, 2000).

Step 4: Construct a matrix of data availability

A matrix of data availability should be constructed for the target endpoint and all other relevant endpoints (see Section R.6.2.7 for an example). The matrix should include the chemical of interest (target chemical) and the analogue(s) (source chemical(s)). If multiple analogues are identified, they should be arranged in a suitable order (e.g. according to molecular weight). The ordering should reflect a trend or progression within the group. The cells of the matrix should indicate whether data are available or unavailable. If possible, the cells should also indicate the available reliable key study results.

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

It is currently only possible to provide limited guidance about how to decide whether data from an analogue can be used to fill a data gap, and the decision remains largely an expert judgement. Similarly, it is not possible to provide definite guidance on how data gaps could be filled quantitatively by read-across.

However, the factors shown in Section R.6.2.2.1 need to be addressed when evaluating the results of a read-across using an analogue approach. The supporting evidence discussed in Section R.6.2.2.1 (subsections f, g and h) should also be considered.

Wherever possible, the relevance of the read-across of other endpoints should be evaluated in the light of the known or suspected mode of action. The applicability of the read-across can also be evaluated in the light of available data for both source and target chemical for other endpoints where the mode of action is likely to be similar. The use of QSAR predictions can also be useful to assess the applicability of the read-across, both by predicting the missing data and comparing the experimental data available and the predictions.

Chemicals that cannot be represented by a molecular formula or structure can be handled on a case-by-case basis, depending on the components of the complex substance and on the data available for the complex substance and/or components.

If the read-across is considered to be suitable, the missing data for the target chemical(s) is evaluated using the data from the source chemical(s) according to the guidance in Section R.6.2.2

If the read-across is not considered to be suitable, three options are possible. It may be necessary to identify alternative analogues – the best analogues may indeed not have the relevant experimental data, so it may be necessary to choose analogues of lower quality in order to obtain data - or the use of a more extended category approach can be considered. It may also be necessary to obtain the information directly by testing (in which case the options provided by Annex XI in REACH are not relevant).

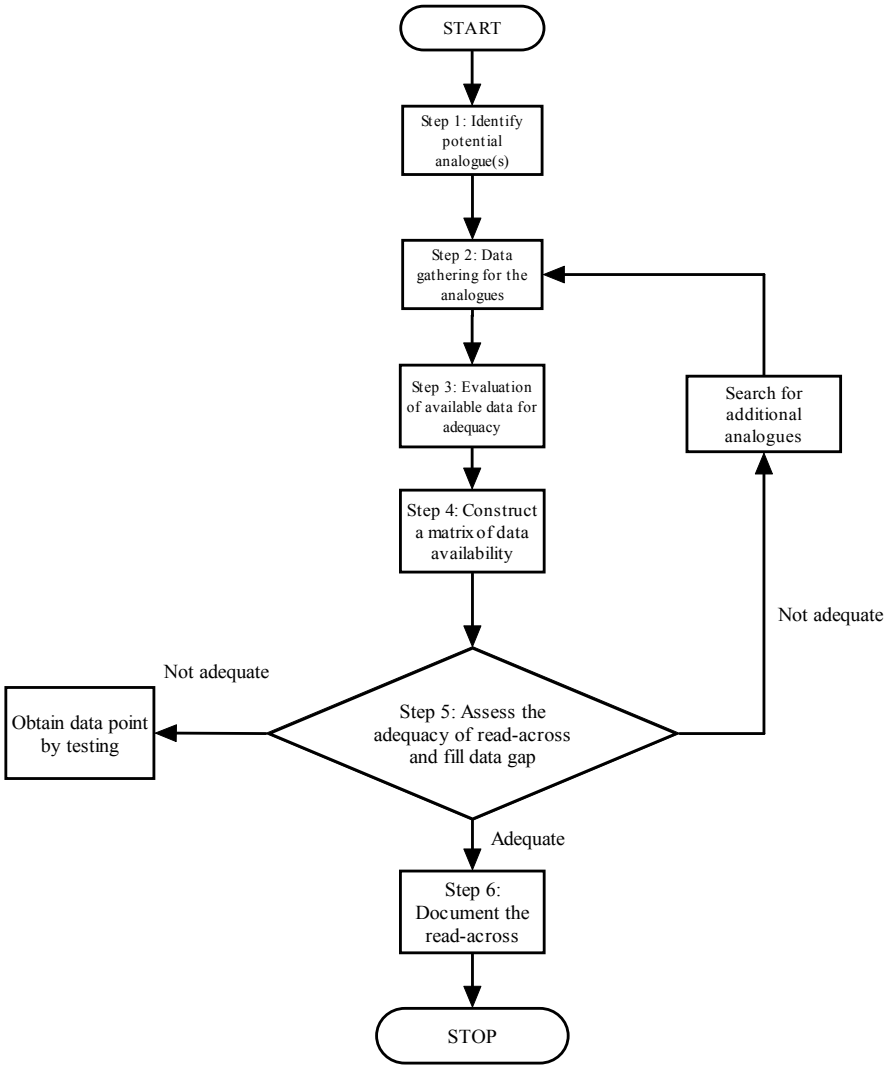
Step 6: Document the analogue approach

If the read-across is considered to be suitable, the approach should be documented according to an appropriate format in order to justify that the approach may be used instead of testing (see Section R.6.2.6). The justification for the read-across should include an explanation of the rationale, as well as the assessment including all relevant supporting information. Ideally examples of unsuitable read-across should also be documented.

Table R. 6-2 Selected internet-based tools for analogue-searching

Internet Tool & Website	Remarks
AIM	US-EPA's Analog Identification Methodology Links to publicly available, experimental toxicity data for target chemical as well as structural analogues Due to be publicly available in early 2007 Contains 31,031 records Searchable by CAS number, SMILES and (sub)structure (see Section R.6.1.8 for further information)
Ambit http://ambit.acad.bg	Chemical databases and functional tools, including a tool for defining applicability domain of QSAR models Developed by IdeaConsult Ltd Publicly available Contains 463,426 records Searchable by chemical name, CAS number, SMILES and (sub)structure (Section R.6.1.8 for further information)
ChemFinder http://www.chemfinder.com	Publicly available and subscription scientific databases Searchable by diverse parameters including chemical name, synonyms, CAS number, formula, chemical structure (exact match, substructure, similarity search), toxicological and physico-chemical properties
ChemID Plus http://chem.sis.nlm.nih.gov/chemidplus	Publicly available database from the US National Library of Medicine (NLM). Contains over 379,000 records Searchable by chemical name and CAS number
Hazardous Substances Database (HSDB) http://toxnet.nlm.nih.gov	Publicly available toxicology database on the National Library of Medicine's (NLM) Toxicology Data Network (TOXNET) More than 4800 peer-reviewed records Searchable by chemical name, fragment name, CAS number, subject terms
Danish (Q)SAR Database http://ecbqsar.jrc.it	Publicly available version of the QSAR database developed by DK EPA, and made available by ECB website Contains 166,000 records Searchable by chemical name, CAS number, endpoint, and (sub)structure Section R.6.1.8. for further information)
Leadscope http://www.leadscope.com	Commercially available databases and (Q)SAR functionalities Searchable by chemical name, (sub)structure, toxic effect, study type, and experimental conditions (Section R.6.1.8 for further information)
SciFinder http://www.cas.org/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, as well as by research topic, author, and company.

Figure R.6-5: Stepwise procedure to the analogue approach



R.6.2.4 General guidance on a stepwise procedure to develop categories

Chemical categories accomplish the goal of obtaining hazard information through the evaluation of all available experimental data for the individual chemicals in the category, so that reliable estimates that are adequate for classification and labelling and/or risk assessment can be made without further testing of the individual members of the category. If there is sufficient experimental data to support the category evaluation that the chemicals in the category behave in a similar or predictable manner, then the relational features described in Table R. 6-5 can be used to assess the chemicals instead of conducting additional testing. If not, it may be necessary to: a) perform limited and targeted testing; b) revise the category hypothesis (and therefore the applicability of the category in terms of members and/or endpoints); or c) as a last resort abandon the category hypothesis.

The review of the use of chemical categories carried out in preparation for the development of this guidance[31] concluded that the main lessons learned with the use of the chemical category approach are:

[31] Modified from Appendix 9 in the Tapir final report (see ECB, 2005)

1. Initial hazard assessments were agreed by OECD member countries for 240 chemicals in 42 different categories as of 2006, by applying the chemical category approach. The chemical category approach can therefore be considered to be widely accepted for regulatory purposes.
2. Currently more than a third of the substances assessed yearly within the OECD HPV Chemicals Programme are assessed through the use of chemical categories and this fraction is estimated to increase significantly over the next few years as experience grows in member countries.
3. As already concluded for the US HPV Challenge Programme, chemical categories can be used to estimate results for both environmental and human health endpoints.

The guidance in this chapter documents a stepwise approach to the formation of categories. The current practice is based on the use of non-computational methods. However, guidance is also included on where computational tools could be used at various steps in this process to support the development of categories. It is emphasised that such computational tools can supplement but do not replace the need for expert judgement, which is required throughout the process. Whilst the use of these tools is considered to be helpful in a category approach, it should be recognised that the use of approaches for which there is little or no regulatory precedence should be used in close collaboration with the relevant regulatory authority.

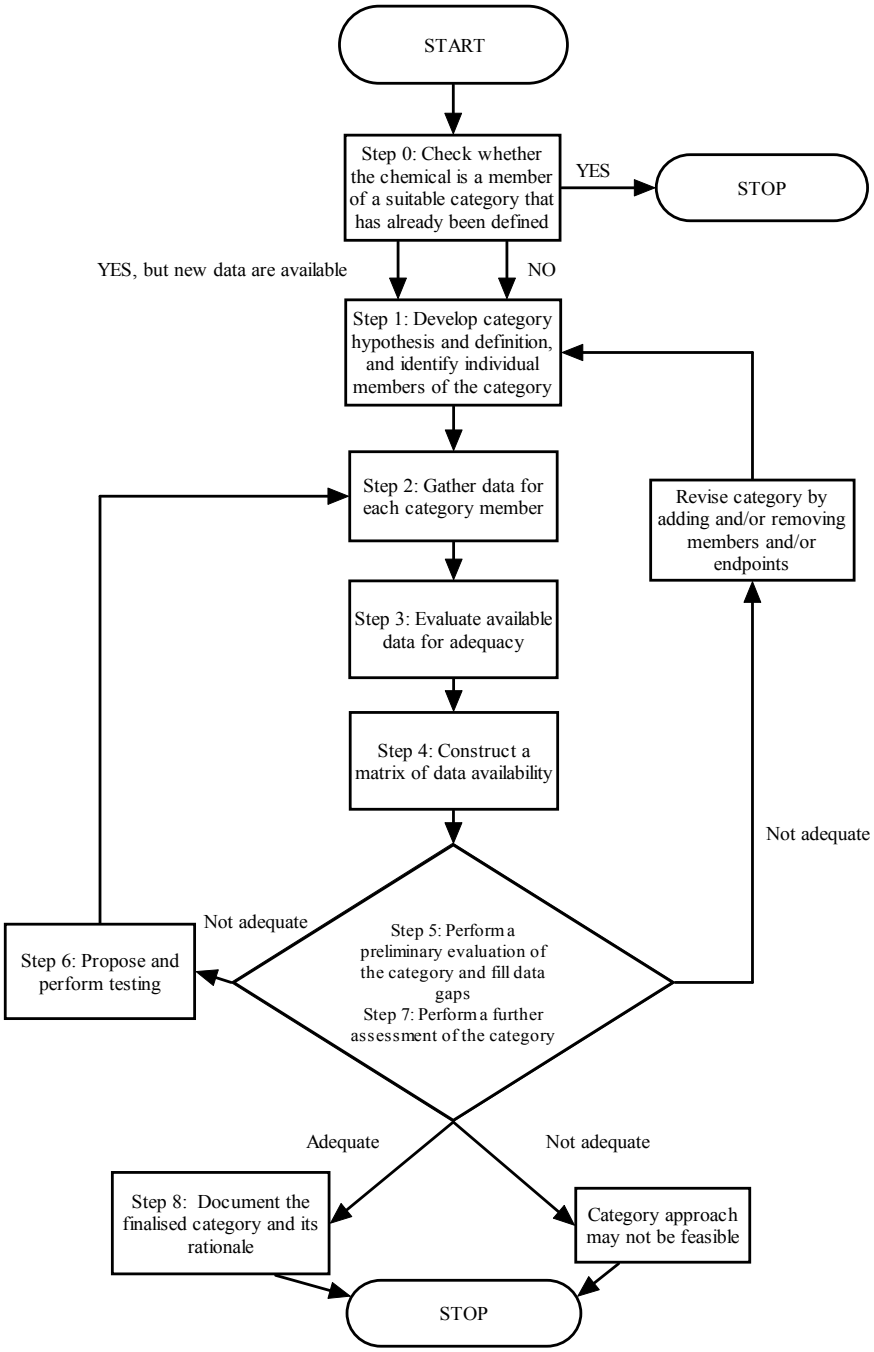
This chapter should be read with the understanding that the formation of categories can be carried out using the expertise routinely used in hazard identification and risk assessment. However, given the large number and diversity of chemicals that exist, and by extension categories that may be formed, guidance on how to develop and evaluate chemical categories cannot be captured in terms of rigid rules. Rather this section describes how information on chemical properties and activities and when available, metabolism and mechanisms of action should be gathered and combined with expert judgement to form robust and well rationalised categories, as well as guidance on how to document the justification for each category. Based on a learning-by-doing approach, the experience gained in application of this approach will lead to further improvements of this guidance in the future.

R.6.2.4.1 Stepwise procedure to the formation of chemical categories

In order to use the results from a category, it is necessary to demonstrate that a chemical category is robust, and to do this, certain types of information should be documented. In order to collect this information in a systematic and transparent manner, it is recommended to follow a stepwise approach (Figure R.6-6). The general scheme should be regarded as flexible, since there may be alternative ways of most efficiently obtaining the information.

One reason for needing flexibility is that there can be different starting points in category formation. For example, it may be desirable to start from a single chemical, or small group of chemicals, and to identify analogues to establish a larger category. Alternatively, it may be desirable to start from a defined set of chemicals (e.g. a set list of already classified substances), and to find ways of grouping them and finding additional analogues relating to them.

Figure R.6-6: Stepwise procedure to category development



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

Before considering whether to develop a category for a group of substances, the first step should be to determine whether the chemical(s) is (are) a named member of a category that has already been evaluated. Information sources on existing categories include:

- 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

A number of industry sectors have applied the principles of *grouping* for use in evaluation of health and environmental hazard properties. Examples, including rationales for grouping, include

petroleum substances (Concawe 2001), dyes and pigments (ETAD, 2001), chlorinated paraffins (CPIA, undated), surfactants (CESIO, 2000, 2003) hydrocarbon solvents (HSPA, 2002), acrylate resins (UV/EB Acrylate Resins, 2003), petroleum additives (ATC, 2000a, b) and bitumen (Eurobitume, 2002) (see also Appendix 9 in the Tapir final report: ECB, 2005).

Categorisation approaches have been applied to flavours and fragrances (Salvito, 2007) under JECFA, USHPV, Environment and Health Canada DSL Program, SPORT, and the safety assessment of fragrance ingredients under RIFM.

Under REACH guidance on the identification of substances and the description of their identity is given in *Guidance on substance identification*.

Under REACH it would be helpful for potential and actual registrants if the Chemicals Agency would collect and make available information on categories that have already been evaluated. In addition, on the basis of the information provided by industry in the pre-registration phase it would be helpful if the Agency would be able to make suggestions for new categories that could be further developed by industry.

If the chemical is a member of a category that has already been evaluated, its inclusion into the new category should be justified. It is usually sufficient to refer to the evaluation of the category when assessing the chemical, and to refer to the results that have been agreed for the category, taking account of the position of the chemical in the category. Where new data are available for some endpoints, these may be used to verify the existing category and could, depending on the results, lead to a revision of the category.

In some cases, a relevant category may exist, but where the chemical of interest has not been specifically included in the category. For example, this can be the case where a category including only a number of HPV chemicals has been evaluated. In this case, it would be appropriate to extend the membership of the currently defined category to include the chemical of interest. For further guidance on the consequences of extending a category in this way see Section R.6.2.1.2

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

The first step in developing a category is to develop a basis for the proposed grouping of chemicals.

The category definition should list all of the substances and endpoints covered. Chemical category definitions have referred to chemical classes with a common functional group (e.g. epoxides) or chemicals with an incremental and constant change across the category (e.g. a chain-length category).

Although the chemical structure is usually the starting point, a category definition could also refer to a group of chemicals related by a mechanism of action (e.g. non-polar narcotics) or a particular property. In practice, this particular property is largely related to the chemical structure. For example, in the case of hydrocarbon solvents, products were separated into categories based on basic hydrocarbon structure - aliphatic or aromatic - and then further separated based on boiling ranges, carbon number, and other properties. In some cases, the aliphatic hydrocarbon categories were further separated into subcategories based on specific aliphatic structure such as normal or branched aliphatics (IHSC, 2004/2005).

Some categories have been defined in terms of a metabolic pathway, i.e. they have a stepwise metabolic pathway producing the different members within the category with each metabolic step. More detailed examples of how these types of categories have been evaluated are shown in Section R.6.2.5.2.

In addition, the category definition should describe the molecular structure a chemical must have to be included in the category, including criteria such as carbon chain length, functionality, and chemical or metabolite equivalence considerations.

It is possible to develop and propose a category for a specific endpoint, or a selection of endpoints, rather than for all of the endpoints required for the substance in question, although this restriction should only be applied where strictly necessary. In particular, all the endpoints that can be expected to be relevant for the category should be included. Since a category is based on an underlying hypothesis of a common mechanism of action, the wider the range of endpoints covered, the more robust the results that are obtained from the category approach.

The category hypothesis should also address:

- the chemical similarities (analogies) and trends in properties and/or activities that collectively generate an association between the members. These features can be regarded as the parameters that hold the category members together.
- the specific instances of read-across and trend analysis (interpolations and extrapolations), and any specific computational methods that have been used.
- the set of inclusion and/or exclusion rules that identify the ranges of values within which reliable estimations can be made for category members for the given endpoint. These rules, can be described as the applicability domain for an endpoint and provide a means of extending the category membership to chemicals not explicitly included in the current definition of a category.

Depending on the basis for the category, the individual members of the category are identified.

In many cases, this is done on an empirical and non-systematic basis. In the OECD HPV and EU ESR programmes, chemicals have frequently been grouped on the basis of their obvious structural similarities (e.g. phthalate esters, groups of oil-derived complex substances, metal compounds).

Since categories have often been developed in the context of a High Production Volume Chemicals programme, the selection of the chemicals that are included in a particular chemical category has normally been guided by the fact that the chemicals in the category are produced in high volumes. However, it should be noted that a category may also contain substances that are not produced in high volumes (or indeed, substances that are not necessarily commercially available). These chemicals are also legitimate members of the category, and may in some cases prove to be relevant candidates for testing in order to evaluate the properties of the category as a whole.

The formation of a category has in many cases also been dependant on which chemicals are manufactured by the consortium of companies sponsoring the category. However, it should be noted that a category may also contain substances that are produced by a number of different companies. It is therefore important for industries wishing to use this approach to consider the formation of a consortium (e.g. based on an Industry sector group) in order to obtain appropriate support and information.

However, when developing a category, the possibility of including additional chemicals that had not been initially selected since they did not meet these pragmatic criteria should be seriously considered. Data may be available for these chemicals that can help in the assessment of the chemicals for which registration dossiers are being prepared. Inclusion of these chemicals will increase the robustness of the category, and reduce the possibility that the addition of these

chemicals at a future date would lead to revision of the conclusions for the chemicals specifically under evaluation.

There are many approaches to making a list of category members from the use of simple manual approaches to the use of automated computer-based analogue searching methods.

In preparing a comprehensive list of ethers to form a category of low molecular weight ethers with carbon numbers from 2 to 6, permutations of the SMILES notation for these compounds was used (see Hart and Veith, 2007). This approach has the advantage of speed and simplicity, but there are also disadvantages associated with the approach. Systematic use of the SMILES notation can ensure that all possible members of a category are included, and the systematic names of the individual members can be derived from the structures. However, it is often difficult to identify the CAS numbers of the substances without additional work. The production process may also vary across the range of a category, leading to the formation of commercial products of varying complexity, and potentially differing impurity profiles, depending on carbon number. Whilst most of the low carbon number ethers are produced as single compounds, many of the higher carbon number ethers are produced as complex substances with varying components. These commercial compounds may have their own separate CAS numbers, and the available data may only be available for the commercially produced complex substance, rather than for the individual compounds identified on the basis of their structure.

In the case of new category proposals, computational methods can help to develop the category hypothesis (rationale) and to define the category in terms of its endpoints and members. The choice of computational method(s) is likely to depend on the starting point of the investigation. For example, the user may start from a single chemical or a small group of chemicals, with the intention of building up a category by drawing on data from multiple sources (bottom-up or systematic approach). Examples of tools that might help include expert systems such as Derek (LHASA Ltd, UK) or other tools such as Leadscope (Leadscope Inc, USA) or AIM (US-EPA) which are described in Section R.6.1.8.. In addition, combinatorial methods exist for identifying, *a priori*, the possible permutations of the substituents on a given substructure. Examples of tools capable of this include TSAR or Cerius2. A variety of computer-based analogue-searching tools have been summarised in Table R. 6-2 in Section R.6.2.3. . In some cases, these techniques may identify compounds which contain more than one isomer, which can give rise to difficulties in estimating the properties of the individual components (see example in Worth *et al*, 2007). However, regulatory experience with the use of these computational tools is still limited, and further guidance will need to be developed in the near future.

In identifying a category, it is important that all potential category members are described as comprehensively as possible. For potential members of a category, all relevant CAS numbers should be selected. For some substances, there may be more than one CAS number, and studies may contain relevant data reported under different CAS numbers. Due to historic reporting errors, a CAS number used to describe a substance may not accurately describe the substance as marketed. The CAS numbers of members of the category should also be checked against different chemical inventories (e.g. TSCA, EU, Customs Inventories) as these inventories may indicate which CAS numbers are used for marketing the substances and hence for which CAS numbers additional data might be available.

It is important that information on the purity and impurity profiles of all potential category members is collected at the same time as details of the molecular structure. Differing purity or impurities could influence the overall toxicity. For example, a category member may contain a particularly toxic impurity that is not present in the other substances making it difficult or impossible to draw conclusions on the toxicity of other substances in the category. It is therefore important that

category members have similar purity profiles or, where they differ, the effect of the differing purity profiles is known.

Step 2: Gather data for each category member

For each member of the category, published and unpublished data should be gathered on physico-chemical property(ies), environmental fate parameter(s), toxicological (human health) and ecotoxicity (environmental species) effect(s). This should include all existing relevant data and not be limited to the endpoints that are mandatory within a given programme (e.g. metabolism and cancer studies are relevant but not part of SIDS in the OECD HPV Chemicals Programme). In some cases where estimated data have been included in an internationally accepted evaluation, these estimates can be included on the same basis as other data that has been critically evaluated.

The computational methods described in Step 2 (Section R.6.2.3) can also be used to identify analogues (and corresponding data) that are included in one or more databases. Having identified a range of possible chemicals, one or more databases could then be searched to identify those chemicals for which data are available. Guidance on data gathering for analogues is also given in Section R.6.2.3.1

Dossiers should be prepared for each category member. Specific guidance on how to prepare Dossiers for chemical categories with the IUCLID software will be developed and made available in a separate guidance document. Reporting formats are described in Section R.6.2.6.

Step 3: Evaluate available data for adequacy

Available data should be evaluated for its adequacy according to Chapter R.4 or by using the OECD Guidance for Determining the Quality of Data for the SIDS Dossier (see Section 3.1 of the OECD Manual for Investigation of HPV Chemicals).

In evaluating the available data for a category, a number of additional factors will apply that are not relevant when evaluating test results for individual compounds.

Different types of data may be available for the same endpoint. It is clear that the scope of the estimated results for a member of a category cannot exceed the scope of the underlying data for the other members of the category, e.g. if for genotoxicity, only in vitro results are available for some members of the category (source chemicals), only conclusions on in vitro genotoxicity can be reached for the members of the category for which experimental results are lacking (target chemical). If the scope of the underlying experimental results for an endpoint vary (e.g. a mix of results from screening tests and higher tier tests), it is necessary to clarify the scope of the estimated results for the category members for which no experimental results are available. It may be possible to apply a Weight of Evidence approach to all the data, which could lead to the same hazard identification for all the members of the category, irrespective of the data available for the individual compounds.

An effect that is defined by a particular numerical cut off may lead to different conclusions for individual compounds. This type of data should be studied carefully to ensure that the compounds are evaluated in a way that reflects the underlying trends across a category. For instance, a series of compounds may give rise to data that shows a borderline positive irritant effect for some members of the category and a borderline negative effect for others. The data should be carefully evaluated to decide whether (a) this reflects accurately a trend across the whole category or whether (b) the uncertainties in the experimental data justify allocating the compounds to different subcategories (in this example, classifying some category members as irritant and not classifying others). If the second option is considered as the most biologically plausible explanation, the conclusion of the

evaluation will lead in some cases to a different conclusion than that based on a simple evaluation of the data taken in isolation. Hence, a borderline positive effect can be interpreted as a negative effect in the light of evidence from other compounds in the category. Similarly, a borderline negative effect can be interpreted as positive evaluated taking into account the data from the whole category.

Where the data suggests possible breakpoints, the data should be evaluated to ensure that these reflect a genuine change in properties or effects and are not due to comparison of results from testing carried out in different laboratories, at different times, with different animal strains, etc.

The data set may contain an apparent outlier, i.e. one category member where there is experimental data that shows the presence of an effect not seen in other category members. This difference can be real, and provide evidence of special conditions relevant to the particular substance (e.g. the chronic and reproductive toxicity of hexane compared to other lower alkanes). Such results need to be evaluated with particular care to establish whether the result reflects a real difference in a mechanism of action across the category or whether the test result should be questioned.

Step 4: Construct a matrix of data availability

A matrix of data availability (category endpoints vs. members) should be constructed with the category members arranged in a suitable order (e.g. according to molecular weight). The ordering of the members should reflect any trends or progression seen within the category. The cells of the matrix should indicate whether data are available or unavailable. If possible, the cells should also indicate the available reliable key study results (see Section R.6.2.7 for an example).

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 whether:

- the category rationale is supported, i.e. the category does in fact exhibit one or more of trends postulated in Step 1; and
- the category is sufficiently robust (i.e. contains sufficient, relevant and reliable information on the category members) for the assessment purpose.
- This assessment should be carried out for each endpoint, as the category rationale may lead to a relevant assessment for some endpoints and not for others.
- This assessment is largely a matter of expert judgement. Assessment of the category rationale and robustness of the category for the particular regulatory purpose is closely related to the approach chosen for filling data gaps for any particular endpoint, and here the guidance in Section R.6.2.2. for read-across, trend analysis and the use of external QSARs should be taken into account.
- If the initial assessment indicates that both criteria are satisfied for a particular endpoint, the data gaps can be filled according to the guidance in Section R.6.2.2 and the chemical category can be finalised and documented.
- In applying these techniques, the background for the basis on which the category is formed should be reflected in the way techniques are chosen and applied. Hence for some effects, where the test data suggest a uniform property across a group, read-across from the existing data would normally be considered appropriate. In other cases, where there is a trend in aquatic toxicity related to a change in log Kow and based on a narcotic mechanism of action, the data gaps may be filled by data from a valid QSAR for the category. Alternatively, the category can be sub-divided into a number of subcategories defined by the breakpoints in the category, and members evaluated within each subcategory.

If the initial category does not satisfy both of these criteria, the following options should be considered:

- If further examination of the data suggests that there is a pattern of effects for a limited number of chemicals in the group, then the analysis might suggest that the category should be modified e.g. divided into subcategories (return to step 1).
- If adequate data do not exist, but the structure-based category is reliable for one or more endpoints, then a category approach may still be proposed for these endpoints. Testing of some chemical category members for some endpoints would still be necessary (go to Step 6). The choice of chemicals and endpoints for testing should be scientifically motivated, but is also likely to involve animal welfare and financial considerations, especially in the case of more expensive endpoints.
- If there are adequate data for a given endpoint, but no apparent pattern, the proposed category may not be appropriate and so testing may be required for all remaining category members for that endpoint (i.e. the category is abandoned).

Step 6: Perform and/or propose testing

If the preliminary assessment supports the category rationale (i.e. a pattern or trend is observed), but the category does not appear to contain sufficient, relevant and reliable information to assess all category members, it may be necessary to perform or propose testing.

In proposing additional testing, a number of factors should be taken into consideration.

- Since a category may contain compounds of different production volumes, the standard information requirements (e.g. those stipulated in Annexes VI to X for REACH or those stipulated in the OECD Manual for Investigation of HPV Chemicals for the OECD HPV Chemicals Programme) may vary from compound to compound within a category. However, there may be strong scientific reasons that the recommended testing should be conducted on lower tonnage category member(s) in order to identify the actual hazards of category. In which case the test plans should be confirmed with the appropriate regulatory authority[32].
- The choice of test will be influenced by the results of the preliminary evaluation of the category.
- If there are no data for any of the members of a category for a particular endpoint, full testing of a limited number of carefully selected category members may be considered appropriate.
- When data are already available indicating the presence or absence of a particular effect, tests may be chosen to provide evidence that compounds selected for testing show the effects that have been predicted from the trend of the property. Hence, for a substance in a category where e.g. skin irritation is predicted, a simple in vitro test would be sufficient to provide confirmation of the effect.

Test plans for chemical categories should include a category definition, rationale, and matrix of data availability and be accompanied by the Dossiers for each category member.

The rationale supporting a category definition should be as simple and transparent as possible, and should explain why the existing data and proposed testing data allow interpolation or extrapolation

[32] Under REACH, the appropriate regulatory authority for approving test proposals is the European Chemicals Agency.

to other members of the category that have no data or proposed testing. The category rationale should be documented in the Category Reporting Format, as described in Section R.6.2.6

The test plan needs to summarise the adequacy of the existing data, and how the proposed testing will adequately characterise the category.

The matrix of data is a useful part of the test plan and provides a tool for consideration and presentation of the available data. The endpoints are rows in the matrix. If toxicity is expected to vary in a regular pattern from one end of the range of category members to the other end (e.g. high toxicity to low toxicity), samples chosen for testing should bracket both ends of toxicity. If the category is large, testing also needs to be performed and/or data should be available for one or more members in the middle of the range of toxicity. Any change in a tendency for a property should be accompanied by data in the adjacent cells in order to define the limits for the resulting subsets of the category or subcategories. Assuming the columns are the category members, there are no rules for the number of columns and cells that must be filled nor the number that can be empty. Acceptability of the matrix will depend on the number of members in the category, the endpoint, and the confidence in the interpolation and extrapolation.

When selecting a sample to test, it should be representative of the substance marketed, including the presence of any manufacturing impurities.

It should be noted that the category test plan is intended to provide information about the properties of the group as a whole rather than the properties of any specific, individual compound. A category test plan may thus identify as key substances for testing substances of little or no commercial importance. Whilst in some cases this may even require the synthesis of chemicals specifically for this purpose, the approach may still prove more economical, both in terms of expense and numbers of animals used for testing, than a more conventional testing strategy based on individual commercially available chemicals.

Under REACH, whether or not testing needs to be proposed (to the Agency) depends on whether the information sought is part of the standard information requirements in Annexes VII or VIII (testing may be performed) or Annexes IX or X (testing must be proposed). The Registrant needs to decide which substances should be included in a category. However, in the case that a testing proposal needs to be submitted, the Agency may decide not to accept a testing proposal for a certain substance if it considers that the substance belongs to a category that already contains the necessary data element.

Step 7: Perform a further assessment of the category

If new test data are generated, the category should be revised and further assessment to determine whether the criteria outlined in Step 5 are satisfied and therefore whether the category can be finalised and documented.

If the results support the category, the testing phase is complete and the chemical category can be finalised and documented. Remaining data gaps can be filled according to the guidance in Section R.6.2.2.

If the results do not support the category, further testing may be carried out, members of the category may be changed (e.g. dividing the category as appropriate), or the category proposal may be dropped altogether. The latter implies that testing will then be done to fill all appropriate endpoints for each category member.

Step 8: Document the finalised category

The finalised category should be documented in the form of a suitable reporting format (see Section R. 6.2.6 for proposed format).

Chemicals that cannot be represented by a molecular formula or structure can be handled on a case-by-case, depending on the components of the substance and on the data available for the substance and/or components.

While a category may be regarded as finalised, it may be revised subsequently in the light of new data and/or experience. For example, the category could be extended by including additional chemicals, or may even be redefined by withdrawing one or more substances.

Under REACH, a category may be revised on the basis of new Registrations. If this leads to new testing proposals, the Agency may decide not to accept a testing proposal for a certain substance if it considers that the substance belongs to a category that already contains the necessary data element.

R.6.2.4.2 IT tools for elaborating dossiers for members of chemical categories

IT tools to build dossiers for members of chemical categories and to document the chemical categories have been developed, e.g. IUCLID 5 or HPVIS.

HPVIS has been developed by the US-EPA in the context of the US HPV Challenge Programme.

IUCLID 5 is the recommended tool for submission of dossiers under REACH as well as under the OECD HPV Chemicals Programme.

Both tools, while focusing on the elaboration of dossiers for single substances, allow for the grouping of substances, either for simple analogues or into more complex chemical categories.

Guidance on how to prepare documentation for chemical categories according to the present guidance document with the above mentioned IT tools will be prepared separately.

R.6.2.5 Guidance on specific types of categories

In this chapter, guidance is provided for some specific types of chemical categories. It should be highlighted that the categories described in this chapter are not the only category types that might ever be formed or created.

R.6.2.5.1 Chain length

Chain-length categories show an incremental, and usually constant, increase in chain length across the category. It is assumed that each category member exhibits the same toxic mode of action unless there is a good scientifically demonstrated reason to believe this is not the case. Examples include the homologous series of alpha-olefins, where each category member differs by a methylene group (–CH_2– unit), and the ethylene glycols, where there is an incremental increase in the number of CH_2CH_2O groups. Examples of chain length categories which have been assessed within the OECD HPV Chemicals Programme are alpha-olefins (CAS Nos: 592-41-6, 111-66-0, 872-05-9, 112-41-4, 1120-36-1), higher olefins (CAS Nos: 25264-93-1, 25339-56-4, 25377-83-7, 27215-95-8, 25339-53-1, 25378-22-7, 85535-87-1, 629-73-2, 112-88-9) or monoethylene glycol ethers (CAS Nos: 2807-30-9, 111-76-2, 112-25-4) (UNEP Chemicals, 2006).

Categories defined by chain length generally show an incremental change in molecular weight and other physico-chemical properties, such as water solubility or log K_{ow}. However, not all properties will necessarily exhibit a linear relationship with chain length and care must be taken in making assumptions about such trends. For many homologous series, increasing log K_{ow} leads to increasing fish toxicity whilst at the same time water solubility decreases. There is usually a point where the solubility is too low to be expressed. For example, in alpha-olefins there is an apparent cut off point between the C_8 and C_{10} chain length at which acute toxicity to fish is no longer observed. Similarly, a trend of increasing molecular weight may lead to decreasing systemic toxicity as absorption decreases. There may be a change of physical state of the category members as chain length increases.

Care should be taken when evaluating a category containing both branched chain chemicals and linear chain chemicals. Whilst there may be no influence of degree of branching on a trend for some endpoints (e.g. aquatic toxicity), significant differences could be expected for other endpoints (e.g. biodegradation). For these endpoints where differences in trend are seen, it may be helpful to divide the category into subcategories in order to provide a robust justification for the assessment.

Careful thought should be given to selecting the boundaries of a chain length category. The cut off points described above may provide useful boundaries. The potential scope and size of a chain length category may be larger than that covered by a particular manufacturer or consortium. Where possible, well-characterised substances which are not necessarily HPV chemicals but which fit into the series should be included. There may be cases when testing the end members of a chain length category is not appropriate. For example if the existing data indicates that the toxicity cut off occurs earlier in the series, it may not be necessary to test the end member for that endpoint.

QSARs can be used to help justify the category and fill data gaps. In general, substances at either end of a chain length category should have all endpoints fulfilled, preferably with test data. This permits interpolation of data to the other category members rather than extrapolation and increases confidence in the estimate. For example, in the category on ethylene glycols, a linear regression was used to predict acute aquatic toxicity, indicating that toxicity decreases with increasing chain length, and further supporting the low toxicity of the category members concluded from available experimental data. For categories where there is more than one variable, such as variation in the length and degree of branching of the chains, more category members are likely to be required to bring confidence to the interpolations being made.

Other examples are oleochemical derivatives which can be grouped in such categories like fatty acids or alkyl sulfates. These categories may contain single-chain chemicals as well as mixtures containing chemicals of distinct chain lengths at varying amounts. The relative amounts of individual chain length molecules in mixtures are usually reflecting the chain length distribution in natural fats and oils from which they are derived. Since the category chemicals differ from each other only by the number of $-CH_2-CH_2-$ units, these categories are very homogenous and exhibit a constant pattern in the changing of the potency of the properties across the category as described below.

R.6.2.5.2 Metabolic pathways

The underlying hypothesis for a metabolic series is a sequential metabolism of a parent chemical to downstream blood metabolites that are chemicals of interest. Hazard identification studies with the parent compound could then be used to identify the hazards associated with systemic blood levels of the downstream primary and secondary metabolites and once quantified, can be used in place of studies using direct exposure to primary and secondary metabolites themselves. In certain instances, the metabolism of the parent compound within barrier tissue (e.g. lung or gut tissue) occurs so

rapidly that the initial primary metabolite is the predominant chemical found within the blood. Under these circumstances data from hazard identification studies conducted with that primary metabolite itself can be used to identify hazards for the parent compound. PBPK or PBPD models may help to define categories. The metabolic pathway approach is usually reserved to some toxicological endpoints. For physico-chemical properties, environmental fate and ecotoxicity, information on the parent compound would need to be available. Examples of metabolic pathway categories which have been assessed within the OECD HPV Chemicals Programme are isobutyl isobutyrate (CAS No 97-85-8) or trimellitic anhydride (CAS No 552-30-7) (UNEP Chemicals, 2006).

The first technical issues faced when forming a metabolic series is to determine if the metabolism that is assumed to occur does occur independently of the requirements of the programme under which the chemical is assessed. This is necessary before moving any further in developing a metabolic category and preferentially should be determined *in vivo*. In certain instances, *in vitro* metabolic studies can be used to help identify metabolic pathways, but the definitive evidence should be conducted in whole animals. The primary and secondary metabolites should be detected either in the blood or tissue. Primary and secondary metabolites that cannot be readily determined in blood or tissue should not be candidates for a metabolic series approach without some limitation placed upon the use of the information.

The second technical issue pertains to the level of evidence required to describe the metabolic processes. Direct measurement of the parent chemical and primary and secondary metabolites in the blood in an *in vivo* exposure is the recommended standard. The level of evidence required to presume that there will be blood-borne levels of primary and secondary metabolites following exposure to parent chemical, will have to be determined on a case-by-case basis. Certain metabolic processes are ubiquitous and well understood and these can be presumed to occur without performing *in vivo* experiments in every instance. Other metabolic processes are not part of normal metabolism or require enzyme induction. These metabolic processes may not be well characterised and should not be assumed without specific *in vivo* evidence of blood levels of primary and secondary metabolites.

The third technical issue provides a limitation for the metabolic approach to forming categories. The metabolic category reasoning is only useful for identifying hazards related to systemic blood levels of the parent compound and/or primary and secondary metabolites. Other endpoints of hazard identification studies that are dependent upon site of contact effects (e.g. eye, skin, respiratory tract irritation, irritation to gastric mucosa) cannot be addressed using the metabolic category logic. These sites of contact effects are often due to the physical chemical property of the chemical in question and therefore may differ considerably between the parent compound and primary and secondary metabolites. In addition, tests that identify unique structural characteristics (e.g. skin or respiratory sensitisation) or are dependant upon physical chemical properties (e.g. volatility and LC_{50} values) should not be considered as part of metabolic category because these properties may not be similar amongst the various members of the metabolic series.

An additional limitation of the metabolic categories approach is that metabolism and toxicokinetics experiments have to be conducted with the parent compound. These types of studies are not requested in most review programmes and therefore would require a sponsor of the chemical to do additional work beyond what is normally considered necessary. However, it should be recognized that the savings involved (numbers of animals used, testing costs) could be considerable compared with generating data for each metabolic category member for each endpoint of systemic toxicity. For screening level assessments that are interested in identifying hazards related to systemic blood levels, it should not become necessary to provide definitive toxicokinetic evidence or develop a

toxicokinetic model for acceptance of hazard identification studies as relevant for the primary and secondary metabolites.

An additional advantage of using the metabolic category toxicity data is that in certain instances, higher systemic blood levels of a chemical can be achieved from metabolic pathways than if the primary or secondary metabolite was administered directly. For example, if a material is corrosive or has limited volatility, higher blood levels may be found following the administration of the parent compound than if the primary or secondary metabolite was administered directly to the animal.

The following specific issues should be taken into account when developing a metabolic pathway category, according to the stepwise procedure described in Section R.6.2.4.1

Step 1: Provide definitive information on the metabolism of the parent chemical to the primary and secondary metabolite. This information should also include, preferably, a time course data for either blood or tissue for both the parent chemical as well as the primary and secondary metabolites.

Step 2: The metabolism experiment should be examined to determine, if in fact, the primary and secondary metabolites are formed, if they achieve appreciable levels within the blood and/or tissues and determine basic toxicokinetic parameters for the parent material. For example, the $T_{1/2}$ for elimination for the parent chemical should be determined if possible. If the metabolism of the parent chemical to the primary metabolite is rapid and is thought to occur within barrier tissues, then it may be appropriate to use hazard identification studies from the primary metabolite to identify hazards associated with exposure to the parent chemical.

Step 3: If there are appropriate hazard identification studies that have been conducted with the parent chemical or primary or secondary metabolites for similar toxicity endpoints, then these studies should be examined to see if these materials have similar toxicity. If data is not available for the metabolic series in question and a study is to be designed and conducted, then the parent compound should be tested, so that blood levels of all category members will be present. The toxicokinetic and metabolic experiments that provide the basis for the metabolic category should have robust summaries prepared and be included in the dossier for the parent chemical, primary and secondary metabolites. A table should be included detailing the relative blood levels of the parent chemical, primary and secondary metabolites.

Step 5: A quantitative analysis between exposures of the parent chemical and the primary and secondary metabolite is usually not necessary if the only objective is hazard identification. It is recognised that in certain cases quantitative differences play an important role in hazard identification (e.g. in the metabolism of C_6 - C_8 alkanes). For risk assessment purposes, a quantitative analysis may become necessary, e.g. additional toxicokinetic analysis (including preparing a model) may be appropriate.

The metabolic approach should not be used for environmental toxicity endpoints unless the metabolism of the parent compound to the primary or secondary metabolite can be demonstrated within the test species in question. Whereas it may be appropriate to extrapolate within mammals, it may not be appropriate to extrapolate between amphibians and fish or insects and other species due to the difference in the metabolic processes and enzymes present within those species.

On the other hand the same concept underlying the metabolic pathways can be used for environmental degradation processes. For example, for a substance which hydrolyses very rapidly

in aquatic test systems (half-life <1 hour), the aquatic toxicity endpoints can be covered by the test results with the degradation product(s) (OECD, 2000).

R.6.2.5.3 Chemical reaction products and multi-constituent substances

Categories can be developed for series of chemical reaction products or multi-constituent substances (MCS) that are related in some regular fashion. As with categories based on discrete chemicals, in a category containing reaction products or MCS some, but not all, of the individual substances may require testing.

A number of categories assessed under the OECD HPV program provide useful case studies on dealing with multi-constituent substances. Further information is available at (http://cs3-hq.oecd.org/scripts/hpv/). For the Ethylene Glycols category, data from PEG 200, a mixture of chain lengths, was used to support the human health assessment. For the Linear alkylbenzene sulfonates category, aquatic toxicity data was available for both commercial products (mixtures) and pure C_{13} and C_{14} homologues. The pure homologues showed higher toxicity than the commercial mixtures but data for the pure homologues was not used to drive the recommendation of the assessment since they were not commercially supplied (Caley *et al*, 2007). The Bicarbonate Special category presented to SIAM 22, and focusing on ammonium bicarbonate, provided an interesting example of assessing a reaction mixture using data from pure components. The commercial material is a reaction mixture of sodium bicarbonate, sodium carbonate and ammonium bicarbonate. Aquatic toxicity data was available for the three components. Ammonium bicarbonate is the most toxic and the evaluation therefore focused on the quantity of ammonium ions released to water from dissolution of bicarbonate special and the impact of pH on the ammonium speciation and toxicity (Caley *et al*, 2007). Effectively, the ammonium ion was used as a marker for aquatic toxicity (see also Section R.6.2.5.5).

Another example is the reproductive toxicity of technical C_7-C_9 phthalate ester mixtures. In case of ortho phthalate esters, there was clear evidence that phthalates with a C_4-C_6 backbone (i.e. the length of the longest branch in the side chain) were toxic to the reproductive system, whereas phthalates with a backbone >C_6 might not be. It was assumed therefore that phthalate ester mixtures which contained both lower and higher homologues, then the reproductive toxicity capacity/potency of the mixture would depend on the amount of the lower homologues (backbone C_4-C_6) present in the mixture. In fact what was observed for some complex mixtures containing a high amount of the lower homologues was similar but fewer reproductive toxicity effects, at higher concentrations and with less severity than the lower homologues. Therefore, when assessing such mixtures, it would not be sufficient to determine just the predominant homologue or different homologues (side-chain, backbone lengths) in the mixture, but also the amount and properties of these different homologues (Fabjan *et al*, 2006).

The composition and physico-chemical properties of substances are useful considerations to take into account when dealing with MCS.

R.6.2.5.4 Isomers

Isomers are chemicals that have identical molecular formulas but different molecular arrangements. Although there are several types of isomers, the two that typically will be considered are structural and geometric.

Structural isomers are molecules with differences in the arrangement of their atoms. Structural isomers can include:

- chain isomers: For example hydrocarbon chains with identical or variable lengths and variable branching patterns (see also Section R.6.2.5.1).
- positional isomers: For example hydrocarbon chains with a functional group that varies in position along the chain. An example is 1-butene and isobutene.
- functional group isomers: These isomers also have identical molecular formulas, but contain different functional groups. Examples are 1-butanal and 2-butanone which both have the molecular formula C4H10O. Each of these isomers contains a carbonyl group (C=O), but are representative of two different chemical families: butanal is an aldehyde whereas butanone is a ketone. This type of structural isomers is less likely to be considered within a category because functional isomers can have very different chemical and biological properties. Functional isomers are not included within the scope of this guidance.

Stereoisomers are isomeric molecules whose atomic connectivity is the same but whose atomic arrangement in space is different. One type of stereoisomerism is geometrical (cis-trans) isomerism.

Geometric (or cis-trans) isomers can occur when a double bond or a ring is present. Bond rotation is restricted in these types of structures, so atoms can be permanently on the same (cis) or on opposite (trans) sides of the bond. For example, *cis*-2-butene and *trans*-2-butene each have carbon groups on either side of a double bond, which cannot rotate, so the carbon groups are arranged on either the same side of the molecule (*cis*) or opposite sides of the molecule (*trans*).

Enantiomers are two stereoisomers that are related to each other by a reflection: they are mirror images of each other. Every stereocentre in one has the opposite configuration in the other. Two compounds that are enantiomers of each other have the same physical properties, except for the direction in which they rotate polarized light and how they interact with different optical isomers of other compounds. In nature, only one enantiomer of most chiral biological compounds, such as amino acids, is present. As a result, different enantiomers of a compound may have substantially different biological effects.

An example showing a profound difference in the effects of enantiomers is the drug thalidomide, The optical “R” isomer is an effective sedative whereas the optical “S”- isomer is a teratogen causing serious birth defects in children to mothers using the drug during pregnancy.

Stereoisomers can have similar or different chemical or toxicological properties. Even though they may behave identically in many chemical reactions, it is for example well known that the enzyme specificity in biological systems may be totally different, so caution is needed in case of such substances. An example of such specificity is certain carbohydrates, which may be metabolised or not depending on the orientation of functional groups. These are examples of *diastereoisomers*, which are defined as stereoisomers that are not enantiomers (i.e. they are not mirror images of each other). Diastereomers can have different physical properties and different reactivity.

There are two general principles for using estimation techniques as they apply to isomers:

- Relatedness: The substance(s) with a data gap as well as substance(s) with data are similar such that their physico-chemical, biological, and toxicological properties would be expected to behave in a predictably similar manner or logically progress across a defined range. This similar manner or logical progress should be demonstrated by the available experimental data. QSAR models and trend analysis can also be used in addition of experimental data to support the estimate.
- Structural Similarity: The substance(s) with data gap possesses a small incremental structural difference from the reference substance(s) or the difference between the two would not be expected to affect the property sufficiently such that it could not be accurately

predicted. This similar property should be demonstrated by the available experimental data. QSAR models and trend analysis can also be used in addition to experimental data to support the estimate.

There can be instances within a category of structural isomers when the estimate for an endpoint is not appropriate. An example is illustrated with two categories of isomers: the pentanes and hexanes. Although the pentanes may be broadly described as isomers, they actually represent three types of hydrocarbons, normal alkanes, branched alkanes, and cyclic alkanes. It is known that n-pentane, 2-methylbutane, 2,2-dimethylpentane, and cyclopentane exhibit distinct differences in potential biodegradability. n-Pentane and 2-methylbutane are readily biodegradable, whereas 2,2-dimethylpentane and cyclopentane are poorly biodegraded. Therefore, it is not possible to assess the biodegradability of the poorly biodegradable pentanes by using the results from the readily biodegradable pentanes, even though the pentane isomers could still be considered a category for other endpoints. In such a case, the potential biodegradability of the two groups of pentanes would each have to be characterised separately within the context of the category. Likewise, the peripheral neurotoxicity in humans associated with exposure to n-hexane has not been demonstrated to occur with exposure to other hexane isomer. Therefore, a discussion of this effect within a hexane isomer category would have to isolate n-hexane from the other isomers.

Based on the category of butenes and their mixtures, the following general principles were derived:

- selected properties of isomers may be read-across to another isomer(s) or to an isomeric mixture within a category if the data are similar and/or if the structure of the isomer(s) without data is similar to the isomers with data.
- extrapolating properties to isomeric mixtures should take into account mode of action, potential additivity and synergy, as well as purity profiles, and mixture composition.
- for toxicological endpoints (e.g. LC50, NOAEL), a range of toxicity or the lowest value in a range of toxicity may be used for read-across.
- read-across from one isomer to another may not be straightforward. Metabolic data may be needed if existing knowledge of category members or related non category members suggests that differences may be expressed within a biological endpoint of interest.

R.6.2.5.5 Complex substances (UVCB)

Complex substances include a diverse range of materials which are defined (see *Guidance on substance identification*) as substances of *Unknown or Variable composition, Complex reaction products or Biological material* (UVCB substances). The range of different types of UVCB is very wide and the specific properties may be diverse, such that the applicability of a common approach needs justification. The following section highlights the key issues, however, it is recognised that in some sectors this approach has been more widely used than others and thus there needs to a cautious approach to defining categories and applying the following recommendations. There are many different types of complex substances, although generally they all have the following characteristics in common:

- they contain numerous chemicals (typically closely related isomers and/or chemical classes with defined carbon number or distillation ranges), and cannot be represented by a simple chemical structure or defined by a specific molecular formula.
- they are not intentional mixtures of chemicals.
- many are of natural origin (e.g., crude oil, coal, plant extracts) and cannot be separated into their constituent chemical species.
- the concept of impurities typically does not apply to complex substances.

- they are produced according to a performance specification related to their physico-chemical properties.

While CAS numbers are important for identifying substances, in the case of complex substances they do not represent a unique chemical and the specificity of the CAS number definition may vary (some CAS number definitions are rather narrow, some are very broad), e.g. CAS numbers for:

- petroleum complex substances are based on a hierarchy of considerations including hydrocarbon type, carbon number range, distillation range and the last processing step,
- coal derived complex substances are based on the applied production process and may include information on the distillation range and the chemical composition, and
- NCS: natural complex substances (e.g., essential oils) are assigned CAS numbers based on their genus and species, in some cases part of plant, extraction method and other processing descriptors.

Due to these numerous considerations, similar products sometimes have different CAS numbers. There are also historical and geographical reasons why similar complex substances may have been assigned different CAS numbers. Further, some CAS numbers have a broad definition that may fit different, but related complex substances that fall into different categories. These complexities lead to the use of physical properties and chemical descriptors (e.g. chain length, chemical class, size of aromatic ring systems) as being the preferred way to define categories of complex substances. In the case of NCS, this categorisation may also occur around the major chemical component(s) present, and might include marker chemicals for toxicity where it is clear that the behaviour of the UVCB is driven by those marker chemicals.

The approach used to define a category of complex substances may vary, although generally the approach will be related to how the category members are manufactured, defined and used.

General guidance on developing categories for complex substances

Stepwise approach to read-across:

The key step is to define the category and identify category members. While initially this may seem repetitive, in fact the steps are different for complex substances. This is best explained by considering the *define analogue(s)* step, which for complex substances means identifying single component substances that represent the range of properties and the matrix being built up by the complex substances. The properties of these analogues are used, often with properties of the complex substances, to develop the data matrix and describe the physico-chemical space.

The following elements are considered to be the main blocks to be used when putting together a category for complex substances.

1. Composition - it is important to clearly characterise the complex substances to the extent it is measurable. In particular, it is necessary to identify which of the following attributes are key and must be specified:

- Cut off ranges
    - Range of chain length or predominant carbon number range or size of condensed ring systems
    - Distillation temperature range
    - Appropriate measures that allow characterisation of category members
- Known or generic composition and description
- Standard index – e.g. Colour Index number

- Chromatographic and other physical "fingerprints"
- Reference to standards
- Information on the production process (especially useful in categorising petroleum or coal derived products)
- For botanical NCS identification of the genus/species, origin should be considered
- If marker chemicals are appropriate, they should be clearly identified and if possible quantified for all category members.

2. Properties of the components of a complex substance can be applied to the complex substance if the properties of the single components are similar, or fall within an expected range, depending on the endpoint.

- it is necessary to identify representative components of the complex substance to cover the carbon range and structure types of members of the complex substance.
- components with outlying properties need to be identified (e.g. specific toxicity of hexane compared to other aliphatic hydrocarbons, higher water solubility of aromatic hydrocarbons compared to aliphatic hydrocarbons).

3. Data gap filling - Read-across/SAR and QSAR: It is possible to fill data gaps within a defined category either using read-across/SAR or establishing a QSAR, which is sometimes best described as a local QSAR. Where the composition of two, or more, complex substances is similar (within boundaries defined by the category description) qualitative properties can be established and data gaps filled. Quantitative read-across is more difficult in such circumstances, although it is possible to establish ranges. Where a valid QSAR is either available or can be established based on components of the complex substance, it can be possible to fill data gaps with either qualitative or quantitative information. When this is done justification for the approach and chosen data needs to be clearly described.
   It is also very important to carefully consider the dose-response relationship for read-across/QSAR versus the nature of the complex substances and the level of components of concern within the complex substances.
4. Data gap filling – testing: Where it is necessary to identify representative complex substances for testing purposes, this should be done bearing in mind the key components of the category definition and the ranges thus defined.

Petroleum complex substances

Petroleum complex substances are generally defined by manufacturing and processing conditions, hydrocarbon chemistry (e.g., aliphatic hydrocarbons, aromatic hydrocarbons), physico-chemical properties such as boiling range or carbon-number range, and common use categories. An example of the grouping of petroleum complex substances, developed for the purposes of the Existing Substances Regulation and also used for classification and labelling purposes, is given in Comber and Simpson (2007). According to this approach, petroleum complex substances are grouped according to the process by which they are manufactured, on the assumption that substances within each group (or sub-group) have similar physico-chemical properties and therefore similar intrinsic hazard properties. Within this approach, two substances and a class of chemicals (DMSO extractable PAHs) were used as markers for carcinogenicity, i.e. the presence of one of these substances at a specified level was used to indicate and classify for carcinogenicity. For other classification endpoints read-across between members of the categories has been used and more recently supported by QSAR.

The approach adopted for the petroleum complex substances has more general applicability to UVCBs and should be considered by other industries for which it may be applicable.

Hydrocarbon solvents

Hydrocarbon solvent categories are based on typical chemistry and carbon-number range. Common use can also contribute to the category definition. Under this approach, those hydrocarbon solvent substances with similar chemistry and carbon-number range are grouped within a category that is generally defined by the predominant constituents of the category members. This approach is practical and has the benefit of ensuring that similar commercial products are grouped together in the same category.

Coal derived complex substances

The principle described in Section R.6.2.5.5 for petroleum derived complex substances also applies to coal derived complex substances. The longer geological history of coal compared to crude oil explains the higher degree of cross-linking of coal derived constituents. This results in a predominance of aromatic ring systems in coal derived complex substances. Longer alkyl chains do not appear. Processing of a coal derived feedstock separates according to volatility (size of condensed ring systems) and/or the extractability of acidic/ alkaline constituents. Formation of categories makes use of the applied processing techniques and of a similar spectrum of intrinsic properties for substances having a similar matrix of physico-chemical properties.

Natural complex substances (NCS)

NCS are botanically-derived substances obtained by subjecting specific parts of the plant to a physical treatment such as extraction, distillation, expression, fractionation, purification, concentration or to fermentation. Their compositions vary depending on the genus, species, the growing conditions and maturity of the crop used as a source, and the process used for its treatment.

NCS constitute a very specific subgroup of UVCBs (substances of unknown or variable composition, complex reaction products or biological materials) and include primarily essential oils and extracts obtained by various separation techniques.

Inclusion in a chemical group is possible based on the constituents of the NCS where the major components can be clearly identified as the same as known chemical substances. An example is provided by Salvito (2007).

Use of toxic equivalency factors or toxic units approach for filling data gaps

The use of toxicity equivalency factors and the estimation of toxic units for mixtures of chemicals which contribute to a biological effect through a common toxicity pathway is a useful approach for filling data gaps in the assessment of chemical mixtures. The techniques are applied to mixtures of compounds in order to express the mixture's toxicity as a single value. The principle requirement is that the chemicals in the mixtures are active in a common toxicity pathway, and so this approach is strictly only applicable for chemical mixtures that have been formally grouped based on mechanistic considerations. Furthermore, toxicity data for the endpoint being assessed must be available for each component in the mixture.

Complex mixtures of PCBs (Clemens *et al*, 1994), furans (Parrott, 1992), dioxins (Safe, 1991; van der Weiden, 1992) and aromatic hydrocarbons (Walker, 1991; Zabel, 1995) have been assessed using toxicity equivalency factors based on Ah receptor binding and joint toxicity models amongst others. Joint toxicity models for calculating the toxic units generally use a strict addition model when a common toxicity pathway is a reasonable approximation. Although synergist effects are conceivable, they are only observed when chemicals in a mixture have different mechanisms, which should not be the case within a chemical category rigorously formed by the principles including toxic mechanistic considerations.

In the Toxic Equivalents (TEQ) approach, the most toxic compound is used as the reference compound. This compound does not necessarily have to be present in the mixture being assessed, but the components of the mixture must all act by the same single toxic pathway and be of the same compound type (structural/functional group similarity) as the reference. The components of the mixture are each assigned toxic equivalency factors (TEFs) such that their individual toxicity is expressed as a fraction of the toxicity of the reference compound (which is given a TEF of 1). This is achieved simply by dividing the effect value of the reference compound by the effect value of the particular component (Equation 6-1).

$$\text{TEF (component A)} = \frac{\text{Reference effect value}}{\text{Component A effect Value}} \qquad \textbf{Equation 6-1}$$

Equation 1

The amount of each component in the mixture is then multiplied by its respective TEF and the values for each component are summed to give the overall toxic equivalency, relative to the reference compound (Equation 6.2).

$$\text{TEQ} = \Sigma \text{ (concentration x TEF)} \qquad \textbf{Equation 6-2}$$

For example in the case of dioxin and furan mixtures, toxicity relative to 2,3,7,8-tetraCDD (2,3,7,8-tetrachloro-*p*-dioxin) was derived, based on mortality of rainbow trout fry following injection of the compounds to eggs. The following table lists TEFs derived from measured toxicity data for some of the compounds found in the literature (Safe, 1991; Walker, 1991; Zabel, 1995):

Dioxin/Furan	**Toxic Equivalency Factor**
2,3,7,8-tetraCDD	1 (reference compound)
1,2,3,7,8-pentaCDD	0.73
1,2,3,7,8,9-hexaCDD	0.1
1,2,3,6,7,8-hexaCDD	0.024

To illustrate the approach using a fictitious example based on these data:

Mixture A contains 20% 2,3,7,8-tetraCDD, 50% 1,2,3,7,8-pentaCDD, 10% 1,2,3,7,8,9-hexaCDD and 20% 1,2,3,6,7,8-hexaCDD.

Therefore, according to equation 6.1:

$(0.2 \times 1) + (0.5 \times 0.73) + (0.1 \times 0.1) + (0.2 \times 0.024) = 0.5798$

So the toxic equivalency of Mixture A relative to the reference compound 2,3,7,8-tetraCDD is 0.5798, the fraction indicating a lower level of toxicity. In order to quote this fraction as an effect value (for example as an acute LC_{50} value) for Mixture A, the effect value of 2,3,7,8-tetraCDD is divided by 0.5798 giving a higher effect value (i.e. lower toxicity) for the mixture.

An adaptation of the method has been applied in the Netherlands EU draft risk assessment of coal tar pitch (under Council Regulation (EEC) 793/93, CAS 65996-93-2 Pitch, coal tar, high-temperature[33]) in which the local concentration (C_{local}) for each component is divided by the component's PNEC, the summation of all expressing the risk characterisation ratio as opposed to toxicity (Equation 6.3). A value greater than 1 indicated a risk.

[33] This draft risk assessment report is available on request.

$$\text{Sum RCR} = \Sigma \frac{C_{local}}{PNEC} \qquad \textbf{Equation 6-3}$$

In another adaptation of the method, the OECD HPV assessment of $C_{6\text{-}22}$ Aliphatic Alcohols (Long Chain Alcohols), measured acute fish toxicity data were not available for all of the alcohols present in these complex mixtures. Therefore (Q)SAR estimation was used to fill toxicity data gaps and so predict the toxicity of the complex mixtures.

In summary, toxic equivalency can be used for complex mixtures when there is a common mode of toxic action such that the effect is additive across the components of the mixture: there is no synergism. In addition, measured toxicity data should be available for each individual component of the mixture. Differences in test protocol for each data point can have a marked effect on the derived TEFs (and so TEQ), therefore if this approach is followed then it is necessary to present all available data and justify the use of the approach. This includes discussion of the shared toxic mechanism of the components in the mixture, choice of data for deriving the TEFs, discussion of the purity of the mixture/presence of impurities and their effects, and any deviations from the method.

R.6.2.5.6 Metals, metal compounds and other inorganic compounds

The concept of chemical categories has traditionally been widely used for hazard assessment for certain endpoints and risk assessment of inorganic substances. The approaches have generally been based on the occurrence of a common metal ion or anion and the use of read-across to fill data gaps.

For example, the chemical category approach based on the metal ion has been extensively used for the classification and labelling of metal compounds in the EU[34]. Other category entries are based on certain anions of concern such as oxalates and thiocyanates. For these EU classifications the category approach has often been applied to certain endpoints of particular concern for the compounds under consideration, and has not necessarily been applied to all endpoints of each individual compound in the category of substances. A category approach has also been used during the categorisation of existing chemicals on Canada's domestic substances list (Environment Canada, 2003).

This approach has also been used for estimating the potency of the effects as well as for their identification. NOAEL(s), NOEC(s) and comparable quantitative estimates have been read-across from data obtained from water-soluble compounds to other water-soluble compounds, including, in the absence of specific data, to compounds of substantially lower water-solubility. One example is the EU risk assessments on nickel (Tsakovska and Worth, 2007).

The application of these concepts has been useful[35]

- to evaluate hazards for substances for which data are limited rather than relying exclusively on conducting tests.
- to evaluate hazards for a range of compounds regarded as difficult substances, as they can present technical difficulties when carrying out standard test protocols (see Section R.6.2.4).
- to evaluate hazards for a number of metal compounds, for which animal models do not

[34] The EU terminology for this type of entry is a "group entry" rather than a category.

[35] The approach of grouping metals and metal compounds in risk assessments has also been applied because it allows addressing together all compounds which potentially lead to exposure to the same metal moiety.

always reliably predict effects on humans. Where the hazard has been identified on the basis of human data the use of read-across provides a method to avoid these difficulties.

The guidance below is based largely on the practice of the EU Technical Committee on Classification and Labelling, the EU Technical Committee on New and Existing Substances and experience gained in other fora (see also Hart, 2007; Schoeters and Verougstraete, 2007). This guidance is intended to supplement the general guidance in the previous chapters with issues specific to metals and inorganic compounds.

Assumptions underlying the grouping of metal compounds

There are a number of assumptions underlying any grouping of metal compounds for estimating their biological properties.

The hypothesis is that properties are likely to be similar or follow a similar pattern as a result of the presence of a common metal ion (or ion complex including a hydrated metal ion). This is a reasonable assumption for the majority of inorganic compounds and some organic compounds (e.g. metal salts of some organic acids). However, it is the bioavailability of the metal ion (or a redox form of this ion) at target sites that in most cases determines the occurrence and severity of the effects to be assessed for the read-across of metal substances. Supporting information to assess the bioavailability of the metal ion at the target site can include information on a number of different factors (e.g. physico-chemical properties such as water solubility, degree of dissociation of the metal –containing compound, particle size and structure[30], *in vitro* solubility, *in vivo* data on systemic effects, toxicokinetics).

Basis for the development of categories or read-across approach of metal compounds

Hazard data is available for some primary metals and some key (high production volume) inorganic compounds. However, for a wide range of inorganic and organic compounds of the same metal, data is usually very limited. Data availability will play an important role in the selection of source chemicals.

As metals occur in a wide and heterogeneous range of substances, including inorganic metal compounds, organic metal salts, organometallic compounds, metals, metal-metal compounds (i.e. compounds containing more than one type of metal), alloys[36] and complex substances, care is needed in order to select those metal compounds for which a category approach is relevant from those where read-across is not applicable.

The following points could alter the assumption of commonality and should be considered:

- Chemical speciation and valency
  When selecting the appropriate source substance, the valence state and its influence on the assumption of commonality should be checked. For some metals (predominantly transition elements), the chemical speciation and in particular the different valencies may result in differences in mechanism of action and a variation in toxicological properties. For example, differences in hazards are seen with Cr3+ and Cr6+ compounds. In some cases, species may be interconvertible, in other cases there is little interconversion between the species.
- Organometallic compounds
  Organometallic compounds will generally have a different mode of action since the metal

[36] Alloys are regarded as special preparations (mixtures) in REACH, and are as such not covered by this guidance. However, some alloys are listed in EINECS and are therefore considered as substances, where this advice may be applicable.

ion is not likely to be present in the same form as for inorganic compounds. In such cases, read-across between inorganic and organometallic compounds is not recommended, although read-across may well be appropriate between different organometallic compounds. On the other hand, especially for environmental risk assessment, if an organometallic compound degrades rapidly to its inorganic metal moiety, it can be assessed together with the inorganic metal moiety.

- Metals
  Particular difficulties have been seen in evaluating the properties of metals on the basis of data for metal compounds. In some cases, read-across of properties from the metal compounds to the metal itself (metallic, zero-valent form) has been agreed (e.g. cadmium oxide to cadmium metal, EC 2007a,b,c, EC 2008), whilst for others it has not (e.g. soluble nickel salts to nickel metal, EC 2006). These need to be evaluated on a case-by-case basis.
- Metal containing UVCBs
  Some metal containing UVCB compounds may not be appropriate for consideration in a category approach, as their effects will not be expected to be adequately described by their metal content. These include compounds such as asphalt, frits and drosses. In cases where read-across is not considered appropriate, clear arguments should be put forward as to why the known hazard profile of the metal is not expected to be relevant (for example very low bioavailability).
- Crystalline structure
  The crystalline structure of insoluble metal compounds could influence the hazard profile. If there is reason to believe that the crystalline structure influences significantly the effects of the compound to be assessed, this must be taken into account in the evaluation. An example is silica of which the crystalline and non-crystalline forms have a different hazard profile (see category for synthetic amorphous silicas assessed within the OECD HPV Chemicals Programme; Silicon dioxide [CAS Nos 7631-86-9, 112945-52-5, 112926-00-8] Silicic acid, aluminum sodium salt [CAS No 1344-00-9] Silicic acid, calcium salt [CAS No 1344-95-2]).

Preliminary evaluation of the category and read-across

The water solubility of the metal compounds is often used as the starting point for establishing a category, as this provides a first indication of the availability of the metal ion in the different compartments of interest. For example, for inorganic nickel a number of sub- categories have been suggested, reflecting different ranges of aqueous solubility (Hart, 2007).

The most simplistic approach to hazard evaluation is to assume that the specific metal-containing compound to be evaluated shows the same hazards as the most water-soluble compounds. This is a conservative approach, since systemic metal ion availability will normally be reduced with decreasing water-solubility and consequently reduced bioavailability.

This simplistic approach can be refined for categories containing many substances by building subcategories based on water solubility, when data is available on trends with water solubility. For example, mixed oxides with limited water solubility can be evaluated by comparison with the hazard profile for the metal oxides (where this is known) rather than for the soluble salts.

This difference in trend is clearly recognised in evaluating the environmental hazards of metals and metal compounds, where the relevant hazards can be evaluated using a transformation/dissolution protocol (OECD 2001).

Information from other endpoints could further support the systemic bioavailability assumptions. For example, the LD_{50} values for the semi-soluble nickel compounds was used to demonstrate

systemic uptake to justify classification for reproductive toxicity for these compounds, but not for the less soluble oxides and sulfides (Hart, 2007). For endpoints where a threshold occurs, estimates of the systemic bioavailability (i.e. toxicokinetics) of the metal ion can be ascertained for representative members of each category in order to ascertain whether the bioavailability exceeds the threshold for the compounds.

In addition to water solubility, phagocytosis, bioaccessibility in synthetic biological fluids, and organ deposition and clearance rates are relevant parameters to be considered (Schoeters and Verougstraete, 2007).

Where toxicokinetic data is available, this should be used as this provides relevant information on whether the source and target chemicals in question behave similarly as expected from read-across or whether there are biologically differences that would bring into question the validity of the category hypothesis.

Other factors may also need to be taken into account.

Counter ions and other metal ions:
The assumption that the metal ion is responsible for the common property or effect implies that the toxicity of the counter ion or of other metals present in the compound will be largely irrelevant in producing the effects to be assessed. This assumption could be affected by interactions between the metal ion and other parts of the substance e.g. the counter ion. It is noted that in certain cases the effect of the counter ion in acute toxicity studies exert another effect than in repeated dose studies using lower dose levels. This could obscure the role of the metal ion in either the acute or repeated dose studies. The influence of the counter ion should be checked for each endpoint. If there is reason to believe that the counter-ion (such as cyanates, oxalates) or other metal ions present in the compound influence significantly the effects of the compound to be assessed and alter the assumption of commonality, this must be taken into account in the evaluation. One option may be to use the additive approach described in the foreword to Annex I, Directive 67/548/EEC, in the guidance to Note A. (see also Section R.6.2.5.6).

Crystalline structure:
The crystalline structure of insoluble metal compounds could influence the hazard profile. If there is reason to believe that the crystalline structure influences significantly the bioavailability and so the effects of the compound to be assessed, this must be taken into account in the evaluation. An example is the low bioavailability of spinels and rutiles.

Particle size information:
Particle size information of the substance influences the deposition behaviour in the respiratory tract and potential toxic effects. Based on particle size distribution data, trends in deposition and potency of effects can be assessed for locally acting substances.

If there is evidence that the crystalline structure and particle size influence significantly the bioavailability and so the severity of the effects of the compound to be assessed, this must be taken into account in a *Weight of Evidence* approach considering all available information (e.g. toxicokinetics).

Considerations of the need for further refinement

As described previously, a preliminary assessment of the read-across or category should be carried out to determine whether the rationale is supported and whether the approach is sufficiently robust for the assessment purpose. If these criteria are satisfied for a particular endpoint, the data gaps can be filled according to the guidance in Section R.6.2.2. .

If these criteria are not satisfied (there is uncertainty or contradictory information), the registrant should consider what additional information may be required. Additional data could include demonstrating a difference in bioavailability/bio accessibility between the substances in a proposed read-across or category.

The following options could be considered:

In vitro data:
In vitro information may be obtained by determining relative solubilities in physiological media (e.g. synthetic gastric juice, synthetic sweat) or by the use of the transformation/dissolution protocol (OECD, 2001) for the endpoints of sparingly soluble metal compounds related to the aquatic environment.
The solubility in alveolar liquids, lysosomal liquid, mucous liquids may provide more relevant information than simple water solubility for argumentation of the extent of availability of the soluble fraction of material during its dwelling time in various regions of the respiratory tract. To test whether slightly soluble, particulate metal compounds are taken up into mammalian cells and release metal ions intracellularly as free metal ions or bound to cellular macromolecules and whether the metal ions reach the cell nuclei, tests in vitro can be carried out using phagocytosing mammalian cells in culture.

In vivo data:
In some cases, in vivo testing may be considered, especially for endpoints where there is uncertainty about the role of the counter-ion. In planning the testing, a starting point for the studies should be confirmation of the effects expected on the basis of a read-across. As an example, if read-across would indicate the skin irritation is expected, an initial test could be carried out in vitro to confirm this effect before in vivo testing is considered.

Toxicokinetic data:
Animal model systems (using rats and mini-pigs) have been successfully used to characterise the speciation-dependent bioavailability differential for metals such as lead, arsenic and cadmium (US Environmental Protection Agency, 2004). Alternative strategies using rare stable isotopes of metals such as lead and zinc have been successfully used for the ascertainment of bioavailability of these metals in humans and animals. These types of studies are not requested in most review programmes and therefore would require a registrant to do additional work beyond what is normally considered necessary. However, where such information is not available, information could be collected for representative members of the category.

General guidance for other compounds

Similar considerations are expected to apply to salts in which the anion is associated with the toxic effects (e.g. cyanides, oxalates, thiocyanates). For categories that cover reactive chemicals, the reaction/degradation products must be of a similar nature for each member of the category to be plausible (Caley *et al*, 2007). One example is the Methanolates category assessed under the OECD HPV programme (http://cs3-hq.oecd.org/scripts/hpv). This consists of 17 potassium and sodium methanolate and both react rapidly in water to form the corresponding hydroxide.

When comparing acids and their salts, differences arising from pH effects should be considered (Caley *et al*, 2007). For example, skin and eye irritation are likely to be different for an acid compared with its salt. This is illustrated by the Phosphonic Acid Compound (Groups 1, 2, 3) categories assessed under the OECD HPV programme (http://cs3-hq.oecd.org/scripts/hpv). For these categories, dermal and irritation studies are considered separately for the acid and salts.

For the Gluconates category assessed under the OECD HPV programme (http://cs3-hq.oecd.org/scripts/hpv), it was found that for categories including ionisable compounds, the effect of the counter-ion needs to be considered (Caley *et al*, 2007). It is possible that the counter-ion(s) may pose hazards of greater concern than the common cation or anion on which the category is based (e.g. metal counter-ions that are inherently hazardous on their own).

Under such circumstances, it may be of limited utility to group and assess substances by the component which is expected to have the least effect. In other cases, it may be concluded that effects of the counter-ion are insignificant and therefore need not be taken into account in the assessment.

R.6.2.6 Reporting Formats for analogue and category evaluations

This chapter provides reporting formats for the analogue and chemical category approaches. The documentation of an analogue or category approach is an integral part of the assessment report and this chapter provides guidance on how to report the analogue and category approach in e.g. Chapter 1 of a SIDS Initial Assessment Report or Chemical Safety Report. An example is given in Section R.6.2.7

For chemical categories the assessment report should address all members of the chemical category and be accompanied for each member of the category by the dossiers containing robust study summaries of the key studies for all relevant endpoints (physical chemical properties, environmental fate and pathways, ecotoxicity, toxicity).

Under REACH, it should be noted that each member of the category has to be registered separately. Therefore a hazard assessment should be developed addressing all members of the category (i.e. Chapters 1-7 of the Chemical Safety Report), while exposure assessments and risk characterisations should be developed individually in separate reports for each chemical in the category. The hazard assessment for the chemical category can then be submitted with each individual registration.

Experience in the OECD HPV Chemical Programme has shown that for a simple analogue approach (read-across), it can be more practical to perform separate assessment reports for the source and target chemicals. In this case, the guidance below is relevant for the target chemical only, provided that the assessment(s) and dossier(s) of the source chemical(s) are referenced. In case no assessment is performed for the source chemical(s), the assessment report and dossier of the target chemical should contain all the relevant information, including robust study summaries from studies performed with the source chemical(s).

Furthermore, when developing an analogue or chemical category with IUCLID 5 or any other similar software having implemented the OECD harmonised templates (OECD, 2006b), dedicated fields are provided in the software where users can insert or append the documentation elaborated with the present formats. Specific guidance on how IUCLID 5 can be used to construct and document an analogue read-across or chemical category can be found in the IUCLID Manual (EC, 2007d).

R.6.2.6.1 Reporting Format for the analogue approach

1.	**Hypothesis for the analogue approach** Describe the molecular structure a chemical must have to be suitable as a source chemical. All functional groups need to be identified. Provide the hypothesis for why the read-across can be performed. If there is a mechanistic reasoning to the read-across, describe the foreseen mode of action for source and target chemicals and if relevant describe the influence of the mode of administration (oral, dermal, inhalation). List the endpoints for which the read-across approach is applied.
2.	**Source chemical(s)** Describe the source chemical(s) as comprehensively as possible. Provide CAS numbers, names and chemical structures of the source chemical(s).
3.	**Purity / Impurities** Provide purity/impurity profiles for the target and source chemicals, including the likely impact on the relevant endpoints. It should be discussed which influence these impurities are thought to have on physico-chemical parameters, fate and (eco)toxicology, and hence on the read-across.
4.	**Analogue approach justification** Based on available experimental data, including basic physico-chemical properties, summarise how these results verify that the read-across is justified. The data should also show that functional groups not common to source and target chemicals do not affect the anticipated toxicity. The available experimental results in the data matrix reported under Section 5. should support the justification for the read-across. More detailed discussion of available test results for individual endpoints (i.e. discussion of the selection of key studies, variability of experimental results between source and target chemicals etc.) should be provided in the corresponding sections of the assessment report (e.g. Chapters 2-4 of the SIDS Initial Assessment Report or Chapters 4-7 of the Chemical Safety Report).
5.	**Data matrix** Provide a matrix of data (endpoints vs. target and source chemicals) (see Table R. 6-3). In each cell in the Data Matrix, the study result type should be indicated in the first line, e.g.: - experimental result - experimental study planned - read-across from supporting substance (structural analogue or surrogate) - (Q)SAR If experimental results are available, the key study results should be shown in the Data Matrix.
6.	**Conclusions per endpoint for C&L, PBT/vPvB and dose descriptor** For the regulatory purposes of REACH, it should additionally be listed and substantiated, per endpoint and substance, whether: - C&L is similar to the source chemical; - PBT/vPvB is similar to the source chemical; - the dose descriptor is similar to the source chemical, or adaptations are necessary; there are uncertainties in the read-across used that need to be addressed

Table R. 6-3 Data Matrix, Analogue Approach

CAS #				
CHEMICAL NAME	[Target chemical]	[Source Chemical 1]	[…]	[Source Chemical n]
PHYSICO-CHEMICAL DATA				
Melting Point				
Boiling Point				
Density				
Vapour Pressure				
Partition Coefficient (log K_{ow})				
Water Solubility				
…				
ENVIRONMENTAL FATE and PATHWAY				
Photodegradation				
Stability in Water				
Transport and Distribution				
Aerobic Biodegradation				
…				
ENVIRONMENTAL TOXICITY				
Acute Toxicity to Fish				
Acute Toxicity to Aquatic Invertebrates				
Toxicity to Aquatic Plants				
…				
MAMMALIAN TOXICITY				
Acute Oral				
Acute Inhalation				
Acute Dermal				
Repeated Dose				
Genetic Toxicity *in vitro* . Gene mutation . Chromosomal aberration				
Genetic Toxicity *in vivo*				
Reproductive Toxicity . Fertility . Developmental Toxicity				
…				

More detailed discussion of how data gaps are filled for individual endpoints should be provided in the corresponding sections of the assessment report (e.g. SIDS Initial Assessment Report or Chemical Safety Report).

R.6.2.6.2 Reporting Format for a chemical category

1.	**Category definition and its members**
1.1.	Category Definition
1.1.a.	Category Hypothesis Describe the molecular structure 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 i.e. the chemical similarities (analogues), purported mechanisms and trends in properties and/or activities 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).
1.1.b.	Applicability domain (AD) 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. For example, the range of log K_{ow} values or carbon chain lengths over which the category is applicable. The justification for the inclusion and/or exclusion rules should be reported under Section 2) *Category justification* below.
1.2.	Category Members Describe all category members as comprehensively as possible. Provide CAS numbers, names and chemical structures of all category members.
1.3.	Purity / Impurities Provide purity/impurity profiles for each member of the category, including their likely impact on the category endpoints. It should be discussed which influence these impurities are thought to have on physico-chemical parameters, fate and (eco)toxicology, and hence on the read-across.
2.	**Category justification** Based on available experimental data (including appropriate physico-chemical data and additional test results generated for the assessment of this 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 reported under 3) below should support the justification for the read-across. More detailed discussion of available test results for individual endpoints (i.e. discussion of the selection of key studies, variability of experimental results between different members of the category etc.) should be provided in the corresponding sections of the assessment report (e.g. Chapters 2-4 of the SIDS Initial Assessment Report or Chapters 4-7 of the Chemical Safety Report).
3.	**Data matrix** Provide a matrix of data (category endpoints vs. members). It should be constructed with the category members arranged in a suitable order (e.g. according to molecular weight) (Table R. 6-4). For example, the ordering of the members should reflect a trend or progression within the category. In each cell in the Data Matrix, the study result type should be indicated in the first line, e.g.: - experimental result - experimental study planned

	- read-across from supporting substance (structural analogue or surrogate) - trend analysis[37] - (Q)SAR If experimental results are available, the key study results should be shown in the Data Matrix.
4.	**Conclusions per endpoint for C&L, PBT/vPvB and dose descriptor** For the regulatory purposes of REACH, the following information should additionally be listed and substantiated, for each individual member in the category and for each endpoint: - C&L - PBT/vPvB - the dose descriptor there are uncertainties in the category approach used that need to be addressed

Table R. 6-4 Data Matrix, Chemical Category

For data-rich substances, the matrix could become very large, and could therefore be broken down into groups of endpoints.

CAS #					
CHEMICAL NAME	[Category member 1]	[Category member 2]	[Category member 3]	[…]	[Category member n]
PHYSICO-CHEMICAL DATA					
Melting Point					
Boiling Point					
Density					
Vapour Pressure					
Partition Coefficient (log Kow)					
Water Solubility					
…					
ENVIRONMENTAL FATE and PATHWAY					
Photodegradation					
Stability in Water					
Transport and Distribution					
Aerobic Biodegradation					
…					
ENVIRONMENTAL TOXICITY					
Acute Toxicity to Fish					

[37] There are slight differences between the terminology used in the OECD Harmonised templates and hence there might be slight differences in a category matrix automatically generated with software using the OECD Harmonised Templates and the present guidance document. For example there is no item “trend-analysis” in the picklist for the data element “study result type”. Instead the item “read-across based on grouping of substances (category approach)”could be used.

Acute Toxicity to Aquatic Invertebrates					
Toxicity to Aquatic Plants					
…					
MAMMALIAN TOXICITY					
Acute Oral					
Acute Inhalation					
Acute Dermal					
Repeated Dose					
Genetic Toxicity *in vitro* . Gene mutation . Chromosomal aberration					
Genetic Toxicity *in vivo*					
Reproductive Toxicity . Fertility . Developmental Toxicity					

More detailed discussion of how data gaps are filled for individual endpoints and individual category members (e.g. interpolation, extrapolation, (Q)SAR) as well as the rationales for the chosen method of filling the data gaps should be provided in the corresponding sections of the assessment report (e.g. Chapters 2-4 of the SIDS Initial Assessment Report or Chapters 4-7 of the Chemical Safety Report).

For UVCBs it may not be feasible to establish a full data matrix, especially where the number of substances in the category is very large. In such circumstances a single data set or template that applies to all members of the category of UVCBs in exactly the same way will be developed. The template will include a clear indication of which members of the category experimental or calculated data exist, and hence maintain complete transparency.

R.6.2.7 Case study using phosphonic acid compounds and alkali metal salts

1.	**Category definition and its members**
1.1.	Category Definition
1.1.a.	Category Hypothesis This category covers 1-Hydroxy-1,1-ethane-diphosphonic acid (HEDP) and various sodium and potassium salts of that acid. The different salts are prepared by neutralising the acid to a specific pH. All category members are based on the HEDP structure, which can be de-protonated up to 5 times. The category hypothesis is that all the members are various ionised forms of the acid 2809-21-4. The main assumption is that sodium and potassium are not significant in respect of all the properties under consideration. In dilute aqueous conditions of defined pH a salt will behave no differently to the parent acid, at identical concentration of the particular speciated form present and will be fully dissociated. Hence some properties (measured or expressed in aqueous media, e.g. ecotoxicity) for a salt can be directly read-across (with suitable mass correction) to the parent acid and vice versa. Where dermal or irritation studies are available the acid and salts are considered separately. The properties of HEDP and its salts are profoundly directed by their ionisation behaviour and complexation of metal ions.
1.1.b.	Applicability domain (AD) of the category The category applies to HEDP and all of its possible sodium and potassium salts.
1.1.c.	List of endpoints covered

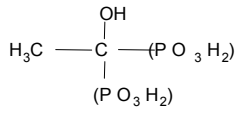
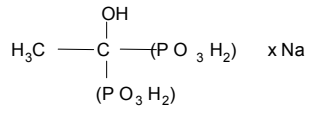
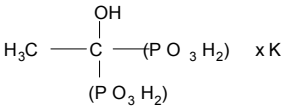
	The category approach was applied to the following endpoints: Dissociation constant and metal complexation Octanol-water partition coefficient Adsorption Biodegradation Stability in water Bioaccumulation Ecotoxicity tests Mammalian toxicity (other than dermal administration) Genotoxicity The category approach was not applied to skin irritation, eye irritation and dermal toxicity since the acid is much more corrosive than its salts.
1.2.	Category Members See Table R. 6-5; Structural formulas:  1-Hydroxy-1,1-ethane-diphosphonic acid CAS # 2809-21-4  1-Hydroxy-1,1-ethane-diphosphonic acid, xNa Salt CAS # 29329-71-3  1-Hydroxy-1,1-ethane-diphosphonic acid, xK Salt CAS # 67953-76-8
1.3.	Purity / Impurities Since the salts are prepared from the acid, the impurity profile for HEDP acid given in Table R. 6-6 below is also typical of the salts in this Category, although acidic impurities would also be present as salts. Exact proportions vary slightly between manufacturers and precise values are not given, to protect commercial interests. All are typical for marketed substance. In addition to those impurities listed in Table R. 6-6, HEDP contains up to 4% of two phosphonic acid components, not unrelated to the main component. Exact details are commercially confidential.
2.	**Category justification** HEDP and its salts all have high water solubility, low Log K_{ow}, and low vapour pressures. Their behaviour in water and biological systems is dominated by their ionisation and complexation of metal ions. Measured data was available for environmental endpoints for HEDP and its 2Na salt and for health endpoints for HEDP, its 2Na salt and 4Na salt. Thus, data is read-across to the remaining Na salts and to all potassium salts. Data for HEDP and the 2Na salt showed low acute toxicity to fish, this result was read-across to the remaining salts. Data for HEDP and the 2Na salt showed low acute toxicity to Daphnia, which was read-across to the other category members. However, the available data indicated that the 2Na salt has a much higher chronic toxicity to Daphnia than HEDP. This result is not consistent with the general pattern of toxicity and therefore a repeat test was requested on the 2Na salt (result not yet available). If the test confirms the chronic toxicity of the 2Na salt, the category may be called into question for aquatic toxicity endpoints. Data for the toxicity of HEDP and its 2Na salt to algae shows toxicity, but evidence shows that these effects are a consequence of complexation of essential nutrients and not of true toxicity. This conclusion applies to the whole category.
3.	**Data matrix** More detailed discussion of how data gaps are filled for individual endpoints and individual category members should be provided in the corresponding sections of the assessment report (e.g. SIDS Initial Assessment Report or Chemical Safety Report). See Table R. 6-7, Table R. 6-8
4.	**Conclusions per endpoint for C&L, PBT/vPvB and dose descriptor**

Table R. 6-5 Category Members

Substance	CAS
1-Hydroxy-1,1-ethane-diphosphonic acid	2809-21-4
1-Hydroxy-1,1-ethane-diphosphonic acid, xNa Salt	29329-71-3
1-Hydroxy-1,1-ethane-diphosphonic acid, Na Salt	17721-68-5
1-Hydroxy-1,1-ethane-diphosphonic acid, 2Na Salt	7414-83-7
1-Hydroxy-1,1-ethane-diphosphonic acid, 3Na Salt	2666-14-0
1-Hydroxy-1,1-ethane-diphosphonic acid, 4Na Salt	3794-83-0
1-Hydroxy-1,1-ethane-diphosphonic acid, 5Na Salt	13710-39-9
1-Hydroxy-1,1-ethane-diphosphonic acid, xK Salt	67953-76-8
1-Hydroxy-1,1-ethane-diphosphonic acid, K Salt	17721-72-1
1-Hydroxy-1,1-ethane-diphosphonic acid, 2K Salt	21089-06-5
1-Hydroxy-1,1-ethane-diphosphonic acid, 3K Salt	60376-08-1
1-Hydroxy-1,1-ethane-diphosphonic acid, 4K Salt	14860-53-8
1-Hydroxy-1,1-ethane-diphosphonic acid, 5K Salt	87977-58-0

Table R. 6-6 Impurity profile for HEDP

CAS-No	EC-No	EINECS-Name	Mol. Formula	Contents % w/w
64-19-7	200-580-7	Acetic acid	$C_2H_4O_2$	< 1
7647-01-0	231-595-7	Hydrogen chloride	HCl	< .1
13598-36-2	237-066-7	Phosphonic acid	H_3PO_3	< 4
7664-38-2	231-633-2	Orthophosphoric acid	H_3PO_4	< 2

Table R. 6-7 Physico-chemical properties and environmental fate

Substance	CAS	Water solubility	Log Kow	Vapour pressure	Melting point	pKa	Vapour pressure	Koc	biodegradability
1-Hydroxy-1,1-ethane-diphosphonic acid	2809-21-4	690 g/l: 60% w/w produced commercially	-3.52	1.24 x 10^{-9} Pa (estimated)	198-199° C; decomposes around 228° C	Four pKa values of HEDP (at 0.1 M ionic strength potassium nitrate): 1.6, 2.7, 6.9, 11.0.	1.24 x 10^{-9} Pa (estimated)	16610	Not readily biodegradable (NRB) (measured)
1-Hydroxy-1,1-ethane-diphosphonic acid, xNa Salt	29329-71-3	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, Na Salt	17721-68-5	465 g/kg solution	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, 2Na Salt	7414-83-7	278 g/kg solution	'low'	'low'	-	-	'low'	'high'	Not readily biodegradable (measured)
1-Hydroxy-1,1-ethane-diphosphonic acid, 3Na Salt	2666-14-0	123 g/kg solution	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, 4Na Salt	3794-83-0	513 g/kg solution	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, 5Na Salt	13710-39-9	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, xK Salt	67953-76-8	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, K Salt	17721-72-1	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, 2K Salt	21089-06-5	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, 3K Salt	60376-08-1	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across

Substance	CAS	Water solubility	Log Kow	Vapour pressure	Melting point	pKa	Vapour pressure	Koc	biodegradability
1-Hydroxy-1,1-ethane-diphosphonic acid, 4K Salt	14860-53-8	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across
1-Hydroxy-1,1-ethane-diphosphonic acid, 5K Salt	87977-58-0	'high'	'low'	'low'	-	-	'low'	'high'	NRB – read-across

Table R. 6-8 Ecotoxicity endpoints

Substance	CAS	Fish acute toxicity 96h LC50 mg/l	Daphnia acute toxicity 48h EC50 mg/l	Daphnia chronic toxicity 22d NOEC mg/l	Algal toxicity 96h EC50	Algal toxicity NOEC mg/l	toxicity to microorganisms 30-min EC0 mg/l
1-Hydroxy-1,1-ethane-diphosphonic acid	2809-21-4	200	167	6.75 (28-day)	3	13 (14d)	>580
1-Hydroxy-1,1-ethane-diphosphonic acid, xNa Salt	29329-71-3	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, Na Salt	17721-68-5	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, 2Na Salt	7414-83-7	360	500	0.1	Nutrient complexation	3- (14d)	960
1-Hydroxy-1,1-ethane-diphosphonic acid, 3Na Salt	2666-14-0	'low'	'low''	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, 4Na Salt	3794-83-0	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, 5Na Salt	13710-39-9	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, xK Salt	67953-76-8	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'

Substance	CAS	Fish acute toxicity 96h LC50 mg/l	Daphnia acute toxicity 48h EC50 mg/l	Daphnia chronic toxicity 22d NOEC mg/l	Algal toxicity 96h EC50	Algal toxicity NOEC mg/l	toxicity to microorganisms 30-min EC0 mg/l
1-Hydroxy-1,1-ethane-diphosphonic acid, K Salt	17721-72-1	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, 2K Salt	21089-06-5	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, 3K Salt	60376-08-1	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, 4K Salt	14860-53-8	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'
1-Hydroxy-1,1-ethane-diphosphonic acid, 5K Salt	87977-58-0	'low'	'low'	Re-testing 2Na salt	Nutrient complexation	Nutrient complexation	'low'

R.6.2.8 References for the guidance on grouping of chemicals

Aptula A.O., Patlewicz G., Roberts D.W. & Schultz T.W. (2006). Non-enzymatic glutathione reactivity and in vitro toxicity: A non-animal approach to skin sensitization. *Toxicology In Vitro* **20**, 239-47.

ATC (2000a). An internationally recognised nomenclature system for petroleum additives. ATC Doc 31, rev 4. Technical Committee of Petroleum Additive Manufacturers in Europe.

ATC (2000b). Classification and user labelling information concerning the health effects of major petroleum additive components. ATC Doc 43, rev 4. Technical Committee of Petroleum Additive Manufacturers in Europe.

Benigni R, Conti L, Crebelli R, Rodomonte A & Vari' MR (2005). Simple and alpha,beta-unsaturated aldehydes: correct prediction of genotoxic activity through structure-activity relationship models. *Environ Mol Mutagen*. **46**, 268-280

Caley J, Henry T, Merckel D, Sonowane M, Benson A (2007). Generic lessons learnt from a review of categories presented at OECD Screening Information Assessment Meetings (SIAM) since April 2004. In: A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar/background/index.php?c=CA

CESIO (2000). CESIO recommendations for anionic and non-ionic surfactants. CEFIC, Brussels.

CESIO (2003). CESIO recommendation for the classification and labelling of surfactants as 'dangerous for the environment'. CEFIC, Brussels.

CONCAWE (2001). Classification and labelling of petroleum substances according to the EU dangerous substances directive (Recommendation, August 2001). Concawe, Brussels.

Clemedson C, Nordin-Andersson M, Bjerregaard HF, Clausen J, Forsby A, Gustafsson H, Hansson U, Isomaa B, Jorgensen C, Kolman A, Kotova N, Krause G, Kristen U, Kurppa K, Romert L & Scheers E. (2002). Development of an in vitro test battery for the estimation of acute human systemic toxicity: An outline of the EDIT project. Evaluation-guided Development of New In Vitro Test Batteries. *Alternatives to Laboratory Animals* **30**, 313-321.

Clemens, J.H., van den Heuvel, M.R., Stegeman, J.J., Dixon, D.G. and Bols, N.C. (1994) A comparison of toxic equivalency factors derived with rainbow trout, RTL-W1 and rat H4IIE cell lines for polychlorinated biphenyls (PCBs). *Society of Environmental Toxicology and Chemistry* Abstracts **14**, 280.

Comber M & Simpson B. (2007). Grouping of Petroleum substances. A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across.. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar/background/index.php?c=CAT

CPIA (undated). Chlorinated Paraffins: A status report. Chlorinated Paraffins Industry Association. Accesssible from: http://www.regnet.com/cpia/publications.htm

Draft OECD HPV Chemicals Programme SIDS assessment (UK/ICCA)

ECB (2005). Scoping study on the development of a technical guidance document on information requirements on intrinsic properties of substances. Report prepared by CEFIC, DK-EPA, Environmental Agency of Wales and England, ECETOC, INERIS, KemI and TNO. European Chemicals Bureau, Joint Research Centre, European Commission, Ispra, Italy. Accessible from: http://ecb.jrc.it/documentation/

EC (2006). European Union Risk Assessment Report, Nickel and Nickel Compounds, Nickel CAS-No.: 7440-02-0, Nickel Sulphate CAS-No.: 7786-81-4, Nickel Carbonate CAS No.: 3333-67-3; Nickel Chloride CAS-No.: 7718-54-9; Nickel Dinitrate CAS-No.: 13138-45-9, unpublished draft.

EC (2007a). European Union Risk Assessment Report, Volume 72, cadmium oxide and cadmium metal, Part I Environment (CAS-No.: 1306-19-0 & Cadmium metal CAS-No.: 7440-43-9), EUR 22919 EN

EC (2007b). European Union Risk Assessment Report, Volume 74, cadmium metal, Part II Human Health (CAS-No.: 7440-43-9), EUR 22767 EN

EC (2007c). European Union Risk Assessment Report, Volume 75, cadmium oxide, Part II Human Health (CAS-No.: 1306-19-0), EUR 22766 EN

EC (2007d). IUCLID Manual. Available via: http://www.echa.europa.eu/

EC (2008), Addendum to European Union Risk Assessment Report, Volume 72, cadmium oxide and cadmium metal, Part I Environment (CAS-No.: 1306-19-0 & Cadmium metal CAS-No.: 7440-43-9) unpublished draft.

Environment Canada 2003. Guidance Manual for the Categorisation of Organic and Inorganic Substances on Canada's Domestic Substances List. Existing Substances Branch, Environment Canada, Gatineau, Canada, 124 p

ETAD (2001). Guidance for the user industry on the environmental hazard labelling of dyestuffs. Ecological and toxicological assocaiation of dyes and organic pigments manufacturers. Basel, Switzerland.

Eurobitume (2002). Guidelines on the Classification of Bitumens & Bituminous Preparations. Eurobitume, Brussels.

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

Gallegos Saliner A et al. (2007). Review of examples of read-across/category approaches discussed by the TC C&L. In: A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar/background/index.php?c=CAT.

Hanway, R.H. & Evans, P.F. (2000). Read-across of toxicological data in the notification of new chemicals. *Toxicology Letters* 116(S1), 61.

Hart J (2007). Nickel compounds – a category approach for metals in EU Legislation. A report to the Danish Environmental Protection Agency. http://cms.mim.dk/NR/rdonlyres/07DB028E-134E-4796-BF6D-97B9AD5F9E82/0/Nikkel.pdf.

Hart J & Veith GD (2007): Applying Chemical Categories to Classification & Labelling: A Case Study of Volatile Aliphatic Ethers. A report to the Danish Environmental Protection Agency. Available at www.qsari.org/documents/aliphaticethers.pdf

HSPA (2002). The classification of petroleum solvent streams and related complex hydrocarbon solvents for aquatic environmental effects under the EU dangerous substances directive. Hydrocarbon Solvents Producers Association. Brussels 2002.

IHSC (2004/2005). Proposed hydrocarbon solvent category approach for the OECD SIDS programme. Presented at the January 2004 OECD SIDS category workshop and published in Section 3.2 of the OECD SIDS Manual (May 2005). Accessible at: http://www.oecd.org/dataoecd/60/47/1947509.pdf

Jäckh R (2007). Investigations of read-across & grouping, including mechanistic considerations and rationales for outliers. In: A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar/background/index.php?c=CAT.

OECD (2000). Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures No. 23, ENV/JM/MONO(2000)6.

OECD (2001). OECD Environment, Health and Safety Publications. Series on Testing and Assessment, No. 29. Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media.

OECD (2004). OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models. http://www.oecd.org/document/23/0,2340,en_2649_34379_33957015_1_1_1_1,00.html

OECD (2005). 3.2 Guidance on the Development and Use of Chemical Categories in HPV Chemicals Programme http://www.oecd.org/document/7/0,2340,en_2649_34379_1947463_1_1_1_1,00.html

OECD (2006a). Report on the Regulatory Uses and Applications in OECD Member Countries of (Q)SAR Models in the Assessment of New and Existing Chemicals. OECD Environment Health and Safety Publications. Series on Testing and Assessment. No. 58

OECD (2006b). OECD Harmonised Templates. http://www.oecd.org/document/13/0,2340,en_2649_34379_36206733-_1_1_1_1,00.html

Parrott, J.L., Hodson, P.V., Dixon, D.G. and Servos, M.R. (1992) Liver EROD induction by several 2,3,7,8- chlorinated dibenzofurans in rainbow trout, *Society of Environmental Toxicology and Chemistry Abstracts* **13**, 113.

OECD (2007a). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. OECD Series on Testing and Assessment No. 69. ENV/JM/MONO(2007)2. Organisation for Economic

Cooperation and Development, Paris, France. 154 pp. http://www.oecd.org/document/30/0,2340,-en_2649_34365_1916638_1_1_1_1,00.html

OECD (2007b). Manual for Investigation of HPV Chemicals. http://www.oecd.org/document/7/0,2340,en_2649-_34379_1947463_1_1_1_1,00.html

Salvito, D. (2007). Approaches to Chemical Categorisation: An Illustrative Example of Approaches Used by the Fragrance Industry. In: : A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar/background/index.php?c=CAT.

Safe, S. (1991) Polychlorinated dibenzo-*p*-dioxins and related compounds. Sources, environmental distribution and risk assessment, *Environmental Carcinogenicity and Ecotox. Reviews* **C9**(2), 261-302.

SCHER (Scientific Committee on Health and Environmental Risks) (2006): Opinion on Classification of Musk ketone. Adopted by the SCHER during the 9th plenary of 27 January 2006; scher_o_022.

Schoeters I & Verougstraete V (2007). Case studies on inorganic metal compounds. In: A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar/background/index.php?c=CAT.

Schultz, T.W., Cronin, M.T.D., Netzeva, T.I. & Aptula, A.O. (2002). Structure-toxicity relationships for aliphatic chemicals evaluated with Tetrahymena pyriformis. *Chemical Research in Toxicology* **15**, 1602-1609.

Tsakovska I & Worth A (2007). Review of examples of read-across/category approaches discussed by the TC NES. In: A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar-/background/index.php?c=CAT.

UNEP Chemicals (2006) Screening Information Datasets for High Volume Chemicals. http://www.chem.unep.ch/irptc/sids/OECDSIDS/sidspub.html

US-EPA (2004). Estimation of relative bioavailability of lead in soil and soil-like materials using *in vivo* and *in vitro* methods. OSWER 9285.7-77, Office of Solid Waste and Emergency Response, Washington, DC.

UV/EB Acrylate Resins (2003). Guide to the classification and labelling of UV/EB Acrylates. CEFIC, Brussels 2003.

van der Weiden, M.E.J., van der Kolk, J., Bleumink, R., Seinen, W. and van den Berg, M. (1992) Concurrence of P450 1A1 induction and toxic effects after administration of a low dose of 2,3,7,8-tetrachloro-*p*-dioxin (TCDD) in rainbow trout (*Oncorhynchus mykiss*), *Aquatic Toxicology* **24**, 123-142.

Walker, M.K. and Peterson, R.E. (1991) Potencies of polychlorinated dibenzo-*p*-dioxin, dibenzofuran and biphenyl congeners, relative to 2,3,7,8-tetrachloro-p-dioxin, for producing early life stage mortality in rainbow trout (*Oncorhynchus mykiss*), *Aquatic Toxicology* **21**, 219-238.

Worth A et al. (2007). Possible applications of QSAR methods. In: A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across. Worth A & Patlewicz G (Eds). EUR report no 22481 EN. Available from ECB website: http://ecb.jrc.it/qsar/background/index.php?c=CAT.

Zabel, E.W., Cook, P.M. and Peterson, R.E. (1995) Toxic equivalency factors of polychlorinated dibenzo-*p*-dioxin, dibenzofuran and biphenyl congeners based on early life stage mortality in rainbow trout (*Oncorhynchus mykiss*), *Aquatic Toxicology* **31**, 315-328.