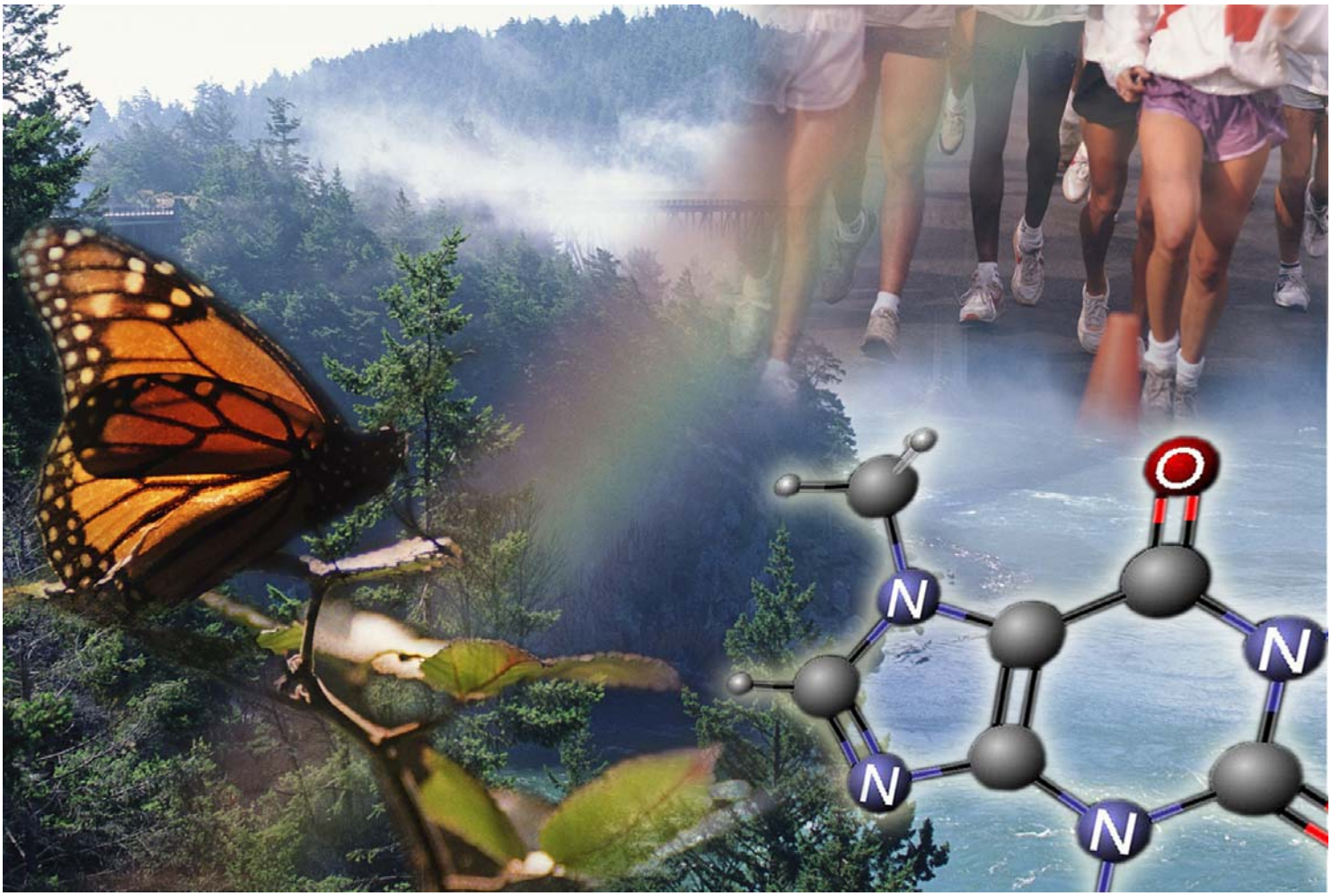
Guidance on information requirements and chemical safety assessment

Chapter R.7a: Endpoint specific guidance



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, uses 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/about/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

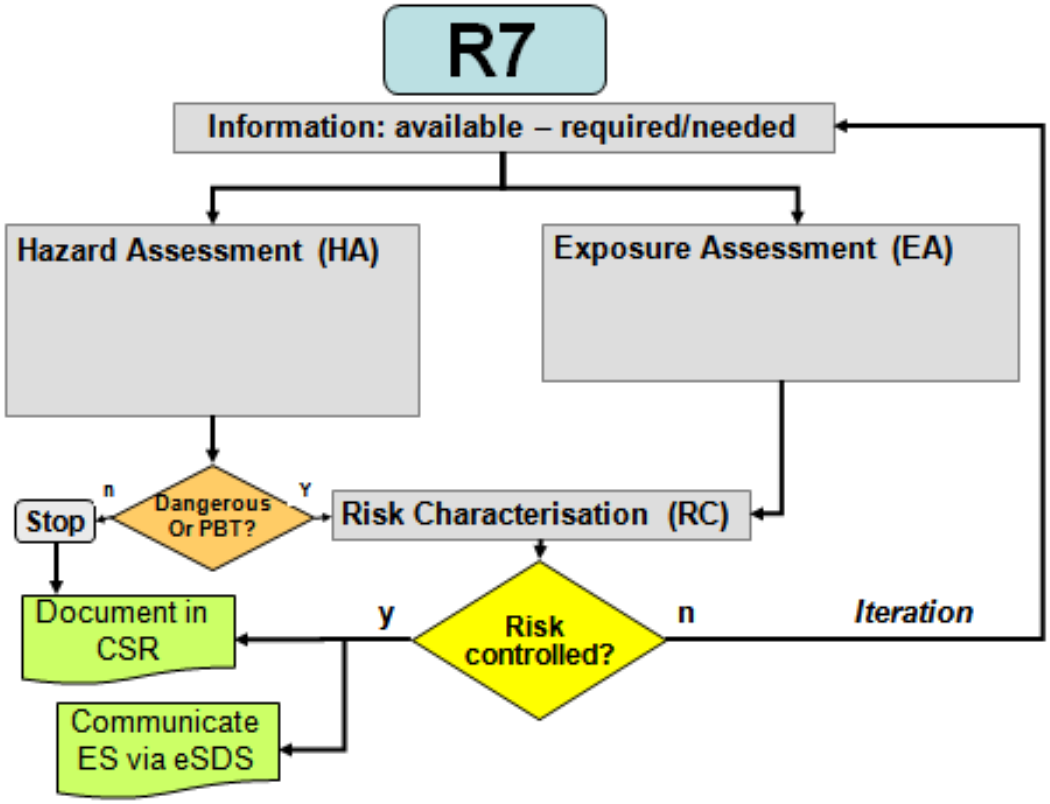
Where the REACH regulation is cited literally, this is indicated by text in italics between quotes.

Table of Terms and Abbreviations

See Chapter R.20

Pathfinder

The figure below indicates the location of part R.7(a) within the Guidance Document



CONTENTS

TABLE OF CONTENTS

TABLES

FIGURES

APPENDICES

R.7 ENDPOINT SPECIFIC GUIDANCE

Introduction

The previous sections provide advice on the interpretation and application of generic aspects of the regulation describing 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. The chapters describe also factors that may modify the information requirements and give advice on how all gathered data from different sources could be integrated and used in a *weight of evidence* (WoE) approach to allow a conclusion on whether or not the available information is sufficient for the regulatory purposes.

The guidance given thus far is applicable across the board and comprises the general rules that should be followed.

In this chapter, specific guidance on meeting the information requirements on physico-chemical properties and the different human health and the environmental endpoints is presented. The guidance for each specified endpoint has been developed as a stand-alone report addressing the aspects of gathering, evaluation and generation of information to help registrants provide adequate and relevant information for registration under REACH.

All data sources, including non-testing data are taken into account. Most of the reports follow a logical common format that complements the generic guidance and the General Decision Making Frameworks detailed above. There are six main sections to the guidance on each endpoint; the first section provides an introduction in which the endpoint is described, defined and an explanation given of its importance in the context of human health or environmental fate and effect of a given substance. The second details the specific information requirements for the endpoint of interest; these will depend on the tonnage band of the chemical, its usage pattern and other considerations including data on other endpoints and on related chemicals. Endpoint specific guidance can be thought of as the four logical steps that should be taken to assemble the information that is detailed under the second section; thus, the third section provides an inventory of all the types of data that could potentially provide useful information on the endpoint of interest and, most importantly the sources of that information.

Guidance is given in the fourth section on how to evaluate the information that might be available for a given substance; this advice focuses on providing the criteria to aid in the judgement and ranking of the available data for their adequacy and completeness. This section also provides an indication of the remaining uncertainty inherent in the different types of data for the given endpoint.

The fifth section describes how conclusions may be drawn for a given substance on the suitability of the available information for regulatory purposes. Guidance is given on how to develop and apply a Weight of Evidence approach for the endpoint in order to establish whether there is a need for further information and if so, what test should be performed. Chemical safety assessment within REACH is fundamentally dependent on an adequate conclusion on classification and PBT/vPvB assessment since exposure assessment and risk characterisation are triggered by classification and fulfilment of PBT/vPvB criteria. Therefore data need to be adequate for both classification & labelling and for chemical safety assessment if the latter is required.

Currently in the EU, dangerous substances and preparations must be classified and labelled according to Directives 67/548/EEC and 1999/45/EC respectively. It should be noted that these

Directives will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS). Guidance to this Regulation is being developed. Thus, where reference is made in the following sections to these documents, the criteria or to guidance on classification and labelling for the specified endpoints, it is advisable to check their status that will be given on the European Chemicals Agency (ECHA) website (http://echa.europa.eu/classification_en.asp).

The final section comprises an Integrated Testing Strategy (ITS) for the given endpoint(s) giving guidance on how to define and generate relevant information on substances in order to meet the requirements of REACH. It is noteworthy that all experiments using vertebrate animals shall be designed to avoid distress and unnecessary pain and suffering to experimental animals (Article 7.4 of Directive 86/609).

The proposed testing strategies are guidance for data generation in a stepwise approach. The strategies build on the concept that if the available information is not sufficient to meet the regulatory needs, further gathering of information at a succeeding step in the testing strategies is needed. On the other hand, if the available information is adequate and the standard information requirements are met, no further gathering of information is necessary. In those cases where the available information is judged to be sufficient to meet the regulatory needs even though the standard information requirements are not met, under certain circumstances, in particular for REACH Annexes IX and X, this might be part of a justification for waiving a certain test that is requested in the standard information requirements.

The following additional considerations apply generally to the endpoint specific guidance given in this chapter:

Use of test data derived from EU or international standardised methods

According to REACH, Article 13(3), tests required for generating information on intrinsic properties of substances shall be conducted in accordance with the test methods included in a Commission Regulation or in accordance with other international test methods recognised by the Commission or the Agency as being appropriate. Toxicological and ecotoxicological tests and analyses shall be carried out in compliance with the principles of Good Laboratory Practice. The new Test Methods Regulation is currently (March 2008) under adoption and contains all the test methods previously included in Annex V to Directive 67/548/EEC. Data generated by any of these methods are *per se* considered adequate for regulatory use. Other internationally standardised test methods are recognised by the Commission or the Agency as being adequate for generating data for regulatory use, most notable being those guidelines developed and published by the OECD. Thus, in the following sections on specific endpoints, references given for each test method will include the OECD TG number and, where available, the EU test method number.

It is the intention of the Commission that the TM Regulation be adapted to technical progress whenever a new test method has been developed, scientifically validated and accepted for regulatory use by the National Coordinators of the Member states..

The following additional considerations apply generally to the endpoint specific guidance given in this chapter:

Inter-relationship of endpoints in hazard assessment

Although guidance is provided for each specific endpoint separately, it should be remembered that different endpoints are related to each other. Information collected within one endpoint may influence hazard/risk assessment of other endpoints, e.g. information on a fast primary degradation

of a parent compound may result in including the degradation products in the overall assessment of the toxicity of a substance. Similarly, information on toxicity/specific mode of action in one endpoint may indicate possible adverse effects for organisms considered for assessment of other endpoints, e.g. endocrine disrupting mode of action in mammals may indicate the same mode of action in fish. Another example may be when data on toxic effects measured in one group of organisms may be directly used in more than one endpoint, e.g. data from a repeated dose toxicity study may also be used in assessment of risk for secondary poisoning of mammals exposed via the food chains.

Adequacy of methods for generating additional information

REACH Article 13.3 states that any new tests should be "conducted in accordance with the test methods laid down in a Commission Regulation or… other international test methods recognised by the Commission or the Agency as being appropriate…. Information on intrinsic properties may also be generated using other test methods provided they meet the conditions set out in Annex XI."

Furthermore, new ecotoxicological and toxicological tests shall be carried out in compliance with the principles of GLP (see Directive 2004/10/EC) or equivalent international standards eventually recognised by the Commission or the Agency. This does not apply to tests for physico-chemical properties.

It is important to emphasise that testing on animals should be seen as the last resort. Testing on animals should only be proposed when the registrant considers it necessary to obtain additional information for assessing and documenting that risks are adequately controlled. Therefore, it is important to first consider all issues that may impact upon this decision, such as:

- testing requirements,
- exposure/use pattern (emissions, yes or no, consumer use etc),
- occurrence (monitoring data),
- indications of the effect/ property based on animal or human data, *in vitro* data and non-testing information,
- any concern e.g. based on toxicokinetics, read-across and (Q)SAR considerations;
- WoE,
- seriousness of the effect,
- other effects of relevance for the endpoint.

All these issues should be considered, not only to design *fit for purpose in vivo* tests, but also for providing evidence for not performing *in vivo* testing under certain circumstances.

Degradation products and metabolites:

In the context of evaluating substances for their effects, it is important to note that when they are released into the environment or taken up by animals, they may be transformed through degradation or metabolism. These processes and their outcome may need to be taken into account in the overall assessment.

Degradation products may be formed as a result of transformation processes in the environment. For distinguishing the substance undergoing degradation from the degradation products, the former is often referred to as *the parent compound*.

Degradation products may be formed as a result of abiotic environmental processes such as hydrolysis, direct or indirect photolysis or oxidation. They may also be formed as a result of aerobic or anaerobic biodegradation, i.e. due to microbial activity. Degradation products require further investigation if the Chemical Safety Assessment indicates the need, i.e. if stable degradation products are formed in the environment within a relevant time frame, as deduced from the test system, and in addition are so stable that they may result in risk, or if they are relevant considering in the PBT-assessment (i.e. are either both very persistent and very bioaccumulative or persistent, bioaccumulative and toxic). Likewise it may be considered to assess whether degradation products fulfil the environmental hazard classification criteria (see Section R.7.9).

Metabolites refer to transformation products, which are formed due to biodegradation (and then the term metabolite is synonymous with the term biodegradation product) or formed as a result of biotransformation (metabolism) within exposed organisms after uptake of the parent compound. Metabolic pathways and hence the identity of many metabolites may or may not be fully known. The latter is frequently the case. Moreover for the same substances metabolic pathways may or may not differ between various organisms belonging to different phyla and/or trophic levels. However, the toxicity of metabolites formed within the duration of laboratory tests will with the exception of delayed effects showing up after the observation time of the tests be reflected in such studies with their parent compound. Knowledge of metabolic pathways and metabolites may increase planning and focussing of toxicity testing, understanding of toxicological findings. (see Section R.7.12) and may in some cases make it possible to use *grouping approaches* for structurally closely related substances, which undergo similar metabolic transformation (see Section R.6.2).

Because many biotransformation processes includes oxidation, metabolites are often less hydrophobic than the parent compound. This is a very general rule of thumb and may not always apply; however, when it does, often this has implications for the hazard profile of the metabolites. More polar metabolites created after oxidation processes have for example normally a lower adsorption potential, and thus the relevance of the metabolites for the soil and sediment compartments is normally lower than that of the parent compound. Such less hydrophobic metabolites also tend to be excreted more rapidly from organisms than the parent compound. Hence their bioaccumulative potential and narcotic toxicity tend to be less.

Similarities in metabolic pathways of structurally similar substances may serve as an alert or waiving for further investigation depending of the case and nature of the metabolites.

Selection of the appropriate route of exposure for toxicity testing

Having comprehensively established the need for additional toxicity testing to meet the requirements of REACH for a given chemical, for certain endpoints, notably acute or repeated dose toxicity but also reproductive toxicity, chronic toxicity and carcinogenicity, a decision must be made on which route of exposure is most appropriate. The overall objective of such testing is to establish the potential hazard of the test chemical to human beings. Humans may be exposed to substances by one or more of three routes: inhalation, dermal or oral. Fundamentally, the use of the inhalation route of administration in animal tests should be considered when inhalation exposure of humans is of relevance. However, the final decision on which route of exposure to use in a particular test should be taken in the light of all available information including physico-chemical properties of the substance; structure-activity relationships (SAR) or the data from available toxicity tests on the substance itself.

Route-to-route extrapolation can be used to assess potential health effects and its threshold in a route other than the one tested. Although toxicity data obtained using the appropriate route of exposure are preferred, REACH stipulates that animal welfare and scientific considerations are to be taken into account before conducting additional animal tests using a more appropriate route of exposure. Route-to-route extrapolation should be considered on a case-by-case basis and may introduce additional uncertainties, especially if the toxicity data were obtained using an administration route that does not correspond to the most relevant route of human exposure. In a subsequent risk assessment the uncertainties introduced through route-to-route extrapolation should be taken into account, for example by adjusting the assessment factor in the determination of the DNEL (see Section R.8.4.3).

Further guidance on this strategic approach to toxicity testing is given in Chapter R.8.

With regard to the evaluation of the environmental impact of a chemical, the interaction of that chemical with the environment is an important consideration. The fate and behaviour of a substance is largely governed by its inherent physicochemical properties. Knowledge regarding the physicochemical properties of the chemical together with results from multimedia fate and transport models (e.g. Mackay level 3 models) enables the identification of the environmental compartment(s) of primary concern. Such information will also inform the prioritisation of higher tiered tests. More extensive guidance and considerations on this aspect are given in R.16 but the phenomenon of substances moving from one environmental compartment to another is also relevant for other environmental endpoints, e.g. transfer from the water column to sediment.

R.7.1 Physico-chemical properties; adsorption/desorption

R.7.1.1 Introduction

The majority of substances registered under REACH will require a full physico-chemical data set according to Annex VIII. These data represent a basic set of information that are used to assess the physical hazards (e.g. flammability) and help predict possible toxicological or environmental hazards, fate and behaviour. It is important therefore, that the data are as accurate as possible. For higher tonnage chemicals, some additional physico-chemical data may be required according to Annex IX

For substances for which new testing needs to be carried out, some consideration should be given to the order in which the tests should be conducted (after taking into account data already available). This is because the results of one test can influence how and/or whether another test should be performed. For example, explosive substances would not normally be tested for flammability. Further details are given in Section R.7.1.1.4 and under specific endpoints.

For some endpoints, more than one technique is described in the standard test method. The one chosen should be suitable for the substance in question and be operating within its validity range.

Several tests are inter-related; for example, water solubility and partition coefficient, vapour pressure and boiling point. The results of these inter-related tests can be used to check the validity of the data. Details of inter-related endpoints are given under each endpoint.

R.7.1.1.1 Information requirements on physico-chemical properties

Physico-chemical data is primarily numeric. For each endpoint, a value or range is required. For some endpoints, a concentration and/or temperature are also required. Where available, details of any observations made during testing should be reported, e.g. decomposition during melting or boiling, emulsion formation during partitioning.

The required physico-chemical endpoints for substances at a supply level of ≥1 t/y are given in Table R.7.1-1.

Table R.7.1-1 Annex VII (required for substances at a supply level of ≥1 t/y)

Melting/freezing point	Flash point
Boiling point	Flammability[2]
Relative density	Explosive properties
Vapour pressure	Self- ignition temperature
Surface tension	Oxidising properties
Water solubility	Granulometry (Particle size distribution)
Partition coefficient n-octanol/water	

Annex VIII (required for substances at a supply level of ≥10 t/y):

- Adsorption/desorption screening

Annex IX (required for substances at a supply level of ≥100 t/y):

- Stability in organic solvents
- Dissociation constant
- Viscosity

The role that these properties and related effects play in human and ecotoxicological risk assessment will be discussed in the subsequent sections and in the endpoint specific chapters.

R.7.1.1.2 Available information on physico-chemical properties

There are many published sources of physico-chemical data. Many of these are compilations of data from other sources. Useful reference books and data compilations containing peer reviewed physico-chemical data include:

- The Merck Index;
- The CRC Handbook of Chemistry and Physics;
- The IUPAC Solubility Data Series;
- Beilstein Database and;

[2] Flammability includes ‘pyrophoric properties’, ‘flammability on contact with water’ and ‘flammability upon ignition’

- Illustrated Handbooks of Physico-chemical Properties and Environmental Fate for Organic Chemicals.

A more complete, but non-exhaustive, list of useful sources of physico-chemical data is included in Table R.7.1-2.

Nowadays, several large reference texts are also available as online databases and not only in printed format. Online databases such as the SRC PhysProp Database[3], ChemFinder[4] and HSDB[5] on the TOXNET network are good sources of data and generally provide a reference for the value that they have selected. These databases often use secondary data sources. In these cases, the original data source should be checked and used rather than directly citing the database.

Data may also be available from the IUCLID database (http://ecb.jrc.it/esis/) or from the OECD HPV programme (http://cs3-hq.oecd.org/scripts/hpv/). It should be noted that the data in IUCLID files has not always been peer reviewed. It should be treated with caution as it may not be reliable.

Another useful online resource is available at the website[6] of the university science libraries at Yale which has hyperlinks for electronic versions of the following handbooks/databases: The Lange's Handbook of Chemistry, Combined Chemical Dictionary, CRC Handbook of Chemistry and Physics, The Beilstein and Gmelin databases, The NIST Chemistry WebBook, Landolt-Bornstein Handbook, and the Yaws Chemical Properties Handbook. These information sources are included in Table R.7.1-2.

Given the wide range of chemical substances available, none of these sources will give information on all possible substances. When a substance is included in any particular publication, a complete data set may not be available and another source may be needed. Many of the sources quoted are updated periodically and thus, a comprehensive review of the scope of each source is not feasible within the realm of this guide.

3 Available on-line at http://esc.syrres.com/interkow/physdemo.h. These data are also used to populate the 'Experimental Database' in the EPIWIN software suite

4 http://chemfinder.cambridgesoft.com/

5 Hazardous Substances Data Bank Available on-line via TOXNET at http://toxnet.nlm.nih.gov/cgi-bin/sis/hlgen?HSDB

6 http://www.library.yale.edu/science/help/chemphys.html

Table R.7.1-2 Sources of physico-chemical data

Source of physico-chemical data	Comments
Merck Index	Physical data are cited as found in the literature. When several alternate data values appear in the literature, the data is evaluated and representative selections are made; values are then reported with the corresponding source.
Hawley's Condensed Chemical Dictionary	This is a compendium of physical data that are taken to be 'reliable'; "where entries are incomplete, it may be presumed that no reliable data were provided by the reference system utilised". [References for values are not provided]
CRC Handbook of Chemistry and Physics	Data for physical constants have been taken from many sources, including both compilations and the primary literature Where conflicts were found, the value deemed most reliable was chosen. [Reference sources are provided for selected properties such as solubility and log K_{ow}; these references are generally authoritative data compilations]. *Online information at*: http://www.crcnetbase.com/)
IUPAC Solubility Data Series	The Solubility Data Series is a project of the International Union of Pure and Applied Chemistry (IUPAC). Publication of the series began in 1979, its goal being to present a comprehensive and critical compilation of data on solubilities in all physical systems, including gases, liquids and solids.
Combined Chemical Dictionary	The Chapman &Hall/CRC Combined Chemical Dictionary is a structured database holding information on chemical substances. It includes descriptive and numerical data on chemical, physical and biological properties of compounds; systematic and common names of compounds; literature references; structure diagrams and their associated connection tables. The Combined Chemical Dictionary online version contains all those compounds published in: Dictionary of Organic Compounds (240,000 records) Dictionary of Natural Products (155,000 records) Dictionary of Inorganic and Organometallic Compounds (100,000 records) Dictionary of Pharmacological Agents (38,000 records) Dictionary of Analytical Reagents (14,000 records) *Online information at*: http://ccd.chemnetbase.com/scripts/ccdweb.exe
Beilstein Database[7]	Beilstein organic substance records contain the critically reviewed and evaluated documents from the Beilstein Handbook of Chemistry as well as data from 176 leading journals in organic chemistry covering the period 1779 to present. [An exhaustive list of values and primary references are provided]
Boethling R.S. and Mackay D. (Eds.), *Handbook of Property Estimation Methods for Chemicals.* Lewis, Boca Raton, FL, 2000	The Handbook of Chemical Property Estimation Methods for Environmental and Health Science reviews methods for estimating Melting point, Boiling Point, Vapour Pressure, Henry's Law Constant, n-octanol/water partition coefficient, Water Solubility, Soil and sediment sorption coefficient.
Illustrated Handbook of Physico-chemical Properties and Environmental Fate for Organic Chemicals (Mackay *et al*)	Physical properties such as melting and boiling point and density are obtained from commonly used handbooks. Other properties such as solubility, vapour pressure, log K_{ow} have been obtained from primary reference sources and handbooks. A range of referenced values is reported for each of these properties. Data have been evaluated and a selected 'best value' is given for each property and used in calculations of environmental distribution.

[7] CrossFire Beilstein, Licensed by MDL Information Systems GmbH

Source of physico-chemical data	**Comments**
SRC PhysProp Database/ EPIWIN Experimental Database	For compounds with abundant data, values have been taken from databases that had already evaluated the data and selected a reliable value. For compounds with less data, values are selected based on a number of factors including the reliability of the source and details of the experimental methodology. [References are provided for all values, except those for melting point and boiling point, and it is clearly indicated whether values are experimental or estimated] *Online information at:* http://www.syrres.com/esc/physdemo.htm
Yaws Chemical Properties Handbook	"Experimental and estimated values are provided in the compilation based on data source publications for organic compounds" [This handbook provides a list of primary references for each property but they are not assigned to particular values or compounds. It is, however, indicated whether data were determined experimentally or estimated]
HSDB on TOXNET	HSDB is peer-reviewed by the Scientific Review Panel (SRP), a committee of experts in the major subject areas within the data bank's scope. All data are referenced and derived from a core set of handbooks, government documents, technical reports and selected primary journal literature". *Online resource at:* http://toxnet.nlm.nih.gov/
ChemIDplus	ChemIDplus is a free, web-based search system that provides access to structure and nomenclature authority files used for the identification of chemical substances cited in National Library of Medicine http://www.nlm.nih.gov/ (NLM) databases, including the TOXNET® system. The database contains over 379,000 chemical records, of which over 257,000 include chemical structures, and is searchable by Name, Synonym, CAS Registry Number, Molecular Formula, Classification Code, Locator Code, and Structure. ChemIDplus gives information on Melting Point ,-Boiling Point, pKa Dissociation Constant, log K_{ow} (octanol-water), Water Solubility, Vapour Pressure, Henry's Law Constant
The Pesticide Manual (currently edited by C Tomlin and previously by CR Worthing).	The introduction to this book (12th Edition) and the discussion of the entries provides no indication that the data has been 'peer reviewed'. There is a brief discussion of vapour pressure (as an example phys-chem property) and it is stated that if there are conflicting values available then the lowest is chosen. A significant proportion of the data is provided directly by manufacturers and is therefore unlikely to have been subject to 'peer review'.
Sax's Dangerous Properties of Industrial Materials	The preface and introduction to this book (10th Edition) provide no indication that the physico-chemical data has been 'peer reviewed'. Physical properties are selected to be useful in evaluating the hazard of a material and designing its proper storage and use procedures. [References for values are not provided]
Bretherick's Handbook of Chemical Reactive Hazards	Several different sources are used. These include primary sources (generally specialist safety journals but also includes general chemical literature), secondary sources (selecting only reactive hazard data) and the direct reporting of incidents to the editors by readers. Full references are given where available. The introduction gives details of the scope and coverage.

Source of physico-chemical data	Comments
CHEMSAFE	CHEMSAFE (Chemical Safety Information) is a factual database that contains evaluated safety characteristics of approximately 2,800 flammable substances and 500 of their mixtures. More than 100 properties, such as flash points, explosion limits, minimum ignition energy, and auto-ignition temperature for gases, liquids, dusts, and hybrid mixtures are included. German and international regulations and classifications, source information, and CAS Registry Numbers are also included. The database is in English. The abstract, controlled terms, mixture descriptors, and substance descriptors are additionally available in German. CHEMSAFE contains two types of documents: factual records and citations. Several factual records may be related to one citation. *Online information at:* http://www.cas.org/ONLINE/DBSS/chemsafess.html
Lange's Handbook of Chemistry	The preface to this book states that "every effort has been made to select the most useful and reliable information and to record it with accuracy" but no references are provided for the data presented and there are no indication as to how they were evaluated. *Online information at:* http://www.knovel.com/knovel2/Toc.jsp?BookID=47
Fire Protection Guide on Hazardous Materials, National Fire Protection Association	No indication is provided on the sources of data or whether they have been 'peer reviewed'. Appendix C of the 12th Edition discusses the preparation of a revised form of the 'Hazardous Chemical Data Sheets' (NFPA 49) contained in this handbook and states that the primary source of information will be safety data sheets. These are not generally regarded as authoritative sources of data for physico-chemical properties. [References for values are not provided]
Dust Explosions in the Process Industries (by R. Eckhoff)	No physico-chemical data are presented in this reference source other than experimental values for median particle diameter and particle size distribution of various dust types and classifications of flammability (these are non-SIDS endpoints).
Verschueren, K. Handbook on Environmental Data on Organic Chemicals.	A useful discussion is provided of the physico-chemical properties that are covered in Verchueren and how they can potentially be used in assessing environmental behaviour but there is no description of sources used to compile the reported data or how they were evaluated. Ranges rather than single values are sometimes presented for parameters such as water solubility and log K_{ow}. [References are not given for phys-chem values but they are provided for entries of biological effect levels, bioaccumulation and degradation rates]
Philip. H. Howard Handbook of Environmental Fate and Exposure Data for Organic chemicals, Lewis Publishers	This series of books outlines in detail how individual chemicals are released, transported and degraded in the environment and how they are exposed to humans and environmental organisms. It is devoted to the review and evaluation of the available data on physico-chemical properties, commercial use and possible sources of environmental contamination, environmental fate and monitoring data of individual chemicals.
Kirk-Othmer Encyclopedia of Chemical Technology. 4th ed. Volume 1: New York, NJ., John Wiley and Sons, 1991	The Kirk-Othmer Encyclopedia of Chemical Technology is designed to present the field of chemical technology to professionals. Written by prominent scholars from industry, academia, and research institutions, the Encyclopedia brings together, and treats systematically, facts on the properties, manufacturing, and uses of chemicals.
ChemFinder	ChemFinder.com is a portal of free and subscription scientific databases. It has a free database that includes: Chemical structures, Physico-chemical properties and links to other websites with information about the query compound. Chemfinder contains information on: Density, Vapour pressure, Flash point, molecular weight, melting and boiling point, water solubility. *Online resource at* : http://chemfinder.cambridgesoft.com/

Source of physico-chemical data	Comments
NIST Chemistry WebBook	Provides information for over 15,000 substances Information in the database includes: thermochemical data (enthalpy of formation, enthalpy of combustion, heat capacity, entropy, phase transition enthalpies and temperatures, vapour pressure), reaction thermochemistry data (enthalpy of reaction, free energy of reaction), IR spectra for over 7,500 compounds, mass spectra for over 10,000 compounds, UV/Vis spectra for over 400 compounds, electronic and vibrational spectra for over 3,000 compounds, constants of diatomic molecules, ion energetics data for over 14,000 compounds (ionisation energy, appearance energy, electron affinity, proton affinity), and thermophysical property data for 16 fluids. Database can be searched using a chemical name, molecular formula, Chemical Abstracts registry number, molecular weight, or selected ion energetics and spectral properties. The information provided is well indexed and provided by a reputable source. The site is free, easy to use, and readily accessible from any computer with Internet access. *Online resource at:* http://webbook.nist.gov/chemistry/
Riddick, J.A., Bunger, W.B., Sakano, T.K. (1986). Organic Solvents: Physical Properties and Methods of Purification. Techniques of Chemistry. 4th Edition. New York, NY, Wiley-Interscience	This reference details physical properties and preparation techniques of previously organic solvents

Available QSAR software for physico-chemicals properties estimation

The following paragraph is an extensive, but non-exhaustive, list of commercially or freely available computer programs that predict physico-chemical properties. These programs are based on Quantitative Structure Property Relationships (QSPRs)[8] and their major features are described below.

ABSOLV-2 is commercial software that calculates various solvation-associated properties from Abraham type equations and predicts Abraham's solvation parameters necessary for those calculations. The Abraham descriptors can be used to model aqueous solubility, octanol-water partition coefficient, vapour pressure and soil adsorption coefficient (K_{oc}) (http://www. ap-algorithms.com).

ACD PhysChem is a commercially available software developed by Advanced Chemistry Development (ACD) laboratories and it can estimate the following physico-chemical properties: boiling point, density, water solubility, octanol-water partition coefficient, acid dissociation constant, surface tension and flash point.

ADMET Predictor predicts all of the important properties critical to oral absorption as well as several pharmacokinetic properties and many aspects of toxicity. ADMET Predictor incorporates

[8] When a mathematical relationship relates chemical structure to a specific physico-chemical property it is referred to as Quantitative Structure Property Relationship: QSPR. The more common expression Quantitative Structure Activity Relationship (QSAR) is generally used for models that predict biological/toxicological effects.

artificial neural network ensemble models. ADMET Predictor predicts the most important properties related to oral absorption and particularly pKa, intrinsic solubility in pure water, n-octanol-water partition coefficient, molal volume (http://www.simulationsplus. com/).

ADMENSA is a commercial platform for computations at all stages of drug discovery. Within this platform it is possible to predict physico-chemical properties such as the n-octanol/water partition coefficient and aqueous solubility (http://www.inpharmatica.com).

CHEMOFFICE is a collection of commercial softwares consisting of different modules. The module ChemProp estimates physical properties of a selected structure, including partition coefficients, boiling and melting points (http://www.cambridgesoft.com/).

ChemProp, developed by the UFZ Centre for Environmental Research, Leipzig, Germany, predicts a number of physico-chemical properties. It has been described by Schüürmann et al (1997). At present its performance and availability are not known (www.ufz.de/index.php?en=6738).

CHEMSILICO is an on line predictor of intrinsic aqueous solubility and n-octanol/water partition coefficient based on topological (Kier and Hall, 1986; Hall and Kier, 1991) and electrotopological (E-state) values (Kier & Hall 1999). It uses an artificial neural network to calculate log S values. It cannot be used in batch mode, so is tedious to use for large numbers of chemicals (http://www.logp.com/).

CLOGP is a commercial software that employs a substructure approach where the final log K_{ow} is determined by summing the single-atom or fragment contributions (www.daylight.com).

EPI Suite. The EPI (Estimation Programs Interface) Suite is a collection of physical/chemical property and environmental fate estimation models developed by the EPA's Office of Pollution Prevention Toxics and Syracuse Research Corporation (SRC). EPI Suite uses a single input to run the following estimation models for physico-chemical parameters: KOWWIN, MPBPWIN, PCKOCWIN, WSKOWWIN and WATERNT. EPI Suite was previously called EPIWIN. According to the American EPA, EPI Suite is a screening level tool and should not be used if representative values are available (http://www.syrres.com/esc/epi.htm). It can be downloaded free of charge from: www.epa.gov/oppt/exposure/pubs/episuitedl.htm

- KOWWIN: Estimates the log octanol/water partition coefficient, log K_{ow}, of chemicals using an atom/fragment contribution method.
- MPBPWIN: Melting point, boiling point, and vapour pressure of organic chemicals are estimated using a combination of techniques.
- PCKOCWIN: The ability of a chemical to sorb to soil and sediment, its soil adsorption coefficient (K_{oc}), is estimated by this program. EPI's K_{oc} estimations are based on the Sabljic molecular connectivity method with improved correction factors.
- WSKOWWIN: Estimates an octanol/water partition coefficient using the algorithms in the KOWWIN program and estimates a chemical's water solubility from this value. This method uses correction factors to modify the water solubility estimate based on regression against log K_{ow}.
- WATERNT is a fragment-based method for the estimation of water solubility.

MOLECULAR MODELLING PRO is a commercial software for molecular modelling by ChemSW that can estimate physico-chemical properties. The methods used are given on the

chemSW web page (http://www.chemsw. com/13052a.htm) and details on their accuracy are given in the user's guide.

PALLAS. This software by Compudrug calculates log K_{ow} values and acidic and basic pK_a values (negative logarithms of acid-base ionisation constants) for organic compounds, in most cases, within an error of 0.25 pK_a units. The calculation can be performed for any organic compound, including aromatics, mono and polyheteroaromatics, and small peptides. The applied logarithm, adapted after Hammett and Taft (Perrin, *et al*. 1981) takes into account all necessary electronic, steric and other effects and relies on an extended database of almost a thousand equations (www.compudrug.com).

Pipeline Pilot is interactive software developed to link with many existing software programs. It is able to calculate octanol/water partition coefficient, aqueous solubility and pKa (www.scitegic.com).

PHARMA ALGORITHMS. The commercial module ADME boxes by Pharma Algorithms is a computational tool for the drug discovery field. It can predict for aqueous solubility, n-octanol/water partition coefficient and pKa (www.ap-algorithms.com).

Pkalc. This software by Compudrug calculates the accurate acidic and basic pK_a values (negative logarithms of acid-base ionisation constants) for organic compounds, in most cases, within an error of 0.25 pK_a units. The calculation can be performed for any organic compound, including aromatics, mono and polyheteroaromatics, and small peptides. The applied logarithm, adapted after Hammett and Taft (Perrin, *et al*. 1981) takes into account all necessary electronic, steric and other effects and relies on an extended database of almost a thousand equations (www.compudrug.com).

PREDICT is a software implemented and commercialised by Dragon Technology, Inc and it computes the following endpoints: boiling point, density, vapour pressure and surface tension (http://www.mwsoftware.com/).

ProPred. The CAPEC (Computer Aided Process-Product Engineering Center) Property Estimation Package has one module (ProPred) that can predict physico-chemical parameters. Property estimation in ProPred is based on a multilevel group-contribution. This multi-level scheme of estimation can provide estimations for several endpoints and particularly for: boiling point, melting point, molar volume, aqueous solubility, octanol/water partition coefficient, acid dissociation constant, surface tension and flash point. CAPEC software is available to CAPEC member companies only (www.capec.kt. dtu.dk).

QikProp is available from Schrödinger Inc. It calculates octanol/water partition coefficient, aqueous solubility and acid dissociation constant (www.schrodinger.com).

SPARC (SPARC Performs Automatic Reasoning in Chemistry) was developed by Karickhoff (Karickhoff *et al*., 1991). It estimates the following physico-chemical properties by means of linear free energy relationships and molecular orbital properties: boiling point, density, vapour pressure, water solubility, octanol/water partition coefficient, Henry's law constant, acid dissociation constant. SPARC is supported by the University of Georgia (http://ibmlc2.chem.uga.edu/sparc/).

VCCLAB is a free on-line predictor that can operate in batch mode. It calculates octanol/water partition coefficient, and aqueous solubility (www.vcclab.org).

(Q)SPR software summary

The physico-chemical properties estimated by the aforementioned software packages are summarised in Table R.7.1-3 and the reliability of the softwares will be discussed in the individual

sections detailing the physico-chemical properties. Complementary information on the performance of some of the mentioned programs can be found in the ECETOC technical report No. 89 (ECETOC, 2003; see Table R.7.1-3).

Many software programs use SMILES (simplified molecular input line entry system) for input of chemical structures. A SMILES tutorial is available (www.daylight.com/smiles/smiles-intro.html).

Table R.7.1-3 Physico-chemical properties estimated by commercially or freely available software

QSPR Software / Endpoint	ABSOLV-2	ACD	ADMET Predictor	ADMENSA	CHEM OFFICE	ChemProp	CHEM SILICO	CLOGP	EPI Suite	Molecular Modeling Pro	Pharma algorithms (ADME boxes)	PALLAS	Pipeline Pilot	PREDICT	ProPred	QikProp	SPARC	VCCLAB
Melting point					X	X			X					X	X			
Boiling point		X			X	X			X	X				X	X		X	
Density		X								X				X	X		X	
Vapour pressure						X			X	X				X	X		X	
Surface tension		X												X	X			
Water solubility		X	X	X		X	X		X	X	X		X		X	X	X	X
log Kow		X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X
Acid dissociation constant		X	X								X	X	X		X	X	X	X
Viscosity						X								X	X			
Organic carbon-water partition coefficient (Koc)	X										X							

(see also the ECETOC Technical report No. 89)

R.7.1.1.3 Evaluation of available information on physico-chemical properties

Experimental data

Test data generated using an appropriate standardised method and to GLP will be acceptable. Tests that have not been done to GLP will also be accepted provided that they have been done using an appropriate test method and there is sufficient documentation about quality procedures (i.e. compliance with ISO 17025). ISO 17025 is the International Quality Assurance system for testing and calibration laboratories. It specifies the general requirements for the competence to carry out tests and/or calibration, including sampling. It covers testing and calibration performed using standard methods, non-standard methods and laboratory-developed methods. Due to the wide range of modifications and variations that are possible, it is difficult to make a generalisation about the

acceptability of data generated using non-standard test methods. In these cases, expert judgement needs to be applied to determine if the method used has produced a valid result.

In general, impurities in the product can have a significant influence on some of the specific physico-chemical endpoints, especially where the impurities and the main component have widely differing values (e.g. volatile components for vapour pressure, highly water soluble components for water solubility). Comparison of the experimentally determined physico-chemical property with a QSAR prediction is often, if not always, recommended to provide reassurance that the experimentally derived value is acceptable and has not been influenced by the presence of impurities in the product.

Non-experimental data

QSPR models exist for some of the physico-chemical endpoints (a non-exhaustive list is given in Section R.7.1.1.2). Details of any specific QSPR models are given under each endpoint.

The majority of QSPR models have been built using *training sets* of substances. The model will have been optimised to calculate values for the *training substances* that most closely match measured ones. The degree of closeness, i.e. how close to the measured value the calculated one is, determines how accurate the model is. Together these factors define the *domain* of the model. Thus, the predictive power of the model will be greatest for substances that are similar to the training set, i.e. they are within its domain. Details of the number of substances in the training set, the range of substance types, the required accuracy, etc are not always available. Therefore, the use of QSPR estimation techniques requires some expert judgement. The calculated values need to be checked to ensure that they are reasonable and that the model used is appropriate.

A simple way to check that a model is appropriate is to check its predictive capability for a set of analogue substances that are similar to your substance and for which measured values exist. A valid model will give values that are in reasonably close agreement with the measured ones for your chosen analogue substances. Thus, the model can be used to provide a predicted value for your substance without the need for testing.

Another check is that the values are reasonable. This can be done by cross-referencing the calculated value to measured values for related endpoints, e.g. a calculated boiling point of >300°C would correlate with a low measured vapour pressure.

Uncertainty evaluation for physico-chemical properties

The ISO norms (ISO, 1995) define uncertainty as a parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the particular quantity subject to measurement. Uncertainty for physico-chemical properties should therefore be assessed in order to demonstrate the quality of the measurement and to document in a transparent way the adopted methodology. Moreover, it allows comparison of results and it is required by ISO 17025 (ISO/IEC, 2005).

The data used as input for the determination of the physico-chemical property should be properly documented by including validation data, manufacturer's specifications, uncertainty assigned to reference data and all relevant experience with the method under scrutiny.

The quantity to be determined should be clearly defined together with the model equation which enables the quantitation of the investigated property. All the possible sources of uncertainty should be discussed and the uncertainties of each input should be evaluated in order to calculate the combined and expanded uncertainty according to the Guide to the expression of Uncertainty in Measurement (GUM) (ISO, 1995).

Instrument response, bias of instrument, variations in repeated observation, instrument resolution, reagent purity, experimental conditions, and uncertainty of standards are among the most common possible sources of uncertainty and they should be properly discussed when assessing and reporting the quality of a measurement.

Quality assurance for the determination of physico-chemical properties

Special care should be given to the quality assessment of data on physico-chemical properties from experiments not carried out according to GLP or the test methods referred to in Article 13 (2). Greater weight should be given to data that meet Klimisch (Klimisch, 1997; see also Section R.4.2) criteria 1 or 2: reliable without restriction and reliable with restriction. Criteria 3 and 4 (not reliable, not assignable) should not be rejected but re-testing could be taken into consideration.

Any evaluation of a chemical should be based on physico-chemical property test data of sufficient quality, rigour and reproducibility. It is critical that the data are scientifically acceptable. It is also important that the studies have been properly managed and are well documented, preferably to GLP standard, or to some other appropriate quality regime or standard that will provide confidence in the management of the study and the acceptability of the data. Indeed, the registrant should establish and maintain procedures for identification, collection, indexing, access, filing, storage, maintenance and disposal of quality and technical records (ISO 17025). A system including technical competence (ISO 17025) is therefore recommended and the adoption of GLP standards would promote an international acceptance of the data.

Use of secondary and historical data sources for physico-chemical properties

The reliability of data must be demonstrated by providing information on the identity and purity of the test substance, the methodology used to make the measurement, and whether or not this was performed to GLP standards. Therefore, the best approach would be to obtain primary references but, in some cases, it may be appropriate to use reliable, authoritative secondary sources of data (OECD[b], 2004).

These secondary sources have to be based on a critical evaluation of peer-reviewed data and a consequent selection of a reliable and representative value for the property under investigation. Information in the review process should be stated in the introduction of the handbook or in the summary information for an electronic database available on Internet. The use of Klimisch codes (Klimisch *et al*., 1997), can be extended to these secondary sources and a reliability code of (2) *valid with restrictions* should be assigned when using an authoritative secondary source. Citations of the original data sources should always be preferred even when using on-line databases. The original data source should be consulted whenever possible.

If the analysis of bibliographic references is only limited to secondary data sources it is essential to create a WoE approach. By means of such an approach, several studies, none of which would in themselves be acceptable because of some deficiency (i.e. missing original reference), could collectively add evidence and support the choice of a specific value. For instance, the physico-chemical data available for vinyl acetate were mostly old and unsupported by test reports (ECETOC, 1998). The use of these data was justified by comparison with predicted values generated by QSPR models (ECETOC, 1998).

Values for physico-chemical properties taken from secondary sources can be supported by manufacturing data and reliable QSPR prediction (Figure R.7.1-2). When using a QSPR model it should be demonstrated that the model is appropriate for the type of chemical under investigation (i.e. correct domain of applicability) and that the model performance has been checked (i.e.

goodness-of-fit, robustness, and predictivity). If the QSPR prediction is considerably different to measured values this difference should be discussed and re-testing should be considered.

The same principles that have been exposed for secondary data sources can be applied to historical data whose quality cannot be unequivocally evaluated because the original test reports are not available or are incomplete. These old data can still be used in a WoE approach (Weed, 2005) even if none of them on its own is acceptable and QSPR or read-across estimations can be used to check their validity and guide the choice of a specific value. If the difference between historical values or between historical values and predictions should prove to be critical (i.e. values close to a regulatory cut off) further investigation must be carried out in order to understand the discrepancies.

Assessing the quality of QSPR models

The European Commission and the OECD member countries adopted in 2004 five principles for the validation of (Q)SAR/(Q)SPR models. According to these principle, a valid (Q)SPR model should have 1) a defined endpoint whose experimental conditions are clearly specified; 2) an unambiguous algorithm; 3) a defined domain of applicability that defines for what kind of chemicals predictions can be made; 4) appropriate measures of goodness-of-fit, robustness and predictivity; and 5) a mechanistic interpretation if possible. These principles are outlined on the ECB website (OECD[a], 2004) and more extensively covered Section R.6.1.3. Moreover, a practical overview of these principles is given in the report from the expert group on (Q)SARs (OECD[c], 2004).

The validity of the OECD principles can only be independently established by the user if the model under investigation is transparent.

A QSPR model can be defined as being transparent if: (1) the chemical structure of the chemicals forming the training and test sets is known: (2) the mathematical or empirical derivation of the descriptors is given (3) the adopted algorithm is clearly explained. Robust summaries of (Q)SPR models and their concordance with the OECD principles should be compiled according to the framework provided by the (Q)SPR Model Reporting Format (QMRF) and the explanation on how an estimate has been derived should be reported within the (Q)SPR Prediction Reporting Format (see Section R.6.1.10).

If such premises are not met (e.g. commercial software protected by copyright) it is possible to benchmark the predictivity of the model only on compounds that are similar to the chemical under investigation and whose physico-chemical properties are experimentally known. The determination of the prediction accuracy is a critical step that can guide the user to apply the most reliable software packages for the analysis of their data as shown in the work published by Tetko *et al*. (2006).

Conflicting predictions from multiple QSPR models:

Consensus modelling

Some QSPR predictions and software programs are better than others. However, even the good ones do not yield perfect predictions. It is therefore always best, provided that it is practicable, to obtain property predictions from at least three different methods. In that way one can see whether one prediction is very different from the others, and should perhaps be discarded. This is exemplified by four separate software predictions of the aqueous solubility (log S, with S in mole/l) of three different chemicals (Table R.7.1-4). All the software programs have been tested (Dearden 2006) and found to give good predictions overall. It should be borne in mind that the average experimental error on aqueous solubility measurements is about +/-0.6 log unit (Katritzky et al., 1998).

Table R.7.1-4 Software predictions of the aqueous solubility of Atropine, Caffeine and Butylparaben.

	Atropine	Caffeine	Butylparaben
Measured	- 2.18	- 1.02	- 2.96
Software no. 1	- 1.87	- 1.87	- 3.09
Software no. 2	- 2.06	- 0.65	- 3.05
Software no. 3	- 1.01	- 0.27	- 3.07
Software no. 4	- 2.03	- 0.56	- 2.58

Predictions were obtained from four different softwares. The measured aqueous solubility is reported as well.

For atropine, it is clear that three programs give similar predictions, well within the experimental error, whereas software no. 3 gives a poor prediction. The mean of the three good predictions is -1.99, which is only 0.2 log unit different from the experimental value of -2.18. However, it may be noted that even if the poor prediction from software no. 3 is included, the mean predicted value is -1.74, which is still within the experimental error of +/-0.6 log unit from the measured value.

For caffeine, there is a considerable divergence of predicted values, indicating that the solubility of this compound is difficult to predict. Only two of the four predictions are within the experimental error of +/-0.6 log unit, but the mean of all four predictions is -0.84, which is well within the experimental error. This example really emphasises the value of consensus modelling.

Butylparaben has a simpler chemical structure than those of atropine and caffeine, and this is reflected in the more accurate predictions of aqueous solubility, with all four being within the experimental error of +/- 0.6 log unit, and the mean of all four being -2.94.

It is recommended that, wherever possible, predictions be obtained from more than one software program and/or QSPR, and that the mean of all the predictions be used, unless one of the predicted values is clearly very different from the others, in which case that prediction should be rejected.

Predictions using artificial neural networks (ANNs)

An artificial neural network (ANN) is software, modelled on the brain's neuronal network, that enables the selection of linear and non-linear (e.g. squares, reciprocals) forms of descriptor values, in contrast to multiple linear regression (MLR), which uses only linear correlations. Thus better correlations can often be obtained with ANN. The drawbacks of ANN are: (1) the network has to be carefully trained, so that it gives a good model and also gives good predictions. Over-training will result in an apparently better model, but one with poor predictivity; (2) the model is not transparent. That is, ANN does not yield a QSAR or QSPR, from which one can see what descriptors contribute, and to what extent, to the model (i.e. it is a *black box* approach).

When using an ANN there is a risk of over-training it, so it has to be used carefully. On the other hand, commercial software that uses the ANN approach has already been trained, so a non-expert user can use it quite safely.

Most, if not all, of the OECD criteria may be met when modelling using this neural network approach as shown by Vracko *et al*. (2006) who published a detailed case study describing the validation of counter propagation neural network models for predictive toxicology according to the OECD principles.

Potential pitfalls in QSPR modelling

The compound of interest should be within the applicability domain of the QSPR/software program. This is generally not easy to determine. Most, but not all, software developers make their training sets available, but even then it is not always obvious whether the compound of interest is within the applicability domain, because software developers do not provide tables of descriptors from which one could check applicability. However, the read-across approach can be used here. That is, one can use the software or QSPR to make predictions for similar compounds whose property values are known. If those predictions are acceptable, then it is reasonable to assume that the prediction for the compound of interest will also be acceptable. Almost invariably, QSPRs and property prediction software are trained on organic compounds, and cannot handle inorganic compounds or metallo-organics (an exception to this is the SPARC software).

The user must be clear as to which endpoint is being predicted. This is particularly important when a software program is able to predict a number of similar endpoints. For example, several commercially available software programs for the prediction of aqueous solubility offer several endpoints, such as solubility in pure water and intrinsic solubility (i.e. the solubility of the undissociated species).

It is essential to check that the units of the property being predicted are known and understood. For example, a predicted log (solubility) value will probably have solubility in moles per litre, whereas the user might think that it is in milligrams per 100 ml.

Walker and de Wolf (2003) have warned against using a predicted property to predict another property. However, sometimes this is unavoidable, e.g. in the case where a compound has not been synthesised. In such cases one must accept that the accuracy of prediction will probably be lower than would otherwise be the case.

When using a QSPR, it is essential to check that it has been validated, preferably by use of external validation, or, failing that, by cross-validation; this is because it is possible to develop a QSPR that models the training set data well, but does not give good predictions. In the case of the former, the prediction errors of the test set should be similar to those of the training set. In the case of the latter, the cross-validated r^2 (q^2) value should not be <0.5 (Eriksson *et al* 2003), and should not be more than 0.3 lower than the r^2 value (Walker *et al* 2003).

The calculation of descriptors for use in a QSPR should always be done using the same software as that used by the workers who developed the QSPR. The reason for this is that different software programs can yield different numerical values for a given descriptor; this is especially so for quantum chemical descriptors.

Major sources of misinterpretation of QSPR endpoints

Selection of wrong endpoint: e.g. intrinsic solubility instead of solubility in pure water.

Use of incorrect units for a property: e.g. g/100 ml instead of mole/litre; use of natural logarithm instead of logarithm to base 10.

Use of a QSPR or software program to make predictions outside its applicability domain.

Placing too much reliance on a single prediction.

Assessing the quality of read-across predictions

This paragraph reports the basic concepts of a read-across approach. Thorough information on this topic can be found in the guidance on the grouping of chemicals (see Section R.6.2).

A read-across/analogue approach assesses the relevance of a given property on one chemical structure and then makes some assessment (qualitative or quantitative) on the relevance of this information for another chemical. Since a read-across can involve only two chemicals[9] it is of paramount importance to detail the reasoning behind the inference on the chemical whose property is unknown. An analogue must:

contain the same major structural features and the same functional groups of the chemical under investigation

have a physico-chemical profile comparable to that of the chemical under examination as far as the known physico-chemical properties are concerned

have comparable values for the relevant molecular descriptors (i.e. excess molar refractivity and hydrogen bond donor and acceptor abilities for water solubility predictions) generally used for the quantification of the property of interest

have approximately the same molecular weight

The interpretative analysis of a read-across is usually the result of an expert judgement evaluation and detailed documentation should therefore always be provided to support the conclusions.

It is important to point out that,, in practice, read-across for physico-chemical properties is not generally recommended, since reliable data should normally be available or easily obtainable.

Remaining uncertainty

Physico-chemical properties have an essential and central role in risk assessment for chemical hazards. They are also used to predict environmental fate and transport and to assess human health and safety issues. Indeed, they give information about the uptake/absorption of chemicals. Absorption of a chemical into the blood stream will depend on its aqueous solubility, its log K_{ow} and its ionisability (pKa). In addition, log K_{ow} is a very important determinant of the partitioning of a solute within a biological membrane (i.e. an increase of log K_{ow} results in a higher membrane permeability and in a reduced aqueous solubility).

Because of their basic function for toxicological studies, physico-chemical properties need to be *fit for purpose* and the need for precision and accuracy becomes extremely important when data are close to a regulatory cut off (e.g. the K_{ow} value used in the PBT screening). On the other hand, if the property under investigation is in a range which is safe the issue of precision and accuracy is less critical. In the ECETOC report 74, there is a comparison between historical data and estimated values for vinyl acetate. In the specific case of vinyl acetate there is a difference between experimental data and estimated values for water solubility. The reasons for this difference do not necessitate further investigation because the effect of the change of solubility at this level is slight on the environmental behaviour and the risk assessment for the substance.

For physico-chemical data, there are often multiple values for the same endpoint. This introduces a degree of uncertainty as to which value to choose. To assist in choosing a valid value the Klimisch code system can be used (Klimisch *et al*. 1997). Test results that have a code of 1 or 2 are usually acceptable, less reliable studies could be used in a *Weight of Evidence* approach. Note that

[9] A read-across can also involve more than two chemicals: .one-to-one (one analogue used to make an estimation for a single chemical) b) many-to-one (two or more analogues used to make an estimation for a single chemical c) one-to-many (one analogue used to make estimations for two or more chemicals) d) many-to-many (two or more analogues used to make estimations for two or more chemicals)

uncertainty should not be confused with *accuracy*. Accuracy is a measure of how good a particular method is of measuring the correct value, e.g. a method may be able to measure a boiling point to within ±5°C.

As far as QSAR/QSPR models are concerned, every model has a degree of uncertainty (Walker *et al*., 2003) for two different reasons: the unavoidable variability of the input data (intrinsic to nature), missing information about actual values, uncertainty of descriptors and experimental endpoints. Moreover, each statistical method varies in its ability to describe reality because of a limited systemic knowledge and the choice of descriptors cannot model all the classes of compounds with the same efficacy.

Confidence intervals for the estimated parameters can be assessed by resampling methods such as Jack-knifing and Bootstrapping (Efron and Gong, 1983, Eriksson *et al*., 2003).

In the first method, the same test is repeated by leaving one subject out each time. The resampling strategy of Bootstrap is more thorough in terms of the magnitude of replication. In Jackknife, the number of resamples is confined by the number of observations but it can be convenient because cross-validation produces results that can be fed to this method.

On the other hand, in bootstrap, the original sample can be duplicated as many times as the computing resources allow, and then this expanded sample is treated as a virtual population. Then samples are drawn (with replacement) from this population to verify the estimators. The basic premise of the method is that the data set is representative of the population from which it was drawn. This statistical methodology is therefore a simulation of what would happen if the population were resampled by randomly resampling the data set.

R.7.1.1.4 General testing strategy for physico-chemical properties

When testing is to be carried out there are advantages to considering the order in which testing should be done. For example:

- Some tests can be omitted when the results from another are available.
- There are safety implications to performing tests on substances that have certain hazardous properties.

  In some tests, better quality results can be obtained through prior knowledge of related properties

Ideally, when a full set of physico-chemical tests is to be performed they should be done according to the following plan:

Tier 1: These tests should be performed before all others (or the results predicted)

Tier 2: These tests should be performed after the Tier 1 is complete. Tier 2 tests are placed in groups. Ideally, the tests within each group should be performed in the order shown. The order in which each group is conducted is not important. However, knowledge of the melting point helps in selecting appropriate flammability and auto-flammability test methods.

Tier 3: These tests can be performed at any point after Tier 1 and in any order.

For selected endpoints, a short explanation of its position in the testing order is given. For some endpoints, knowledge of the properties from another test group can be helpful but is not critical. For example, auto-flammability does not need to be conducted for solids that melt at <160°C.

Tier 1

Pyrophoric Properties: A positive result removes the need for further flammability tests and makes testing for other endpoints extremely difficult.

Tier 2

Group 1

1) Flammability (contact with water): A positive result removes the need for further water-based physico-chemical testing on the substance. Consideration should be given to providing toxicological or ecotoxicological data on the reaction products.

2) Water Solubility.

3) Surface Tension: This test requires knowledge of the water solubility and the time taken to generate a saturated solution.

4) Dissociation Constant: This property is only relevant to ionisable substances. The partition coefficient should, where possible, be measured for the substance at a pH at which it is in a non-ionised state. Therefore, knowledge of the acid dissociation constant (pKa) is required before measuring the partition coefficient. *

5) Partition Coefficient: This test should not be performed on surface-active substances as they interfere with partitioning. Thus, knowledge of surface tension is required prior to testing.

6) Adsorption/desorption (K_{oc}): It is useful to have information on water solubility, dissociation constant, surfactant properties (surface tension) and partition coefficient before conducting an adsorption/desorption test. It is often possible to calculate a value, which will usually be sufficient at 10 t/y. Experimental studies may need to be considered at higher tonnage depending on the needs of the CSA.

* Dissociation constant is not a testing requirement until a substance reaches the 100 t/y supply level. Therefore, for lower volume substances, it will not be possible to follow this idealised testing strategy. In these cases, it is acceptable to move on to test 5 in the series.

Group 2

Explosivity: A positive result would remove the need for any further tests in this group.

Flammability: Liquids with very high flash points (>200°C) do not need to be tested for auto-flammability.

Auto-flammability: The physical form of the test substance must be known in order to select an appropriate method. For solids, the test method does not work well for substances with a melting point of <160°C and testing is not required.
Substances that fall into any of the flammable classifications should not be tested for oxidising potential.

Oxidising Properties: This test should not be performed on flammable or explosive substances.

Group 3

Melting Point: If the melting point is

- >300°C, testing for vapour pressure is not required.

- between 200°C and 300°C a limit value for the vapour pressure or a calculated value is sufficient.

If there is decomposition on melting, a boiling point is not required. Differential Scanning Calorimetry (DSC) can be used to evaluate the melting and boiling point in a single test.

Boiling Point: This is closely related to vapour pressure and for some substances, a single test can provide both endpoints.

Vapour Pressure: Testing should be conducted across a temperature range in which:

- there is no phase transition (e.g. melting).
- the substance is stable (i.e. at least 20°C below any decomposition temperature).
- the substance is in the same physical state as under standard temperature and pressure.

Tier 3

Relative Density

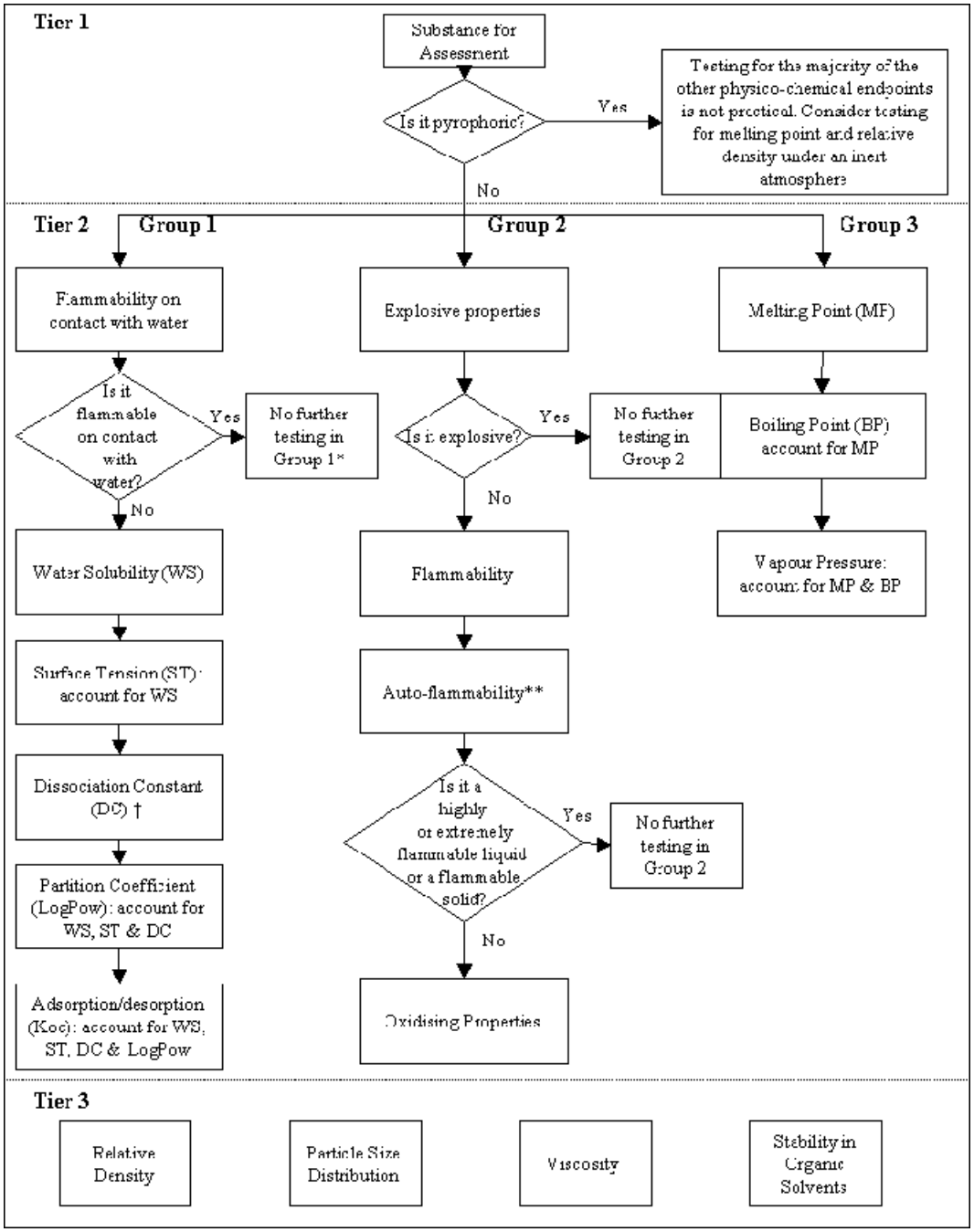
Particle Size Distribution

Viscosity

Stability in Organic Solvents

As far as the quality is concerned, new studies for substances that have been conducted under GLP standards will enable the registrant to prove that the study uses a Study Director, study plans, Quality Assurance Programme and well-defined archiving policies and procedures. Methods and practices conform to GLP standards will therefore promote the transparency and credibility of the submitted data by ensuring their quality and integrity.

Figure R.7.1-1 Tiered testing scheme on physico-chemical testing



* Consideration should be given to providing toxicological or ecotoxicological data on the reaction products.

** Testing is not required for solids with a melting point of <160°C, for liquids with a flash point of >200°C and for gases without a flammable range in air.

† Dissociation constant is not a testing requirement until a substance reaches the 100 tonne/annum supply level. Therefore, for lower volume substances, it will not be possible to follow this idealised testing strategy. In these cases, it is acceptable to move on to the next test in the series

R.7.1.1.5 General assessment of the available information on physico-chemical properties

The flow chart in Figure R.7.1-2 exemplifies the general assessment of all the possible available information (test results, literature, QSPRs and read-across) with respect to the determination of physico-chemical properties. The major criteria that characterise the analysis of the available information are:

- *Experimental data*. When assessing physico-chemical properties, priority is given to first hand experimental results or to primary references provided that the methods are suitable for the substance under investigation and that they operate within their validity range. Proper documentation on the methods and the inherent uncertainty of the measurements should also be provided

- *Non-testing information*. If the information described in point (a) is not available QSPRs, read-across or secondary data sources can be used instead.

- *QSPR transparency*. Predictions from transparent QSPR models can be accepted if they are supported by adequate and reliable documentation. Robust summaries of (Q)SPR models and their concordance with the OECD principles should be compiled according to the framework provided by the (Q)SAR Model Reporting Format (QMRF). The explanation on how an estimate has been derived should be reported within the (Q)SPR Prediction Reporting Format (see paragraph 5 of the cross-cutting guidance on (Q)SARs for further information). QSPR model can be regarded as transparent if it meets the criteria outlined in Section 0) which allow for a detailed and clear understanding of the logic underpinning the model (i.e. algorithm, adopted descriptors and chemical structure of the substances forming training and test sets).

- If the algorithm and the training set of the QSPR model are proprietary a result from such a model can be accepted provided that the software yields reliable predictions for chemical compounds whose structure is similar (see criteria for read-across in Section 0) to the chemical under investigation. The chemicals selected for this benchmark must also have an experimentally determined value for the physico-chemical property of interest.

- *Prediction errors*. If a result from a QSPR prediction is close to a regulatory cut off value, the prediction can obviate the need for experimental testing only if its inherent error is close to the experimental error. Clear explanation on how the prediction error has been derived should be given. For flash-points, calculated values can be adopted if the method used can be shown to be valid and the result is clearly outside of any classification range.

- *Secondary data sources/historical data*. The physico-chemical property of interest can be determined by using only secondary data sources/historical data, provided that they can collectively add and support the choice of a specific value. QSPR and read-across predictions can also be used in a WoE approach together with these sources of information and they can be very useful in re-establishing the validity of historical data.

Lastly, as a general practice, it is suggested to check values that have been experimentally measured or retrieved from literature by means of one or several QSPR predictions whenever such models are available. A significant discrepancy between QSPR prediction(s) and values determined by means of other methodologies should prompt further investigation in order to understand the reasons for such a difference.

R.7.1.1.6 Overall consistency of the physico-chemical profile

The physico-chemical profile for a given chemical cannot contain incompatible values for two or more properties (i.e. high boiling point and high vapour pressure at normal temperature) This consistency check should be always done and it can turn out to be particularly useful when measured values are significantly at odds with predictions from QSPR models. Indeed, in this case a wider assessment of the known physico-chemical properties should be performed in order to determine the possible cause of the inconsistencies.

For example, if the prediction for the vapour pressure (at normal temperatures) is much lower than the experimental counterpart then the experimental vapour pressure and boiling point should be checked. This is because for a given substance a high vapour pressure (i.e. the substance is volatile) is consistent only with a low boiling point. This is shown by the Clausius-Clapeyron equation, from which the following equation can be derived:

$\log VP = a/T_b + k$

where VP is the vapour pressure of a chemical at a given temperature, T_b is its normal boiling point (i.e. at atmospheric pressure), and a and k are constants. The equation shows that the logarithm of vapour pressure is inversely proportional to boiling point. If the boiling point of a chemical with a high experimental vapour pressure is high, then the experimental value for the vapour pressure could be anomalously high due to a volatile impurity.

Similarly, a high n-octanol/water partition coefficient is generally consistent only with a low aqueous solubility (S_{aq}) and a high K_{oc} (organic carbon/water partition coefficient), as shown by the following QSPR correlations:

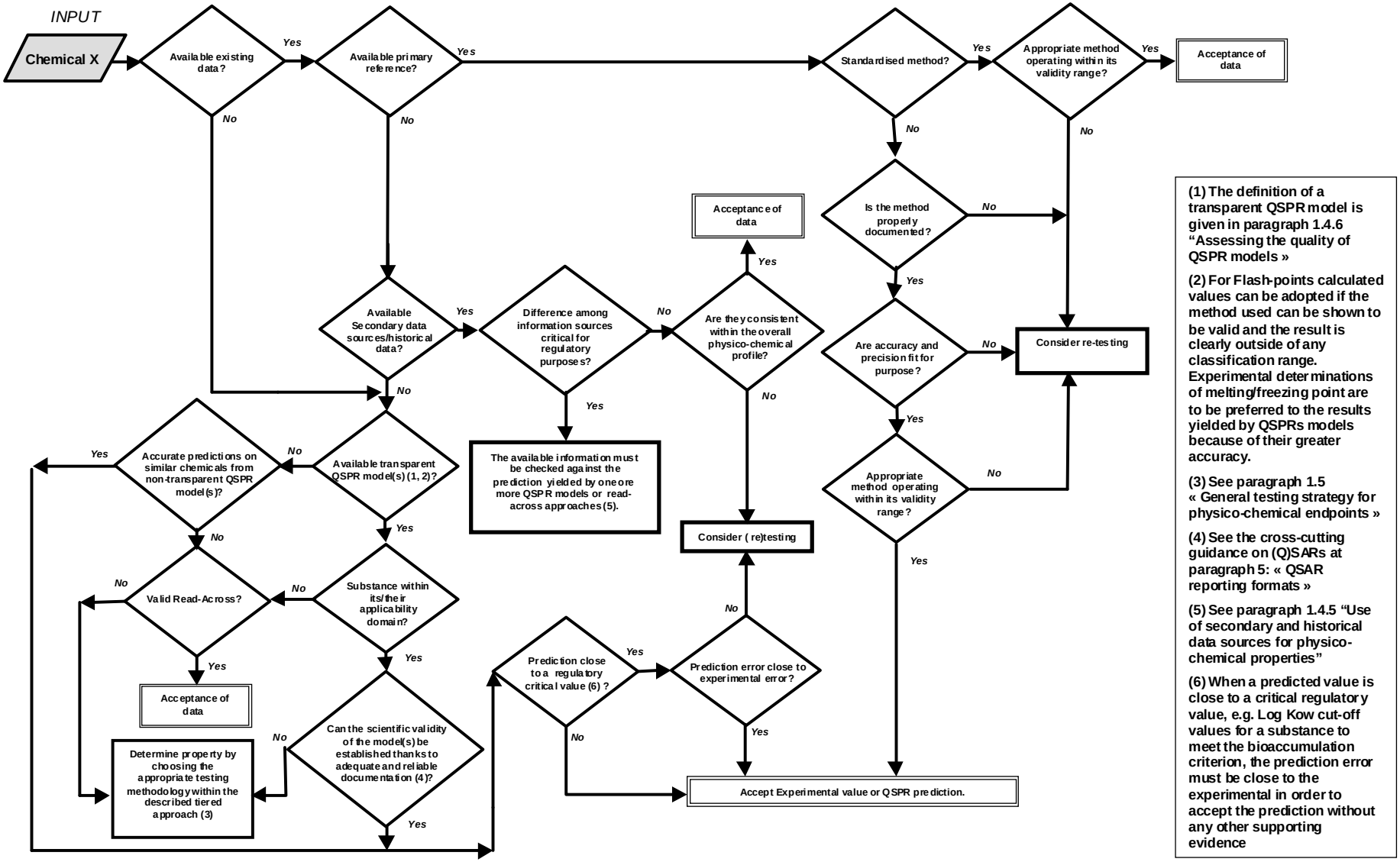
- Relationship between log S_{aq} and log K_{ow} for diverse liquids (Hansch et al., 1968):
  $\log S_{aq} = -1.214 \log K_{ow} + 0.850$ ($n = 140$ $r^2 = 0.912$ $s = 0.344$)

- Relationship between log S_{aq} and log K_{ow} for diverse solids (Ran et al., 2002):
  $\log S_{aq} = -1.022 \log K_{ow} - 0.0096 (MP - 25) + 0.381$ ($n = 1096$ $r^2 = 0.96$ MAE = 0.496)

  where MP = melting point (°C), and MAE = mean absolute error. The (MP – 25) term corrects for the entropy of fusion of solid solutes.

- Relationship between log K_{oc}- log K_{ow} for diverse chemicals (Baker et al., 1997)
  $\log K_{oc} = 0.903 \log K_{ow} + 0.094$ ($n = 72$ $r^2 = 91$ $s = 0.397$)

- Relationship between log K_{oc} and log S_{aq} for urea (Gerstl,1990):
  $\log K_{oc} = -0.381 \log S_{aq} + 1.177$ ($n = 57$ $r^2 = 0.616$ s not given)
  It should be noted that a chemical can have a relatively low K_{ow} and a relatively high water solubility, but still a high K_{oc} (e.g. if sorption is caused by ionic interactions).



(1) The definition of a transparent QSPR model is given in paragraph 1.4.6 "Assessing the quality of QSPR models »

(2) For Flash-points calculated values can be adopted if the method used can be shown to be valid and the result is clearly outside of any classification range. Experimental determinations of melting/freezing point are to be preferred to the results yielded by QSPRs models because of their greater accuracy.

(3) See paragraph 1.5 « General testing strategy for physico-chemical endpoints »

(4) See the cross-cutting guidance on (Q)SARs at paragraph 5: « QSAR reporting formats »

(5) See paragraph 1.4.5 "Use of secondary and historical data sources for physico-chemical properties"

(6) When a predicted value is close to a critical regulatory value, e.g. Log Kow cut-off values for a substance to meet the bioaccumulation criterion, the prediction error must be close to the experimental in order to accept the prediction without any other supporting evidence

Figure R.7.1-2 Global assessment of the available information for the determination of physico-chemical properties.

Conclusions on physico-chemical properties

It is very important that physico-chemical properties are determined in an accurate way because of their central role within the registration process. The "hazardous" physico-chemical properties (flash point, flammability, explosive and oxidising properties and self ignition temperature) are used mainly for the purposes of safe handling and hazard communication (classification and labelling). Results from other tests are used in designing the appropriate toxicological and ecotoxicological test packages. A summary of the overall significance of the tests is given in Table R.7.1-5.

The order in which the tests are carried out should be given consideration, as the results of each test can influence how the others are conducted. The global consistency of the physico-chemical profile for a given chemical should also be taken into account (see Section R.7.1.1.6).

The experimental tests should be carried out according to recognised test methods and preferably under a Quality Assurance regime (possibly, under conditions of Good Laboratory Practice, although this is not a requirement for REACH). Above all, it must be proven and documented that the chosen tests are (1) scientifically valid; (2) suitable for the substance in question and (3) operating within their validity range. This consideration is especially important if a non-standard method is used.

QSPR models can replace the need for testing if their scientific validity and the logic that underpins the derivation of estimates can be established thanks to adequate and reliable documentation ((Q)SAR Model Reporting Format and (Q)SAR Prediction Reporting Format). A thorough check of the compliance of a QSPR model with the OECD principles is not possible if a commercial QSPR model is not fully transparent. In this case, its prediction can be used provided that the registrant proves that the QSPR model can successfully predict properties for chemicals whose structure is similar to that of the investigated compound.

Read-across/analogue approaches can be used as well if their quality (see Section 0) can be well documented.

Secondary data sources can replace the need for testing if they are used in a *Weight of Evidence* approach and they can collectively support the choice of a specific value for the property of interest.

Concluding on C&L and chemical safety assessment

The results of the physico-chemicals tests can be used to determine the physical hazards that the substance possesses for example the flammability. This should be done by comparing the data against the criteria given in the classification and labelling rules according to Directive 67/548/EEC [10]. Physico-chemical properties relevant for C&L are given below.

Results of the hazardous physico-chemical tests are used directly for classification of substances for flammability, explosive properties or oxidising properties.

The results of the water solubility and K_{ow} tests, along with the biodegradation study, are important in assigning classification and labelling of substances for environmental hazards. K_{ow} is used as an indication of bioaccumulation potential in organisms. The environmental classification is dependent on these physico-chemical test results: therefore, it is vital to ensure that accurate determinations are

[10] The Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

made. As far as human health is concerned, it is worth noting that log K_{ow}, water solubility, vapour pressure and particle size distribution are essential parameters in order to assess the bioavailability of a substance after oral, inhalative or dermal exposure.

Physico-chemical data is also used in the environmental and human health sections of the chemical safety assessment. The assessment determines the risk posed to humans and the environment from all stages of its lifecycle. This includes manufacture, transport, use and disposal of the substance. Toxicology results are used to judge other hazards to human health, but the vapour pressure and particle size determination are required to estimate the likely exposure that will occur to humans both in the workplace and in consumer use. Viscosity is a key parameter in determining aspiration hazards. Exposure estimates calculated using a model use volatility (vapour pressure) or the size and nature of particles to estimate inhalation exposure. Particle size is also important for determining the likely dermal exposure. The physical state of a substance at the process temperature is also an important consideration for determining likely hazards and therefore melting and boiling temperatures are needed.

Table R.7.1-5 Summary of use of physico-chemical properties

Test	Impact on other Physico-chemical tests	Impact on toxicology	Impact on Ecotoxicology	Impact on Risk assessment
Melting/freezing point	Choice of method for flash point/flammability, auto-flammability, oxidising properties, explosive properties (sensitivity to friction). If decomposition occurs during the melting point study, a boiling point need not be measured.			It indicates (together with the boiling point) the physical state of a compound.
Boiling point	Related to vapour pressure. Affects classification as "highly flammable" or "extremely flammable".			Consider process temperature for risk assessment.
Relative density			To decide if an immiscible compound floats in water or sinks to the bottom.	Fire-fighting measures: H_2O extinguishers may not be suitable if $D^{20}_4 < 1$
Vapour pressure	Extra care needed to minimise vapour losses. Related to boiling point.	Choice of dermal or inhalation exposure route for acute toxicity test. Route of exposure for sub-acute toxicity test. Exposure and excretion routes for toxicokinetic assessment. It enables a statement about inhalative absorption in the living organism.	Choice of test method for biodegradation test. Closed/covered vessels for ecotoxicity tests	Health risk assessment. Calculation of Predicted Environmental Concentrations (PECs) for environmental risk assessment- vapour pressure is a key parameter in determining environmental fate and behaviour. Determination of atmospheric behaviour as for exposure of man via the environment calculations.
Surface tension	Suitability of methods for K_{ow} and K_{oc} determination for surface-active substances. Can occasionally interfere with measurement of water solubility.	Cellular disruption. Surface active substances have a higher local irritant or corrosive effect. As a consequence of local corrosion the dermal uptake of a substance can be enhanced		Environmental fate.

Test	Impact on other Physico-chemical tests	Impact on toxicology	Impact on Ecotoxicology	Impact on Risk assessment
Water solubility	Surface tension test not applicable for WS < 1 mg/l. Need to prepare 90% saturated solution (up to a maximum of 1 g/l) for surface tension test. Time to achieve saturation can be relevant to solution preparation for surface tension test. Water solubility affects concentration used in hydrolysis test	Toxicokinetic behaviour.	Sample preparation for ecotoxicity tests.	Environmental classification and labelling. PEC calculations – water solubility is a key parameter in determining environmental fate and behaviour.
Partition coefficient n-octanol/water	Generally, substances with a high log K_{ow} will be hydrophobic and have low water solubilities. Substances with negative log K_{ow} will be hydrophilic and have high water solubilities.	Toxicokinetic behaviour: It indicates the potential for absorption across biological membranes and for passive diffusion. It provides information on the potential for accumulation in the body. Suitable vehicle for toxicity studies. Prediction of dermal absorption.	Choice of test method for biodegradation test (some are not suitable for substances which have high adsorption) High log K_{ow} may lead to losses in ecotoxicity tests through adsorption. Bioaccumulation and adsorption potential. Toxicity prediction.	Environmental classification and labelling. PEC calculations – log K_{ow} is a key parameter in determining environmental fate and behaviour and is used as a surrogate for bioaccumulation potential in the absence of bioaccumulation tests.
Granulometry		Computation of inhalable, thoracic and respirable fractions as a function of size of particles.		
Adsorption/ desorption			As for log K_{ow}	log Koc is a key parameter in determining environmental fate and behaviour.
Dissociation constant	Interpretation of results for surface tension, K_{ow} and K_{oc} tests and water solubility.	Exposure to hydrolysis products *in vivo*. It indicates the potential for absorption from the gastrointestinal tract, because ionised compounds are thought not to cross biological membranes.	Preparation of test solutions for ecotoxicity tests. Interpretation of ecotoxicity results.	Influences PEC calculation.

Test	Impact on other Physico-chemical tests	Impact on toxicology	Impact on Ecotoxicology	Impact on Risk assessment
Viscosity	Choice of methods for the determination of density.	Parameter for aspiration hazard		Assessment of spreadibility of liquids

R.7.1.1.7 References for introduction of Physico-Chemical properties

Baker J.R., Milhelcic J.R., Luehrs D.C. and Hickey J.P. Evaluation of estimation methods for organic carbon normalised sorption coefficients. *Water Environ. Res.* (1997) 69, 136-144.

Gerstl Z. Estimation of organic chemical sorption by soils. *J. Contaminant Hydrology* (1990) 6, 357-375 .

Bretherick (1999) Bretherick's Handbook of Chemical Reactive Hazards: An Indexed Guide to Published Data, 6th Edition (2 volume set). P Urben and L Bretherick (Authors). Butterworth Heinemann

CRC (2000) CRC Handbook of Chemistry and Physics 81st Edition. Editor in Chief, D. Lide. CRC Press

Dearden J.C. *In silico* prediction of aqueous solubility. *Expert Opinion on Drug Discovery* 1 (2006) 31-52.

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

Eckhoff RK (1997) Dust Explosions in the Process Industries 2nd Edition. Butterworth-Heinemann, Oxford, UK

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

Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., Gramatica, P.(2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. *Environ. Health Perspect*. 111, 1361-1375.

Hall, L.H., Kier, L.B. "The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Relations", in Reviews of Computational Chemistry, Chap. 9, pp 367-422, eds. Donald Boyd and Ken Lipkowitz, VCH Publishers, Inc. (1991).

Hansch C., Quinlan J.E. and Lawrence G.L. The linear free-energy relationship between partition coefficients and aqueous solubility of liquids. *J. Org. Chem.* (1968) 33, 347-350.

Hawley (2001) Hawley's Condensed Chemical Dictionary 14th Edition. Edited by GG Hawley. 14th Edition revised by RJ Lewis Sr. Wiley, New York, USA

ISO (1995) Guide to the expression of uncertainty in measurement. ISO. Geneva, Switzerland

ISO/IEC 17025 (2005) General requirements for the competence of testing and calibration laboratories. Geneva, Switzerland

IUPAC Solubility Data Series (1994-). A database containing solubilities originally published in the International Union for Pure and Applied Chemistry-National Institute of Standards and Technology (NIST) Solubility Data Series can be viewed on-line at http://srdata.nist.gov/solubility/ Published as journal series by Oxford University Press.

Karickhoff SW, McDaniel VK, Melton C, Vellino AN, Nute DE, Carreira LA. (1991) Predicting chemical reactivity by computer. *Environ. Toxicol. Chem*.;10:1405-16.

Katritzky AR, Wang Y, Sild S and Tamm T. QSPR studies on vapour pressure, aqueous solubility, and the prediction of water-air partition coefficients. *J. Chem. Inf. Comput. Sci.* 38 (1998) 720-725.

Kier, L.B., Lowell, H. Hall Molecular Connectivity in Structure-Activity Analysis", Research Studies Press Ltd., Hertfordshire, England and John Wiley and Sons, New York (1986)

Kier L.B. and Hall L.H. *Molecular Structure Description: the Electrotopological State.* Academic Press, San Diego, CA, 1999.

Klimisch H.J., Andreae E. and Tillmann U. (1997) - A systematic approach for evaluating the quality of experimental and ecotoxicological data. *Reg. Tox. and Pharm.,* 25, 1-5.

Lange (1998) Lange's Handbook of Chemistry 15th Edition. Edited by JA Dean and NA Lange. McGraw Hill, New York and London.

Mackay D, Shiu WY and Ma KC (1991-1997) Illustrated Handbook of Physico-chemical Properties and Environmental Fate for Organic Chemicals (Volumes 1-5). Lewis Publishers, Boca Raton, USA

Merck (2001) Merck Index 13th Edition. Edited by S Budavari *et al*. Merck & Co, Inc, USA

National Fire Protection Association (1997) Fire Protection Guide to Hazardous Materials 12th Edition. National Fire Protection Association, Quincy, MA, USA

OECD[a] (2004) Principles for the Validation of (Q)SARs http://ecb.jrc.it/QSAR/home.php?CONTENU=/QSAR/background/background_oecd_principles.php

OECD[b] (2004) Manual for Investigation of HPV Chemicals. Section 3.1 Guidance for Determining the Quality of Data for the SIDS Dossiers: (Reliability, relevance and adequacy). [http://www.oecd.org/dataoecd/13/15/36045203.pdf]

OECD[c] (2004) series on testing and assessment Number 49 The report from the expert group on (quantitative) structure activity relationships [(Q)SARs] on the principles for the validation of (Q)SARs. 2nd Meeting of the ad hoc Expert Group on QSARs http://appli1.oecd.org/olis/2004doc.nsf/linkto/env-jm-mono(2004)24

Perrin, D. D., Dempsey, B. and Serjeant, E. P., *pKa Prediction for Organic Acids and Bases,* 1st ed., Chapman and Hall, London (1981), pp 27-52 (and references therein).

Ran Y., He Y., Yang G., Johnson J.L.H. and Yalkowsky S.H. Estimation of aqueous solubility of organic compounds by using the general solubility equation. *Chemosphere* (2002) 48, 487-509.

Sax (2000) Sax's Dangerous Properties of Industrial Materials 10th Edition (3 volume set). Edited by RJ Lewis Sr. John Wiley & Sons Inc, New York, USA

Schüürmann G., Kühne R., Kleint F., Ebert R.-U., Rothenbacher C. and Herth P. A software system for automatic chemical property estimation from molecular structure, in *Quantitative*

Structure-Activity Relationships in Environmental Sciences – VII, Chap. 7, pp. 93-114, eds. Chen F. and Schüürmann G, SETAC Press, Pensacola, FL (1997)

Sedykh A.Y. and Klopman G. A structural analogue approach to the prediction of the octanol-water partition coefficient. *J. Chem. Inf. Model.* 46 (2006),

Tetko, I.V., Bruneau, P., Mewes, HW, Roher, D.C., Poda, G.I. (2006). Can we estimate the accuracy of ADME-Tox predictions ? *Drug Disc. Today* 15/16, 700-707

Tomlin CDS (2000) The Pesticide Manual: A World Compendium 12th Edition. Editor: CDS Tomlin. British Crop Protection Council, Surrey, UK

Verschueren K (2001) Handbook of Environmental Data on Organic Chemicals 4th Edition (2 volume set). John Wiley & Sons Inc, New York, USA

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, and Worth, A. (2006). Validation of counter propagation neural network models for predictive toxicology according to the OECD principles: a case study. *SAR and QSAR in Environ. Res.* 3, 265-284.

Walker, J.D., Carlsen, L., Jaworska, J. (2003) Improving opportunities for regulatory acceptance of QSARs: the importance of model domain, uncertainty, validity and predictability. *QSAR Comb. Sci.* 22, 346-350.

Walker J.D., Jaworska J., Comber M.H.I., Schultz T.W. and Dearden J.C. Guidelines for developing and using quantitative structure-activity relationships. *Environ. Toxicol. Chem.* 22 (2003) 1653-1665.

Walker J.D., de Wolf W. QSARs on the world wide web: a need for quality assurance to prevent misuse. In Walker J.D. (Ed.), *QSARs for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications.* SETAC Press, Pensacola, FL, 2003, pp. 199-201.

Weed, D.L., (2005) *Weight of Evidence*: a review of Concept and Methods. Risk Analysis, 25:1545-1557.

Yaw CL (1999) Chemical Properties Handbook: Physical, thermodynamic, environmental, transport, safety, and health related properties for organic and inorganic chemicals. McGraw Hill, New York and London.

R.7.1.2 MELTING/FREEZING POINT

The melting point of a chemical is an environmentally relevant property:

- it determines the temperature at which a substance changes its physical state from solid to liquid and thereby gives an indication of the distribution of the substance within and between the environmental media: water, soil and air.
- as a measure of the substance's purity it can give an indication of impurities which may have environmental relevance.
- the melting point is important for identification reasons.

The melting point serves as an indicator for the physical state (liquid or solid) of a substance.

Definition of melting/freezing point

The melting point is defined as the temperature at which the phase transition from solid to liquid state at normal atmospheric pressure takes place. This temperature ideally corresponds to the temperature at the solidification point or the freezing point.

As the phase transition of many substances takes place over a large temperature range, it is often described as the melting range.

Conversion of units (K to ºC)

t: Celsius temperature, degree Celsius (ºC)

T: thermodynamic temperature, Kelvin (K)

R.7.1.2.1 Information requirements on melting/freezing point

The study does not need to be conducted below a lower limit of –20°C. Information on other physical properties that should be known before the test is conducted, the classes of compounds for which the test is not required, the tonnage triggering, and the information that should be reported is given in the general introduction.

R.7.1.2.2 Available information on melting/freezing point

Testing data on melting/freezing point

The methods described can be applied to solids, provided that these do not undergo chemical reaction at temperatures below the melting point (for example, auto-oxidation, rearrangement, degradation, etc.).

Most of the methods to determine the melting point described here are based on international and national standards (OECD 102, EU A.1). The fundamental principles are given by IUPAC (1975) and Weissberger (1959).

The various methods for determining the temperature (temperature range) of the phase transition from the solid to the liquid state are shown. In practice, the temperatures at the initial melting and the final stage of melting are determined while heating a sample of the test substance at atmospheric pressure.

Table R.7.1-6 Capillary methods

Method of measurement	Substances which can be pulverised	Substances which are not readily pulverised	Temperature range	Estimated accuracy *	Existing standards
Melting point devices with liquid bath	yes	only a few	273 to 573 K	± 0.3 K	JIS K 0064
Melting point devices with metal block	yes	only a few	293 to >573 K	± 0.5 K	ISO 1218 (E)
Photocell detection Not suitable for some highly coloured substances	yes	Several, with appliance devices	253 to 573 K	± 0.1 K	

* dependent on type of instrument and degree of purity of the substance

Table R.7.1-7 Hot stages and freezing methods

Method of measurement	Substances which can be pulverised	Substances which are not readily pulverised	Temperature range	Estimated accuracy *	Existing standards
Kofler hot bar	yes	no	283 to >573 K	± 1.0 K	ANSI/ASTM D3451-76
Melt microscope	yes	only a few	273 to >573 K	± 0.2 K	DIN 53736
Meniscus method	no	specifically for polyamides	293 to >573 K	± 0.5 K	ISO 1218(E)
Freezing point methods	partially	partially	223 to 573 K	± 0.5 K	e.g. BS 4695

* dependent on type of instrument and on degree of purity of the substance

Table R.7.1-8 Thermal analysis

Method of measurement	Substance which can be pulverised	Substances which are not readily pulverised	Temperature range	Estimated accuracy *	Existing standards
Differential Thermal Analysis	yes	yes	173 to 1273 K	up to 600 K ± 0.5 K up to 1273 K ± 2.0 K	ASTM E 737-76
Differential Scanning Calorimetry	yes	yes	173 to 1273 K	up to 600 K ± 0.5 K up to 1273 K ± 2.0 K	ASTM E 737-76

* dependent on type of instrument and on degree of purity of the substance

Table R.7.1-9 Pour point

Method of measurement	Substance which can be pulverised	Substances which are not readily pulverised	Temperature range	Estimated accuracy *	Existing standards
Pour point.	for petroleum oils and oily substances	for petroleum oils and oily substances	223 to 323 K	± 0.3 K	ASTM D 97-66

* dependent on type of instrument and on degree of purity of the substance

Reference substances do not need to be employed in all cases when investigating a new substance. They should primarily serve to check the performance of the method from time to time and to allow comparison with results from other methods. Details of some typical calibration standards are (Jucker and Suter, 1968/69; European Pharmacopoeia, 1974; Bervenmark *et al*., 1963; IUPAC, 1976).

Published data on melting/freezing point

Most physical properties, such as molecular weight, melting point, boiling point and density can be obtained from commonly used environmental handbooks, such as Verschueren's Handbook of Environmental Data on Organic Chemicals (1983), Howard's Handbook of Environmental Fate and Exposure Data, Vol. I and II (1990), Lide's CRC Handbook of Physics and Chemistry, Lange's Handbook of Chemistry, the Merck Index, the Aldrich Catalogue, Kirk-Othmer Encyclopaedia of Chemical Technology and other handbook compilations such as Riddick *et al*. (1986).

Alternatively, searching on various environmental databases, such as HSDB (http://www.toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB), will provide summaries of chemical and physical properties of substances.

R.7.1.2.3 Evaluation of available information on melting/freezing point

Experimental data on melting/freezing point

The methods described in the various OECD and EC test guidelines are appropriate for the determination of the melting point of most chemical substances without any restriction in respect to their degree of purity, but in the case of commercial grade substances, potential impurity effects on measurement results have to be expected.

Examples of the various stages of melting are provided in the test guidelines. Temperatures are recorded at the beginning of melting and at the final stage. If the difference is within the limits of accuracy of the method, the temperature at the final stage of melting is taken as the melting point, otherwise the two temperatures should be reported.

Some substances will decompose or sublime before the melting point is reached. In these circumstances the decomposition temperature or sublimation temperature should be reported. Information and remarks relevant to the interpretation of results must be reported, especially in regard to impurities and physical state of the substance.

Non-experimental data on melting/freezing point

The melting point of a crystalline compound is controlled largely by two factors – intermolecular interactions and molecular symmetry. For example, 3-nitrophenol, which can hydrogen-bond via its

–OH group, melts at 97°C, whereas its methyl derivative, 3-nitroanisole, which cannot hydrogen-bond with itself, melts at 39°C. The symmetrical 1,4-dichlorobenzene melts at 53°C, whilst the non-symmetrical 1,3-dichlorobenzene melts at -25°C. These and other effects have been discussed in detail by Dearden (1999).

There have been many attempts to predict the melting point of organic chemicals, and these have been reviewed by Horvath (1992), Reinhard and Drefahl (1999), Dearden (1999, 2003) and Tesconi and Yalkowsky (2000). It may be noted that in the 19th century Mills (1884) developed a QSPR based on carbon chain length for melting points of homologous series of compounds that was accurate to +/-2°.

Essentially two approaches have been used in the prediction of melting point – the physico-chemical/structural descriptor approach and the group contribution approach. The former is exemplified by the work of Katritzky *et al* (1997), who used 9 of their CODESSA descriptors to model a diverse set of 443 aromatic chemicals with $r^2 = 0.837$ and s = 30.2°.

This is a complex QSPR, with descriptors that are not easy to comprehend, and reflects the difficulty of modelling the melting points of diverse data sets. Even for a set of 58 PCB congeners with 1-10 chlorine atoms, a 5-term QSPR was required (Abramowitz & Yalkowsky 1990), with $r^2 = 0.83$ and s = 22.1°.

The group contribution approach was used by Simamora and Yalkowsky (1994) to model the melting points of a diverse set of 1690 aromatic compounds. Using a total of 41 group contributions and 4 intramolecular hydrogen bonding terms they found a standard error of 37.5°. Marrero and Gani (2001) used two levels of group contributions to predict the melting points of 1103 diverse chemicals with a standard error of 25.3°.

There are a few software programs that predict melting point; they all use one or more group contribution approaches. Dearden (2003) used a 100-compound test set to compare the performances of three of these programs (Table R.7.1-10).

Table R.7.1-10 Examples of software programs that predict melting point

Software	Website	Availability	Mean absolute error
MPBPVP	www.epa.gov/oppt/exposure/docs/episuitedl.htm	Freely downloadable	26.3^o
ChemOffice	www.cambridgesoft.com	Commercial	27.0^o
ProPred	www.capec.kt.dtu.dk	Purchase	25.8^o

It can be seen that there is little to choose between the programs in terms of accuracy of prediction. They can all operate in batch mode. It is therefore recommended that the MPBPVP software be used to calculate melting point. The method requires the input of a chemical structure using SMILES. The SMILES method is simple to use, and a tutorial can be found at www.daylight.com/smiles/smiles-intro.html.

It should be noted that currently both QSPR methods and software programs have prediction errors well in excess of the error on experimental measurement of melting point, which is usually <2°. Experimental determinations of melting/freezing point are therefore to be preferred to the results yielded by QSPRs models and predictive software because of their greater accuracy.

Remaining uncertainty on melting/freezing point

There is little difference between the various methods in the accuracy of determination of melting point. The selection of the melting point method is dependent on the nature of the substance to be tested. In consequence the limiting factor will be according to, whether or not the substance can be pulverised easily, with difficulty, or not at all. The melting point of a pure substance is always higher than the melting point of that substance when a small amount of impurity is present. The more impurity the lower the melting point. Melting points are not normally carried out for mixtures.

R.7.1.2.4 Conclusions on melting/freezing point

For some substances, the determination of the freezing or solidification point is more appropriate and details of this method have also been included. Where due to the particular properties of the substance, none of the above parameters can be conveniently measured, a pour point may be appropriate.

Information on the melting point will impact the choice of method for flash point, flammability, autoflammability, oxidising properties and explosive properties. If decomposition occurs during the melting point study, a boiling point need not be carried out.

Concluding on C&L and Chemical Safety Assessment

It is not used as a Classification & Labelling criterion or to define PBT properties. It indicates (together with the boiling point) the physical state of a compound, e.g. a liquid (according to the GHS) is defined as a substance or mixture that is not a gas and which has a melting point or initial melting point of 20ºC or less at standard pressure of 101.3 kPa.

R.7.1.2.5 Integrated testing strategy (ITS) for melting/freezing point

Testing for melting/freezing point is normally carried out on all liquids and solids.

Examples and Case studies on melting/freezing point

For most substances, melting and freezing points are equal. For example, the melting point and freezing point of the element mercury is 234.32 Kelvins (-38.83ºC or –37.89ºF). However, certain substances possess differing solid-liquid transition temperatures. For example, agar melts at 85ºC (185ºF) and solidifies from 32ºC to 40ºF (89.6ºF to 104ºF); this process is known as hysteresis. Certain materials, such as glass, may harden without crystallising; these are called amorphous solids. Unlike the boiling point, the melting point is relatively insensitive to pressure. The chemical element with the highest melting point is tungsten, at 3695ºK (3422ºC, 6191ºF). Carbon does not melt at ambient pressure but sublimates at about 4000ºK; a liquid phase only exists above pressures of 10MPa and estimated 4300-4700ºK. At the other end of the scale, helium does not freeze at all at normal pressure, even at absolute zero.

R.7.1.2.6 References on melting/freezing point

Abramowitz R. and Yalkowsky S.H. Estimation of aqueous solubility and melting point of PCB congeners. *Chemosphere* (1990) 21, 1221-1229.

Budavari, S., Ed. (1989) The Merck Index – Encyclopedia of Chemicals, Drugs and Biologicals. Rahway, NJ:Merck and Co., Inc.

Dearden J.C. Quantitative structure-property relationships for prediction of boiling point, vapour pressure, and melting point. *Environ. Toxicol. Chem.* (2003) **22**, 1696-1709.

Dearden J.C. The prediction of melting point. In Charton M. and Charton B. (Eds.), *Advances in Quantitative Structure-Property Relationships*, Vol. 2, JAI Press, 1999, pp. 127-175.

EC Method A1. "Melting/Freezing Temperature", Dir 92/69/EEC, Official Journal of the European Communities, O.J. L383 A)

Horvath A.L. Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds. Elsevier, Amsterdam, 1992.

Howard, P.H., Ed. (1990). Handbook of Fate and Exposure Data for Organic Chemicals, Vol. I and II, Lewis Publishers, Chelsea, Michigan.

IUPAC (1975) "Experimental thermodynamics", Le Neindre, B. and Vodar, B. (eds), vol. II, pp 803-834, Butterworths, London, 1975

IUPAC (1976) "Physico-chemical Measurements: Catalogue of Reference Materials from National Laboratories", Pure and Applied Chemistry, 48, 505-515.

Katritzky A.R., Maran U., Karelson M. and Lobanov V.S. Prediction of melting points for the substituted benzenes. *J. Chem. Inf. Comput. Sci.* (1997) 37, 913-919.

Kirk-Othmer Encyclopedia of Chemical Technology. 4th ed. Volume 1: New York, NJ,. John Wiley and Sons, 1991-Present.

Lange's Handbook of Chemistry, Editor, Speight, James G. McGraw-Hill.Lide, D.R., Ed. (2005). CRC Handbook of Chemistry and Physics, 86th edition, CRC Press, Taylor and Francis, Boca Raton, FL.

Marrero J. and Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibr.* (2001) 183-184, 183-208.

Mills E.J. On melting point and boiling point as related to composition. *Phil. Mag.* (Series 5) (1884) 17, 173-187.

OECD Guidelines for Testing of Chemicals – Method 102 "Melting Point/Melting Range"

Riddick, J.A., Bunger, W.B., Sakano, T.K. (1986). Organic Solvents: Physical Properties and Methods of Purification. Techniques of Chemistry. 4th Edition. New York, NY, Wiley-Interscience

Simamora P. and Yalkowsky S.H. group contribution methods for predicting the melting points and boiling points of aromatic compounds. *Ind. Eng. Chem. Res.* (1994) 33, 1405-1409.

Tesconi M. and Yalkowsky S.H. Melting point. In Boethling R.S. and Mackay D. (Eds.), *Handbook of Property Estimation Methods for Chemicals.* Lewis, Boca Raton, FL, 2000, pp. 3-27.

Verschueren, K. (1983) Handbook of Environmental Data on Organic Chemicals. Fourth Edition. John Wiley & Sons, Inc., New York.

Weissberger, R. (ed), "Techniques of Organic Chemistry, Physical Methods of Organic Chemistry", 3rd ed., vol. I, Part I, Chapter VII, Interscience Publ., New York, 1959.

R.7.1.3 BOILING POINT

The boiling point of a substance is an environmentally relevant physical chemical property because it is important for identification purposes and is one factor influencing the states in which the substance will exist in the environment.

Besides being an indicator for the physical state (liquid or gas) of a substance, boiling point serves as an indicator of volatility, with higher boiling points indicating lower volatility at ambient temperatures. Evaporation rates generally have an inverse relationship to boiling points; i.e. the higher the boiling point, the lower the rate of evaporation. The boiling point is a key input in equations that provide estimates of a chemical's vapour pressure as a function of temperature.

The boiling point value is also useful for the identification of pure substances and along with melting point and refractive index, as criteria of purity.

Definition of boiling point

The standard boiling point is described as the temperature at which the pressure of the saturated vapour of a liquid is the same as the standard atmospheric pressure, 101.325 kPa.

The measured boiling point is dependent on the atmospheric pressure. This dependence can be described quantitatively by the Clausius-Clapeyron equation as follows:

$$\log p = - \Delta H_v / 2.3\ RT + \text{constant}$$

where p = vapour pressure of the substance

ΔH_v = the heat of vaporisation of the substance

R = the universal molar gas constant (8.31441 J mol^{-1} K^{-1})

T = temperature expressed in K

The temperature at the boiling point (boiling temperature) is stated in K, with regard to the ambient pressure during the measurement. If no pressure is given, the result refers to a standard pressure of 101.325 kPa.

R.7.1.3.1 Information requirements on boiling point

Column 2 of REACH Annex VII provides the following exemptions. A study does not need to be conducted:

- for gases;
- for solids which either melt above 300°C or decompose before boiling;
- for substances which decompose before boiling.

R.7.1.3.2 Available information on boiling point

Published data on boiling point

Most physical properties, such as molecular weight, melting point, boiling point and density can be obtained from commonly used environmental Handbooks, such as Verschueren's Handbook of

Environmental Data on Organic Chemicals (1983), Howard's Handbook of Environmental Fate and Exposure Data, Vol. I and II (1990), Lide's CRC Handbook of Physics and Chemistry, Lange's Handbook of Chemistry, the Merck Index, Kirk-Othmer Encyclopaedia of Chemical Technology and other handbook compilations such as Riddick *et al*. (1986).

Alternatively, searching on various environmental databases, such as HSDB (http://www.toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB), will provide summaries of chemical and physical properties of substances.

Testing data on boiling point

Test methods are applied to liquids, provided that they do not undergo chemical reaction at temperatures below the boiling point (for example, auto-oxidation, rearrangement, degradation, etc.) Table R.7.1-11 lists the methods for determining the boiling point.

Table R.7.1-11 Methods for determining the boiling point

Method of measurement	Estimated accuracy		Existing standard
Ebulliometer	± 1.4 K ± 2.5 K	(up to 373 K) (1,2) (up to 600 K) (1,2)	ASTM D 1120-72 (1)
Dynamic method	± 0.5 K	(up to 600 K) (2)	
Distillation process (boiling range)	± 0.5 K	(up to 600 K)	ISO/R 918, DIN 53171, BS 4591/71
According to Siwoloboff	± 2 K	(up to 600 K) (2)	Based on JIS K 0064-1966.
Photocell detection	± 0.3 K	(up to 373 K) (2)	
Differential thermal analysis	± 0.5 K ± 2.0 K	(up to 600 K) (up to 1273 K)	ASTM E 537-76
Differential scanning calorimetry	± 0.5 K ± 2.0 K	(up to 600 K) (up to 1273 K)	ASTM E 537-76

(1) This accuracy is only valid for the simple device as for example described in ASTM D 1120-72; it can be improved with more sophisticated ebulliometer devices.
(2) Only valid for pure substances. The use in other circumstances should be justified.

R.7.1.3.3 Evaluation of available information on boiling point

The results obtained for mixtures or impure samples are to be interpreted with care. With an impure sample, for instance, the emergence of a low boiling component will be registered as the boiling point. Repeated determinations with the same impure sample can change the composition from measurement to measurement, due to the volatilisation of low boiling components; continuously increasing values are obtained in these circumstances.

Liquids with a tendency to superheat can yield incorrect results. The values obtained are usually too high. This happens frequently at higher temperatures. Distillation methods or the dynamic vapour pressure method are more suitable for these types of compound.

Non-experimental data on boiling point

Lyman (2000) has discussed seven recommended methods for the prediction of boiling point. The methods are based on physico-chemical and structural properties and group contributions. Rechsteiner (1990), Reinhard and Drefahl (1999) and Dearden (2003) have reviewed the QSPR prediction of boiling point.

Many studies of boiling point prediction have dealt with specific chemical classes, and very good correlations have generally been obtained. In 1884 Mills (1884) modelled the boiling points of a number of homologous series with QSPRs based on carbon chain length. Ivanciuc *et al* (2000) used 4 topological descriptors to model the boiling points of 134 alkanes with a standard error of 2.7°C, whilst Gironés *et al* (2000) used only one quantum chemical descriptor (electron-electron repulsion energy) to model the boiling points of 15 alcohols with a standard error of 5.6°C

Models based on diverse training sets are, however, more widely applicable. Katritzky *et al* (1996) used 4 CODESSA descriptors to model the boiling points of 298 diverse organic compounds, with r^2 = 0.973 and standard error = 12.4°C. Basak *et al* (2001) used 8 topochemical, topological and hydrogen bonding descriptors to model the boiling points of 1015 diverse organic compounds, with a standard error of 15.7°C.

A group contribution approach was used by Marrero and Gani (2001) to model the boiling points of 1794 organic compounds with a standard error of 8.1°C, whilst Labute (2000) used 18 atomic contributions on a set of 298 diverse organics, to give a standard error of 15.5°C. Simamora and Yalkowsky (1994) used 36 group contributions and 4 intramolecular hydrogen bonding terms to model the boiling points of a diverse set of 44 aromatic compounds, with a standard error of 17.6°C. There are several software programs available for the prediction of boiling point, and Dearden (2003) compared the performance of six of these using a 100-compound test set (Table R.7.1-12).

Table R.7.1-12 Software programs for the prediction of boiling point

Software	Website	Availability	Mean absolute error
ACDLabs	www.acdlabs.com	Purchase	1.0°
SPARC	ibmlc2.chem.uga.edu/sparc	Free on line	6.3°
MPBPVP	www.epa.gov/oppt/exposure/ docs/episuitedl.htm	Freely downloadable	13.8°
ChemOffice	www.cambridgesoft.com	Purchase	13.8°
ProPred	www.capec.kt.dtu.dk	Members only	16.1°
Molecular Modelling Pro	www.chemsw.com	Purchase	21.7°

The ACDLabs result is based on the 54 chemicals in the test set that were not included in the ACDLabs training set. All the programs except SPARC can be run in batch mode.

Clearly the ACDLabs software gives by far the best predictions, but has to be purchased. SPARC is freely accessible, but operates only in manual mode, with SMILES input. MPBPVP can be freely downloaded, but its standard error of prediction is more than twice that of SPARC.

Other software programs, namely ASTER, ChemProp and PREDICT (mwsoftware.com/dragon/) also predict boiling point, but no indication of their performances is available.

The performance of QSPR models for boiling points should be carefully and thoroughly checked (e.g. estimation of the predictive power for several analogs not belonging to the training set) because predictions errors for boiling points tend to be greater than experimental accuracy.

Remaining uncertainty on boiling point

In the literature, different boiling points are sometimes quoted for the same substance. These differences are due to such variables as the dimensions of the apparatus (e.g. the fit of the thermometer), the type of thermometer, the stem correction, the pressure correction and the accuracy of the pressure measurement. Therefore the above mentioned international and national standardised methods contain precise requirements for these specified conditions.

The influence of impurities on the determination of the boiling point depends greatly upon the kind of impurity. Thus, the effect can be considered if a highly volatile solvent is present in the sample. Impurities will usually increase/decrease the measured boiling temperature.

R.7.1.3.4 Conclusions on boiling point

Boiling point is one of the most useful properties for the characterisation of organic compounds and, subsequently, experimentally derived boiling points are preferred over predicted methods. A boiling point should be a mean of two measurements, which are in the range of approximate accuracy indicated in the table above.

The measured boiling points and their mean should be stated (in K), and the pressures at which the measurements were made should be recorded (in kPa). Where a test substance boils over a temperature range, this range should be provided. The measured values should also be corrected to standard pressure.

Concluding on C&L and Chemical Safety Assessment

The boiling point is one of the criteria used in assigning a substance to an appropriate flammability category. Under GHS, substances are classified as Category 1 - *extremely flammable liquid and vapour* if they have a flash point of <23ºC and boiling point <35ºC (e.g. diethyl ether, carbon disulphide) and as Category 2 - *highly flammable liquid and vapour* if they have a flash point of <23ºC and boiling point >35ºC (e.g. petrol, acetone).

R.7.1.3.5 Integrated testing strategy (ITS) for boiling point

Information on the melting point is useful before undertaking any boiling point determination. The summary should include the numerical value or range for the boiling point, the accuracy and the method. Comments should refer to any decomposition and list any volatile impurities known to be present.

Examples and case studies for boiling point

The boiling point of members of a homologous series will increase uniformly with increasing molecular weight (or size). If a hydrogen atom of a hydrocarbon is replaced by another atom or group, then the boiling point will increase. Thus, alkyl halides, aldehydes, ketones, acids, etc have higher boiling points than their respective hydrocarbons. If a group is introduced that can promote association, then a marked rise in boiling point arises. This is pronounced in acids and alcohols, where hydrogen bonding can occur.

R.7.1.3.6 References on boiling point

Basak S.C. and Mills D. Use of mathematical structural invariants in the development of QSPR models. *Commun. Math. Comput. Chem.* (2001) No. 44, 15-30.

Budavari, S., Ed. (1989) The Merck Index – Encyclopedia of Chemicals, Drugs and Biologicals. Rahway, NJ:Merck and Co., Inc.

Dearden J.C. Quantitative structure-property relationships for prediction of boiling point, vapour pressure, and melting point. *Environ. Toxicol. Chem.* (2003) **22**, 1696-1709.

Gironés X., Amat L., Robert D. and Carbó-Dorca R. Use of electron-electron repulsion energy as a molecular descriptor in QSAR and QSPR studies. *J. Comput.-Aided Mol. Des.* (2000) 14, 477-485.

Howard, P.H., Ed. (1990). Handbook of Fate and Exposure Dta for Organic Chemicals, Vol. I and II, Lewis Publishers, Chelsea, Michigan.

Ivanciuc O., Ivanciuc T., Cabrol-Bass D. and Balaban A.T. Evaluation in quantitative structure-property relationship models of structural descriptors derived from information-theory operators. *J. Chem. Inf. Comput. Sci.* (2000) 40, 631-643.

Katritzky A.R., Mu L., Lobanov V.S. and Karelson M. Correlation of boiling points with molecular structure. 1. A training of 298 diverse organics and a test set of 9 simple organics. *J. Phys. Chem.* (1996) 100, 10400-10407.

Kirk-Othmer Encyclopedia of Chemical Technology. 4th ed. Volume 1: New York, NJ,. JohnWiley and Sons, 1991-Present.

Labute P. A widely applicable set of descriptors. *J. Mol. Graph. Model.* (2000) 18, 464-477.

Lange's Handbook of Chemistry, Editor, Speight, James, Publ. G. McGraw-Hill.

Lide, D.R., Ed. (2005). CRC Handbook of Chemistry and Physics, 86th edition, CRC Press, Taylor and Francis, Boca Raton, FL.

Lyman W.J. Boiling point. In Boethling R.S. and Mackay D. (Eds.), *Handbook of Property Estimation Methods for Chemicals*, Lewis, Boca Raton, FL, 2000, pp. 29-51.

Marrero J. and Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilbr.* (2001) 183-184, 183-208.

Mills E.J. On melting point and boiling point as related to composition. *Phil. Mag.* (Series 5)(1884) 17, 173-187.

OECD Guidelines for Testing of Chemicals – Method 103 "Boiling Point/Boiling Range"

Official Journal of the European Communities L225/1, Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances

Riddick, J.A., Bunger, W.B., Sakano, T.K. (1986). Organic Solvents: Physical Properties and Methods of Purification. Techniques of Chemistry. 4th Edition. New York, NY, Wiley-Interscience

Simamora P. and Yalkowsky S.H. Group contribution methods for predicting the melting points and boiling points of aromatic compounds. *Ind. Eng. Chem. Res.* (1994) 33, 1405-1409.

Verschueren, K. (1983) Handbook of Environmental Data on Organic Chemicals. Fourth Edition. John Wiley & Sons, Inc., New York.

R.7.1.4 RELATIVE DENSITY

The density of a substance is environmentally relevant because it is helpful in estimating the distribution of the substance within and between water, soil and air.

For gaseous materials, density is of value in determining the tendency to settle or to disperse when discharged at high concentrations into the atmosphere. The density of gaseous substances can be calculated from molecular weight using the gas law.

For insoluble liquids and solids, density will be a determining factor in the settling of the substance.

Definition of relative density

Density (ρ) of a substance is the quotient of the mass m and its volume V

$\rho = m/V$ SI units (kg/m^3)

The relative density, D_4^{20}, of solids or liquids is the ratio between the mass of a volume of substance to be examined, determined at 20ºC, and the mass of the same volume of water, determined at 4ºC (at which temperature, water has its maximum density, i.e. 0.999975 $kg/m^{3)}$. The relative density has no dimension.

R.7.1.4.1 Information requirements on relative density

For liquids, it is useful to have some indication of the dynamic viscosity as this can affect the choice of method.

The study does not need to be conducted if:

- the substance is only stable in solution in a particular solvent and the solution density is similar to that of the solvent. In such cases, an indication of whether the solution density is higher or lower than the solvent density is sufficient; or
- the substance is gaseous at room temperature. In this case, an estimation based on calculation shall be made from its molecular weight and the Ideal Gas Laws.

The summary should include the numerical value and temperature at which it was measured, purity of the sample used, physical state, method used and reference substance (if any).

R.7.1.4.2 Available information on relative density

Testing data on relative density

Test methods for determining density are applicable to solids and liquids. Table R.7.1-13 lists the respective test methods.

Table R.7.1-13 Test methods for determining density

Method	Applicability	Maximum Dynamic Viscosity (Liquids Only)/Pa.S
Hydrometer	Liquids	5
Hydrostatic balance	Solids and Liquids	5
Immersion ball	Liquids	20
Pycnometer	Solids and Liquids	500
Air comparison pycnometer	Solids	-
Oscillating densitimeter	Liquids	5

The hydrometer method, the immersed ball method and the oscillating density meter methods are applicable for liquids only.

A number of suitable reference substances, to be used primarily for performing calibration of the method, have been recommended by IUPAC (1983).

Published data on relative density

Most physical properties, such as molecular weight, melting point, boiling point and density can be obtained from commonly used environmental Handbooks, such as Verschueren's Handbook of Environmental Data on Organic Chemicals (1983), Howard's Handbook of Environmental Fate and Exposure Data, Vol. I and II (1990), Lide's CRC Handbook of Physics and Chemistry, Lange's Handbook of Chemistry, the Merck Index, the Aldrich Catalog, Kirk-Othmer Encyclopedia of Chemical Technology and other handbook compilations such as Riddick *et al*. (1986).

Alternatively, searching on various environmental databases, such as HSDB (http://www.toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB), will provide summaries of chemical and physical properties of substances.

R.7.1.4.3 Evaluation of available information for relative density

Experimental data on relative density

The relative density related to seawater allows a judgement on whether a substance will float or sink based on the standard phrases of the European Classification System criteria of the Bonn Agreement (1983). Assuming that the water solubility is relatively low (<50 g/l) the substance is likely to float with a density <1.0 and sink if the density is >1.03.

Non-experimental data on relative density

For this endpoint there are often experimental measurements and therefore QSPR models for this property did not receive special attention in the environmental literature. Several softwares programs can calculate the density of a given chemical but the documentation and validation of the methods is limited.

Nelken (1990) and Reinhard and Drefahl (1999) have reviewed the prediction of relative density, ρ_L. A related property is molar volume, V_M (the volume in cm^3 occupied by 1 gram mole of a compound), and the two are related thus:

$$\rho_L = M/V_M$$

where M = molecular weight, and ρ_L has the units of g cm^{-3}.

The statistics and descriptors of the methods for the determination of relative density reported in Reinhard and Drefahl (1999) are listedTable R.7.1-14.

Table R.7.1-14 Computational models for the estimation of vapour pressure

Compounds	Authors	Methodology	Statistics
Alkanes C_5-C_9	Kier and Hall (1976)	Molecular descriptors	n=46, s=0.0024, r=0.9971
Dialkyl methylphosphonates	Xu, Wang and Su (1992)	Molecular descriptors	n=14, s=0.006, r=0.97
Organic, inorganic and metal-organic liquids	Girolami (1994)	Group contribution approach	r^2=0.982 for correlation between observed and estimated densities
Alkanes	Dubois and Loukianoff (1993)	Group contribution approach	R=0.9952; s=0.4783; F=3258; n=355

(reviewed in Reinhard and Drefahl, 1999) n is the sample size, s the standard deviation and r the correlation coefficient

Correlations between density or molar volume and molecular surface area (Grigoras 1990), molecular connectivities (Kier & Hall 1976) and group contributions (Girolami 1994) have been reported. The Girolami method is very simple, and is based on the following equation:

$$\rho_L = M(5V_{scal})^{-1}$$

where M = molecular weight, and V_{scal} = scaled volume calculated as the sum of the atom contributions of the constituent atoms. The method is claimed to be accurate to within 0.1 g cm^{-3}.

The variation of density with temperature can be estimated using the method of Grain (reported in (Nelken 1990):

$$\rho_L = M\rho_{Lb}[3 - 2(T/T_b)]^n$$

where M = molecular weight, subscript "b" refers to the boiling point, and n = a constant that depends on chemical class (n = 0.25 for alcohols, 0.29 for hydrocarbons and 0.31 for other organics).

Abraham and McGowan (1987) reported a very simple method for the calculation of characteristic volume, which is closely correlated with molar volume. Atomic and bond contributions are: C 16.35, H 8.71, O 12.43, N 14.39, F 10.48, Cl 20.95, Br 26.21, I 34.53, S 22.91, P 24.87; for each bond, irrespective of type, subtract 6.56. Thus for NH_2COCH_3 the value is (2 x 16.35 + 12.43 + 14.39 + (5 x 8.71) – (8 x 6.56) = 50.59 cm^3 mol^{-1}; the experimental value of its molar volume is 50.86 cm^3 mol^{-1}.

There are five software programs that predict liquid density, namely SPARC (ibmlc2.chem.uga.edu/sparc), ACDLabs (www.acdlabs.com), Molecular Modeling Pro (www.chemsw.com), ProPred (www.capec.kt.dtu.dk) and PREDICT (mwsoftware.com/dragon/). SPARC is freely accessible on line, whereas the others have to be purchased. The ACDLabs website reports a standard error of 0.028 g cm^{-1} for the densities of a test set of 671 liquids. PREDICT is reported to yield errors of <2%. The performance of the others is not known.

It is recommended that one of the software programs or the Abraham and McGowan method (1987) be used for the calculation of liquid density and/or molar volume.

Remaining uncertainty on relative density

All the methods are capable of greater precision than is likely to be required for environmental and human health assessment.

R.7.1.4.4 Conclusions on relative density

Relative density will usually be determined experimentally. The temperature measured in the test should be reported.

Concluding on C&L and chemical safety assessment

Relative density is not used for classification and labelling.

R.7.1.4.5 Integrated testing strategy (ITS) for relative density

The tiered approach to testing (Section R.7.1.1.4) in conjunction with the choice of an appropriate test method represents an integrated testing strategy for this endpoint.

Information on the density of a test material is used to provide the conversion between dynamic and kinematic viscosity (as required for the classification criteria for aspiration hazard).

Examples and case studies on relative density

There are no real examples and case studies. The density of a material decreases with increasing temperature. It is therefore important to measure and report the temperature during the test.

R.7.1.4.6 References for relative density

Abraham M.H. and McGowan J.C. The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. *Chromatographia* (1987) 23, 243-246.

Budavari, S., Ed. (1989) The Merck Index – Encyclopedia of Chemicals, Drugs and Biologicals. Rahway, NJ:Merck and Co., Inc.

Dubois, J.E., and M. Loukianoff (1993) DARC "Logic method" for molal volume prediction. SAR. QSAR *Environ. Res.* 1, 63-75

Girolami, G.S., (1994) A simple "Back of the envelope" method for estimating the densities and molecular volumes of liquids and solids. *J. Chem. Educ*. 71, 962-964.

Grigoras S. A structural approach to calculate physical properties of pure organic substances: the critical temperature, critical volume and related properties. *J. Comput. Chem.* (1990) 11, 493-510.

Howard, P.H., Ed. (1990). Handbook of Fate and Exposure Dta for Organic Chemicals, Vol. I and II, Lewis Publishers, Chelsea, Michigan.

I.U.P.A.C., Recommended Reference Materials for Realization of Physico-Chemical Properties, Pure and Applied Chemistry, 48, 508 (1976).

Kier, L.B., and Hall, L.H., (1976) Molecular Connectivity in Chemistry and Drug research. San Diego, CA. *Academic press*

Kirk-Othmer Encyclopedia of Chemical Technology. 4th ed. Volume 1: New York, NJ,. JohnWiley and Sons, 1991-Present.

Lange's Handbook of Chemistry, Editor, Speight, James, Publ. G. McGraw-Hill.Lide, D.R., Ed. (2005). CRC Handbook of Chemistry and Physics, 86th edition, CRC Press, Taylor and Francis, Boca Raton, FL.

Mobility standards based on the European Classification System criteria of the Bonn Agreement (1983), http//www.bonnagreement.org/eng/html/welcome.html

Nelken L.H. Densities of vapours, liquids and solids. In Lyman W.J., Reehl, W.F. and Rosenblatt D.H. (Eds.), *Handbook of Chemical Property Estimation Methods*. American Chemical Society, Washington DC, 1990, pp. 19.1-19.24.

OECD Guidelines for Testing of Chemicals. Method 109 "Density of Liquids and Solids"

Official Journal of the European Communities L225/1, Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances

Reinhard M, Drefahl A. (1999) Handbook for estimating physico-chemical properties of organic compounds. New York: Wiley; 1999

Riddick, J.A., Bunger, W.B., Sakano, T.K. (1986). Organic Solvents: Physical Properties and Methods of Purification. Techniques of Chemistry. 4th Edition. New York, NY, Wiley-Interscience

Verschueren, K. (1983) Handbook of Environmental Data on Organic Chemicals. Fourth Edition. John Wiley & Sons, Inc., New York.

Xu, L., Wang, H.Y., Su, Q., (1992) A newly proposed molecular topological index for the discrimination of Cis/Trans isomers and for the studies of QSAR/QSPR. Comput. Chem., 16: 187-194.

R.7.1.5 VAPOUR PRESSURE

Vapour pressure is environmentally relevant because:

- the vapour pressure gives an indication of the probability of the phase transitions liquid/gas and solid/gas.
- the vapour pressure, together with the solubility in water, is the major auxiliary variable for calculating the volatility of a substance from an aqueous solution. These provide estimates of air-water partition coefficients K_{AW} or Henry's law constants H ($Pa.m^3/mol$) (see Appendix R.7.1-4 Henry's law constant and evaporation rate), and thus the relative air-water partitioning tendency (e.g. volatilisation from surface water, stripping in the aerator tank of a sewage treatment plant).
- vapour pressure is a significant factor for predicting atmospheric concentrations. Volatilised material is available for airborne transport and as such can give rise to the distribution of a chemical over a wide area and into bodies of water (e.g. in rainfall) far from the site of release.
- the vapour pressure of a substance can furthermore be useful as a basis for deciding whether or not a photochemically induced degradation (in the homogeneous gas phase or in an absorbed phase) is necessary.

Vapour pressure data is required as a pre-requisite for animal as well as environmental studies. It can inform whether a substance may be available for inhalation as a vapour and whether occlusive conditions are necessary for dermal studies (to limit evaporation from skin). Such data will also be used to determine the approach for ecotoxicological testing (e.g. use of closed vessels to help maintain test concentrations) and in the choice of biodegradation test method.

Definition of vapour pressure

The vapour pressure of a substance is defined as the saturation pressure above a solid or liquid substance. At thermodynamic equilibrium, the vapour pressure is only a function of temperature.

The SI unit of pressure which should be used is the Pascal ($Newton/m^2$). Units which have been employed historically, together with their conversion factors, are

1 Torr (mm Hg)	$= 1.333 \times 10^2$ Pa
1 atmosphere (physical)	$= 1.013 \times 10^5$ Pa
1 atmosphere (technical)	$= 9.81 \times 10^4$ Pa
1 bar	$= 10^5$ Pa

The chemical's tendency to partition into the atmosphere is controlled by its vapour pressure, which is essentially the maximum vapour pressure that a pure chemical can exert in the atmosphere. This can be viewed as a kind of *solubility* of the chemical in the atmosphere. The gas law (PV = nRT) applies, where

P = vapour pressure (Pa)	R = gas constant (8.314 $Pa.m^3/mol.K$)
T = absolute temperature (K)	n = number of moles (m/mw)

and allows conversion of the vapour pressure into a solubility C (mol/m^3).

R.7.1.5.1 Information requirements on vapour pressure

The study does not need to be conducted if the melting point is above 300°C. If the melting point is between 200°C and 300°C, a limit value based on measurement or a recognised calculation method is sufficient. Vapour pressure testing is also not required for chemicals with a standard boiling point of <30ºC, as these substances will have vapour pressures above the limit of measurement (i.e. 10^5 Pa).

R.7.1.5.2 Available information on vapour pressure

Testing data on vapour pressure

Many articles summarising and evaluating vapour pressure data have been published. Among these are those of Stull (1947), Driesbach and Strader (1949), Driesbach and Martin (1949) and Zwolinski and Wilhoit (1971). The most extensive are those of Stull, who lists 1200 compounds, and of Jordan (1954).

Published data on vapour pressure

Important sources of vapour pressure are scientific journals or environmental Handbooks. The Handbooks include Verschueren's Handbook of Environmental Data on Organic Chemicals (1983), Howard's Handbook of Environmental Fate and Exposure Data, Vol. I and II (1990), Lide's CRC Handbook of Physics and Chemistry, Lange's Handbook of Chemistry, The Handbook of Vapour Pressures and Heats of Vaporisation of Hydrocarbons and Related Compounds (Zwolinski and Wilhoit, 1971), the Vapour Pressure of Pure Substances (Boublik *et al*., 1984) and the Handbook of the Thermodynamics of Organic Compounds (Stephenson and Malanowski, 1987).

Alternatively, searching on various environmental databases, such as HSDB (http://www.toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB), will provide summaries of chemical and physical properties of substances.

R.7.1.5.3 Evaluation of available information on vapour pressure

Experimental data on vapour pressure

There is no single vapour pressure measurement procedure applicable to the entire range of vapour pressures. Therefore, several methods are recommended (in EU A.4 and OECD TG 104; see Table R.7.1-15) for the measurement of vapour pressures across the range from $<10^{-4}$Pa to 10^5Pa. However, vapour pressure measurement is an area where care must be taken over the selection of a suitable test method, as the techniques are applicable over only certain ranges of vapour pressure (see Table R.7.1-16). If the wrong technique is selected, the result may be deemed invalid and the notification delayed while the test is repeated. The gas saturation method may allow measurements of considerably lower vapour pressure (as low as approximately 10^{-5}Pa). Vapour pressure testing is not required for chemicals with a standard boiling point of <30ºC, as these substances will have vapour pressures above the limit of measurement (i.e. 10^5Pa).

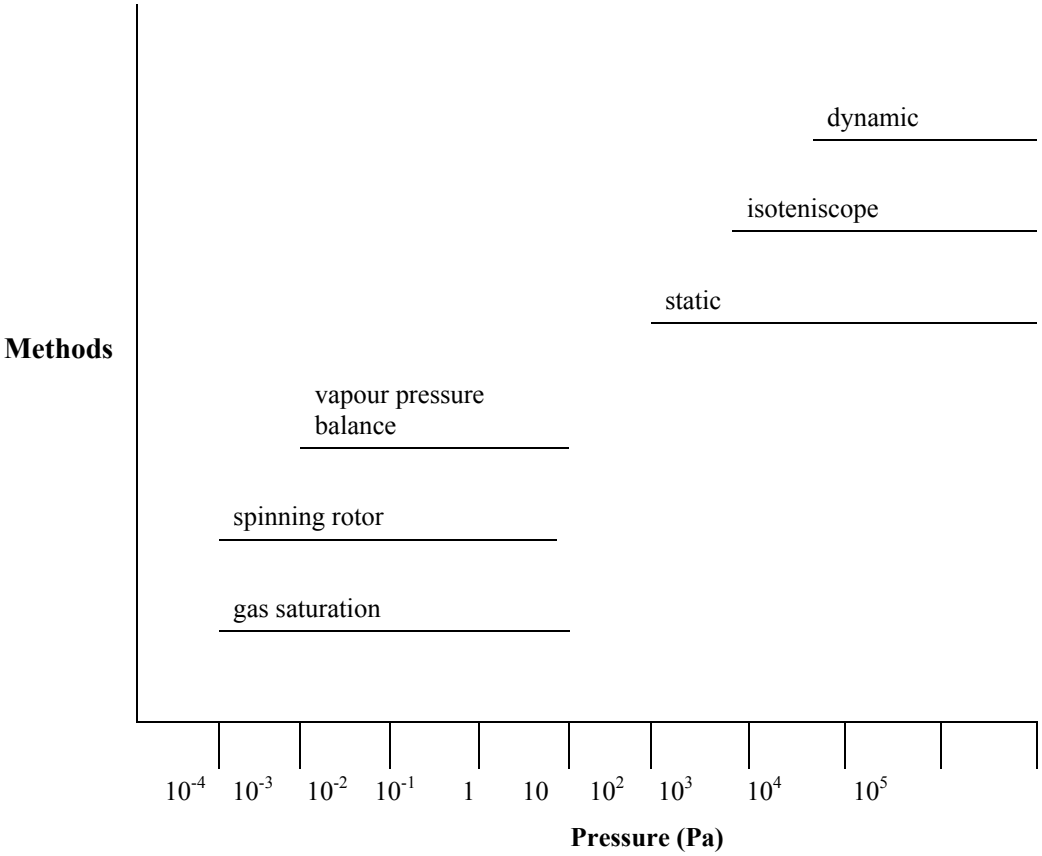
The dynamic method, static method and isoteniscope method can be applied to pure and commercial grade substances although impurities will affect the results. The vapour pressure balance method and the gas saturation method can only be applied to pure substances.

For substances that sublime (e.g. iodine, ammonium chloride), the gas saturation method or vapour pressure balance method should be used.

Table R.7.1-15 Methods for the measurement of vapour pressure

Method	Repeatability/%	Reproducibility/%	Range/Pa
Dynamic This method measures the boiling temperature which pertains to a specified pressure. Recommended for boiling point determination and is useful for that purpose up to 350ºC	up to 25 1 - 5	up to 25 1 - 5	10^3 to 2 x 10^3 2 x 10^3 to 10^5
Static At thermodynamic equilibrium, the vapour pressure established in a closed system is determined at a specified temperature. Suitable for one-component and multicomponent solids and liquids	5 - 10	5 - 10	10 to 10^5
Isoteniscope This standardised method (ASTM-D 2879-75) is also a static method but is not usually suitable for multicomponent systems.	5 - 10	5 - 10	10^2 to 10^5
Vapour pressure balance (effusion methods) The quantity of substance leaving a cell per unit time through an aperture of known size is determined under vacuum conditions in such a way that return of the substance into the cell is negligible (e.g. by measuring the pulse generated on a sensitive balance by a vapour jet or by measuring the weight loss)	5 - 10	up to 50	10^{-3} to 1
Gas Saturation A stream of inert carrier gas is passed over the test substance in such a way that it becomes saturated with its vapour and the vapour is then trapped. Measurement of the amount of material transported by a known amount of carrier gas is used to calculate the vapour pressure at a given temperature.	10 - 30	up to 50	10^{-4} to 1
Spinning Rotor In the spinning rotor gauge, the actual measuring element is a small steel ball which is suspended in a magnetic field and rotates with high speed. The gas pressure is deduced from the pressure-dependent slow down of the steel ball.	10 - 20	_	10^{-4} to 0.5

Table R.7.1-16 Measurement of vapour pressure – ranges of various methods



The vapour pressure using any of the above methods should be determined for at least three temperatures in the range 0º–50ºC. If the chosen method has required measurement at temperatures above this range, the vapour pressure curve (log p versus 1/T) should be extrapolated to these temperatures. Care must be taken when extrapolating over large temperature ranges. Vapour pressures at 20ºC or 25ºC should be reported. This value should preferably be an experimental one, but may be interpolated or extrapolated if necessary.

Reference substances are used for calibration from time to time of the method and to offer the chance to compare results when another method is applied.

There are also convenient indirect methods that are based on evaporation measurement (Dobbs and Cull, 1982) or chromatographic retention times (Hinckley *et al*., 1990).

Non-experimental data for vapour pressure

Grain (Grain, 1982) derived from thermodynamic principles two different estimation methods for vapour pressure that can be respectively applied for compounds that are liquid or gaseous at the temperature of interest (Equ. 1) and for solid and liquid compounds (Eq. 2). The first method is derived from the Antoine equation, which describes the temperature dependence of vapour pressure and it is formalised by the following equation:

$$\ln P_V = K_F (8.75 + R \ln T_b) (T_b - C)^2/(0.97 RT) [1/(Tb - C) - 1/(T - C)]$$

(Equation 7-1)

where:

P_V = vapour pressure [atm] K_F = compound class specific constant

R = gas constant [cal/mol * K] T_b = boiling point [K]

T = environmental temperature [K] $C = -18 + 0.19\ T_b$

The second method makes use of the Watson correlation, which describes the temperature dependence of the heat of vaporisation and it is quantified by this equation:

$$\ln P_V = K_F\ (8.75 + R \ln T_b)/(0.97\ R) * [1 - (3 - 2T^*)^m/T^* - 2m(3 - 2T^*)^{m-1} \ln T^*]$$

(**Equation 7-2**)

where:

P_V = vapour pressure [atm] K_F = compound class specific constant

R = gas constant [cal/mol * K] T_b = boiling point [K]

T = environmental temperature [K] $T^* = T/T_b$

m = empirical factor depending on T^* and the physical state of the compound at the temperature of interest.

Liquids: m = 0.19

Solids: $T^* > 0.6$: m = 0.36

$0.6 > T^* > 0.5$: m = 0.8

$T^* < 0.5$: m = 1.19

The compound specific constant K_F is assumed to describe the polarity of the compound. K_F values are reported for simple compound classes by Grain (Grain, 1982) and its value is varies between 0.97 and 1.23.

For compound classes not included in the tables, a K_F value of 1.06 is recommended. The factor K_F is derived for monofunctional compounds, but it is also applicable for polyfunctional compounds if assuming the respective highest K_F value.

A third method described by Mackay *et al*. (1982) is applicable only for hydrocarbons and halogenated hydrocarbons:

$$1n\ P_V = -(4.4 + \ln T_b)(1.803\ (T_b/T-1) - 0.803 \ln (T_b/T) - 6.8\ (T_m/T-1)$$

where:

P_V = vapour pressure [atm] T_b = boiling point [K]

T_m = melting point [K] T = environmental temperature [K]

The OECD guideline 104 reports that the Watson correlation is applicable over the pressure range from 10^5 Pa to 10^{-5} Pa. It should in any case be pointed out that estimated values for vapour pressure can be subjected to great uncertainty if the computed pressure is lower than 1 Pa, especially when the boiling point has not been experimentally determined (OECD monograph 67). The uncertainty is even greater if the estimated value is used together with water solubility in order to estimate the Henry's Law constant.

The environment monograph 67 of the OECD describes all of the above mentioned methods and the OECD guideline 104 supports the use of the Watson correlation for the calculation of vapour pressure, but does not specifically reject other calculation methods.

The handbook for estimating the physico-chemical properties of organic compounds (Reinhard and Drefahl, 1999) reports another method based on thermodynamic properties and elaborated by Mishra and Yalkowsky that discussed the application of the method of Mackay (Mackay *et al*., 1982):

$$\ln P_v = -\frac{T_m - T}{T}\,(8.5 - 5.0\log\sigma_{sym} + 2.3\log\phi_{flx}) - \frac{T_b - T}{T}\,(10 + 0.08\log\phi_{flx}) + \left(\frac{T_b - T}{T} - \ln\frac{T_b}{T}\right)\,(-6 - 0.9\log\phi_{flx})$$

where:

P_V = vapour pressure [atm] T_b = boiling point [K]

T_s = melting point [K] T = environmental temperature [K]

σ_m = rotational symmetry number ϕ_{flx} = conformational flexibility number

The parameter σ_m is the equal to the number of ways in which the molecule under investigation can be brought in positions that are identical with a reference position. The parameter is the number of reasonable conformations in which the molecule can exist (for further details, see Mishra and Yalkowsky (1991).The equation by Mishra and Yalkowsky give significantly better estimates than the method of Mackay on the same data set (Mishra and Yalkowsky, 1991).

Another methodology that proved to be effective in estimating vapour pressure relies on group contribution approaches. Several models using this strategy have been proposed (Reinhard and Drefahl, 1999; see Table R.7.1-17).

Table R.7.1-17 Group contribution approach and vapour pressure

Compounds	Authors	Methodology	Statistics
Alkyl aromatic compounds	Amidon and Anik	Group contribution approach	Standard error … 1.1 kJ on the estimation for the free energy of vaporisation
Mono-, di-, tri- and tetra substituted	Hoshino *et al*.	Group contribution approach	Average error 3.7% Max. Error 30.9%
Perfluorinated saturated hydrocarbons	Kelly *et al*.	Group contribution approach	Arithmetic mean deviation < 0.5%

Numerous other models are available for the estimation of vapour pressure, and Grain (1990), Schwarzenbach *et al* (1993), Delle Site (1996), Sage and Sage (2000) and Dearden (2003) have reviewed many of these. The descriptors used in vapour pressure QSPRs include physico-chemical, structural and topological descriptors, and group contributions. Katritzky *et al* (1998) used 4

CODESSA descriptors to model the vapour pressure (in atmospheres at 25°C) of 411 diverse organic chemicals, with $r^2 = 0.949$ standard error = 0.331 log unit.

Liang and Gallagher (1998) used polarisability and 7 structural descriptors to model the vapour pressure of 479 diverse organic chemicals, using both multiple linear regression and an artificial neural network. There was little difference between the two methods with MLR giving a standard error of 0.534 log unit and ANN yielding 0.522 log unit.

Tu (1994) used a group contribution method to model the vapour pressure of 1410 diverse organic chemicals. Using 81 group contributions, 2 hydrogen bonding terms and melting point he obtained a standard error of 0.36 log unit.

The vapour pressures of 352 hydrocarbons and halohydrocarbons were modelled by Goll and Jurs (1999), using 7 of their ADAPT descriptors. Vapour pressure was recorded in pascals, and the data covered the log VP range –1.016 to +6.65; they obtained $r^2 = 0.983$ and RMS error = 0.186 log unit.

The ADAPT descriptors are difficult to interpret, but have been found to give good correlations of a number of physico-chemical properties. The very low standard error reflects the fact that there is little chemical diversity within the compounds used.

A number of studies (Andreev *et al* 1994, Kühne *et al* 1997, Yaffe & Cohen 2001) allow of the estimation of vapour pressures over a range of temperatures.

There are several commercially available software programs that will calculate vapour pressure; one of them (ACDLabs) will allow the calculation of vapour pressure over a temperature range. Using a 100-compound test set of organic chemicals with vapour pressures measured at 25°C, Dearden (2003) compared the performance of four software programs that calculate log (vapour pressure); see Table R.7.1-18.

Table R.7.1-18 Software programs that calculate vapour pressure

Software	Website	Availability	Mean absolute error (log unit)
SPARC	http://ibmlc2.chem.uga.edu/sparc	Free on line	0.105
ACDLabs	Www.acdlabs.com	Purchase	0.107
MPBPVP	www.epa.gov/oppt/exposure/ docs/episuitedl.htm	Freely Downloadable	0.285
Molecular Modelling Pro	www.chemsw.com	Purchase	0.573

The programs can operate in batch mode, except for SPARC. The ACDLabs result was determined on only 42 compounds; 46 test set compounds that were used in the ACDLabs training set were deleted, and in addition the ACDLabs software did not give a vapour pressure at 25°C for 18 very volatile compounds. Other software programs that calculate vapour pressure, but were not tested by Dearden (2003), are Absolv-2, ASTER, ChemProp, ProPred and PREDICT (mwsoftware. com/dragon/); its prediction errors are reported to be 2–5%, depending on the method of calculation. The performances of the other software are not known.

It is recommended that either SPARC, MPBPVP or ACDLabs software be used for the calculation of vapour pressure. Chemical structures are inputted into the software as SMILES strings. The

SMILES approach is simple to follow, and a tutorial can be found at www.daylight.com/smiles/smiles-intro.html.

Remaining uncertainty on vapour pressure

Vapour pressure, being a specific property of a compound, is widely used for practical calculations in physical chemistry and chemical engineering. Any error in vapour pressure measurement usually relates to the determination of pressure rather than of temperature. The greatest difficulty and uncertainty arises when determining the vapour pressures of chemicals with low volatility, i.e. those with vapour pressures below 1.0 Pa. The major cause of inaccurate vapour pressure measurements is due to the presence of impurities in the sample which have a higher vapour pressure than that of the major component.

Experimental vapour pressures usually do not cover the full pressure-temperature range. Also vapour pressure may, for some substances, change considerably according to temperature even within a temperature range of only 10°C. In such cases an estimated vapour pressure at the relevant temperature should be obtained either from interpolation from vapour pressures at 10°C and 30°C or by use of extrapolation methods (Schwartzenbach *et al*., 1993).

As vapour pressure is related to boiling point, it is sensible to check boiling point and vapour pressure results for consistency. For example, a high melting point-point solid is unlikely to have a high vapour pressure at ambient temperatures. The temperature at which the extrapolated $\log_{10}$ P versus 1/T graph reaches atmospheric pressure (or reduced pressure) should approximately correlate with the measured boiling point. If they do not then either the vapour pressure or boiling point determinations (or both) may be incorrect.

Care should be taken when handling a substance with a high vapour pressure to minimise any vapour losses during any physico-chemical, toxicology and ecotoxicology tests, e.g. the use of closed/covered vessels during ecotoxicity testing. For many volatile substances, nominal concentrations are often not appropriate and additional information may be necessary in order to verify the actual exposure concentrations. This will affect the route of exposure for sub-acute animal toxicity tests and the choice of dermal or inhalation exposure route for acute animal toxicity tests, as well as the choice of biodegradation test method.

R.7.1.5.4 Conclusions on vapour pressure

For single organic substances experimentally derived vapour pressures, or values which have been evaluated in reviews and assigned *recommended values*, are preferred over predicted values. For mixtures of components (e.g. resins and polymer mixes) the experimentally derived vapour pressure value will represent the component with the highest vapour pressure; in this case prediction methods may be used to generate vapour pressure data on the individual components.

Concluding on C&L and Chemical Safety Assessment

Vapour pressure is not used as a Classification and Labelling criterion or to define PBT properties. Vapour pressure data may be useful in indicating the physical state of a compound, e.g. a gas (in Carriage of Dangerous Goods) is defined as a substance at 50°C which has a vapour pressure of >300 kPa or is completely gaseous at a standard pressure of 101.3 kPa.

In terms of risk assessment, estimates of emission factors (the fractions released) are provided in the Guidance Document for various industries. The higher the vapour pressure (in Pa) or lower the boiling point (°C) the greater the emission factors for air. Regional exposure concentrations in all

environmental compartments are computed using multimedia fate models based on the fugacity concept. Vapour pressure is a key parameter in determining the environmental fate and behaviour, such as calculations of Predicted Environmental Concentrations (PECs) for environmental risk assessment, vapour exposure for human health risk assessment and determining atmospheric behaviour as for exposure of man via the environment.

R.7.1.5.5 Integrated testing strategy (ITS) for vapour pressure

It is important to have knowledge of both the melting and boiling points prior to measuring the vapour pressure, to ensure that an appropriate temperature range is chosen, i.e. one in which there is no phase transition and the substance is stable. The result should correlate with the boiling point.

Ideally, two vapour pressures at different temperatures in the range 0–50°C should be reported. The purity should also be stated. Comments should indicate if any physical or chemical transition occurred. If a transition occurs, the temperature of transition at atmospheric pressure should be stated along with the vapour pressure 10° and 20°C above and below this temperature (except for sublimation). The nature of the transition should also be recorded, together with the effects of any impurities. Where measurements are made at elevated temperatures, a value for the vapour pressure at 25°C should be determined graphically.

Melting and boiling point results must also be considered when selecting the temperature range over which vapour pressure measurements are made, to ensure that no phase transitions occur during the determination.

Examples and case studies on vapour pressure

The vapour pressure of a chemical provides considerable insight into the transport and partitioning of a chemical in the environment and in commercial settings. The volatility of a pure chemical is dependent upon the vapour pressure, and volatilisation from water is dependent upon the vapour pressure and water solubility. The form in which a chemical will be found in the atmosphere is dependent upon the vapour pressure. Water surface condition and wind speed will have a significant effect on any evaporation of chemicals.

Table R.7.1-19 Common Examples

Chemical	Vapour Pressure at 25°C	
	in Pa	in mm Hg
Water	3200	24
Acetone	30800	232
Dichloromethane	58100	437
Diethyl ether	716009	538
Benzene	127004	95
Toluene	3800	29
Xylene	1100	8.2
Phenol	55	0.41
n-Pentane	68300	513
n-Hexane	20200	152
n-Heptane	6100	46
n-Decane	170	1.3
n-Dodecane	16	0.12
n-Tridecane	5	0.048
n-Pentadecane	0.5	0.0038
n-Heptadecane	0.03	0.00022
Eicosane	0.000567	0.0000042

Chemicals with relatively low vapour pressures, high adsorptivity onto solids or high solubility in water are less likely to vaporise and become airborne than chemicals with high vapour pressures or less affinity for solution in water or adsorption to solids and sediments. In addition, chemicals that are likely to be gases at ambient temperatures and that have low water solubility and low adsorptive tendencies are less likely to transport and persist in soils and water. Such chemicals are less likely to biodegrade or hydrolyse, but are prime candidates for photolysis.

R.7.1.5.6 References on vapour pressure

Andreev N.N, Kuznetsov S.E, Storozhenko S.Y. (1994) Prediction of vapour pressure and boiling points of aliphatic compounds. *Mendeleev Commun.* 173-174.

Amidon, G.L., and Anik, S.T. (1981) Application of the surface area approach to the correlation and estimation of aqueous solubility and vapour pressure: alkyl aromatic hydrocarbons. *J. Chem. Eng. Data* 26. 28-33

Boublik, T., Fried, V, Hala, E. (1984). The Vapour Pressure of Pure Substances, 2nd revised edition, Elsevier, Amsterdam

Grain C.F., (1982) Handbook of chemical property estimation methods. New York, Mc Graw-Hill

Lide, D.R., Ed. (2005). CRC Handbook of Chemistry and Physics, 86th edition, CRC Press, Taylor and Francis, Boca Raton, FL.

Delle Site A. (1996) The vapour pressure of environmentally significant organic chemicals: a review of methods and data at ambient temperature. *J. Phys. Chem. Ref. Data* 26:157-93.

Dearden JC. (2003) Quantitative structure-property relationships for prediction of boiling point, vapour pressure, and melting point. *Environ Toxicol Chem* 22(8):1696-709.

Dobbs, A.J., Cull, M.R. (1982). Volatilization of chemical relative loss rates and the estimation of vapour pressures. *Environ. Toxicol*. Chem. 8, 499-512

Driesbach, R.R. and Martin, R.A. (1949). *Ing. Eng. Chem*., 41, 2875

Driesbach, R.R. and Shrader, S.A. (1949). *Ing. Eng. Chem*., 41, 2879EC Method A4. "Vapour Pressure", Dir 92/69/EEC, Official Journal of the European Communities, O.J. L383 A)

Goll ES, Jurs PC. (1999) Prediction of vapour pressures of hydrocarbons and halohydrocarbons from molecular structure with a computational neural network model. J. *Chem. Inf. Comput. Sci.* 39:1081-9.

Hinckley, D.A., Bidleman, T.F., Foreman, W.T. (1990). Determination of vapour pressures for nonpolar and semipolar organic compounds from gas chromatographic retention data. *J. Chem. Eng. Data*, 35, 232-237.

Hishino, D., *et al*., (1985) Prediction of vapour pressures for substituted benzenes by a group-contribution method. *Ind. Eng. Chem. Fundam*. 24, 112-114

Howard, P.H., Ed. (1990). Handbook of Fate and Exposure Data for Organic Chemicals, Vol. I and II, Lewis Publishers, Chelsea, Michigan.

Jordan, T.E (1954). Vapour pressure of organic compounds, Interscience, New York

Katritzky AR, Y. W, Sild S, Tamm T, Karelson M. (1998) QSPR studies on vapour pressure, aqueous solubility, and the prediction of water-air partition coefficients. *J. Chem. Inf. Comput. Sci.* 38:720-5.

Kühne R, Ebert RU, Schüürmann G. (1997) Estimation of vapour pressures for hydrocarbons and halogenated hydrocarbons from chemical structure by a neural network. *Chemosphere* 34:671-86.

Lange's Handbook of Chemistry, Editor, Speight, James, Publ. G. McGraw-Hill.

Liang CK, Gallagher DA. (1998) QSPR prediction of vapour pressure from solely theoretically-derived descriptors. *J. Chem. Inf. Comput. Sci.* 38:321-4.

Mackay D, Bobra A, Chan W, Shiu WY. (1982) Vapour pressure correlation for Low-Volatility Environmental Chemicals. *Environ. Sci. Technol.* 16:645-9.

Mishra DS, Yalkowsky SH. Estimation of vapour pressure of some organic compounds. *Ind. Eng. Chem. Res.* 1991;30:1609-12.

OECD Guidelines for Testing of Chemicals, Method 104 "Vapour Pressure Curve"

PREDICT (2002). Chemical Thermodynamic & Transport Properties of Interest to Chemical Engineers and chemists. Free download http://www.mwsoftware.com/dragon/manual.pdf. In; 2002.

Reinhard M, Drefahl A. Handbook for estimating physico-chemical properties of organic compounds. New York: Wiley; 1999.

Sage M.L, Sage G.W.(2000) Handbook of Property Estimation Methods for Chemicals. Boca Raton, FL: Lewis;

Stephenson, R.M., Malanowski, A. (1987). Handbook of the Thermodynamics of Organic Compounds. Elsevier, New York

Schwartzenbach, R.P., Gswend, P.M., Imboden, D.M. (1993). Environmental Organic Chemistry. John Wiley and Sons.

Stull, D.R. (1947). *J. Amer. Chem. Soc*., 59, 2796.

Tu CH.(1994)Group-contribution method for the estimation of vapour pressures. *Fluid Phase Equilibr* 99:105-20.

Verschueren, K. (1983) Handbook of Environmental Data on Organic Chemicals. Fourth Edition. John Wiley & Sons, Inc., New York.

Yaffe D, Cohen, Y (2001) Neural network based temperature-dependent quantitative structure property relationships (QSPRs) for predicting vapour pressure of hydrocarbons. *J. Chem. Inf. Comput. Sci.* 41:463-77.

Zwolinski, B.J., Wilhoit, R.C. (1971). Handbook of Vapour Pressures and Heats of Vaporisation of Hydrocarbons and Related Compounds. API-44, TRC Publication No. 101, Texas A&M University, College Station, TX.

R.7.1.6 SURFACE TENSION

When a liquid is in contact with air, solid or an immiscible liquid, the energetics of the surface will minimise the area of the interface. Surface tension is the measure of the force that tends to diminish the area of the interface. Compounds known as surface-active agents (surfactants) reduce the surface tension of a liquid, the interfacial tension between two liquids or between a liquid and a solid. Surfactant molecules typically contain both polar (hydrophilic) and non-polar (lipophilic) moieties[11]. They are also called *amphiphilic* compounds. Such a chemical structure is used as an initial indicator (*alert*) of surface-active properties (Porter, 1994; Hartland, 2004; Rosen, 2004). Another alert is the occurrence of foaming of aqueous solutions. Experimental surface tension determinations are performed to confirm the surface-active properties of the molecule.

Surface tension measurements of aqueous solutions are significant since decreasing the surface tension of water may impact properties of the solution and other physico-chemical measurements. Surface active agents may co-solubilise other less soluble chemicals in the environment. Decreasing the surface tension of water will increase dispersion of hydrophobic materials, increase wetting of surfaces, stabilise aerosol formation and increase bubble formation and foaming. Some surfactants may show irritant- to corrosive effects on human and animal skin, and may increase dermal penetration of other chemicals.

Partition coefficient measurements may be impacted since decreased surface tension of the water will increase solubilisation and emulsification of the hydrophobic phase. OECD TG 107 (EU A.8) for partition Coefficient Determination by Shake Flask Method specifically excludes surface active materials.

Chromatographic properties used to estimate other physico-chemical parameters may be altered since solid:liquid interactions introduce secondary partitioning effects.

Adsorption of the surface-active agent at the air:water interface may alter transport across the interface and impact the measurement of Henry's Law constants.

Surface tension data should therefore be collected for surfactants to verify the suitability of other physical chemical testing methods.

Surface active substances have a higher irritant or corrosive effect than substances with the same solubility characteristics. Skin permeability (dermal uptake) of a substance may be enhanced.

Definition of surface tension

The surface tension of a liquid is the surface free energy per unit area and corresponds to the minimum work required to expand the surface by a unit area (Sprycha & Krishnan, 2000). Appropriate SI units for surface tension are N/m or 10^{-3}N/m. 1 N/m is equivalent to 10^3 dynes/cm in historical units. Pure water has a surface tension of 72.75 mN/m at 20°C (CRC, 2006).

When the surface tension of a liquid is less than or equal to the *critical surface tension* of a solid surface, the liquid will freely spread or *wet* on the solid surface.

Surface-active agents are compounds that reduce the surface tension of a liquid, the interfacial tension between two liquids or between a liquid and a solid. Surfactants molecules have a structure with a hydrophilic headgroup (e.g. $-SO_3^-$, $-SO_4^-$, $-(OCH_2CH_2)_n$, $-N^+(CH_3)_3$) and a hydrophobic tail

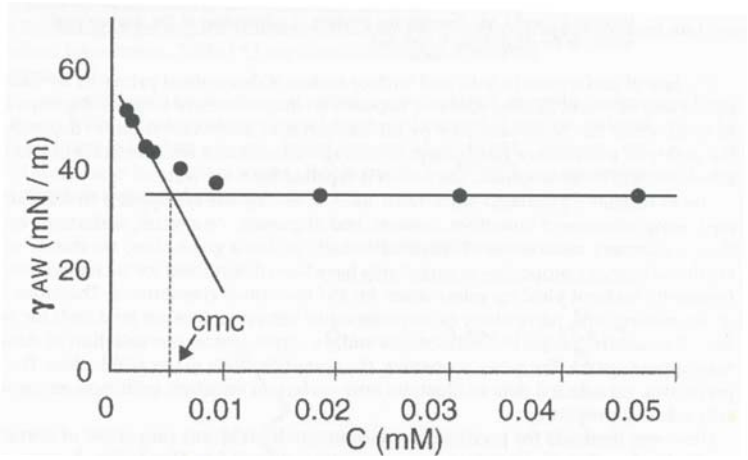
[11] Exception to this rule are some polymers (like polyols) which can have weak surface active properties

(e.g. a linear or branched alkyl chain). As such, presence or absence of surface active properties can often be deduced from the molecule's structure (Schwartz & Perry, 1978).

NB: According to EU Directive 648/2004 (CEC, 2004): *Surfactant* means any organic substance and/or preparation used in detergents, which has surface-active properties and which consists of one or more hydrophilic and one or more hydrophobic groups of such a nature and size that it is capable of reducing the surface tension of water, and of forming spreading or adsorption monolayers at the water-air interface, and of forming emulsions and/or microemulsions and/or micelles, and of adsorption at water-solid interfaces.

The surface tension of a solution of surfactant generally decreases with increasing concentration until the *critical micelle concentration* (CMC) is achieved (Figure R.7.1-3). At concentrations higher than the CMC, the surface tension remains essentially unchanged. In general, for surfactant concentrations below the CMC, the surface tension is inversely related to the logarithm of the molar concentration.

Figure R.7.1-3 Idealised curve of the surface tension (γ) against surfactant concentration (C)



(from: Tolls & Sijm, 2000)

A *tensiometer* measures the surface tension of a liquid by measuring the force required to increase the area of the interface or to separate a ring or plate from the interface.

The *Hydrophilic-Lipophilic Balance* (HLB) of a surfactant is a crude but simple measure of the degree to which it is hydrophilic or lipophilic. It is determined by calculating values for the different moieties of the molecule, as described by Griffin (1949). It can serve as an alert for surface tension lowering activity. Other methods have been suggested, but will not be discussed here.

HLB = 20 * Mh/M,

Where Mh is the molecular mass of the hydrophilic portion of the molecule, and M is the total molecular mass. The HLB value provides an indication of the surfactant properties:

- HLB from 3-6 indicates an W/O emulsifier
- HLB from 7-9 indicates a wetting agent

- HLB from 8-12 indicates an O/W emulsifier
- HLB from 12-15 indicates a typical detergent surfactant
- HLB of 15-20 indicates a solubiliser of hydrotrope.

R.7.1.6.1 Information Requirements on surface tension

The surface tension study needs only be conducted if:

- based on structure, surface activity is expected or can be predicted; or
- surface activity is a desired property of the material.

Concentrations should be consistent with the OECD guidelines (OECD 1995) and an estimate of aqueous solubility is required prior to testing. If the water solubility is below 1 mg/l at 20°C the test does not need to be conducted.

It sometimes occurs that chemical test samples contain surface active impurities, which may bias the measurement results.

R.7.1.6.2 Available Information on surface tension

Testing data on surface tension

ISO Standard 304-1985 (ISO, 1985) and more recently EN 14370:2004 (EN, 2004), ASTM D1331 (ASTM, 2001) and OECD TG 115 (EU A.5) standards describe methods for measuring surface tension of aqueous solutions. All of the methods involve measuring the force required to vertically lift an object of well-defined geometry (plate, stirrup or ring) from the interface. All measurements are performed using either a manual or an automated tensiometer capable of measuring the force with 0.1mN/m readability. Alternate methods such as capillary rise, drop weight, pendant drop and maximum bubble pressure techniques (Sprycha & Krishnan, 2000) have not achieved a comparable level of standardisation and are less preferred (NB: reference is made to the general ITS on data quality evaluation to assess use of non-standard data). Guidelines for a harmonised approach to surface tension measurements of aqueous solutions based on the ring geometry have been adopted by OECD (1995).

Surface tension of a solution depends on the concentration of the surface-active agent, temperature of the system, and equilibration time of the solution. For large surfactant molecules present in solution at concentrations below the critical micelle concentration, equilibration times may vary from minutes to hours. Repeated measurements with varied equilibration times are appropriate to assess the time dependence of the measurement. Measurements should be repeated over a time period until a constant surface tension is reached.

Surface tension measurements require a test material that is stable against hydrolysis during the experiment period and soluble in water at concentrations of >1mg/L. Measurements should be performed on a solution at either 90% of the solubility limit or 1 g/L, whichever is smaller (OECD, 1995). Dynamic viscosity of the solution to be tested should not exceed 200mPa.s (remark: viscosity information may not be available at low tonnage levels).

Tensiometers require periodic calibration and specific correction factors depending on the geometry of the ring, stirrup or plate. Procedures for calibration and geometry-specific apparatus correction factors are described in detail in the respective guidelines.

Surface tension measurements require extreme cleanliness of the apparatus, accurate temperature control and access to high purity, particle-free water. Residual material must be removed from all equipment between tests using a combination of acid cleaning, organic solvents and heating followed by extensive rinsing with purified water. Obtaining accurate and precise measurements requires attention to details described in the procedures to avoid errors introduced by buoyancy, rapid separation of the test apparatus from the interface and contamination.

Published data on surface tension

Surface tension measurements have been published in peer reviewed literature for over 80 years. A comprehensive review of critical micelle concentrations for surfactants compiled by Mukerjee & Mysels (1971) provides an excellent reference to historical data.

R.7.1.6.3 Evaluation of available information on surface tension

Experimental data on surface tension

Available data should be evaluated based on the composition and concentration of the test chemical and method used to perform the test. One should select those data that have documented the purity of the test material, demonstrate that solution equilibrium was achieved and that the test was performed in the concentration range of 90% of saturation, but less than 1 g/L.

Highest quality measurements include various equilibration times demonstrating the test solution has reached equilibrium. Availability of measurements of reference materials performed by the same operator on the same equipment will verify the accuracy and precision of the data. Assuming all other documentation is comparable, data generated using the ring, stirrup or plate method is preferred over alternate methods.

Non-Experimental data on surface tension

At present, there are no QSPR/QSAR tools available for accurately predicting surface tension of *aqueous solutions* for a broad class of molecules. Therefore the property will need to be experimentally determined.

Remaining uncertainty on surface tension

For the measurement of surface tension the Ring- or Plate tensiometer methods are preferred. The error on the measurement is in the order of 0.1–0.3 mN/m. Use of the standard protocols and GLP procedures are recommended.

Surface active impurities in chemicals may in some cases lead to false-positive surface tension measurements.

Difficult to test substances: this is not generally a problem, but sufficient attention should be given to ensure that solubility, volatility, viscosity, stability issues and potential chemical impurities do not significantly disturb the measurements.

R.7.1.6.4 Conclusions on surface tension

Testing for surface active properties is triggered by structural alerts (distinct polar and non-polar parts of the molecule, eventually combined with calculation of the HLB value) and/or by the observation of foaming, and will therefore be needed for a small fraction of chemicals only. There are no QSARs to be considered for aqueous solutions. Several testing methods exist, of which the plate- and ring tensiometer methods are most commonly used and standardised. Surface activity may influence the approach for testing other physico-chemical properties, such as log K_{ow}. Surface active agents have environmental relevance since some of them are used in very high volumes in various industrial and domestic applications. Their preference to concentrate at interfaces influences their environmental behaviour and partitioning.

Concluding on C&L and Chemical Safety Assessment

Surface tension is not used as a classification & labelling criterion, to define PBT properties, or as a specific property in chemical safety assessment.

The surface tension measurements can be used to provide guidance as to whether a chemical would be considered a surfactant under EU Directive 648/2004 (CEC, 2004).

R.7.1.6.5 Integrated testing strategy (ITS) for surface tension

Testing for surface tension is to be performed only for those substances that have structural alerts for surface tension reducing properties (i.e. presence of both hydrophilic and hydrophobic parts in the molecule/foaming), and/or if the molecule was designed to be used as a surfactant. Surface tension is typically measured in the first tier of physico-chemical properties testing. It requires information about solubility, hydrolytic stability, and possible impurities in the sample. Information about the viscosity is also desirable, but may not always be available.

The ITS for surface tension is straightforward (Figure R.7.1-4). Since there are no non-testing methods, surface tension for aqueous solutions of chemicals with designed or expected surface active properties will be experimentally determined.

Examples and case studies on surface tension

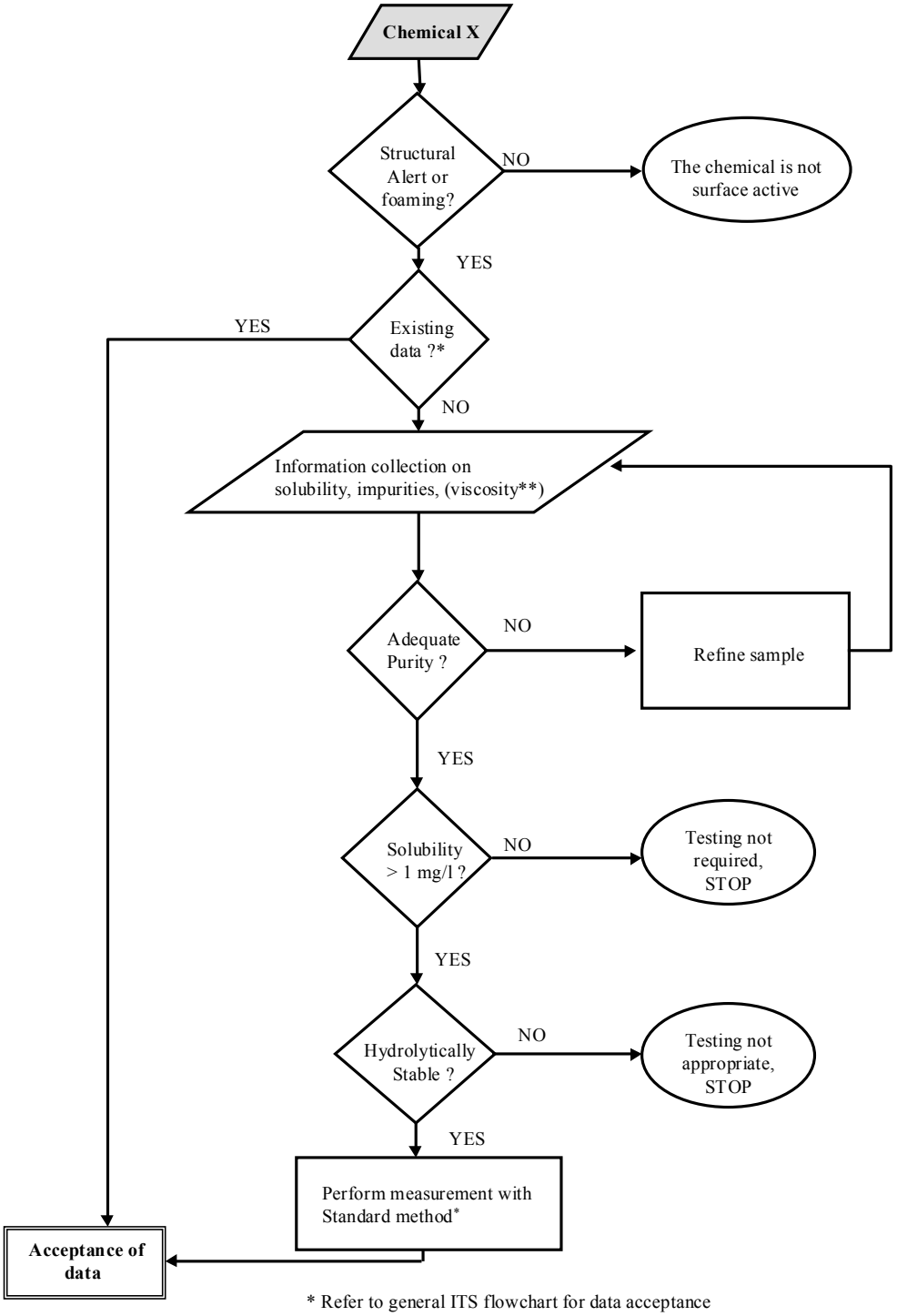
Sodium Dodecyl Sulfate:

The publication by Elworthy and Mysels (1966) provides an example of surface tension measurements for aqueous solutions of sodium dodecyl sulfate (SDS) where a plate method was used to measure the surface tension of sodium dodecyl sulfate with various degrees of purity. The solubility of SDS is well above 1 mg/l. Increasing the concentration of SDS from $5x10^{-3}$ M to the critical micelle concentration ($8.2\ x10^{-3}$ M) decreases the solution surface tension from 49 to approximately 39 mN/m. At concentrations above the CMC, the change in surface tension was less significant, decreasing from approximately 39mN/m to 37mN/m at $6x10^{-2}$M.

Considering the high solubility of sodium dodecyl sulfate, current guidelines would test the material at a concentration of 1 g/L, corresponding to $3.47\ x\ 10^{-3}$M (MW = 288 g/M). This concentration is below those tested by Elworthy and Mysels. Based on the observed linear decrease in surface tension with the logarithm of increasing surfactant concentration and the known surface tension of pure water, one would predict a surface tension measurement of approximately 50mN/m at 25°C for

a 1g/L solution. Using a graphical extrapolation of the data in Elworthy, the value is slightly higher at approximately 55mM/m. This difference is comparable to the variability reported for different samples of sodium dodecyl sulfate and reflects the importance of both purity of the material and strong dependence of surface tension on solution concentration below the CMC.

Figure R.7.1-4 Integrated Testing Strategy for surface tension



* Refer to general ITS flowchart for data acceptance

** Viscosity data may not be available at low tonnage triggers

R.7.1.6.6 References on surface tension

ASTM D1331-89 (2001). Test methods for surface and interfacial tension of solutions of surface active agents.

ASTM D459-00. Standard Terminology Relating to Soaps and Other Detergents.

CEC (2004). Regulation (EC) No 648/2004 of the European Parliament and of the Council of 31 March 2004 on detergents. O.J. EU L104, 1-34.

CRC (2006). "Properties of Water in the Range 0 – 100°C", in CRC Handbook of Chemistry and Physics, 86th Edition, Taylor and Francis, Boca Raton, FL, 2006.

Elworthy, P.E. and K.J. Mysels (1966). The surface tension of sodium dodecyl sulfate solutions and the phase separation model of micelle formation. J. Colloid Interface Sci. 21, 331-347.

EN (2004). I.S. EN 14370:2004. Surface active agents – Determination of surface tension.

Griffin, W.C. (1949). Classification of surface active agents by HLB. J. Soc. Cosmetic Chemists, 1, 311.

Hartland, S. (2004) Surface and Interfacial Tension - Measurement, Theory and Applications Surfactant Science Series volume 199.

I.S. EN 14370:2004 (2004). Surface active agents – Determination of surface tension.

ISO (1985). ISO-304-1985 Surface active agents – Determination of surface tension by drawing up liquid films.

Mukerjee, P. and K.J. Mysels (1971). Critical Micelle Concentrations of Aqueous Surfactant Systems, NSRDS-NBS 36, U.S. Department of Commerce, Washington DC 1971.

OECD (1995). OECD Guidelines for Testing of Chemicals, Method 115: Surface tension of aqueous solutions, OECD, Paris. .

Porter, M.R. (1994). Handbook of Surfactants. Blackie Academic, London.

Rosen, M.J. (2004) Surfactants and Interfacial Phenomena 3rd Ed. Wiley Interscience, New York, NY USA.

Stanton D.T. and Jurs P.C. Computer-assisted study of the relationship between molecular structure and surface tension of organic compounds. *J. Chem. Inf. Comput. Sci.* (1992) 32, 109-115.

Tolls, J. and Sijm, D.T.H.M. (2000). Estimating the properties of surface active chemicals. In: R.S. Boethling and D. Mackay, eds. Handbook of Environmental and Health Sciences; Property Estimation Methods for Chemicals.

R.7.1.7 WATER SOLUBILITY

Water solubility is a significant parameter for a number of reasons:

- the mobility of a test substance is largely determined by its solubility in water. In general, highly soluble chemicals are more likely to be distributed by the hydrological cycle.
- it can affect adsorption and desorption on soils and volatility from aquatic systems. Substances that are highly water soluble are more likely to desorb from soils and less likely to volatilise from water. Water solubility can also affect possible transformations by hydrolysis, photolysis, oxidation and reduction.
- water soluble substances gain access to humans and other living organisms
- knowledge of the water solubility is a prerequisite for setting up test conditions for a range of fate (e.g. biodegradation, bioaccumulation) and effects studies.
- it is also used to derive other environmental parameters, such as K_{ow}, K_{oc} and Henry's Law Constant.

Definition of water solubility

The solubility of a substance in water is specified by the saturation mass concentration of the substance in water at a given temperature. The solubility in water is specified in units of mass per volume of solution. The SI unit is kg/m^3 (grams per litre may also be used).

Mixtures of organic compounds, e.g. petroleum substances, behave differently from their single constituent compounds when brought into contact with water. Petroleum substances are typically hydrophobic and exhibit low solubility in water. However, reflecting the range of structures, constituent hydrocarbons will exhibit a wide range of water solubility. When adding incremental amounts of a complex petroleum substance to water, a point will be reached where the solubility limit of the least soluble component is exceeded and the remaining components will partition between the water and the undissolved hydrocarbon phases. Consequently, the composition of the total dissolved hydrocarbons will be different from the composition of the parent substance. Water solubility measurements for these substances are loading rate dependent due to their complex composition. This water solubility behaviour impacts on both the conduct and interpretation of aquatic toxicity tests for these complex substances. The complex composition, and generally low water solubility, impact on the choice and conduct of biodegradation studies.

The same testing issues also apply to inorganic compounds. Solubility of metal compounds will depend on the species, the particle size and the characteristics of the aqueous medium. As with petroleum products, the solubility properties of a given complex substance can differ substantially from the individual constituents. Further guidance for inorganic substances is detailed in Section R.7.1.7.6.

So the above definition for solubility of a single substance in water is not applicable to substances which are multi-component, i.e. complex substances. The term *complex substance* refers to a mixture of compounds produced by a reaction process or by deliberate mixing of various compounds. The usually accepted meaning of *solubility* in such cases is *the composition of the aqueous solution formed at equilibrium under a defined set of conditions*. Temperature and the amount of substance added per unit volume of water (i.e. the *loading*) are the main factors to consider. It may not always be possible to establish that equilibrium of all components has been achieved; in these cases, time and type of agitation of the test vessels must also be described.

It should be noted that there are various ways that aqueous solubilities can be reported: in pure, usually distilled, water, at a specified pH, at a specified ionic strength, as the undissociated species (intrinsic solubility), or in the presence of other solvents or solutes. Solubilities are also reported in different units, for example g/100 ml, mole/litre, mole fraction. The use of mole/litre is recommended, as this provides a good basis for comparison. However, most text books and QSARs quote solubility data as g/l or mg/l.

R.7.1.7.1 Information requirements on water solubility

The study does not need to be conducted if:

- the substance is hydrolytically unstable at pH 4, 7 and 9 (half-life less than 12 hours);
- the substance is readily oxidisable in water;
- the substance is flammable in contact with water.

If the substance appears *insoluble* in water, a limit test up to the detection limit of the analytical method shall be performed.

For ionising substances, the pH-dependence of the water solubility should be known. At least the pH of the test water needs to be identified. In the context of marine risk assessment, when the pK_a is close to 8 it may be necessary to obtain realistic measurements using seawater.

For volatile compounds, it can be useful to have information on the vapour pressure.

R.7.1.7.2 Available information on water solubility

It is useful to have preliminary information on the structural formula, the vapour pressure, the dissociation constant and the hydrolysis (as a function of pH) of the substance to perform the test. Whenever feasible, an estimation of the water solubility of pure substances (or components of a mixture) should first be made using estimation programmes or from read-across from structurally similar compounds. A *preliminary test* may then be carried out to ascertain/confirm which of the particular methods is suitable, and to determine the amount of test substance and aqueous solution required for the flask method.

The three properties, solubility, hydrolytic stability and acid dissociation constant (see Section R.7.1.17) are inter-related. It is not possible to measure any of these without some knowledge of the other two. A schematic way of determining which solubility measurements should be performed for a given substance is presented in Figure R.7.1-5. For example, if the test substance is an acid or base with a dissociation constant (pKa) value between 3 and 10, it is advisable to use pH buffers in the water to enable the water solubility of the ionised and/or non-ionised (neutral) forms of acids and bases to be determined over an environmentally relevant range (e.g. pH 4-9). For many substances, however, knowledge of the composition may be enough, e.g. it may be obvious to the chemist that no hydrolytically unstable or ionisable groups are present. Conversely, if the half-life is less than 12 hours then solubility studies are not realistic. If it is not intended to perform a full hydrolytic stability test, then procedures for performing a reduced test of stability should be included as part of the solubility test.

Testing of substances for hydrolysis is relevant to their persistence. Hydrolysis is one of the most common reactions controlling abiotic degradation and is therefore one of the main degradation paths of substances in the environment. A procedure, such as EU C.7 or OECD TG 111, to

determine rate of hydrolysis is employed and will indicate whether other, subsequent testing should be carried out on the parent compound (if the parent compound proves to be hydrolytically stable) or on its hydrolysis products (if unstable), which may be more relevant to the risk assessment.

When it is known that significant biodegradation of the test substance may occur during the performance of a solubility test, particularly where a long equilibration period is required, then consideration should be given to the modification of the test procedures in order to minimise this effect. Such modifications can include sterilisation of the water before the equilibration period for the water solubility measurements or on collection (if the samples are not going to be analysed immediately).

An appropriate substance-specific analytical method is required to determine the mass concentration of the test substance in the aqueous solution and which will avoid detection of small amounts of soluble impurities which can cause large errors in the measured solubility. Examples of such methods are: gas or liquid chromatography, titration methods, photometric methods, voltametric methods.

Testing data on water solubility

The approximate amount of the sample and the time necessary to achieve the saturation mass concentration should be determined in a simple *preliminary test*. The preliminary test uses a visual assessment and should be used with care. In this test, increasing volumes of distilled water at room temperature are added to a small amount (approximately 0.1g) of the test substance (solid substances must be pulverised) in a glass-stoppered 10 ml graduated cylinder. If the test substance is still apparently insoluble at ~1g/l, further dilution should be undertaken to ascertain whether the column elution or flask solubility method should be used.

No single method is available to cover the whole range of solubilities in water, from relatively soluble to very low soluble chemicals. General test guidelines (OECD Method 104; EU Method A6) include two test methods which cover the whole range of solubilities but are not applicable to volatile substances. The methods should be applied to essentially pure substances that are stable in water. Details of suitable methods are shown in Table R.7.1-20.

The tests are preferably run at 20ºC ±0.5ºC. If temperature dependence is suspected in the solubility (>3% per ºC), two other temperatures should be used – both differing from each other and the initially chosen temperature by 10ºC.

Table R.7.1-20 Test methods for the determination of water solubility

Method details	Applications and requirements	Repeatability and sensitivity
Column elution method Based on elution of the test substance with water from a micro-column which is charged with an inert carrier material such as glass beads, silica gel or sand and an excess of test substance. The water solubility is determined when the mass concentration of the eluate is constant. The mass concentration of the test substance is determined analytically	Applicable to essentially pure substances only Used for low solubilities (< 10^{-2} g/l) Organic substances, but not mobile oils or liquids	< 30% ; down to 1µg/l
Flask method The test substance is dissolved in water at a temperature somewhat above the test temperature. When saturation is achieved the mixture is cooled and kept at the test temperature, stirring as long as necessary to reach equilibrium The mass concentration of the test substance is determined analytically	Applicable to essentially pure substances and also complex substances. Use of fast stirring techniques (300-400 rpm) appropriate for higher solubility (> 10^{-2} g/l) test substances. Use of slow-stirring techniques (<100 rpm) appropriate for low solubility (< 10^{-2} g/l) test substances (Letinski et al, 2002) Requires equilibration study to determine the time taken to equilibrate the test substance and water	<15 %; down to 1µg/l

An *equilibration study* is undertaken to determine the time taken to equilibrate the test substance and water. The time to equilibration will depend upon the octanol-water partition coefficient (P_{ow}) and the rate of stirring employed. Appropriate amounts of test substance (e.g. the loading rate is ~5 times the expected water solubility) are added to a known amount of water in a series of flasks. All flasks (plus control) are stirred in a temperature controlled water bath using magnetic stirrers to give a vortex one-third of the way down from the initial level. Flasks are removed at appropriate time intervals. The contents of the flasks are sub-sampled and these aliquots are centrifuged (e.g. 10 minutes at 8000 rpm). If traces of excess test substance are present then the surface layer is removed under suction using a Pasteur pipette and the final supernatant analysed. The quantity of test substance necessary to saturate the desired volume of water is determined from the results of the equilibration study. Normally, for a pure substance, about five times this quantity of material is used for the *definitive water solubility study*.

The required amounts of test substance (as determined from the equilibration study) are weighed separately into a number of flasks containing water (100 ml). The flasks are stirred until equilibrium has been reached (as identified from the equilibration study) under the same temperature and stir conditions as before. The contents of the flasks are sub-sampled and aliquots removed for subsequent analysis.

The whole test should be repeated if the results show that the amounts of test substance added is not significantly greater than the amount dissolved. If information on temperature dependence of

solubility is needed, then two or three additional flasks can be prepared, run at different temperatures and processed in the same run.

For a complex substance the measured solubility is dependent on the amount of test substance added. At least two loading rates (e.g. 100 mg/l and 1000 mg/l) are normally chosen to study the effect of loading rate on the measured water solubility of the test substance. The flask method is the appropriate test method. The composition of the aqueous sample produced is unlikely to be the same as the test substance itself and therefore final results may need to be qualified with information about the composition. All the components might be analysed in a single analytical approach (e.g. HPLC) or, more commonly, using a non-discriminating method (e.g. TOC) in order to quantify the total amount of material solubilised.

The stirring rate used in the equilibration study and definitive water solubility study may be greatly reduced (<100 rpm) in order to deter the formation of micro-emulsions or suspensions of the test substance, which could result in erroneously high water solubility results. Both slow-stir (Letinski et al., 2002) and head-space diffusion (Urrestarazu Ramos et al., 1997) methods have been used. However, if this is done, it is not unusual for the equilibration period to be much longer than normal (e.g. >96 hours). Therefore consideration has to be given to the possible breakdown of the test item during the extended test period. Addition of preservatives (e.g. mercuric (II) chloride or sodium azide at 50 mg/l) can be used to inhibit any biodegradation of the test substance during the study.

A turbidity meter can be employed to semi-quantify the point at which a test substance might no longer be deemed to be truly in solution. When using this procedure, an arbitrary figure (e.g. 10 N.T.U's) can be selected as a figure above which the solution is deemed to be visibly cloudy and the test substance is no longer fully dissolved.

Reference substances do not need to be employed in all cases when investigating a new substance. They should primarily serve to check the performance of the method from time to time and to allow comparison with results from other methods.

Published data on water solubility

Most physical properties, such as molecular weight, melting point, boiling point, density and water solubility can be obtained from commonly used environmental Handbooks, such as Verschueren's Handbook of Environmental Data on Organic Chemicals (1983), Howard's Handbook of Environmental Fate and Exposure Data, Vol. I and II (1990), Lide's CRC Handbook of Physics and Chemistry, Lange's Handbook of Chemistry, the Merck Index, the Aldrich Catalog, Kirk-Othmer Encyclopaedia of Chemical Technology and other handbook compilations such as Riddick *et al*. (1986).

Alternatively, searching on various environmental databases, such as HSDB (http://www.toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB), will provide summaries of chemical and physical properties of substances.

The sixth edition of the AQUASOL database (www.pharmacy.arizona.edu/ outreach/aquasol/) of Aqueous Solubility contains almost twenty thousand solubility records for almost six thousand compounds. These data were extracted from hundreds of scientific references. Several chemical classes including many pharmaceuticals, pollutants, nutrients, herbicides, pesticides, agricultural, industrial, and energy related compounds.

R.7.1.7.3 Evaluation of available information on water solubility

Water is a polar and protic compound and as such a poor solvent for hydrocarbons but a good one for salts. Acids and amines are more water-soluble in their ionised form than their neutral form. Compounds with polar functions (e.g. ethers, esters, ketones, alcohols, nitriles, amides, acids and amines) will decrease in solubility as the hydrocarbon (non-polar) part of the molecule increases. In general, an increase in molecular weight will lead to an increase in intermolecular forces on a solid. For example, formaldehyde, methyl acrylate and glucose are soluble, but their polymers are insoluble). Branching increases water solubility for paraffin, olefin and acetylene hydrocarbons but not for cyclic structures (cycloparaffins, cyclo-olefins and aromatic hydrocarbons). For a given carbon number, ring formation increases water solubility. Addition of unsaturations*** imparts increased water solubility.

Water solubility determinations are normally run at 20ºC in distilled water according to standard test guidelines (OECD Method 104; EU Method A6). Solubility data determined using these standard physico-chemical guidelines may differ if the test material is solubilised in either aqueous solutions containing salts or at different test temperatures (or both) (e.g. ecotoxicological test media).

Experimental data on water solubility

A number of standardised methods are available for the determination of single substances and complex mixtures of liquids and solids. The test methods are not applicable to volatile substances. Care should be taken to ensure that the test substances examined are as pure as possible and their solubility levels are determined analytically using a specific analytical method wherever possible. Precautions should be taken to minimise degradation of the test substance, in particular if long periods of equilibration are required (e.g. *slow stir* methods)

It is not unusual to find in the literature a wide range of solubilities for the same product. The oldest literature generally yields the highest solubility values: this is due to the fact that products were originally not as pure as they are nowadays and also non-specific methods were used which would not differentiate between the dissolved product and any impurities. Measurement of water solubility does not usually impose excessive demands on chemical techniques. However, measurement of the solubility of sparingly soluble compounds requires extreme care to generate saturated solutions of the material without the introduction of dispersed material; invariably specific methods of analysis are able to determine the low levels (sub ppb-ppm) in solution. Reported water solubility data for such compounds can often contain appreciable errors.

Non-experimental data on water solubility

Aqueous solubility depends not only on the affinity of a solute for water, but also on its affinity for its own crystal structure. Molecules that are strongly bound in their crystal lattice require considerable energy to remove them. This also means that such compounds have high melting points, and in general high-melting compounds have poor solubility in any solvent.

Removal of a molecule from its crystal lattice means an increase in entropy, and this can be difficult to model accurately. For this reason, as well as the fact that the experimental error on solubility measurements can be quite high (generally reckoned to be about 0.5 log unit), the prediction of aqueous solubility is not as accurate as is the prediction of partition coefficient. Nevertheless, many papers (Dearden 2006) and a book (Yalkowsky & Banerjee 1992) have been published on the prediction of aqueous solubility, as well as a number of reviews (Lyman 1990, ECETOC 1998, Reinhard & Drefahl 1999, Mackay 2000, Schwarzenbach *et al* 2003, Dearden 2006). There are also a number of commercial software programs available for that purpose (ECETOC 2003, Dearden

2006). Livingstone (2003) has discussed the reliability of aqueous solubility predictions from both QSPRs and commercial software.

It should be noted that there are various ways that aqueous solubilities can be reported: in pure water, at a specified pH, at a specified ionic strength, as the undissociated species (intrinsic solubility), or in the presence of other solvents or solutes. Solubilities are also reported in different units, for example g/100 ml, mole/litre, mole fraction. The use of mole/litre is recommended, as this provides a good basis for comparison.

Hansch *et al* (1968) first reported the inverse correlation between the aqueous solubilities of liquids and their octanol-water partition coefficients (K_{ow});

$$\log S_{aq} = - 1.339 \log K_{ow} + 0.978$$

$$n = 156 \quad r^2 = 0.874 \quad s = 0.472$$

Lyman (1990) lists 18 ($\log S_{aq}$ *vs.* $\log K_{ow}$) QSPRs for various classes of chemicals. So, for example, for a liquid ketone one could use the QSPR for ketones developed by Hansch *et al* (1968):

$$\log S_{aq} = - 1.229 \log K_{ow} + 0.720$$

$$n = 13 \quad r^2 = 0.960$$

The $\log K_{ow}$ value could be either a measured or a calculated value (see section on octanol-water partition coefficient).

However, for solids work has to be done to remove molecules from their crystal lattice, and the simplest way to account for this is to use what Yalkowsky and co-workers have termed the general solubility equation (GSE), which incorporates a melting point term to account for the behaviour of solids (Sanghvi *et al* 2003):

$$\log S_{aq} = 0.5 - \log K_{ow} - 0.01(MP - 25)$$

where MP is the melting point (^oC). The melting point term is taken as zero for compounds melting at or below 25^oC. Calculated $\log K_{ow}$ and MP values can be used in the GSE, although measured values are preferred. Aqueous solubilities of 1026 non-electrolytes, with a $\log S_{aq}$ range of – 13 to + 1 (S in mole L^{-1}), calculated with the GSE had a standard error of 0.38 log unit.

Yalkowsky and co-workers have also developed the AQUAFAC group contribution method for calculating aqueous solubility (Myrdal *et al* 1995). They calculated the ideal solubility or fugacity ratio F as:

$$\log F = - (56.5 - 19.2 \log \sigma)(MP - 25)/5706$$

where σ = a symmetry number, i.e. the number of indistinguishable positions in which a molecule can be oriented, and the units in the equation are SI units.

For liquids, $\log S_{aq} = - \log \gamma_m$, and for solids $\log S_{aq} = \log F - \log \gamma_m$, where γ_m is the molar activity coefficient, which itself is given by:

$$\log \gamma_m = \Sigma n_i q_i$$

where n_i is the number of times a group appears in a molecule and q_i is the contribution of that group. Mackay (2000) lists a large number of the group contribution values. For a set of 97 diverse chemicals, the AQUAFAC mean absolute error of prediction was 0.41 log unit, whilst that using the $\log K_{ow}$ approach was 0.61 log unit. As usual, there is a trade-off between accuracy and ease of use.

Good predictions for a large diverse data set have been obtained by the use of linear solvation energy descriptors (Abraham & Le 1999):

$$\log S_{aq} = 0.518 - 1.004\ R + 0.771\ \pi^H + 2.168\ \Sigma\alpha^H + 4.238\ \Sigma\beta^H - 3.362\ \Sigma\alpha^H.\Sigma\beta^H - 3.987\ V_X$$

$$n = 659 \quad r^2 = 0.920 \quad s = 0.557$$

where R = excess molar refractivity (a measure of polarisability), π^H = a polarity/polarisability term, $\Sigma\alpha^H$ and $\Sigma\beta^H$ = sums of hydrogen bond donor and acceptor abilities respectively, and V_X = McGowan characteristic molecular volume. All of these terms can be calculated with the Absolv 2 software (www.ap-algorithms.com). It can be seen from the Abraham and Le equation that the main factors controlling aqueous solubility are hydrogen bond acceptor ability and molecular size.

Electrotopological state descriptors (Kier & Hall 1999), hydrogen bonding and nearest-neighbour similarities (Raevsky *et al* 2004) and group contributions (Klopman & Zhu 2001) have also been used to model the aqueous solubilities of large diverse data sets of organic chemicals.

There are relatively few studies of solubility prediction within specific chemical classes. Hawker and Connell (1988) obtained the following QSPR for polychlorinated biphenyls (PCBs) with 1-10 chlorine atoms:

$$\log S_{aq} = (-4.13 \times 10^{-2})\ TSA + (23.8/R)(1 - T_m/T) + 3.48$$

$$n = 17 \quad r^2 = 0.901 \quad s = 0.464$$

where TSA = total surface area, R = universal gas constant, T_m = melting point in K and T = temperature at which solubility is required, in K.

Huuskonen *et al* (1997) used artificial neural network modelling to predict the aqueous solubilities of steroids and other drug classes. For a set of 28 steroids, with a log S_{aq} range of –5.4 to –2.6 (S in mole L^{-1}), they obtained a standard error of 0.29 log unit, using 5 molecular connectivity descriptors.

Solubility can vary considerably with temperature, and it is important that solubility data are reported at a given temperature.

Dearden *et al* (2003) compared 11 commercial software programs for aqueous solubility prediction (as log S), and found considerable variation in performance against a 113-chemical test set of organic chemicals that included 17 drugs and pesticides. The best four programs performed as given in Table R.7.1-21.

Table R.7.1-21 Software programs for aqueous solubility prediction (a)

Software	Website	% Predicted within +/- 0.5 Log unit	Standard error (Log unit)
ChemSilico	www.logp.com	75.0%	0.49
ADMET	www.simulationsplus.com	74.3%	0.50
Predictor ACDLabs	www.acdlabs.com	72.6%	0.50
WSKOWWI	www.epa.gov/oppt/exposure/docs/episuitedl.htm	69.9%	0.56

Dearden (2007) also tested the new WATERNT module in Episuite, and found it to perform better than all of the software tested previously; 79.6% of the 113 chemicals were predicted within +/-0.5 log unit of the measured log S values, and the standard error was 0.44 log unit.

Dearden (2006) tested 16 commercially available software programs for their ability to predict the aqueous solubility of a 122-compound test set of drugs with accurately measured solubilities in pure water. Again there was considerable variation in performance. The best five programs performed as given in Table R.7.1-22.

Table R.7.1-22 Software programs for aqueous solubility prediction (b)

Software	Website	% Predicted within +/- 0.5 Log unit	Standard error (Log unit)
Admensa	www.inpharmatica.com	72.1%	0.65
ADMET Predictor	www.simulationsplus.com	64.8%	0.47
ChemSilico	www.logp.com	59.8%	0.73
Pharma Algorithms	www.ap-algorithms.com	59.0%	0.62
ACDLabs	www.acdlabs.com	59.0%	0.66

ChemSilico is an on line predictor of intrinsic solubility based on electrotopological state (E-state) values (Kier & Hall 1999); it uses an artificial neural network to calculate log S values. Chemicals are inputted as SMILES strings, and a SMILES tutorial is available (www.daylight.com/smiles/smiles-intro.html). It cannot be used in batch mode, so is tedious to use for large numbers of chemicals.

ADME Predictor and ACDLabs software predicts intrinsic solubility, solubility in pure water and solubility at user-specified pH. They can operate in batch mode, and various input formats can be used.

WSKOWWIN, freely downloadable from the U.S. E.P.A. website, is based essentially on the relationship between aqueous solubility and partition coefficient, first demonstrated by Hansch *et al* (1968). It uses SMILES input to calculate intrinsic solubility, and can be used in batch mode.

ADMENSA is relatively new software. It accepts SMILES strings and Sdfiles, and can run in batch mode.

Pharma Algorithms ADME Boxes predicts solubility in pure water. It accepts SMILES notation, and can run in batch mode.

Dearden (2007) also tested the new WATERNT module in Episuite, using his 122-drug test set, and found it to perform very poorly, in contrast to its excellent performance with the 113-chemical test set. The reason appeared to be that WATERNT missed some fragments and/or correction factors in the more complex drug molecules.

The value from WSKOW is generally preferred, especially if a melting point has been inputted into the programme as the programme then uses a modified QSAR equation (with a term included for Mpt) to give a better estimate. WATERNT uses a fragment based approach to calculating the substance's water solubility, and by all accounts is less validated than the WSKOW model

Abshear *et al* (2006) showed that a consensus approach gave better aqueous solubility predictions than did four separate models, using the same test set as Dearden *et al* (2003).

It is recommended that at least one of the above software programs be used for the prediction of aqueous solubility as log S_{aq}. If possible, the average of several predictions should be taken.

Remaining uncertainty on water solubility

The water solubility of the test substance can be considerably affected by the presence of impurities.

For a complex substance, the measured solubility is dependent on the amount of test substance added. In practical terms, solubility data are generated using at least two loading rates (e.g. 100 mg/l and 1000 mg/l). Accuracy in determining water solubility decreases as the water solubility of a test substance is reduced (e.g. as shown for reference substance data in OECD Method 105). When dealing with test substances with water solubilities of the order of <10 µg/l, precautions need to be taken to avoid the introduction of dispersed material into the final extract.

R.7.1.7.4 Conclusions on water solubility

Water solubility will usually be determined experimentally. For single substance, it is helpful to have an estimated water solubility value before carrying out the water solubility experiment.

Concluding on C&L and chemical safety assessment

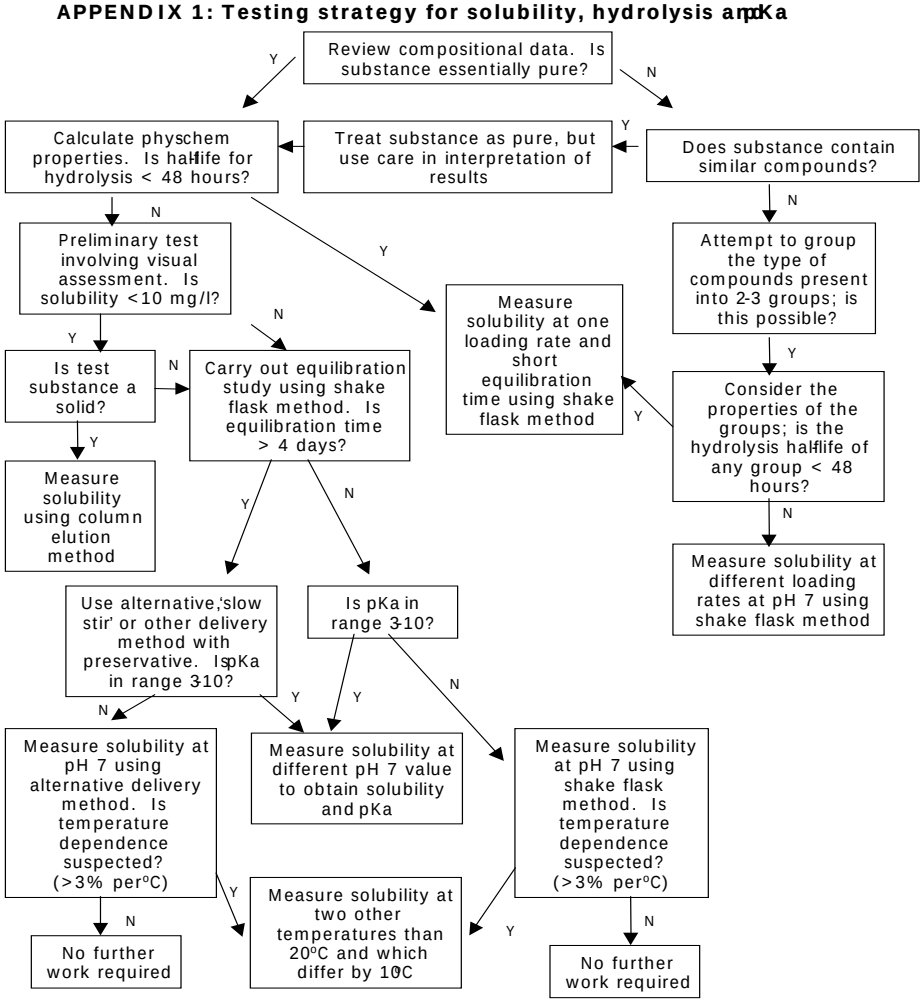
Water solubility is not a classification and labelling endpoint, as such. However, poorly water-soluble substances, i.e. substances with a solubility of less than 1 mg/l, are classified as "R53 – May cause long-term adverse effects in the aquatic environment" if they are also not readily degradable and have a log P_{ow} ≥3.0 (unless the experimentally determined BCF ≤100).

This criterion applies to substances unless there exists additional scientific evidence concerning degradation and/or toxicity sufficient to provide an adequate assurance that neither the substance nor its degradation products will constitute a potential long-term and/or delayed danger to the aquatic community. Such additional scientific evidence should normally be based on the studies required at level 1 (Annex VIII) or studies of equivalent value. These could include a proven potential to degrade rapidly in the aquatic environment or an absence of chronic toxicity effects at the solubility limit.

R.7.1.7.5 Integrated testing strategy (ITS) for water solubility

It is important to have preliminary information on the vapour pressure, the dissociation constant, and the hydrolysis rate (as a function of pH) of the substance before any testing is carried out. The three properties, i.e. solubility, hydrolytic stability and acid dissociation, are inter-related. It is not possible to measure any one of these without some knowledge of the other two. The flow diagram (Figure R.7.1-5) presents a schematic way of determining which solubility measurements should be performed for a given substance. For example, if the test substance is an acid or base with a dissociation constant (pKa) value between 3 and 10, it is advisable to use pH buffers in the water to enable the water solubility of the ionised and/or non-ionised (neutral) forms of acids and bases to be determined over an environmentally relevant range (e.g. pH 4-9). For many substances, however, knowledge of the composition may be enough (e.g. it may be obvious to the chemist that no hydrolytically unstable or ionisable groups are present. Conversely, if the half-life is less than 12 hours then solubility studies are not realistic. If it is not intended to perform a full hydrolytic stability test (OECD TG 111; EU C.7), then procedures for performing a reduced test of stability should be included as part of the solubility test.

Figure R.7.1-5 The three properties, solubility, hydrolytic stability and acid dissociation constant are inter-related. It is not possible to measure any of these without some knowledge of the other two



Examples and case studies on water solubility

It is well known that poorly water-soluble compounds dissolve slowly (Yalkowsky and Banerjee, 1992). Solution formation rate is greatly increased by vigorous shaking due to the formation of small droplets or emulsions. However, removal of undissolved chemical from the water phase may be problematic. Centrifugation may not be effective in separating emulsions, particularly if the chemical is a liquid with a specific gravity near to that of water. Filtration may lead to losses of dissolved chemicals to surfaces or to air.

Letinski et al (2002) have employed a *slow stir* method for water solubility determinations of C_8-C_{15} aliphatic alcohols and phthalate and adipate diesters that overcomes the emulsion problem. In their experiments, water was stirred quiescently or with little or no visible vortex with addition of a mercuric chloride (50 mg/l) to inhibit microbial degradation of the test substances. An equilibration time of 7 days was usually sufficient to ensure the maximum solubility of the test substances, which, for undecanol and pentadecanol, were 7.97 and 0.0061 mg/l, respectively. Alcohols with methyl branching are more soluble than the pure linear alcohols. Longer equilibration times (9 days) were required for the phthalate and adipate diesters, which had measured water solubilities three orders of magnitude lower (70–0.044 µg/l). Comparison of the measured water solubility of phthalate and adipate diesters with literature values, confirms that many of the older literature values are much higher than the slow stir water solubility results, most likely due to emulsion formation and the use of non-specific chemical methods of analysis.

For complex substances such as alcohol ethoxylates, which have a number of carbon chain lengths (typically C_{12}-C_{18}) with each chain length ethoxylated with up to 20 ethoxylate (EO) units, the measured solubility is dependent on the amount of test substance added. At least two different loadings (e.g. 100 mg/l and 1000 mg/l) are normally chosen to study the effect of loading on the measured water solubility of test substances such as NEODOL 1-5E and NEODOL 245E (Shell internal reports).

R.7.1.7.6 Special guidance on water solubility of inorganic substances

Water solubility among compounds of the same metal may differ by several orders of magnitude. Differences in the solubility of metal compounds are related to the metal species, the particle size, and the characteristics of the aqueous medium. Highly soluble inorganic metal compounds can be assessed through the normal procedures. For sparingly soluble metal compounds, a solubility product can be calculated thermodynamically (e.g. by using the Facility for Analysis of Chemical Thermodynamics ("F*A*C*T", FACT-Win version 3.05).

Although metals are generally insoluble, metals in the elemental state may react with water or a dilute aqueous electrolyte to form soluble or sparingly soluble cationic or anionic products. During this process the metal will oxidise, or transform, from the neutral or zero oxidation state to a higher oxidation state. The OECD Test Guidance on transformation/dissolution of metals and sparingly soluble metal compounds (OECD, 2001) can be used to determine the rate and extent to which metals and sparingly soluble metal compounds can produce soluble bioavailable ionic and other metal-bearing species in aqueous media under a set of standard laboratory conditions representative of those generally occurring in the environment. The outcomes of the transformation/dissolution tests are to be used for aquatic classification and labelling purposes and can also be used for the integration of fate parameters into the Predicted Environmental Concentration (PEC) calculations.

Complex inorganic substances like flue dusts, slags or alloy manufacturing products are not simple mixtures and, justified by the intricate production process, the solubility properties of a given complex substance can differ substantially from what is observed for each individual constituent of that complex substance. All these materials are typically not readily soluble in any aqueous

medium. In addition, this material is often heterogeneous in size and composition on a microscopic/macroscopic scale. Therefore, a representative result is hardly achievable, especially when small amounts of that material are weighed out. This should also be kept in mind when water accommodated fractions (WAF) are prepared in order to get elutions for ecotoxicity tests. Adequate amounts of the material should then be weighed out per volume of the solvent. Additional guidance is given in the OECD Guidance Document 23 (aquatic toxicity testing of difficult substances and mixtures).

R.7.1.7.7 References on water solubility

Abraham M.H. and Le J. (1999) The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. *J. Pharm. Sci.* 88, 868-880.

Abshear T., Banik G.M., D'Souza M.L., Nedwed K. and Peng C. (2006) A model validation and consensus building environment. *SAR QSAR Environ. Res.* **17**, 311-321.

Budavari, S., Ed. (1989) The Merck Index – Encyclopedia of Chemicals, Drugs and Biologicals. Rahway, NJ:Merck and Co., Inc.

Dearden J.C. (2006) *In silico* prediction of aqueous solubility. *Expert Opinion on Drug Discovery* **1**, 31-52.

Dearden J.C. (2007). Unpublished work.

Dearden J.C., Netzeva T.I. and Bibby R. A comparison of commercially available software for the prediction of aqueous solubility. In Ford M., Livingstone D., Dearden J. and van de Waterbeemd H. (Eds.), *Designing Drugs and Crop Protectants: Processes, Problems and Solutions.* Blackwell, Oxford, 2003, pp. 169-171.

EC Method A6. "Water Solubility", Dir 92/69/EEC, Official Journal of the European Communities, O.J. L383 A)

EC Method A7. "Hydrolysis ", Dir 92/69/EEC, Official Journal of the European Communities, O.J. L383 A)

ECETOC Technical Report No. 74: *QSARs in the Assessment of the Environmental Fate and Effects of Chemicals*. ECETOC, Brussels, 1998.

ECETOC Technical Report No. 89: *(Q)SARs: Evaluation of the Commercially Available Software for Human Health and Environmental Endpoints with Respect to Chemical Management Applications.* ECETOC, Brussels, 2003.

Hansch C., Quinlan J.E. and Lawrence G.L. (1968) The linear free-energy relationship between partition coefficients and aqueous solubility of liquids. *J. Org. Chem.* 33, 347-350.

Hawker D.W. and Connell D.W. (1988) Octanol-water partition coefficients of polychlorinated biphenyl congeners. *Environ. Sci. Technol.* 22, 382-387.

Howard, P.H., Ed. (1990). Handbook of Fate and Exposure Data for Organic Chemicals, Vol. I and II, Lewis Publishers, Chelsea, Michigan.

Huuskonen J., Salo M. and Taskinen J. (1997) Neural network modelling for estimation of the aqueous solubility of structurally related drugs. *J. Pharm. Sci.* 86, 450-454.

Kier L.B. and Hall L.H. *Molecular Structure Description: the Electrotopological State.* Academic Press, San Diego, CA, 1999.

Kirk-Othmer Encyclopedia of Chemical Technology. 4th ed. Volume 1: New York, NJ,. JohnWiley and Sons, 1991-Present.

Klopman G. and Zhu H. Estimation of the aqueous solubility of organic molecules by the group contribution approach. *J. Chem. Inf. Comput. Sci.* (2001) 41, 439-445.

Lange's Handbook of Chemistry, Editor, Speight, James, Publ. G. McGraw-Hill.

Letinski, D.J., Connolly, M.J., Peterson, D.R. and Parkerton, T.F. (2002) "Slow-stir water solubility measurements of selected alcohols and diesters", *Chemosphere*, 48, 257 – 265).

Lide, D.R., Ed. (2005). CRC Handbook of Chemistry and Physics, 86th edition, CRC Press, Taylor and Francis, Boca Raton, FL.

Livingstone D.J. (2003) Theoretical property predictions. *Current Topics in Med. Chem.* 3, 1171-1192.

Lyman W.J. Solubility in water. In Lyman W.J., Reehl W.F. and Rosenblatt D.H. (Eds.) (1990), *Handbook of Chemical Property Estimation Methods*. American Chemical Society, Washington DC, pp. 2.1-2.52.

Mackay D. Solubility in water. In Boethling R.S. and Mackay D. (Eds.), *Handbook of Property Estimation Methods for Chemicals: Environmental and Health Sciences.* Lewis, Boca Raton, FL, 2000, pp. 125-139.

Meylan W.M., Howard P.H. and Boethling R.S. Improved method for estimating water solubility from octanol/water partition coefficient. *Environ. Toxicol. Chem.* (1996) 15, 100-106.

Myrdal P., Manka A.M. and Yalkowsky S.H.(1995) AQUAFAC 3: aqueous functional group activity coefficients; application to the estimation of aqueous solubility. *Chemosphere* 30, 1619-1637.

OECD Environmental Health and Safety Publications, 2000. Number 23. Guidance document on aquatic toxicity testing of difficult substances and mixtures.

OECD Environment, Health and Safety Publications, 2001. Series on Testing and Assessment, No. 29, Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media

OECD Guidelines for Testing of Chemicals – Method 105 "Water Solubility"

OECD Guidelines for Testing of Chemicals – Method 111 "Hydrolysis as a Function of pH"

Raevsky O.A., Raevskaja O.E. and Schaper K.-J. (2004) Analysis of water solubility data on the basis of HYBOT descriptors. Part 3. Solubility of solid neutral chemicals and drugs. *QSAR Comb. Sci.* 23, 327-343.

Reinhard M. and Drefahl A. *Handbook for Estimating Physico-chemical Properties of Organic Compounds.* Wiley, New York, 1999.

Riddick, J.A., Bunger, W.B., Sakano, T.K. (1986). Organic Solvents: Physical Properties and Methods of Purification. Techniques of Chemistry. 4th Edition. New York, NY, Wiley-Interscience

Sanghvi T., Jain N., Yang G. and Yalkowsky S.H. (2003) Estimation of aqueous solubility by the general solubility equation (GSE) the easy way. *QSAR Comb. Sci.* 22, 258-262.

Schwarzenbach R.P., Gschwend P.M. and Imboden D.M. (2003) *Environmental Organic Chemistry*, 2nd edition, Wiley, Hoboken, NJ.

Tetko I.V., Tanchuk V.Yu., Kasheva T.N. and Villa A.E.P. (2001) Estimation of aqueous solubility of chemical compounds using E-state indices. *J. Chem. Inf. Comput. Sci.* 41, 1488-1493.

Urrestarazu Ramos, E., Vaes, W.H.J., Verhaar, H.J.M., Hermans, J.L.M. (1997) "Dissolving low soluble chemicals", *Environ. Toxicol. Chem*., 16, 2229-2231.

Verschueren, K. (1983) Handbook of Environmental Data on Organic Chemicals. Fourth Edition. John Wiley & Sons, Inc., New York.

Yalkowsky S.H. and Banerjee S. Aqueous Solubility: Methods of Estimation for Organic Compounds. Marcel Dekker, New York, 1992.

R.7.1.8 PARTITION COEFFICIENT N-OCTANOL/WATER

The n-octanol/water partition coefficient (K_{ow}) is one of the key physico-chemical parameters, and it is used in numerous estimation models and algorithms for environmental partitioning, sorption, bioavailability, bioconcentration/bioaccumulation and also human- and ecotoxicity. As such K_{ow} is a critical parameter for chemical safety assessment (CSA), classification and labelling (C&L), and PBT assessment.

The generation of a K_{ow} value is required at all tonnage bands (i.e. > 1 t/y; Annex VII - IX).

Definition of partition coefficient n-octanol/water

The n-octanol/water partition coefficient (K_{ow}) is defined as the ratio of the equilibrium concentrations of a dissolved substance in a 2-phase system consisting of the largely immiscible solvents n-octanol and water (OECD Methods 107, 117, 123 and 122 (draft)). The property is moderately temperature-dependent and typically measured at 25°C).

$$K_{ow} = [X]\text{n-octanol} / [X]\text{water}$$

(where [X] indicated the concentration (mass/volume) in the specific solvent)

Octanol is regarded as a model solvent that mimics lipids in organisms and humans, and organic carbon in oils and sediments. The partition coefficient is commonly expressed as $\log_{10} K_{ow}$ (also known as log P).

Since log K_{ow} has been shown to be a key parameter in risk assessment and regulatory schemes, it needs to be determined with the greatest possible accuracy.

The ability to measure an accurate log K_{ow} is also related to, or influenced by, other physico-chemical parameters, such as water solubility, vapour pressure, surface tension, ionisable properties (dissociation constant(s)), hydrolytic stability and redox stability.

R.7.1.8.1 Information requirements on partition coefficient n-octanol/water

The generation of the K_{ow} datapoint cannot be waived since it is essential for CSA, C&L and PBT assessments. It can be determined either by an appropriate *in silico* estimation method based on the molecule's structure (see above), or by a laboratory test. K_{ow} does not need to be determined if the substance is purely inorganic.

A computational (*in silico*) method shall be used in case the test cannot be technically performed. This can be the case for example if the substance decomposes, has a high surface activity, reacts violently during the performance of the test, or does not sufficiently dissolve in water or in octanol, or if it is not possible to obtain a sufficiently pure substance. QSARs can also be the solution in case of absence of sufficiently sensitive or selective analytical methods. In a situation where the test substance is very labile or reactive, the K_{ow} value of the parent material may not be relevant for further risk- or PBT assessment, and the degradation/reaction products may need to be considered instead.

R.7.1.8.2 Available information on partition coefficient n-octanol/water

Testing data on partition coefficient n-octanol/water

(Measured) log K_{ow} is a commonly documented property in chemical databases, such as IUCLID (http://ecb.jrc.it**)**, or MedChem (http://www.daylight.com) which includes approximately 61,000 measured log K_{ow} values. The PHYSPROP database of the Syracuse Research Corporation contains log K_{ow} data for over 25000 chemicals (http://www.syrres.com). Another source is the Canadian National Committee for CODATA (CNC/CODATA) database with evaluated log K_{ow} values for over 20000 chemicals (http://logKow/cisti/nrc/ca).

Published data on partition coefficient n-octanol/water

Ample** of scientific publications and QSAR development studies report measured log K_{ow} values for a variety of chemicals that have been used in the training- or validation sets. Some key references are: Sangster (1989); Hansch & Leo (1995); Mackay *et al*. (1997); ECETOC Technical Report # 89 (2003).

R.7.1.8.3 Evaluation of available information on partition coefficient n-octanol/water

Experimental data on partition coefficient n-octanol/water

Four experimental methods (Table R.7.1-23) are described to measure log K_{ow}, of which three are direct measurement methods (Shake Flask Method, Generator Column Method, and Slow- Stirring Method), and one indirect measurement method (Reverse Phase HPLC Method). Highly accurate measurements of log K_{ow} >~5 are complicated by the fact that small amounts of octanol are entrained in the aqueous phase, leading to a potential underestimation of the measured log K_{ow} values. All of the direct methods for measuring log K_{ow} require quantifying the test material in either octanol or water and preferably in both matrices. Since most of the experimental protocols are fully documented as Standard Methods, only a short summary will be provided here, with emphasis on their respective merits and drawbacks. Guidance and recommendations for selecting the appropriate procedure are also given in ANNEX V of CEC (1993).

The focus here is on OECD methods, but test data derived with equivalent methods from other organisations such as EPA-FIFRA, ASTM, ISO, etc. should equally be accepted.

Table R.7.1-23 Methods for determination of partition coefficient n-octanol/water

Method details	**REPEATABILITY**	**APPLICABILITY RANGE**
Shake Flask Method The Shake Flask method is the default procedure. It is considered to give accurate results for low to medium hydrophobic chemicals. For chemicals with a high expected log K_{ow}, alternative methods are recommended. A suitable analytical method is needed to determine the concentration of the test material in the octanol and water phases. By applying mass balance considerations, it may be possible to measure the test material in only the less-soluble phase. However, this approach significantly decreases the reliability in the reported value. This technique is not suitable for surface active compounds (surfactants), or compounds that hydrolyse rapidly.	Three replicates should fall within +/- 0.3 log K_{ow}	$-2 < \log K_{ow} < 4$
Computational Approach in OECD Guideline 107 (EU A.8) This method enables partition coefficients to be calculated based on the ratio of the solubility of the material in octanol and water. For some substances (e.g. some surfactants and pigments) it is technically not feasible (or good practice) to measure an octanol-water partition coefficient by OECD 107, and other methods (OECD 117 and QSAR) may be unsuitable as they are based on extrapolations of measured values. For such substances it may be possible to obtain a ratio of the saturated water solubility (OECD 105) and saturated octanol solubility (no guideline currently available but based on the principles of OECD 105). The calculation method however has the drawback that it does not include the interaction between the water and solvent phase (i.e. a chemical with high Kow is rather 'pushed out of the water' than 'pulled into octanol"). This explains the poor correlation typically observed between octanol solubility and K_{ow} (Dearden, 1990, Sijm et al., 1999. The ratio was found to be somewhat more representative if one uses octanol/saturated water and water/saturated octanol. As such, a ratio calculation would be a less preferred yet acceptable alternative for the octanol/water partition coefficient (Kow), but must be treated with caution as it would not have been derived in the same manner as other K_{ow}s (OECD 107).		
Generator Column Method (Woodburn et al., 1984) The Generator Column method is suitable for measuring the K_{ow} of more hydrophobic chemicals. The principle of the method is that water-saturated octanol, containing the test compound, is coated on a column, and then eluted with a mobile phase of octanol/saturated water. Concentrations in both phases are measured at the end of the test. This approach minimises formation of emulsions and allows direct coupling to an HPLC or concentrator column for analysis of the aqueous phase.	Not documented in the literature	Suitable for measuring the K_{ow} of more hydrophobic chemicals.

No formal OECD guidelines or other standards for this approach have been published. A disadvantage of the generator column method is that it requires sophisticated equipment. A description of the generator column method is presented in USEPA (1985).		
Slow-Stirring Method (OECD Draft Guideline 123, 2003) This is a more recent method developed as an alternative to the shake flask procedure. The advantage of slow stirring versus shaking is that emulsion formation will be reduced. The method requires a few days to reach equilibrium. The method may be difficult to adapt to a high throughput approach. As with the other direct methods, a suitable analytical method is needed to measure the concentration of the test material in the octanol and water phases. NB: Radiolabelled substances –which may be synthesised for use in other tests- can be very useful for accurate log K_{ow} determination.	Intralaboratory median standard deviation from 0.15 – 0.3 Log K_{ow} (Tolls et al, 2003).	Validation has shown that this method can also be used for very hydrophobic chemicals, up to Log K_{ow} 8.3 (OECD 2003, Tolls et al, 2003).
Reverse Phase HPLC Method (OECD Guideline 117, EU A.8) This is a relatively quick way of estimating log K_{ow}. It is not measured directly, but from a correlation between log k (capacity factor) and log K_{ow} for a series of reference chemicals. It therefore depends on the quality of the log K_{ow} measurement of reference chemicals (often measured by the shake flask method). A series of reference compounds with similar chemical functionality to the test material should be used to generate the log k: log K_{ow} correlation. In general, the HPLC method is less sensitive to impurities than the shake flask method. The RP-HPLC is not recommended for strong acids and bases, metal complexes or surface active agents, or for measurements across very different classes of chemicals. The HPLC method is also very suitable for measuring the K_{ow} of mixtures of chemical homologues.	Three replicates should fall within +/- 0.1 log K_{ow}	This method enables log K_{ow} to be measured between 0 and 6.

Non-experimental data on partition coefficient n-octanol/water

Many publications have dealt with the estimation of log K_{ow} values from molecular structure, and Lyman (1990), Schwarzenbach *et al* (1993), Nendza (1998), Reinhard and Drefahl (1999), Leo (2000), Livingstone (2003) and Klopman and Zhu (2005) have reviewed prediction methods for log K_{ow}; Livingstone (2003) in particular gives a detailed critical analysis of available methods. The main prediction methodologies are based on physico-chemical, structural and/or topological descriptors or on atomic or group contributions. For example, Bodor *et al* (1989) developed a QSPR with 14 physico-chemical and quantum chemical descriptors to model log K_{ow} of a diverse set of 118 organic chemicals, with $r^2 = 0.882$ and a standard error of 0.296 log unit. The method of Ghose *et al* (1988) used atomic contributions, and on a set of 893 compounds the standard error was 0.496 log unit. Klopman and Wang (1991) used their MCASE group contribution approach to predict the log K_{ow} values of 935 organic compounds with a standard error of 0.39 log unit. This error is close to the experimental error on log K_{ow}.

A method of predicting log K_{ow} values that provides mechanistic insight is that of Abraham *et al* (1994). Using their solvatochromic descriptors they developed the following QSPR:

$$\log K_{ow} = 0.088 + 0.562\ R - 1.054\ \pi^H + 0.034\ \Sigma\alpha^H - 3.460\ \Sigma\beta^H + 3.814\ V_x$$

$$n = 613 \quad r^2 = 0.995 \quad s = 0.116$$

where R = excess molar refractivity, π^H = a polarity term, $\Sigma\alpha^H$ and $\Sigma\beta^H$ = hydrogen bond donor and acceptor abilities respectively, and V_x = the McGowan characteristic molecular volume. Since the descriptors are approximately autoscaled, the magnitudes of the coefficients give an indication of the relative contribution of each descriptor to log K_{ow}. Thus it can be seen that hydrogen bond acceptor ability and molecular size make the most important contributions to log K_{ow}; on the other hand the contribution of hydrogen bond donor ability is negligible, and this is attributed to the hydrogen bond acceptor abilities of both water and octanol being very similar, while in contrast the hydrogen bond donor ability of water is very strong, accounting for the high negative coefficient on the $\Sigma\beta^H$ term. The standard error is very low, and may indicate some over-fitting of the data.

Although measured values of the Abraham descriptors are not available for all compounds, they can be calculated using the Absolv-2 software (www.ap-algorithms.com).

There are also numerous software programs[12] available for the estimation of Log K_{ow} of organic chemicals, and some of these give good predictions. A recent comparison of 14 such programs (Dearden *et al* 2003) found that, using a 138-chemical test set, the percentage of chemicals with log K_{ow} predicted within +/-0.5 log unit of the measured log K_{ow} value ranged from 94% to 50%. The performances of the top six programs are given in Table R.7.1-24.

Table R.7.1-24 Software programs for the estimation of log K_{ow}

Software	Website	Availability	Batch Operation	% Predicted within 0.5 Log unit	Standard Error
ADMET	www.simulationsplus.com	Purchase	Yes	94.2	0.27
ACDLabs	www.acdlabs.com	Purchase	Yes	93.5	0.27
ChemSilico	www.logp.com	Free on line	No	93.5	0.30
KOWWIN	www.epa.gov/oppt/exposure/pubs / episuitedl.htm	Free to download	Yes	89.1	0.34
SPARC	ibmlc2.chem.uga.edu/sparc	Free on line	No	88.5	0.33
ClogP	www.daylight.com	Purchase	Yes	88.4	0.29

Some of the software programs not tested by Dearden *et al* (2003) are ChemOffice (www.cambridgesoft.com), Admensa (www.inpharmatica.co.uk) and AUTOLOGP (Devillers *et al* 1995). ChemOffice is reported to yield a standard error of 0.43 log unit, but that rises to 0.83 log

[12] It should be noted that a calculated log K_{ow} value relates to the unionised form of a chemical. Some software (such as ACD and Pallas) will also calculate log D_{ow}, where D_{ow} is the distribution coefficient taking into account both ionised and unionised species; this requires a knowledge of pKa, which the software also calculates. log D_{ow} is related to Log K_{ow} as follows:
For acids: $\log D_{ow} = \log K_{ow} - \log(1 + 10^{(pH-pKa)})$
For bases: $\log D_{ow} = \log K_{ow} - \log(1 + 10^{(pH-pKa)})$

unit for compounds possessing intramolecular hydrogen bonding. Admensa is reported to yield a test set RMS error of 0.44 log unit. AUTOLOGP is reported (Devillers *et al* 1995) to yield a standard error of 0.39 log unit for a heterogeneous set of 800 organic compounds.

It is recommended that at least one of the above software programs be used for the prediction of log K_{ow}. If possible, the average of several predictions should be taken.

Users may also have access to log K_{ow} predicting software other than the ones described above, such as e.g. SCILOGP, Molinspiration, Chemaxon, TERRQSAR-logP, ALOGPS (http://146.107.217.178/lab/alogps/start.html), XLOGP, etc. For their performance, no broad comparison statistics were found; as for all QSPR models, their applicability should be carefully considered on a case-by-case basis. The recommendations of the developers should be considered prior to use of the models.

It is important to point out that, contrary to what might be thought, solubility in octanol is of no value as a measure of lipophilicity. When a chemical is taken up by lipid *in vivo*, it is always from an aqueous phase, and so it is the distribution between aqueous and lipid phases that is important, and not the absolute solubility in lipid. In fact, the term hydrophobicity is preferable to lipophilicity, because the driving force for transfer from water to lipid comes largely from the aqueous phase; that is, a chemical is pushed from water to lipid, rather than being pulled by lipid from water. The driving force has a large entropic component (Dearden & Bresnen 2005) because of water-structuring. Octanol tends to behave much as an ideal solvent and solubility in octanol (S_o) is inversely correlated with melting point, but not with octanol-water partition coefficient. Dearden (1990) showed that the correlation between log K_{ow} and log S_o is very poor.

It is also pointed out that the calculation of log K_{ow} from the ratio of solubilities in octanol and water is rather inaccurate, as the results below show (Yalkowsky et al 1983):

Solute	**log (S_o/S_w)**	**log K_{ow}**
Antipyrine	-0.73	0.26
Ethyl 4-aminobenzoate	1.86	1.96
Caffeine	-0.75	-0.20
Theophylline	-0.57	-0.09

Remaining uncertainty on partition coefficient n-octanol/water

An overview of various experimental techniques and the factors which may influence measurement of log K_{ow} is provided in a review by Dearden and Bresnen (1988). The GHS Guidelines (GHS, 2003) also contain clear guidance on log K_{ow} determination, including issues of difficult-to-test substances or conflicting data.

Particular attention should be paid to the higher log K_{ow} values that fall around cut off values of different regulatory and PBT schemes (i.e. where the cut off point is within the uncertainty/error margin of the measurement or the QSAR model). If a log K_{ow} around 3 or 4 has been obtained for a substance, then the data of the highest quality and the best documentation should be used for determining the K_{ow}. If differences still exist, generally the highest reliable value should take precedence.

On the other hand, the impact of errors on model predictions for a low log K_{ow} value (<1) is usually less critical. Also, for very high log K_{ow} values (>8) the measurement itself is generally difficult, and the environmental models are often less sensitive and predictive in this log K_{ow} range. For a

further discussion on the need for accuracy in log K_{ow} determination and error propagation, the reader is referred e.g. to the paper by Renner (2002) and references therein.

Difficult to test substances

There are certain structural or physico-chemical properties, which can make the accurate determination of K_{ow} or its measurement difficult. Difficult to test substances include poorly soluble, volatile, surface active, ionisable chemicals, mixtures of chemicals, as well as chemicals subject to rapid degradation due to such processes as phototransformation, hydrolysis, oxidation, or biotic degradation.

Guidance on regulatory compliant K_{ow} determination for ionisable substances and salts

The K_{ow} is typically defined as the partition coefficient of the neutral, undissociated form of a substance. However, the relative extent to which an ionisable substance is likely to be dissociated in the environment (with pH usually in the range 5-9) can have a marked effect on its physico-chemical properties, especially the octanol-water partition coefficient and water solubility, which in turn affect fate and behaviour. As log K_{ow} is routinely used to predict bioconcentration/bioaccumulation potential, this aspect is especially important in a PBT context. For substances which dissociate within an environmentally relevant pH range (pKa 5-9), values for K_{ow} shall be derived for the neutral form, and preferably also for the dissociated form. In some cases a factor 4-5 has been recorded between the log K_{ow} of both species. The value for the dissociated molecule determined around a pH of 7 (sometimes referred to as D_{ow}) is considered more realistic for PBT and chemical safety assessment.

Based on practical experience the following guidance is provided:

Simple acids and bases in the normal pH range:

- The HPLC method is to be applied to acids and bases in their non-ionised forms, although the pH should be kept in the range 2 to 9 (however pH 5 to 9 is preferred).
- For the shake-flask method, the approach must be followed in which the study is conducted at a pH where the substance is not ionised, if possible, or at a pH where the extent of ionisation is minimised.
- Validated QSAR estimations may be useful for acids and bases.

Zwitterionic substances:

- For zwitterions, the shake-flask method should be used to develop a valid K_{ow} value. Even if the ionic charge pattern of the compound in octanol is not known, the value represents a practical and useful parameter. It is not justifiable to expect a full description of all the equilibria in both water and octanol. The pH of such a study should be 7 or the iso-electric point, as long as that point is in the range pH 5 to 9, so as to maximise the possibility of partition into octanol. There is no need to ask for both pH values.
- The HPLC method must not be used. The usual estimation methods should be valid, but particular care should be exercised.
- QSAR estimations may be useful provided they are validated.

Salts of organic compounds:

- The shake-flask method should be used, usually at pH 7, or at any pH in the range 5 to 9 which maximises the potential for partition into octanol. For salts, the nature of the analytical method compared to the chemical composition will have to be considered. The ideal is to monitor cation and anion** individually in both phases. When only one half can be analysed, then the result must be understood as partial, even if it is the best that is achievable.
- Estimation by HPLC is not valid for the whole salt.
- QSAR methods will be valuable in assessing the properties of each half of the salt. Current estimation methods cannot estimate the K_{ow} of the ion pair.

Guidance on regulatory compliant K_{ow} determination for surfactants

In many cases a calculated K_{ow} value will be the first choice for surfactants. It is also useful to compare a calculated with a measured value. For the calculation approaches, one needs to consider the pH of the system (which determines the ionisation of the surfactant – see Section R.7.1.17).

None of the experimental methods is very well suited for determining the K_{ow} of surface active chemicals. The shake flask method is the least suitable experimental method for surfactants. HPLC methodology may fail due to secondary interactions, and is sensitive to fluctuations of ionic strength. The slow stirring method in theory is the best, but still not demonstrated to be perfect. If using slow stir, one needs to demonstrate a consistent result when starting with the surfactant in either phase, not just in the octanol.

Another approach for surfactants can be the comparison of measured solubilities in octanol and water. However, it is prudent to take the critical micelle concentration in water (CMC) as a solubility limit, in order to avoid unrealistically low K_{ow} values.

Guidance on regulatory compliant Kow determination for mixtures

It is possible that different components of mixtures have significantly different behaviour in the physico-chemical tests and therefore also *in vivo* and in the environment. It is therefore important to ensure that the results presented for the physico-chemical tests represent each component rather than the mixture being treated as a single component. For simple mixtures where the components are known and easily identifiable, this may mean presenting individual values for K_{ow}. For complex mixtures, the HPLC method is ideal for determination of K_{ow}, and a range of values should be presented, with an indication of the proportion of substance within a given range (e.g. >90% of components have log K_{ow} >6), to allow the significance of these results to be reflected in the risk assessment. The HPLC method is also recommended for petroleum products, which are typically mixtures.

Situations with multiple Kow values – Weight of Evidence

In case of situations where multiple log K_{ow} data are available for the same substance, the possibility of conflicting results might arise. Measured K_{ow} values are generally given precedence over estimated values.

R.7.1.8.4 Conclusions on partition coefficient n-octanol/water

For organic substances experimentally derived high-quality K_{ow} values, or values which are evaluated in reviews and assigned *recommended values*, are preferred over other determinations of K_{ow}. When no experimental data of high quality are available, or if experimental methods are

known to be unreliable, validated quantitative structure activity relationships (QSARs) for log K_{ow} may be used. Such validated QSARs may be used if they are restricted to chemicals for which their applicability is well characterised. There is a broad availability of free and commercial QSAR models. This also allows checking of consistency between predictions obtained with different softwares.

R.7.1.8.5 Integrated testing strategy (ITS) for partition coefficient *n*-octanol/water

The text below and associated ITS (Figure R.7.1-6) describe a high-level scheme for log K_{ow} determination, i.e. for the purpose of simplicity, detailed individual decision points were not listed. The reader is also referred to the generic ITS scheme on data quality that provides further guidance on the acceptance/rejection of test data.

Step 1: Information collection on structure and other PC properties

Collect information on structure and PC properties that may provide an indication of the hydrophilic/lipophilic nature of the compound (e.g. structure, presence of halogen atoms, solubility, pKa), or properties that may cause testing difficulties (e.g. surface activity, salts or ionising compounds, mixtures, etc.).

Step 1* (optional): range-finding QSAR assessment to obtain a first indication of log Kow value

Determine if the existing QSAR models are considered suitable for this type of compound. Is the log K_{ow} expected to be either very low (<1), around regulatory cut off points, or very high (>8)? Will the log K_{ow} affect the method of choice? Consider this information in the next steps of the scheme.

Step 2: Decide if the compound belongs to a class of difficult-to-test substances.

If NO, THEN move to step 3.

If YES, THEN move to step 3D. Consider the complexities of testing and associated error ranges when reviewing existing data.

Step 3: Evaluation of available experimental data for quality and consistency

If YES (valid measured data exist), THEN accept the results and STOP (NB: if multiple data exist, expert judgement will be needed to select the most reliable one, or an average can be made. In case of doubt, the highest value would take precedence).

If NO (no valid measured data), THEN move to step 4.

Step 3D: Difficult substances: Evaluation of available experimental and/or QSAR data for quality and consistency

If YES (valid measured or QSAR data exist), THEN accept the results and STOP (NB: if multiple data exist, expert judgement will be needed to select the most reliable one, or an average can be made. In case of doubt, the highest value would take precedence. Measured data will typically take precedence over QSAR data).

If NO (no valid measured or predicted data) move to step 4D.

Step 4: Determination of the availability of valid QSARs

If YES, THEN run the QSAR(s) and move to step 7. Preferably several QSARs are used and their predictions compared.

If NO, THEN move to step5 (testing).

Step 4D: Determination of an appropriate strategy for difficult substances – based on all available relevant information.

This may result in 1) the generation of (new) QSARs data only, 2) the use of a (non-perfect) test method only, or 3) the generation and comparison of both test and QSAR data (preferred).

Step 5: Determination of the most suitable test methods based on the available information

If YES (suitable test method available), THEN execute test and move to step 6.

If NO, THEN it must be considered that the compound is not amenable to testing, and only QSAR predictions should be used.

Step 6: Assessment of the experimental test results. Are experimental results of good quality and do they match with QSAR data (where available)?

If YES, the data are considered consistent and of good quality, THEN accept the result and STOP.

If NO, the test has revealed issues or is inconsistent with QSAR estimations THEN it is worth considering entering into a refinement loop, where alternative test methods and or QSARs can be tried until a more robust value has been obtained.

Step 7: Assessment of the agreement between QSARs (if several valid models are used).

If YES (good agreement), THEN select the average number as most representative. Move to step 8.

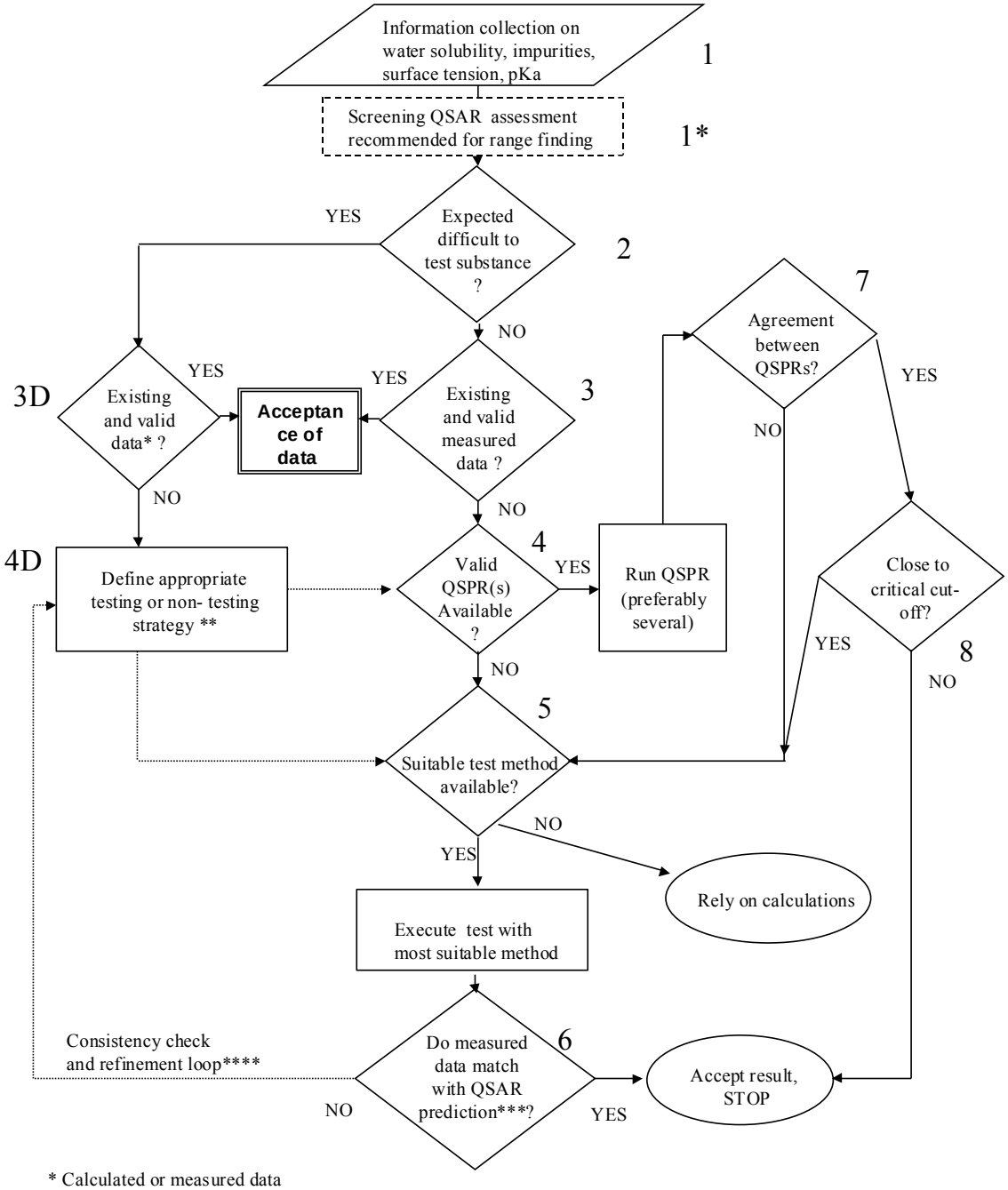
If NO (poor agreement), THEN move to step 5.

Step 8: Are the QSAR predictions close to a critical cut off point (in C&L or PBT assessments)? Would uncertainty around QSAR prediction affect the conclusions?

If YES, go to step 5 (testing).

If NO: accept result, STOP.

Figure R.7.1-6 Integrated testing strategy for the partition coefficient *n*-octanol/water



* Calculated or measured data

** Consider all other known properties of the substance. For difficult to-to-test substances, the ideal strategy may entail a comparison of calculated versus measured data (from one or more methods/QSPRs) - or an approach based on the ratio of measured water- and octanol solubility.

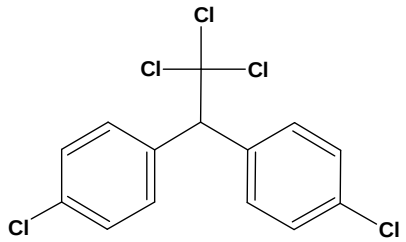
*** It is good practice to compare all measured log Kow values with predicted values.

**** for log Kow determination is not unusual to follow an iterative procedure to obtain a more accurate end result.

Examples and Case studies on the partition coefficient n-octanol/water

Kow Example: High Kow Material - p,p' DDT

Partitioning behaviour of organochlorine pesticides has been widely studied using both experimental and QSAR techniques. p,p' DDT has been studied extensively and literature reports include all common methods of determining the octanol-water partition coefficient. Pontolillo and Eganhouse (2001) have critically reviewed existing data and recognised that p,p' DDT is a difficult substance to evaluate since it is analytically difficult to measure and is expected to have a relatively high log K_{ow}. Finizio et al. (1997) applied the RP-HPLC approach to a series of pesticides and compared the results to previously published data. More recently, Shen and Wania summarised experimental data on p,p' DDT including results from slow stir (OECD 121), shake flask (OECD 107) and RP-HPLC (OECD 117) methods. These reports and the references within provide the foundation for the example.



SMILES: Clc1ccc(cc1)C(c2ccc(Cl)cc2)C(Cl)(Cl)Cl

p,p' DDT

1,1'-(2,2,2-TRICHLOROETHANE-1,1-DIYL)BIS(4-CHLOROBENZENE)
CASNO 50-29-3 EINECS 200-024-3

Both experimental data and suitable QSAR tools are available in the literature.

An Initial prediction using cLogP software indicates that the chemical will have a log P value greater than 6, suggesting the material will be difficult to test using a shake-flask approach (OECD TG 107, EU A.8) where octanol emulsion formation may produce a low bias. For a direct measurement of octanol/water partitioning either the slow stir approach or the RP-HPLC method (OECD TG 117, EU A.8) would be preferred.

Summary of Experimental Data

A subset of the experimental data reported in the literature is provided in Table R.7.1-25.

Table R.7.1-25 Subset of the experimental data for p,p'DDT

Method	Reported Result	Reference
QSAR CLogP	6.76	Daylight cLogP v4.82
ACDlabs logP	5.94	ACDlabs v4.5
ChemSilico	5.74	ChemSilico v1.6.1
SPARC	6.91	SPARC v3
Shake Flask	3.98	Kapoor (1977)
	6.19	O'Brien (1975)
	6.36	Chiou (1982)
Slow Stirring	6.91	De Bruijn (1989)
	6.20	Brooke (1986)
	6.24	Tolls (2003)
RP-HPLC	5.50	Finizio (1997)
	5.60	Brooke (1986)
	5.13	Rapaport (1984)
	6.38	Hammers (1982)
	6.21	Eadsforth (1986)

QSAR techniques are based on training sets of data and the precision of the predicted log K_{ow} from the four models selected is comparable to the variability in any single laboratory test method.

For the shake flask method, some early reported data (Kapoor, 1977) shows a low bias typically attributed to octanol emulsion formation. When emulsion formation is an issue, applying the OECD 107 guideline with different octanol:water phase ratios would typically produce highly variable results within a study, indicating octanol emulsion formation is an issue.

The slow stir method minimises this issue and provides consistent results for the highly lipophilic molecule (Tolls, 2003). Using the slow stir approach, inter-laboratory variation for p,p' DDT was higher than the 2% coefficient of variation typically observed for other compounds, however the mean result reported from eight determinations was 6.24 +/-0.05 (mean +/- standard deviation).

Reverse phase HPLC methods rely on the correlation of retention time with log K_{ow} and the availability of reference compounds to calibrate the correlation. In the case of highly lipophilic materials this method is generally reliable, especially if a calibration set of chemically similar materials is available. The reverse phase approach tends to be less reliable when highly polar functional groups are present on the test chemical but not represented within the calibration set of compounds.

R.7.1.8.6 References on *n*-octanol/water partition coefficient

Abraham M.H., Chadha H.S. and Mitchell R.C. Hydrogen bonding. 32. An analysis of water-octanol and water-cyclohexane partitioning and the log P parameter of Seiler. *J. Pharm. Sci.*(1994) 83, 1085-1100.

Bodor N, Gabanyi N.Z. and Wong C.-K. A new method for the estimation of partition coefficient. *J. Am. Chem. Soc.* (1989) 111, 3783-3786.

Brooke, D.N., Dobbs, A.J., Williams, N. "Octanol:water partition coefficients (P): Measurement, estimation, and interpretation, particularly for chemicals with P > 10^5" *Ecotoxicology and Environmental Safety* (1986) 11, pp. 251-260.

CEC (1993). Directive 67/548/EEC; Annex V Test Methods: A8: Partition Coefficient.

Chelsea MI. Nendza M. *Structure-Activity Relationships in Environmental Sciences*. Chapman & Hall, London, 1998.

Chiou, C.T., Schmedding, D.W. Manes, M. "Partitioning of organic compounds in octanol-water systems *Environ. Sci. Technol.* (1982), 16, 4-10.

Dearden JC. Molecular structure and drug transport. In Ramsden CA (Ed.), *Comprehensive Medicinal Chemistry*, vol. 4, Pergamon Press, Oxford, 1990, pp. 375-411.

Dearden J.C., Netzeva T.I., Bibby R. A comparison of commercially available software for the prediction of partition coefficient, in Ford M., Livingstone D., Dearden J. and van de Waterbeemd H. (Eds.), *Designing Drugs and Crop Protectants: Processes, Problems and Solutions*, Blackwell, Oxford, 2003, pp. 168-169.

Dearden, JC, Bresnen, GM (1988) The measurement of partition coefficients. *Quant. Struct.-Act. Relat.* 7, 133-144

Dearden JC, Bresnen GM (2005). Thermodynamics of water-octanol and water-cyclohexane partitioning of some aromatic compounds. *Int. J. Mol. Sci.* 6, 119-129.

De Bruijn, J.H.M., Busser, F., Seinen, W., Hermens, J. (1989). Determination of octanol/water partition coefficients with the 'slow stirring' method. *Environ. Toxicol. Chem.* 8, 499-512.

ECETOC (2003). QSARs: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications. Technical Report 89. ECETOC, Brussels.

Finizio, A., Vighi, M., Sandroni, D. "Determination of n-octanol/water partition coefficient (Kow) of pesticide critical review and comparison of methods" *Chemosphere,* (1997) 34, 131-161.

Ghose A.K., Pritchett A. and Crippen G.M. Atomic physico-chemical parameters for three dimensional structure directed quantitative structure-activity relationships: III. Modelling hydrophobic interactions. *J. Comput. Chem.* (1988) 9, 80-90.

GHS (2003). Globally Harmonised System of Classification and Labelling of Chemicals (GHS). United Nations, New York & Geneva.

Hansch, C. and Leo, A. (1995). Electronic Effects on Organic Reactions. In: Exploring QSAR, S. Heller, Ed., vol 1, ACS Reference Books, Washington D.C. ;

Kapoor, I.; Metcalf, R.; Hirwe, A.S.; Coats, J.R.; Khalsa, M.S.; "Structure activity correlations of biodegradability of DDT analogs" *J. Agr. Food Chem.* (1973) 21, 310-315.

Klopman G. and Wang S. A computer automated structure evaluation (CASE) approach to calculation of partition coefficient. *J. Comput. Chem.* (1991) 12, 1025-1032.

Klopman G. and Zhu H. Recent methodologies for the estimation of n-octanol/water partition coefficients and their use in the prediction of membrane transport properties of drugs. *Mini-Rev. Med Chem.* (2005) **5**, 127-133.

Leo A. Octanol/water partition coefficients. In Boethling R.S. and Mackay D. (Eds.), *Handbook of Property Estimation Methods for Chemicals.* Lewis, Boca Raton, FL, 2000, pp. 89-114.

Livingstone D.J. Theoretical property predictions. *Current Topics in Med. Chem.* (2003) **3**, 1171-1192.

Lyman W.J. Octanol/water partition coefficient. In Lyman W.J., Reehl W.F. and Rosenblatt D.H. (Eds.), *Handbook of Chemical Property Estimation Methods*. American Chemical Society, Washington DC, 1990, pp. 1.1-1.54.

Mackay, D., Shiu, W.-Y. and Ma, K.-C. (1997) "Illustrated Handbook of Physico-chemical Properties and Environmental Fate of Organic Chemicals" Lewis Publishers.

Nendza M. *Structure-Activity Relationships in Environmental Sciences*. Chapman & Hall, London, 1998.

O'Brien, R.D., "Environmental Dynamics of Pesticides" (1975) pp.336 Haque and Freed Eds, Plenum Press, New York, NY, USA.

OECD (1989). OECD Guideline for the Testing of Chemicals: 117, Partition Coefficients. By Reverse Phase Chromatography. OECD, Paris.

OECD (1995) OECD Guideline for the Testing of Chemicals: 107, Partition Coefficients. OECD, Paris.

OECD (2000). LogK_{ow} pH-metric Method for Ionisable Substances. Organisation for Economic Cooperation and Development (OECD), OECD Guideline for the Testing of Chemicals, Proposal for a new Test Guideline 122 (draft November 2000), Paris.

OECD (2003) . OECD Draft Guideline for the Testing of Chemicals: 125, Partition Coefficients. Slow-stirring method. OECD, Paris.

Pontolillo, J., Eganhouse R.P. "The search for reliable aqueous solubility (SW) and octanol-water partition coefficient data for hydrophobic organic compounds: DDT and DDE as a case study. Water Resources Investigations Report 01-4201. (2001) U.S. Geological Survey, Reston, VA USA.

Rapaport, R.A., Eisenreich, S.J. "Chromatographic determination of octanol-water partition coefficients (Kow's) for 58 polychlorinated biphenyl congeners" *Environ. Sci. Technol.* (1984) 18, 163-170.

Reinhard M. and Drefahl A. Estimating Physico-chemical Properties of Organic Compounds. Wiley, New York, 1999.

Renner, R. (2002). The K_{ow} controversy. Env. Sci. Technol., November 1, 411-413A.

Sangster, J. (1989) "Octanol-water partition Coefficients of Simple Organic Compounds" J. Phys. Ref. Data 18(3) 1111-1129.

Schwarzenbach R.P., Gschwend P.M. and Imboden D.M. Environmental Organic Chemistry. Wiley, New York, 1993.

Shen, L. and Wania, F. "Compilation, evaluation and selection of physico-chemical property data for organochlorine pesticides" *J. Chem. Eng. Data* (2005), 50, p742-768

Tolls, J., Bodo, K., De Felip, E., Dujardin, R., Kim, Y.H., Moeller-Jensen, L., Mullee, D., Nakajima, A., Paschke, A., Pawliczek, J-B., Schneider, J., Tadeo, J-L., Tognucci, A.C., Webb, J., Zwijzen, A. C.

"Slow-stirring method for determining the n-octanol/water partition coefficient (Pow) for highly hydrophobic chemicals: Performance evaluation in a ring test" *Environmental Toxicology and Chemistry* (2003), 22, pp. 1051-1057

USEPA (1985). U.S. Environmental Protection Agency. Office of Toxic Substances. Toxic Substances Control Act Test Guidelines. 50 FR 39252.

Woodburn, K.B., Doucette, W.J., Andren, A.W. (1984). Generator column determination of octanol/water partition coefficients for selected polychlorinated biphenyl congeners. Environmental Science and Technology 18:457–459.

Yalkowsky SH, Valvani SC, Roseman TJ. (1983) Solubility and partitioning. 6. Octanol solubility and octanol-water partition coefficients. *J. Pharm. Sci.* 72, 866-870.

R.7.1.9 FLASH POINT

The flash point is a key measure of the flammability of a liquid. It measures the lowest temperature at which the vapour/air mixture above the liquid can be ignited. This gives some indication of how easy it is to initiate the burning of this substance. Generally, substances with low flash points are considered to present a higher flammable risk than those with higher flash points.

Unlike, for example a melting point, the flash point is not a definitive value. Instead, it is a function of many variables. These include: the size of the sample, the heating rate, the use of an open or closed sample cup, the presence or absence of stirring, the energy and type of ignition source (e.g. spark or flame), etc. Thus, in any one standard method, these parameters are fixed. It is normal for the flash point of the same substance to vary when measured using different methods or equipment.

Definition of flash point

The flash point is the lowest temperature of the test portion (as measured in the prescribed manner), corrected to a barometric pressure of 101,3 kPa, at which application of a test flame causes the vapour of the test portion to ignite momentarily and the flame to propagate across the surface of the liquid under the specified conditions of test (standard ISO definition).

R.7.1.9.1 Information requirements on flash point

No other physical properties need be known in advance. However, for substances that are very viscous (comparable to paints, varnishes, etc) specific methods should be used. The flash point is only a relevant property for liquids, thus it does not need to be done for substances that are solids or gases at room temperature.

R.7.1.9.2 Available information on flash point

Testing data for flash point

There are many tests methods available for flash point. They are all standardised according to national or international standards. For new tests, only closed-cup methods are acceptable. However, where data already exists generated using open cup methods, the result may be acceptable if it shows a flash point significantly above the cut off values for classification into one of the flammable classes. Table R.7.1-26 lists some of the available methods.

Table R.7.1-26 Methods for measuring flash point (non-exhaustive list):

Type of method	Apparatus	Standards
Equilibrium methods	-	ISO 1516 ISO 1523 ISO 3679 ISO 3680
Non-equilibrium methods	Abel	BS 2000 part 170 NF M 07-011 NF T 66-009.
	Abel-Pensky	EN 57 DIN 51755 part 1 (temperatures from 5-65°C) DIN 51755 part 2 (temperatures below 5°C) NF M07-036.
	Pensky-Martens	ASTM D 93 BS 2000-34 DIN 51758 EN 11 ISO 2719 NF M07-019
	Setaflash	ASTM D 3278
	Tag	ASTM D 56.

To determine the flash point of viscous liquids (paints, gums and similar) only apparatus and test methods suitable for determining the flash -point of viscous liquids may be used. These are ISO 3679, ISO 3680, ISO 1523, DIN 53213 part 1.

Full details of the testing procedure are documented in the standards.

Published data on flash point

As with the other physico-chemical endpoints, information may be available from the commonly used handbooks detailed in the introduction.

R.7.1.9.3 Evaluation of available information for flash point

Experimental data on flash point

Where test data is available, this should be evaluated against the set criteria for classification and labelling. Where a result obtained with a non-equilibrium method is within 2°C of a classification cut off value then a more accurate value should be obtained using an equilibrium method. When multiple values of equal validity are available, in the interests of safety, the lowest one should be used.

Non-experimental data on flash point

The flash point is not a truly intrinsic property; it depends on the conditions under which it was carried out. Consequently, it is difficult to make a prediction based solely on the chemical structure. In theory, the flash point will be the temperature at which the vapour concentration is equal to the lower explosive limit (LEL). Thus, with prior knowledge of the LEL (and its temperature dependence) and the vapour pressure (and its temperature dependence) a flash point can be calculated. However, as the flash point is affected by the conditions under which it is measured, these calculated values rarely correlate exactly with measured values. The situation is further complicated as the LEL itself is not a truly intrinsic value and it too depends on the conditions under which it is measured.

There have been some attempts to correlate flash point with structure and/or to some other property of the substance (such as boiling point). Sometimes however, the correlations are only applicable to a narrowly defined set of substances. Examples of the different approaches can be found in publications by Butler *et al* (1956); Ellis (1976); Gramatica *et al* (2004); Hshieh (1997), Katritzky *et al* (2001); Katritzky *et al* (2001); Patil, (1988); Satyanarayana and Rao (1992); Suzuki *et al* (1991); Tetteh *et al* (1999) & Zhokhova *et al* (2003). Hagopian (1990) and Vidal *et al* (2004) have reviewed the published work.

Some of the approaches used in these publications give a mean error of around 11-20°C. The most successful methods use a molecular fragment approach to model flash point. Of course, such an approach means that predictions cannot be made for compounds that do not contain the molecular fragments used to train the model.

Good predictions were obtained by means of the fragmental approach by Zhokhova *et al* (2003). Indeed, for a training set of 266 diverse compounds and using 9 molecular fragments, they obtained $r^2 = 0.872$ and $s = 18.8°$.

Several methods have been reported for the calculation of the flash points for mixtures (McGovern 1992:1; McGovern 1992:2, Wickey & Chittenden, 1963). These methods were developed to address particular industry sectors and to reduce the need to test every possible preparation. They are not generally applicable to all substances.

*There do not appear to be any software programs available for prediction of flash point.

There are two software programs available for the prediction of flash point,* namely ACD/Labs and ProPred. No indications of their performance are available.

Calculated values can be used where the method used can be shown to be valid and the result is clearly outside of any classification ranges.

Remaining uncertainty on flash point

In the literature, different flash points are often quoted for the same substance. These differences could be due to the use of different methods etc. Where possible, the method used should be found. Values obtained using closed cup methods are preferred over open cup methods, where disagreements are significant e.g. in the vicinity of classification threshold temperatures, equilibrium procedures are preferred.

R.7.1.9.4 Conclusions on flash point

The flash point is only relevant to liquids and low melting point solids. Therefore, for many substances it is not required. This would form the basis of a suitable justification for non-testing. For those substances that are tested, the results can be used to assign a suitable hazard class. Non-experimental methods exist, but they often require data from related endpoints or are difficult to apply.

For substances within the scope of REACH, literature data or measured values may be easier to obtain than calculated values.

Concluding on C&L and Chemical Safety Assessment

The flash point is used to allocate a substance into the appropriate flammability class. Substances with flash points that fall outside the classification limits should not be designated as non-flammable as this can be misleading.

For those substances that are flammable, this will need to be taken account of in the chemical safety report.

R.7.1.9.5 Integrated testing strategy (ITS) for flash point

The tiered approach to testing (Section R.7.1.1.4) in conjunction with the choice of an appropriate test method represents an integrated testing strategy for this endpoint.

R.7.1.9.6 References on flash point

Butler R.M., Cooke G.M., Lukk G.G. and Jameson B.G. Prediction of flash points of middle distillates. *Ind. Eng. Chem.* (1956) 48, 808-812.

Ellis (1976): William H. Ellis, "Solvent Flash Points – Expected and Unexpected", *Journal of Coatings Technology,* 48**,** 44-57, (1976)

Gramatica P., Battaini F. and Papa E. QSAR prediction of physico-chemical properties of esters. *Fresenius Environ. Bull.* (2004) 13, 1258-1262.

Hagiopan J.H. Flash points of pure substances. In Lyman W.J., Reehl W.F. and Rosenblatt D.H. (Eds.), *Handbook of Chemical Property Estimation Methods*. American Chemical Society, Washington DC, (1990), pp. 18.1-18.14.

Hshieh F. Note: correlation of closed-cup flash points with normal boiling points for silicone and general organic compounds. *Fire Mat.* (1997) 21, 277-282.

Katritzky A.R., Petrukhin R., Jain R. and Karelson M. QSPR analysis of flash points. *J. Chem. Inf. Comput. Sci.* (2001) 41, 1521-1530.

McGovern (1992:1): John L McGovern, "Four Methods for Predicting the Flash Point of Alkyd Paints Containing VM&P Naphtha and Mineral Spirits – Part I", *Journal of Coatings Technology,* 64**,** 33-38, (1992)

McGovern (1992:2): John L McGovern, "A Method for Estimating the Flash Points of Coatings Containing Mixtures of Oxygenated and hydrocarbon Solvents and Petroleum Distillates – Part II", *Journal of Coatings Technology,* 64**,** 39-44, (1992)

Patil (1988): G.S. Patil, "Estimation of Flash Point", *Fire and Materials*, 12, 127-131, (1988)

Satyanarayana K. and Rao P.G. Improved equation to estimate flash points of organic compounds. *J. Hazardous Materials* (1992) 32, 81-85.

Suzuki T., Ohtaguchi K. and Koide K. A method for estimating flash points of organic compounds from molecular structures. *J. Chem. Eng. Japan* (1991) 24, 258-261.

Tetteh J., Suzuki T., Metcalfe E. and Howells S. Quantitative structure-property relationships for the estimation of boiling point and flash point using a radial basis function neural network. *J. Chem. Inf. Comput. Sci.* (1999) 39, 491-507.

Vidal M., Rogers W.J., Holste J.C. and Mannan M.S. A review of estimation methods for flash points and flammability limits. *Process Safety Progress* (2004) 23, 47-55.

Wickey & Chittenden (1963): R.O. Wickey & D.H. Chittenden "Flash Points of Blends Correlated", *Hydrocarbon Processing & petroleum REFINER",* 42 **(**6**),** 157-158, (1963)

Zhokhova N.I., Baskin I.I., Palyulin V.A., Zefirov A.N. and Zefirov N.S. Fragmental descriptors in QSPR: flash point calculations. *Russian Chem. Bull. Intl. Edn.* (2003) **52**, 1885-1892.

R.7.1.10 FLAMMABILITY

Knowledge of the flammability of a substance is an important safety consideration. Special precautions need to be taken during the handling, use and storage of flammable substances to avoid fires or explosions. Flammability is usually seen as the ease with which a substance can burn or be ignited. However, it also includes spontaneously flammability (pyrophoricity) and ignition on contact with water. The flammable properties of solids, liquids and gases are evaluated using separate procedures.

For liquids, both the pyrophoricity and flammability on contact with water are considered in this Section. However, for liquids the primary value for ease of ignition is the flash point, this is dealt with in Section R.7.1.9 and will not be discussed further in this section.

Definitions of flammability

Pyrophoricity

A substance is considered pyrophoric if, under the conditions of a standardised test, it spontaneously ignites within five minutes of being exposed to air.

Flammability on contact with water

Substances, which, in contact with water, are liable to become spontaneously flammable or emit flammable gases in dangerous quantities.

Gases: A flammable gas is a gas having a flammable range with air at 20°C and standard pressure (101.3 kPa).

Liquids: A flammable liquid is one with a flash point below the upper limit set in the classification and labelling criteria (see Section R.7.1.9 for further details on flash point).

Solids: A flammable solid is one that is readily combustible. A readily combustible solid is a powdered, granular or pasty substance that can be easily ignited by brief contact with an ignition source (such as a burning match) and the flame spreads rapidly.

It is especially difficult to extinguish a fire in metal powders. For this reason, the cut off value for the rate of propagation of burning at which they are considered flammable is lower than that for other substances (i.e. it takes longer for the burning to spread).

R.7.1.10.1 Information requirements on flammability

Pyrophoricity

For the majority of substances, pyrophoricity is not a concern and testing can be waived based on a consideration of the structure and experience in handling and use (see Section R.7.1.10.3 for examples of the type of substances that exhibit pyrophoric properties). Gases do not need to be tested.

Flammability on contact with water

For the majority of substances, flammability on contact with water is not a concern and testing can be waived based on a consideration of the structure and experience in handling and use (see Section

R.7.1.10.3 for examples of the type of substances that are flammable on contact with water). Gases do not need to be tested.

Flammability

Gases: The lower explosive limit (LEL) and upper explosive limit (UEL) should be determined or a statement that the gas is non-flammable over a full range of mixtures with air. These LEL and UEL are usually expressed as % of gas in air by volume.

Liquids: See Section R.7.1.9 for full details.

Solids: It is useful to know of any explosive properties before testing is carried out. Testing is likely to be hazardous for highly sensitive or explosive substances. Pyrophoric substances should not be tested.

The fastest burning rate should be recorded, together with the purity, physical state and moisture content of the test substance.

R.7.1.10.2 Available information on flammability

Testing data on flammability

Pyrophoricity

Experience in handling and use may provide sufficient information to indicate whether a substance is pyrophoric. For many substances, comparison with the example substances listed in Section R.7.1.10.3, which are known to be pyrophoric, would indicate that the substance is highly unlikely to display this property. Taken together, experience in handling and use and the theoretical assessment, could form the basis of a suitable justification for non-testing. When this screening procedure highlights that a substance may possess pyrophoric properties, in the interests of safe handling it is better to apply a cautious approach and perform testing.

There are several tests for pyrophoric properties. They are all essentially variations of the same basic procedure. The tests contain a separate procedure for solids and liquids.

Tests can be done according to:

Liquids:

- Test N3 Part III, sub-section 33.3.1.5 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A13 of Annex V of 67/548/EEC

Solids:

- Test N2 Part III, sub-section 33.3.1.4 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A13 of Annex V of 67/548/EEC

Flammability on contact with water

Experience in handling and use may provide sufficient information to indicate whether a substance is flammable on contact with water. For many substances, comparison with the example substances listed in Section R.7.1.10.3, which are known to be flammable on contact with water, would indicate that the substance is highly unlikely to display this property. Taken together, experience in handling and use and the theoretical assessment could form the basis of a suitable justification for non-testing. When a screening procedure highlights that a substance may be flammable on contact with water, in the interests of safe handling it is better to apply a cautious approach and perform testing.

There are several tests for flammability on contact with water. They are all essentially variations of the same basic procedure. The tests contain the same procedure for solids and liquids.

Tests can be done according to:

- Test N5 Part III, sub-section 33.4.1.4 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A12 of Annex V of 67/548/EEC

Flammability

For both gases and solids there are several tests for flammability. For obvious practical reasons the tests for gases and solids are not the same.

Gases: Tests can be done according to:

- ISO 10156-Part1
- Test A11 of Annex V of 67/548/EEC
- European Standard CEN 1839
- American Standard ASTM E 681

Solids: Tests can be done according to:

- Test N1, sub-section 33.2.1.4 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A10 of Annex V of 67/548/EEC

Full details of the testing procedure are documented in the appropriate sources.

Published data on flammability

As with the other physico-chemical endpoints, information may be available from the commonly used handbooks detailed in the introduction. Bretherick's Handbook of Reactive Chemical Hazards gives data on pyrophoric properties and flammability on contact with water.

No electronic databases that are specific to flammable properties could be found at the time of publication. The general physico-chemical sources outlined in Section R.7.1.1.2 should be used.

R.7.1.10.3 Evaluation of available information on flammability

Experimental data on flammability

The tests for pyrophoric properties, flammability on contact with water are designed as *pass or fail* tests; i.e. a positive result in one of these tests indicates that the substance has this property. The flammability tests for solids provide a burning rate. This is used by the classification criteria to assign a hazard class. For gases, any gas that shows a flammable range in air is considered flammable.

Non-experimental data on flammability

Pyrophoricity

The vast majority of substances do not display pyrophoric properties. Those that do display pyrophoric properties fall into a number of broad categories. For example, alkyl aluminium derivatives, alkyl boranes, alkyl haloboranes, alkyl halophosphines, alkyl metals, alkyl phosphines, alkyl silanes, aryl metals, boranes, metal hydrides, complex acetylides freshly-produced finely-divided metal powders (aged powders can develop a coating of oxide that prevents further reaction). A more comprehensive list can be found in Bretherick's Handbook of Reactive Chemical Hazards.

Flammability on contact with water

The vast majority of substances do not ignite or liberate flammable gases on contact with water. Those that do display these properties fall into a number of broad categories. For example, alkali metals, alkyl aluminium derivatives, alkyl metals, metal hydrides, metal phosphides, certain metal powders. A more comprehensive list can be found in Bretherick's Handbook of Reactive Chemical Hazards.

Flammability

For both solids and gases, the flammability is not a truly intrinsic property; it depends on the conditions under which the testing is carried out. Consequently, it is difficult to make a prediction based solely on the chemical structure.

Gases: In theory, any gas that can react with the oxygen in air to form an oxide could be flammable. This includes lower oxides reacting to form higher oxides, e.g. carbon monoxide to carbon dioxide. However, certain gases are known to be inert to combustion, e.g. the noble gases, carbon dioxide.

Care should be taken not to confuse the inability to support combustion with the inability to burn.

Solids: As any organic solid is capable of combustion, there are no methods available for screening out those that will give negative results. Inorganic oxides in which the inorganic element is in its highest possible oxidation state are incapable of further reaction with oxygen and can thus be designated as non-flammable.

Remaining uncertainty on flammability

The procedures for assessing pyrophoric properties, flammability of contact with water and general flammability are generally *pass or fail* tests. There is generally little uncertainty in the results. The only uncertainty comes from the chemical structure screening procedures for pyrophoric properties and flammability of contact with water. Here, if the result is uncertain then, in the interests of safety, testing should be conducted.

R.7.1.10.4 Conclusions on flammability

For many substances, the absence of structural alerts will mean that testing is not necessary for pyrophoric properties and flammability of contact with water. This would form the basis of a suitable justification for non-testing. For flammability, the results of testing can be used to assign a suitable hazard class.

Concluding on C&L and chemical safety assessment

The flammable properties tests are designed to allocate a substance into the appropriate hazard class. The results of testing can be used to allocate a suitable hazard class.

R.7.1.10.5 Integrated testing strategy (ITS) for flammability

The screening procedures above represent an intelligent testing strategy for some aspects of flammability. If applied correctly then only substances it is suspected will give a positive result in either the pyrophoric properties or flammability of contact with water tests will need to be tested. In the absence of a suitably justified read-across to another substance, flammability testing for solids and gases will usually have been carried out already or will need to be conducted.

Examples and case studies

Examples of substances with different types of flammable properties (non-exhaustive list) are given in Table R.7.1-27.

Table R.7.1-27 Examples of substances with different types of flammable properties

Name	Flammability
aluminium phosphide	Contact with water
calcium	Contact with water
calcium hydride	Contact with water
diethyl(ethyldimethylsilanolato)aluminium	Pyrophoric, Contact with water
diethylzinc	Pyrophoric
dimethylzinc	Pyrophoric
di-n-octylaluminium-iodide	Pyrophoric
ethyl propoxy aluminium chloride	Contact with water
Lithium	Contact with water
lithium tetrahydridoaluminate	Contact with water
magnesium (powder or turnings)	Contact with water
n-hexyllithium	Pyrophoric, Contact with water
phosphine	Pyrophoric
potassium	Contact with water
potassium mu-fluoro-bis(triethylaluminate)	Contact with water
sodium	Contact with water
sodium hydride	Contact with water
tert-butylarsine	Pyrophoric
tricalcium diphosphide	Contact with water
trichlorosilane	Pyrophoric
trimagnesium diphosphide	Contact with water
trizinc diphosphide	Contact with water
white phosphorus	Pyrophoric
aluminium phosphide	Contact with water
cadmium (pyrophoric)	Pyrophoric
calcium	Contact with water
calcium hydride	Contact with water
diethyl(ethyldimethylsilanolato)aluminium	Pyrophoric, Contact with water
diethylzinc	Pyrophoric
dimethylzinc	Pyrophoric
di-n-octylaluminium-iodide	Pyrophoric
ethyl propoxy aluminium chloride	Contact with water
Lithium	Contact with water
lithium tetrahydridoaluminate	Contact with water
magnesium (powder or turnings)	Contact with water
magnesium powder (pyrophoric)	Pyrophoric
n-hexyllithium	Pyrophoric, Contact with water
phosphine	Pyrophoric

Name	Flammability
potassium	Contact with water
potassium mu-fluoro-bis(triethylaluminate)	Contact with water
sodium	Contact with water
sodium hydride	Contact with water
tert-butylarsine	Pyrophoric
tricalcium diphosphide	Contact with water
trichlorosilane	Pyrophoric
trimagnesium diphosphide	Contact with water
trizinc diphosphide	Contact with water
white phosphorus	Pyrophoric
zinc powder or dust (pyrophoric)	Pyrophoric, Contact with water

R.7.1.10.6 References on flammability

Bretherick (1999) Bretherick's Handbook of Chemical Reactive Hazards: An Indexed Guide to Published Data, 6th Edition (2 volume set). P Urben and L Bretherick (Authors). Butterworth Heinemann

R.7.1.11 EXPLOSIVE PROPERTIES

Knowledge of explosive properties is an important safety consideration. If explosive substances are handled incorrectly then there can be serious consequences. Some substances are designed to be explosive, for these, data on their explosive properties will usually be available. Other substances however, may possess explosive properties, even though this was not intended. Therefore, screening for explosive properties should be carried out for all substances.

Definition of explosivity

Explosivity can be defined as the tendency of a substance to undergo violent and rapid decomposition, under appropriate conditions, to produce heat and or gas. Whether or not a substance with explosive properties can cause an explosion depends on a number of factors. These include: the degree of confinement, the strength of the container it is in, the rate of heating, the nature of the initiation source, etc. To overcome these variables standard tests have been devised in which the parameters are fixed.

R.7.1.11.1 Information requirements on explosivity

For the majority of substances, explosivity is not a concern and testing can be waived based on a consideration of the structure. Gases do not need to be tested and liquids do not need to be tested for sensitivity towards friction.

R.7.1.11.2 Available information on explosivity

Electronic databases on explosivity

No electronic databases that are specific to explosive properties could be found at the time of publication. The general physico-chemical sources outlined in Section R.7.1.1.2 should be used.

Testing data for explosivity

There are many tests for explosive properties that provide information on the degree of sensitivity of the substance and the consequences of initiation (severity of the event). These cover different aspects of the explosive properties, for example: detonability, deflagration and the nature of the sensitivity (i.e. thermal, impact or friction sensitive).

Tests can be done according to:

- Test Series 1 to 8, Part I of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A.14 of Annex V of 67/548/EEC

Thermal analysis can provide supplementary data (Differential Thermal Analysis, Differential Scanning Calorimetry).

Full details of the testing procedure are documented in the appropriate sources.

Published data for explosivity

As with the other physico-chemical endpoints, information may be available from the commonly used handbooks detailed in the introduction.

No electronic databases that are specific to explosive properties could be found at the time of publication. The general physico-chemical sources outlined in Section R.7.1.1.2 should be used.

R.7.1.11.3 Evaluation of available information on explosivity

Where test data is available, this should be evaluated against the set criteria for classification and labelling. When the screening procedure highlights that a substance may possess explosive properties, in the interests of safe handling, it is better to apply a cautious approach and perform testing.

Experimental data on explosivity

The tests for explosivity are designed to give results that can be evaluated directly against the criteria for classification & labelling[13]. Thermal analysis (Differential Thermal Analysis, Differential Scanning Calorimetry) can provide supplementary data such as, evidence of exothermic decomposition, rate of energy release etc. This data can be used to help interpret the thermal behaviour of a substance.

Full details of the testing procedure are documented in the appropriate sources.

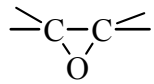
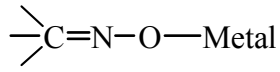
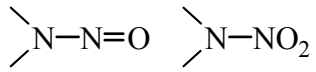
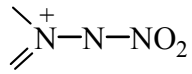
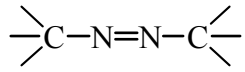
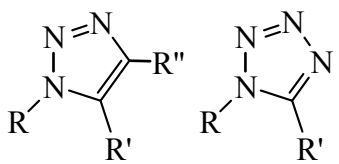
[13] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS). see Chapter R.7, Introduction.

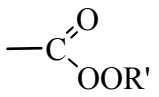
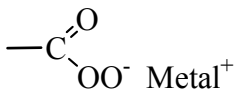
In all cases the first step should be the screening procedure detailed in the subsequent chapter. When the screening procedure highlights that a substance may possess explosive properties, in the interests of safe handling, it is better to apply a cautious approach and perform testing.

Non-experimental data on explosivity

Experience with explosive substances has provided us with a list chemical groups that are known to contribute to explosivity, some of these are shown in Table R.7.1-28. Substances that contain multiple groups from the list below will be more likely to be explosive. The list is not exhaustive but contains some of the more common groups.

Table R.7.1-28 Chemical groups associated with explosive properties

Chemical group	Chemical Class
-C≡C- -C≡C-Metal -C≡C-Halogen	Acetylenic Compounds Metal Acetylides Haloacetylene Derivatives
	Diazo Compounds
-N=O -NO_2	Nitroso and Nitro Compounds,
R-O-N=O R-O-NO_2	Acyl or Alkyl Nitrites and Nitrates
	1,2-Epoxides
	Metal Fulminates or *aci*-Nitro Salts
	N-Metal Derivatives (especially heavy metals)
	N-Nitroso and N-Nitro Compounds
	N-Azolium Nitroimidates
	Azo Compounds
Ar-N=N-O-Ar	Arene Diazoates
$(ArN=N)_2O$, $(ArN=N)_2S$	Bis-Arenediazo Oxides and Sulfides
RN=N-NR'R''	Triazines
	High-nitrogen Compounds: e.g. Triazoles, Tetrazoles

Chemical group	Chemical Class
[1] ROOR', [2] 	Peroxy Compounds: [1] Alkyl hydroperoxides (R'=H), Peroxides (R'=organic); [2] Peroxo acids (R'=H), Peroxyesters (R'=organic)
[1] ROOMetal, [2] 	Metal peroxides, Peroxoacids salts
$-N_3$	Azides e.g. PbN_6, CH_3N_3
^{-}O——C–N_2^{+}	Arenediazonium oxides i.e. inner diazonium salts in which the counter ion is an oxide
Ar-N=N-S- Ar-N=N-S-Ar	Diazonium sulfides and derivatives, Arenediazo Aryl Sulfides
XO_n	Halogen Oxide: e.g. percholrates, bromates, etc
NX_3 e.g. $NC1_3$, $RNC1_2$	N-Halogen Compounds

Adapted from Bretherick (Bretherick's Handbook of Reactive Chemical Hazards 6th Ed., 1999, Butterworths, London).

Screening Procedure:

The screening procedure for explosive properties is based on the use of both theoretical considerations and, where necessary, experimental data.

STEP 1

The first stage is to examine the chemical structure and check for the presence of groups associated with explosive properties. If the substance does not contain any groups associated with explosivity then a negative result is likely. It should be noted that certain groups could be said to be:

- Directly concerned with the explosive property e.g. nitrate ester, aromatic nitro, aliphatic nitro, nitramine, azide, nitroso, perchlorate, acetylides etc.
- Able to contribute to the explosive property, when present alongside groups directly associated with explosivity e.g. hydroxyl, carbonyl, ether, amino, sulphonic acid, etc.
- Able to contribute to the explosive property e.g. hydroxyl, carbonyl, ether, amino, sulphonic acid, etc.

For substances that contain one or more groups associated with explosivity, then further evaluation is required and testing should be considered.

When the substance contains chemical groups associated with explosive properties, and if oxygen is present in the molecule, calculate the oxygen balance (OB) according to the chemical reaction and mathematical equation below (Lothrop *et al*):

$$C_xH_yO_z + [x + (y/4)-(z/2)].\ O_2 \rightarrow x\ CO_2 + (y/2)\ H_2O$$

using the formula:

$$OB = -1600\ [2x +(y/2) -z] / \text{molecular weight.}$$

Where: x = Number of carbon atoms y = Number of hydrogen atoms

z = Number of oxygen atoms

(The number of nitrogen atoms present is neglected.)

For example, for Trinitrotoluene (TNT) $C_7H_5N_3O_6$

Mol. Wt.: 227.13

OB = -1600 [14 + 2.5 –6] / 227.13 = -74

If groups associated with explosive properties are present but the oxygen balance is less than –200 then testing does not need to be conducted and a negative result can be predicted.

Although oxygen balance (OB) is a good indicator of potential explosive instability, it should not be used in isolation. For example, the OB for water is 0, yet it is clearly not an explosive, nitro-glycerine on the other hand has an OB of +3.5 and is well known for its explosive properties. For an OB calculation to be valid, the substance must contain some oxygen. If the molecule contains groups that have multiple nitrogens, e.g. azides, tetrazoles etc. then an oxygen balance calculation is not always reliable. This is because these substances decompose to form high volumes of nitrogen gas that can lead to a pressure-induced explosion, i.e. the rapid expansion of the gas causes the container to blow apart.

STEP 2

If the substance contains chemical groups associated with explosive properties and is not excluded by the theoretical examination in step 1 then examine any experimental data which may give a qualitative or quantitative indication of possible explosive behaviour. For example, differential scanning calorimetry (DSC) or differential thermal analysis (DTA)) can give information on the decomposition energy and the decomposition temperature. If the exothermic decomposition energy is more than 500 J/g and the onset of exothermic decomposition is below 500ºC. Where calorimetry is used, the procedure should involve a relatively slow heating rate, e.g. 5 K/min or less.

If the screening procedure identifies the material as having the potential to possess explosive properties, or there is any doubt, then testing should be carried out.

Remaining uncertainty

Some of the tests for classification as explosive, especially the UN tests, are performed on the substance *as packaged*. Packaging can vary greatly and this will have an influence on the test results. For example, a substance that is relatively safe when packaged in soft packaging, could be explosive when packaged in hard packaging or if under high confinement in a plant or a large pile. This introduces a certain degree of uncertainty in the result. Standardised tests, in which all parameters are set and the only variable is the substance, give the most reliable result.

R.7.1.11.4 Conclusions on explosivity

For many substances, the absence of structural alerts will mean that testing is not necessary. This would form the basis of a suitable justification for non-testing. For those substances that are tested, the results can be used to assign a suitable hazard class.

Concluding on C&L and Chemical Safety Assessment

The explosive properties tests are designed to allocate an explosive substance into the appropriate hazard class. Substances for which it is justified not to test can be designated as non-explosive. There are some substances, which although they do not fall into one of the explosive hazard classes, are on the borderline of being explosive. Consideration should be given to applying a suitable warning phrase to these substances, (e.g. "R44 Risk of explosion if heated under confinement" from Directive 67/548/EEC)

For those substances that are explosive, (either in the packed or unpacked state) this will need to be taken account of in the chemical safety report.

R.7.1.11.5 Integrated testing strategy (ITS) for explosive properties

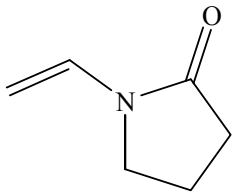
The screening procedures above represent an intelligent testing strategy for explosive properties. If applied correctly, only substances that will give a positive result in one of the explosive properties tests will need to be tested. In addition, Test Series 2 to 8 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria" is designed such that it contains a structured approach to testing. Together with thermal analysis techniques, this constitutes the testing strategy

Examples and Case studies on explosivity

The substance 1-vinyl-2-pyrrolidone [A] (CAS: 88-12-0; EC: 201-800-4) contains none of the groups associated with explosive properties. Testing for explosive properties would be not need to be carried out. A suitable justification statement could be:

"Examination of the structure indicates that there are no groups associated with explosive properties. Therefore, negative results can be predicted and no testing for explosive properties has been carried out."

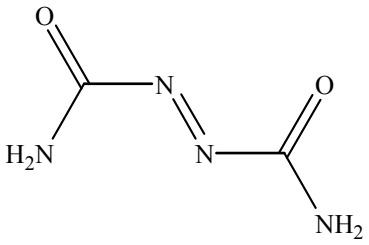
[A] Structure of 1-vinyl-2-pyrrolidone



Example 2: A substance with marginal explosive properties

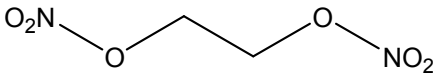
The substance azobiscarbonamide [B] (CAS: 123-77-3; EC: 204-650-8) contains an azo group; one of the alert groups for explosive properties. Thus, following the screening procedure, one continues to calculate the oxygen balance. The molecular formula is $C_2H_4N_4O_2$ and the molecular weight is 116.08. These give an oxygen balance of –55.13. This is more positive than the –200 threshold. Testing according to method A14 of Annex V of Directive 67/548/EEC gave a negative result [IUCLID 1]. However, testing of material from a different manufacturing source did give some indication of explosive properties. These differences may have been due to differences in particle size (smaller particles decompose more rapidly that larger ones, thus increasing any instability). The substance has been given the EU R phrase "R44 Risk of explosion if heated under confinement" from Directive 67/548/EEC.

[B] Structure of azobiscarbonamide



Example 3: A substance with clear explosive properties

The substance ethylene dinitrate [C] (CAS: 628-96-6; EC: 211-063-0) contains multiple structural alerts for explosive properties. Thus, following the screening procedure, one continues to calculate the oxygen balance. The molecular formula is $C_2H_4N_2O_6$ and the molecular weight is 152.06. These give an oxygen balance of 0 (zero). This is more positive than the –200 threshold and in theory the substance is able to fully combust in the absence of air. Testing according to method A14 of Annex V of Directive 67/548/EEC gave positive results for thermal and mechanical sensitivity (both shock and friction) [IUCLID 2].



R.7.1.11.6 References on explosive properties

Bretherick (1999) Bretherick's Handbook of Chemical Reactive Hazards: An Indexed Guide to Published Data, 6th Edition (2 volume set). P Urben and L Bretherick (Authors). Butterworth Heinemann

W. C. Lothrop and G. R. Handrick. "The relationship between performance and constitution of pure organic explosive compounds." *Chemical Reviews*, 44:419–445, 1949.

IUCLID 1: IUCLID data sheet for azobiscarbonamide; available at http://ecb.jrc.it/IUCLID-DataSheets/123773.pdf

IUCLID 2: IUCLID data sheet for ethylene dinitrate; available at http://ecb.jrc.it/IUCLID-DataSheets/628966.pdf

R.7.1.12 SELF-IGNITION TEMPERATURE

The self-ignition temperature of a substance can be used to assess situations in which a substance can spontaneously catch fire; for example, a liquid dripping onto a hot surface.

Unlike, for example, the melting point, the self-ignition temperature is not a definitive value. Instead, it is a function of many variables. These include, the size of the sample, the nature of the substance (solid or liquid), the way the substance is confined. Thus, in any one standard method, these parameters are fixed. For solids the self-ignition temperature will also depend on the particle size. For liquids and gases, the value is known as the auto-ignition temperature; for solids, it is the relative self-ignition temperature.

Definition of self-ignition temperature

Gases & Liquids: The auto-ignition temperature is the lowest temperature at which a substance will ignite when mixed with air under the conditions defined in the test method.

Solids: The relative self-ignition temperature is the minimum temperature at which a certain volume of a substance will ignite under defined conditions.

R.7.1.12.1 Information requirements on self-ignition temperature

Testing should not be conducted for explosive or pyrophoric substances. For solids, the melting point should be known beforehand, as substances that melt at <160°C do not need to be tested. Liquids with a flash point above 200°C do not need to be tested. Gases that have no flammable range in air do not need to be tested.

R.7.1.12.2 Available information on self-ignition temperature

Testing data for self-ignition temperature

Gases & Liquids: There are several tests for gases and liquids. They are all essentially variations of the same basic procedure.

Tests can be done according to:

- Test A15 of Annex V of Directive 67/548/EEC
- International Electrotechnical Commission: IEC 79-4
- German Standard: DIN 51794
- American Standard: ASTM-E 659-78
- British Standard: BS 4056
- French Standard: NF T 20-037

Solids: The UN transport tests include some tests for self-heating substances. However, these are designed to assess the suitability of a substance for transport purposes and to assign packing groups. Information from these tests may be useful for REACH, especially when self-ignition has taken place. The standard test for relative self-ignition temperature in Annex V of Directive 67/548/EEC

measures the temperature at which self-ignition occurs by continuous heating of the substance until either the substance ignites or the limit value is reached.

Tests methods:

- Test N4 Part III, sub-section 33.4.1.6 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A.16 of Annex V of Directive 67/548/EEC

Published data on self-ignition temperature

As with the other physico-chemical endpoints, information may be available from the commonly used handbooks detailed in the introduction.

No electronic databases that are specific to self-ignition temperature could be found at the time of publication. The general physico-chemical sources outlined in Section R.7.1.1.4 should be used.

R.7.1.12.3 Evaluation of available information for self-ignition temperature

Experimental data on self-ignition temperature

Test data for auto-ignition temperature or relative self-ignition requires little evaluation. The temperature at which self-ignition takes place should be reported. For solids, observations of the sample behaviour should be made, notably decomposition or melting.

Non-experimental data on self-ignition temperature

There have been only a very few publications concerning the prediction of self-ignition temperature (SIT). Taskinen and Yliruusi (2003) have reviewed the available literature. Tetteh *et al* (1996, 1998) used radial basis function neural networks to model a set of 232 organic chemicals with 13 different functional groups. They obtained a mean test set error of 33°C.

Mitchell and Jurs (1997) used their ADAPT software to model SIT values of a data set of 327 diverse organic chemicals. They were unable to obtain good correlations for the whole data set, but found improved correlations when the chemicals were divided into hydrocarbons, nitrogen compounds, oxygen/sulphur compounds and alcohols/ethers. For example, for alcohols and ethers they obtained a 6-descriptor QSPR with r^2 = 0.854 and RMS error = 35.0°C. The authors commented that their prediction errors were in the range of experimental errors.

Neither of the two approaches given above is very amenable to general use, so it is unfortunately the case that there is at present no simple method available for the prediction of auto-ignition temperature. No software is available for the prediction of auto-ignition temperature.

Remaining uncertainty on self-ignition temperature

In general, due to the lack of non-experimental models, self-ignition temperature data will have been generated by a standard test method. The only uncertainty is therefore the inherent measurement error of the methods.

R.7.1.12.4 Conclusions on self-ignition temperature

The self-ignition temperature will usually be determined experimentally. The temperature measured in the test should be reported.

Concluding on C&L and chemical safety assessment

Both the relative self-ignition for solids and the auto-ignition temperature for liquids are not used directly for classification and labelling. However, they can be used in the chemical safety assessment in considering risks associated with processing the substance. Information from the UN tests can be used for transport purposes; however, this is outside of the context of this guide.

R.7.1.12.5 Integrated testing strategy (ITS) for self-ignition temperature

The tiered approach to testing (Section R.7.1.1.4) in conjunction with the choice of an appropriate test method represents an integrated testing strategy for this endpoint.

R.7.1.12.6 References on self-ignition temperature

Mitchell B.E. and Jurs P.C. Prediction of autoignition temperatures of organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* (1997) **37**, 538-547.

Taskinen J. and Yliruusi J. Prediction of physico-chemical properties based on neural network modelling. *Adv. Drug. Delivery Rev.* (2003) **55**, 1163-1183.

Tetteh J., Metcalfe E., Howells S. Optimisation of radial basis and backpropagation neural networks for modelling auto-ignition temperature by quantitative structure-property relationships. *Chemometr. Intell. Lab. Syst.* (1996) **32**, 177-191.

Tetteh J., Howells S., Metcalfe E. and Suzuki T. Optimisation of radial basis function neural networks using biharmonic spline interpolation. *Chemometr. Intell. Lab. Syst.* (1998) **41**, 17-29.

R.7.1.13 OXIDISING PROPERTIES

Knowledge of the oxidising properties of a substance is an important safety consideration. If oxidising substances are handled or stored incorrectly then there can be serious consequences, such as a fire or explosion.

Chemical oxidation always takes place alongside a reduction process. This means that the *strength* of a substance's oxidising properties is relative; it depends upon the relative ease with which each reactant can be oxidised or reduced respectively. Not all substances that have oxidising properties are hazardous; some will be mildly oxidising only and present very little hazard. To distinguish those that are hazardous, a substance's oxidising properties are compared to those of a standard reference substance.

Definition of oxidising properties

An oxidising substance is one that, while in itself not necessarily combustible, may cause or contribute to the combustion of other material.

Oxidising substances generally act by yielding oxygen; however, those that release active halogens can also act as oxidants.

R.7.1.13.1 Information requirements on oxidising properties

For the majority of substances, oxidising properties are not a concern and testing can be waived based on a consideration of the structure. For solids, testing should not be performed on explosive or highly flammable substances. Organic peroxides form a separate class of substances that are always oxidising.

R.7.1.13.2 Available information on oxidising properties

Testing data on oxidising properties

For practical reasons the tests for oxidising properties are different for solids, liquids and gases.

Solids: Tests can be done according to:

- Test O.1 in Part III, sub-section 34.4.1 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A.17 of Annex V of Directive 67/548/EEC

Full details of the testing procedures are documented in the appropriate sources.

Liquids: Tests can be done according to:

- Test O.2 in Part III, sub-section 34.4.2 of the "UN Recommendations on the Transport of Dangerous Good, Manual of Tests and Criteria"
- Test A.21 of Annex V of Directive 67/548/EEC

Full details of the testing procedures are documented in the appropriate sources.

Gases: Assessments can be done according to:

- ISO 10156: 1996

A full detail of the assessment procedure is documented in the appropriate source.

Published data on oxidising properties

As with the other physico-chemical endpoints, information may be available from the commonly used handbooks detailed in the introduction.

No electronic databases that are specific to oxidising properties could be found at the time of publication. The general physico-chemical sources outlined in Section R.7.1.1.4 should be used.

R.7.1.13.3 Evaluation of available information on oxidising properties

Where test data are available, this should be evaluated against the set criteria for classification and labelling. When the screening procedure highlights that a substance may possess oxidising properties, in the interests of safe handling, it is better to apply a cautious approach and perform testing.

Experimental data on oxidising properties

The tests for oxidising properties are designed as *pass or fail* tests; i.e. a positive result in one of these tests indicates that the substance has this property. For oxidising properties, the results are always compared to a standard reference substance.

Non-experimental data on oxidising properties

Experience with oxidising substances has provided us with a list of chemical groups and elements that are known to confer oxidising properties. A non-exhaustive list of these is given in Table R.7.1-29. In general terms, one can perform a screening procedure to filter out substances that are unlikely to be oxidising and for which testing can be waived.

For organic substances (with the exception of peroxides) testing does not need to be carried out for substances if:

- The substance does not contain oxygen, fluorine or chlorine; or
- The substance contains oxygen, fluorine or chlorine and these elements are chemically bonded only to carbon
- Inorganic substances that do not contain oxygen of halogens do not need to be tested.

Organic peroxides form a special class of substance. They are always treated as oxidisers, whereas in reality, they tend to be self-reactive and to be explosive by auto-oxidation (i.e. they oxidise themselves).

If the screening procedure identifies the material as having potential oxidising properties, or there is any doubt, then, in the interests of safety, testing should be carried out.

Table R.7.1-29 Chemical groups associated with oxidising properties*

Chemical group	Chemical Class
Nitrates (salts or esters)	NO_3-M^+
	O_2N-O-R
Nitrites (salts or esters)	NO_2-M^+
	ON-O-R
Organic nitro compounds	
Nitroalkyl	NO_2-R
Nitroaryl	NO_2-Ar
Fluorodinitro	$(NO_2)_2$-C-(F)-
Metal oxides	MO_n
Metal oxometallates	$M^+MO_n^-$
N - Halogen compounds	N-X
N – Haloimides	-C(O)-NX-C(O)-
Difluoroamino	- NF_2
Difluoroaminopolynitroaryl	(NO_2)n-Ar-NF_2
Oxohalogen compounds:	
Acyl hypohalites	R C(O)-OX
Hypofluorites	FO^-
Bis(fluoroxy)perhaloalkanes	$F_3CCI(OF)_2$ etc
Perchlorates	ClO_4^-
Chlorates	ClO_3^-
Chlorites	ClO_2^-
Hypochlorites	ClO^-
Perbromates	BrO_4^-
Bromates	BrO_3^-
Bromites	BrO_2^-
Hypobromites	BrO^-
Periodates	IO_4^-
Iodates	IO_3^-
Difluoroperchloryl salts	$F_2ClO_2^+Z^-$
Dioxygenyl polyfluoro salts	O_2^+ $[MF_n]^-$ or O_2^+ $[EF_n]^-$
Interhalogen compounds:	
Metal polyhalohalogenates	M+ [XX'n]$^-$

* Adapted from Bretherick's Handbook of Chemical Reactive Hazards

Remaining uncertainty on oxidising properties

The test procedure for solids involves an assessment of enhanced burning. In some instances, testing appears to give a positive result when, in reality, the substance has no oxidising character. This is often due to the substance burning or some form of chemical reactivity of the test substance. In these cases, the test should be repeated using an inert substance, e.g. diatomite (kieselguhr), to replace the cellulose (or other combustible material). If a positive result is still obtained then the

substance does not possess oxidising properties (as diatomite cannot be oxidised, any observed reaction cannot be due to oxidation).

R.7.1.13.4 Conclusions on oxidising properties

For many substances, the absence of structural alerts will mean that testing is not necessary. This would form the basis of a suitable justification for non-testing. For those substances that are tested, the results can be used to assign a suitable hazard class.

Concluding on C&L and chemical safety assessment

The oxidising properties tests are designed to allocate an oxidising substance into the appropriate hazard class by comparison to one or more reference substance. Substances for which it is justified not to test can be designated as non-oxidising.

For those substances that are oxidising, this will need to be taken account of in the chemical safety report.

R.7.1.13.5 Integrated testing strategy (ITS) for oxidising properties

The screening procedures above represent an intelligent testing strategy for oxidising properties. If applied correctly, only substances which, it is suspected, will give a positive result in one of the oxidising properties tests will need to be tested. Together with the choice of an appropriate test method, this constitutes the testing strategy.

R.7.1.13.6 References on oxidising properties

Bretherick (1999) Bretherick's Handbook of Chemical Reactive Hazards: An Indexed Guide to Published Data, 6th Edition (2 volume set). P Urben and L Bretherick (Authors). Butterworth Heinemann

R.7.1.14 GRANULOMETRY

The CEN document, EN 481 "Workplace Atmospheres – size fraction definitions for measurement of airborne particles" (CEN 1993) provides satisfactory definitions of the inhalable, thoracic and respirable size fractions, and target specifications (conventions) for sampling instruments to measure these fractions. This standard will shortly be revised by CEN and based on calm air situations. The current standard defines sampling conventions for particle size fractions which are to be used in assessing the possible health effects resulting from inhalation of airborne particles in the workplace. The different particle sizes defined in EN 481 are:

- inhalable fraction (the mass fraction of particles that can be inhaled by nose and mouth. Particles >100 μm are not included in the inhalable convention
- thoracic fraction (the mass fraction of particles that passes the larynx). It has been shown that 50% of the particles in air with an aerodynamic diameter of 10 μm belong to the thoracic fraction.
- respirable fraction (the mass fraction of particles that reaches the alveoli) It has been shown that 50% of particles with an aerodynamic diameter of 4 μm belong to the respirable fraction

The requirement for the result from this test is linked closely to the inhalation toxicity strategy and the need to decide which route of administration is most appropriate for the acute toxicity and 28-day base set studies. The strategy states that an important argument in favour of the performance of the inhalation toxicity studies is the following *"substance as used containing particles in the inhalable size range (i.e. may be deposited anywhere in the respiratory tract; the inhalable size range of particles is important in determining not only if the situation poses an inhalation problem, but also where in the respiratory tract the particles may deposit)"*. Therefore, the particle size distribution can be used as an argument in favour of inhalation testing. Particle size is also a factor in environmental exposure assessment.

Methods capable of particle size distribution measurement can determine the appropriate fractions as defined in EN481 (CEN 1993), using the aerodynamic diameter of a particle, which is the measure of its behaviour in air, as the basis of the measurement. Aerodynamic diameter is a form of "equivalent diameter" often used in aerosol measurement. The aerodynamic diameter of a particle is defined as the diameter of a sphere of density 1g cm^{-3} with the same terminal velocity (falling speed) due to gravitational force in calm air as the particle under the prevailing conditions of temperature, pressure and relative humidity (CEN, 1993). For particles of aerodynamic diameter less than 0.5 μm, the particle diffusion diameter should be used instead of the particle aerodynamic diameter. In particular this holds for nanoparticles <100 nm. It is important to recognise that workers may be exposed to nanomaterial, during its production, use and disposal; their potential risks have been reviewed (Borm et al., 2006). In general, occupational hygiene has largely focussed on exposure from the inhalable route based on the general belief that this was generally the highest in terms of risk. Hence an understanding of aerosol behaviour is necessary and the particle size distribution is important information. Although aerosol science is well understood, the commonly used granulometric tests used to determine particle size ranges may not be appropriate for nanoparticles (Aitken et al., 2004). For diffusion, the appropriate *equivalent diameter* is the diffusion (mobility) diameter. This is defined as the diameter of a sphere with the same diffusion coefficient as the particle under the prevailing conditions of temperature, pressure and relative humidity.

Definition of granulometry

Details of methods for determining particle size distribution and for fibre length and diameter distributions are outlined in OECD TG 110 and HSE Guidance document on methods for measuring particle size distribution (1996).

The parameter of interest is the effective hydrodynamic radius, or effective Stoke's radius R_s. The terminal velocity of a small sphere falling under the influence of gravity in a viscous fluid is given by:

$$\upsilon = 2_g R_s^2 (d_1 - d_2) / 9\eta$$

where υ=velocity (m/sec)

g=gravitation constant (m/sec^2)

R_s=Stokes radius (m)

d_1=density of sphere (kg/m^3)

d_2=density of fluid (kg/m^3)

η=dynamic viscosity (N sec/m^2 = Pa s) of the fluid

In other situations, similar relationships apply. Particle size is usually measured in micrometers (= 10^{-6} m).

R.7.1.14.1 Information requirements on granulometry

The study does not need to be conducted if the substance is marketed or used in a non solid or non granular form. Particle size distribution (effective hydrodynamic radius) requires information on water insolubility. Fibre length and diameter distributions require information on the fibrous nature of the product and on stability of the fibrous shape under electron microscope conditions.

R.7.1.14.2 Available information on granulometry

Testing data on granulometry

Many methods are available for particle size measurements, but none of them is applicable to the entire size range (see Table R.7.1-30). Sieving, microscopic sedimentation and elutriation techniques are most commonly employed. Methods for determining particle size distribution are designed to provide information on the transportation and sedimentation of insoluble particles in water and air.

These methods are generally applicable and frequent in use. They are used to calculate the effective hydrodynamic radius of both fibrous and non-fibrous particulates without prior inspection indirectly from other measurements of particle size and density. If applied properly, they represent an estimate of the aerodynamic property and mass fractions present and as such can indicate whether or not respirable particles may be present. They are applicable to water insoluble (i.e. water solubility < 10^{-6} g/l) substances and cover the range 2-100 μm.

In the case of materials which can form fibres; which is initially confirmed using light microscopic examination to determine the approximate nature of the particles (e.g. plates, needles, etc.), an additional set of measurements is recommended to help identify the potential health hazards arising from inhalation or ingestion. This is comparatively specialised, infrequently required and involves specialised microscopic examination (e.g. TEM, SEM). A fibre is a water insoluble particle with an aspect ratio (length/diameter > 3) and diameter < 100 μm. Fibres of length < 5 μm need not be considered.

Image analysis of particle size and shape measurements can be used to determine the aspect ratios of fibrous particles. Image analysis generates data by capturing direct images of each particle. This provides users with the ultimate sensitivity and resolution as subtle differences in particle size and particle shape can be accurately characterised. Images of each individual particle are also recorded, providing a further visual verification of the data and also enabling detection of important phenomena such as agglomeration, breakage and foreign particles. A range of industries (e.g. pharmaceuticals, biotechnology, abrasives, ceramics, polymers, explosives and toners) are increasingly using image analysis systems in order to characterise their products. This method provides a vast amount high-quality information through image analysis, such as:

- Particle size
- Particle shape
- Particle count and foreign particle detection

- Captured images providing information on optical density, homogeneity and particle structure

An integrated testing strategy (ITS) detailing the appropriate methods for determination of particle size distribution of respirable and inhalable particles is shown in Figure R.7.1-7.

Figure R.7.1-7 Integrated testing strategy for granulometry

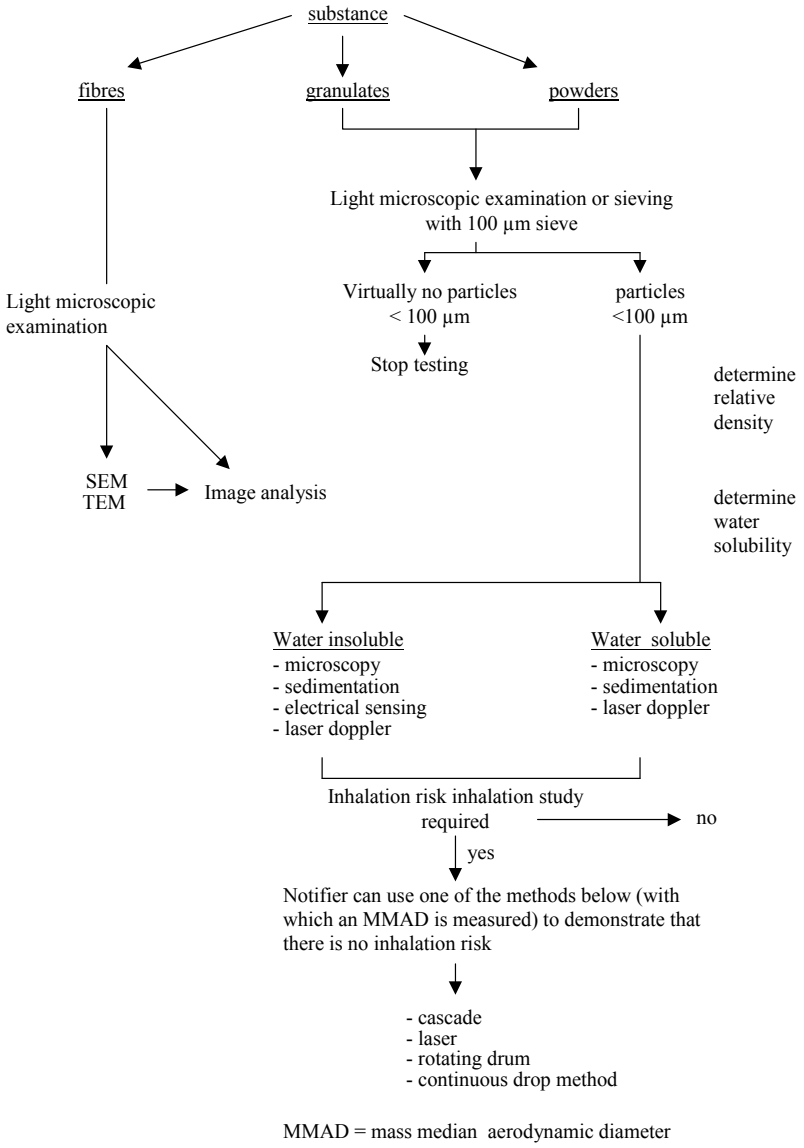


Table R.7.1-30 Methods to determine particle size distribution of the material as it is

Method and details	Material and size range	MMAD
Microscopic examination It is preferable to prepare samples directly in order not to influence shape and size of the particles. This method determines distribution of particles of respirable and inhalable size and does not refer to airborne dust or dispersed or nebulised particles.	Particles of all kinds Size range: 0.5–5000 microns (light microscope) and <0.1–10 microns (SEM/TEM)	MMAD cannot be determined
Sieving Sieving using wire-mesh sieves and perforated sheet metal sieves is not suitable to determine the distribution of particles of respirable and inhalable size since their range is only 100-10,000 microns. Micro mesh sieves (range 5-100 micron) may give better results. However, since these sieves are generally operated in combination with mechanical or ultrasonic vibration, modification of median size and form may result. Sieving not suitable to determine distribution of particles of respirable size, but might be suitable to determine particles of inhalable size.	Dry powders/granulates Size range: 100–10,000 microns (wire mesh/metal sieves) and 5-100 (micromesh)	MMAD cannot be determined
Sedimentation (gravitational settling) Method is based on gravitational settling of particles in liquid and the effective hydrodynamic radius is determined. Effective hydrodynamic radius distribution should be measured 3x with no two values differing by >20%. Requires sufficient numbers of radius intervals be used to resolve the radius distribution curve. Binary or ternary mixtures of latex spheres (2-100 microns) are recommended as calibration material. Method might be suitable to determine the distribution of particles of respirable and inhalable size.	Dry powders/granulates Size range: 2-200 microns	MMAD cannot be determined
Electrical Sensing Zone (e.g. Coulter) method Samples are suspended in an electrolytic solution. As the particle is drawn through an aperture, the change in conductance gives a measure of particle size. The important parameter is the settling velocity in the liquid phase, which depends on both density and diameter. Particles having a density of several g/cm^3 can be determined. Applicable to particles that are complete electrical isolators in the fluid. Difference in density between particles and fluid must not be too large. Method might be suitable to determine the distribution of particles of respirable and inhalable size	Dry powders/granulates (non-conducting) Size range: 1-1000 microns	MMAD cannot be determined
Phase Doppler Anemometry Expensive technique. Particle size distribution can be measured either in air or in liquid. The method presupposes that the particles are spherical with	Dry powders/granulates Size range: 0.5-80 microns (in air); 0.5-1000 microns	MMAD cannot be determined

known refractive index. Method might be suitable to determine the distribution of particles of respirable and inhalable size	(in liquid)	
Determination of fibre length and diameter distributions Light microscopy used to examine likelihood of fibres present by comparing similarities to known fibrous or fibre releasing substances or other data. Extreme care required during ample preparation to avoid fibre breaking and clumping. Care should also be taken to avoid contamination by airborne fibres. Samples might be prepared by (a) producing suspensions in water by gentle hand agitation or vortex mixing or (b) transfer of dry material onto copper tape either directly or by spraying of the dry fibres by use of atomiser or pipette. Length and diameter distributions should be measured independently at least twice and at least 70 fibres counted. No two values in a given histogram interval should differ by > 50% or 3 fibres, whichever is larger. The presence of long thin fibres would indicate a need for further, more precise measurements. This method might be suitable to determine the distribution of fibres of respirable and inhalable size	Fibrous products Size range: diameters as small as 0.1 micron and as large as 100 micron and lengths as small as 5 micron and as large as 300 micron	

Using the methods listed in Table R.7.1-30, the following information should be presented

- Expected % change of reported values in the future (e.g. variations between production batches)
- Sample preparation methods and analysis methods used
- Approximate information on particle shape (e.g. spherical, platelike, needle shaped)
- Lot number, sample number
- Suspending medium, temperature, pH
- Concentration
- Stoke's (effective hydrodynamic) radius R_s, distribution for $2 < R_s < 200$ micron
- Mean value and approximate *area* (%) of any resolvable peak in R_s distribution
- % of particles with $R_s < 2$ micron and $R_s > 200$ micron

It is advantageous to have accurate information about the propensity of materials to produce airborne dust (the *dustiness* of the material). No single method of dustiness testing is likely to represent and reproduce the various types of processing and handling used in industry. The measurement of dustiness depends on the test apparatus used, the properties of the dust and various environmental variables. There are a number of methods for measuring the dustiness of bulk materials, based on the biologically relevant aerosol fractions defined in EN 481. Two methods (the rotating drum method and the continuous drop method) are detailed in EN 15051 "Workplace

atmospheres – Measurement of the dustiness of bulk materials – Requirements and reference test methods" (CEN, 2006). The methods (Table R.7.1-31) are used to determine the distribution of respirable particles and (to a lesser extent) the distribution of inhalable particles. These methods require the generation of representative test atmospheres using suitable generation equipment and correct sampling techniques. These methods are preferred since they measure particles in the air and as such the mass median aerodynamic diameter (MMAD) and geometric standard deviation (GSD). These methods are only applied if light microscopic examination indicates a likelihood that fibres are present.

The two methods in EN 15051, however, provide quite different results. A recent comparison of dustiness results for a range of minerals based on the two methods revealed a difference in classification for the respirable fraction for 50% of the tested materials. Considering the inhalable fraction, classification was different for 60% of the tested materials. There was no trend in the data. Consequently, a recommendation has been given within CEN to revise the standard. It is not recommended to use these methods for classification and labelling purposes. However, an order of relative dustiness could be achieved by applying the same method to a range of materials.

The particle size distribution of a dust cloud may be different from the powder source. Studies on dust generation by free falling powders have demonstrated that the manner in which the powder is handled may be as important as the dust generating capacity of the bulk material, in terms of the resulting exposure (e.g., Heitbrink et al., 1992). Falling height has an important influence on dust generation and release for more than one reason. The higher the impact, the more dissemination of dust there is. Moreover, the greater the falling height, the greater flow of entrained air, which favours dust dissemination. This shows the importance of process design and adequate work practices.

There have been many interesting studies on material flow which demonstrate that the influence of the various factors is not so obvious. For example, it is sometimes erroneously assumed that a powdered material with a larger proportion of coarse particles offers less dust hazard; however, a higher proportion of coarse particles in the bulk material may actually increase dustiness due to a *decrease in the cohesion of the material as the proportion of coarse particles increases* (Upton et al., 1990), and also due to the agitation of the fine particles as there are more collisions with large particles. The higher the impact between particles, the more dissemination of dust there is.

Table R.7.1-31 Measurement of airborne dispersed or nebulised particles

Method and details	Material and size range	MMAD
Cascade impaction Cascade impactors can be used to obtain the size distribution of an aerosol (or a dust cloud). Air samples are drawn through a device which consists of several stages on which particles are deposited on glass or glass fibre. Particles will impact on a certain stage depending on their size. The cut off size can be calculated from the jet velocities at each stage by weighing each stage before and after sampling and the MMAD derived from these calculations. A well established techniques to measure the distribution of particles of respirable or inhalable size	Particles of all kind Size range: 0.1-20 and 0.5-80 microns	MMAD can be determined
Laser scattering/diffraction In general, the scattering of the incident light gives distinct pattern which are measured by a detector. This technique is particle property dependent – i.e. material has unique scattering and diffraction properties which are also particle size dependent. It is important to calibrate the instrument with similar material (of the same size range as the material to be measured). Laser scattering techniques are suitable for geometric particles, viz spheres, cubes and monocrystals. Particle size will be established optically. The MMAD can be calculated by means of a calculation correction. The method is suitable to determine the distribution of particles of respirable and inhalable size.	Particles of all kind Size range: 0.1-100 microns	MMAD can be determined
Rotating drum method (EN 15051) This method is based on size selective sampling of an airborne dust cloud produced by the repeated lifting and dropping of a material in a rotating drum. Air drawn through the drum passes through a specially designed outlet and a 3-stage fractionating system consisting of two porous polyurethane foams and a membrane filter. The mass of dust collected on each collection stage is determined gravimetrically to give a direct measure of the biologically relevant size fractions. This method simulates a wide range of material handling processes in industry and determines the biologically relevant size functions of a material in the airborne	Dry powders/granulates/friable products Size range: 0.5-10,000 microns	MMAD can be determined

state. Full size distributions can be obtained by analysing the contents on the dust collection stages. This method is suitable to determine the distribution of particles of respirable or inhalable size.		
Continuous drop method (EN 15051) This method is based on the size selective sampling of an airborne dust cloud produced by the continuous single dropping of material in a slow vertical air current. The dust released by dropping material is conducted by the airflow to a sampling section where it is separated into the inhalable and respirable fractions. This method is suitable to determine the distribution of particles of respirable or inhalable size.	Dry powders/granulates/friable products Size range: 0.5-10,000 microns	MMAD can be determined

Using the methods listed in Table R.7.1-31, the following information should be presented:

- Sample description, method description
- Number of particles per field
- Total number of fibres measured
- Histograms of distributions based on length and diameter of fibres (at least 50 each). For diameters the ranges should be 0.1-0.5, 0.5-1.0, 1-2, 2-3, 3-5 micron and >5 micron. For lengths they should be 0-5, 5-10, 10-15, 15-20 (etc) micron
- Mean value and approximate *area* (%) of any resolvable peak in the R_s distribution

Other methods that measure inhalable fractions only or that give no detailed distributions are detailed in Table R.7.1-32. They have not been included in the flowsheet.

Table R.7.1-32 Methods that measure inhalable fractions only or that give no detailed distributions

Method and details	Material and size range	MMAD
Elutriation Particles are drawn out on a column at varying velocity. The velocity is used to calculate particle size and the weight of the remaining sample at a particular velocity is used to calculate the distribution. The method is limited to particles >15 microns. The method is not suitable to determine the distribution of particles of respirable size, but might be suitable to determine the distribution of particles of inhalable size	Dry powders/granulates Size range: 15-115 microns	MMAD cannot be determined
Air jet sieve Air is aspirated through a weighted sample on a fine sieve and the weight loss measured. The method is capable of estimation of the non-floatable fraction of the material under investigation. Aggregation of he particles will result in unreliable values. In addition, since the lower detection limit is only 10 micron, this method is not suitable to determine the distribution of particles of respirable size. The method is not suitable to determine the distribution of particles of respirable size, but might be suitable to determine the distribution of particles of inhalable size.	Particles of all kind Size range: 10-10,000 microns	MMAD cannot be determined
Cyclons The use of a cyclone is a simple approach to determining whether respirable and/or inhalable particles are present in the test atmospheres by constructing the cyclone cut off points at 4.25 and 100 microns. By measuring the weight of particles which pass through the cyclone it can be decided whether more sophisticated methods have to be applied to determine the size distribution of the particles smaller than 10 micron. This method is suitable to determine the fraction of particles of respirable and inhalable size.	Particles of all kind Size range: 0.1-200 microns	MMAD cannot be determined

Reference substances:

Five reference substances of defined particle size covering the overall range 0.35 to 650 micron (excepting the 50 to 200 micron region) have been certified with respect to the cumulative mass

distribution of particles versus equivalent settling rate diameter or equivalent volume diameter. The materials will be available from the Community Bureau of Reference of the European Economic Community and will be issued with certificates of measurement. The certification report will also be available (Community Bureau of Reference, 1979).

Published data on granulometry

There are a number of web sites and electronic databases that include compilations of and evaluations of data on particle properties. However, there appear to be a limited number of reference books that provide particle size data (Brandrup and Immergut, 1989).

R.7.1.14.3 Evaluation of available information on granulometry

The particle size distribution is carried out on the material under investigation and not as airborne dust. It is important to note that the original particle size distribution is highly dependent on the industrial processing methods used and care should be taken to avoid changing the particle size distribution by subsequent environmental or human transformations. The small quantities used as samples must be representative of product batches comprising many kilograms. Great care should be taken on* the fact that non-conducting particles in non-conducting liquid may be electrically charged resulting in non-representative settling of particles of certain size. In addition, in the process of particle size distribution determination, it is very important to take the electrostatic charge of the particles into account. Electrostatically charged particles behave different and may influence sampling.

Methods which determine the MMAD need the generation of representative test atmospheres using suitable generation equipment and correct sampling techniques. They can be used in case of airborne particles (dusts, smokes, fumes), nebulised particles (wet aerosol) or dispersed particles (dry aerosol).

Experimental data on granulometry

Particle size is not a specific physico-chemical property of a substance. The original particle size distribution is highly dependent on the industrial processing methods used and can also be affected by subsequent environmental or human transformations. In that respect any published data on particle size analysis will only be pertinent to that particular sample or process.

Non-experimental data on granulometry

There are no QSPR/QSAR tools available for predicting particle size and the data will therefore need to be experimentally determined.

Remaining uncertainty on granulometry

The equivalence of the various national and international standard methods for particle size distribution has not been tested and is not known. There is a particular problem in relation to sedimentation and Coulter counter measurements. The effect of impurities on particle shape should be considered when measuring fibre length and diameter distributions.

The small quantities used as samples must be representative of product batches comprising many kilograms; therefore sampling and sample handling require great care.

It is useful to distinguish between aggregates and agglomerates. While an aggregate may be considered to be permanent, agglomerates may break up under certain circumstances. As small

particles often form agglomerates, sample pre-treatment (e.g. the addition of dispersing agents, agitation or low-level ultrasonic treatment) may be required before the primary particle size can be determined. However, great care must be taken to avoid changing the particle size distribution.

R.7.1.14.4 Conclusions on granulometry

The particle size distribution is needed in order to decide which route of administration is most appropriate for the acute toxicity and 28-day base set animal studies. A number of methods are provided for determining the particle size fractions which are then used to assess the possible health effects resulting from inhalation of airborne particles in the workplace. A number of methods covering different ranges of particle sizes is available though none of them is applicable to the entire size range.

R.7.1.14.5 Integrated testing strategy (ITS) for granulometry

Testing for particle size analysis is not required for those substances which are marketed or used in a non solid or non granular form. A testing strategy detailing which methods to use to determine particle size distribution of respirable and inhalable particles is provided.

Examples and Case studies on granulometry

Particles visible to the naked eye are 40 μm in diameter. Many substances will have a wide range of particulate sizes. Some examples include beach sand and granular activated carbon (ranging from 100 μm to >1000 μm), coal dust and milled flour (typically 1 μm to 100 μm) and asbestos (<0.1 μm to 1 μm).

R.7.1.14.6 References on granulometry

Aitken, R.J., Creely, K.S., Tran, C.L., (2004). Nanoparticles: An occupational hygiene review, Institute of Occupational Medicine, Research Park North, Research Report 274, Edinburgh.

Borm, P.J.A., Robbins, D., Haubold, S., Kuhlbusch, T., Fissan, H., Donaldson, K., Schins, R., Stone, V., Kreyling, W., Lademann, J., Krutmann, J., Warheit, D., Oberdorster, E. (2006) "The potential risks of nanomaterials: a review carried out for ECETOC", Particle and Fibre Toxicology, 3, 11.

Brandrup, J. and Immergut, E.H. (1989) eds " Polymer Handbook", 3rd edition, Wiley

CEN (1993). EN 481 document "Workplace atmospheres - size fraction definitions for measurement of airborne particles.

CEN (2006), EN 15051 document "Workplace atmospheres – Measurement of the dustiness of bulk materials – Requirements and reference test methods"

Community Bureau of Reference (1979) "Certification Report on Particles of Defined Particle Size", Brussels

Hazard Prevention and Control in the Work Environment: Airborne Dust, WHO/SDE/OEH/99.14 Chapter 1 – Dust: Definitions and Concepts

Heitbrink WA, Baron PA, Willeke K (1992). An investigation of dust generation by free falling powders. American Industrial Hygiene Association Journal,53:617-624.

HSE (1996) The Notification of New Substances (NONS): Guidance Document on methods for measuring Particle Size Distribution

OECD Guidelines for Testing of Chemicals – Method 110 "Particle Size Distribution/Fibre Length and Diameter Distributions"

Upton, S.L., Hall, D.J., Marsland, G.W. (1990). Some experiments on material dustiness. Aerosol Society Annual Conference 1990, Surrey, UK.

R.7.1.15 ADSORPTION/DESORPTION

This property indicates the binding capacity (or *stickiness*) of a substance to solid surfaces, and so is essential for understanding environmental partitioning behaviour.

Information on adsorption/desorption is an essential input to environmental exposure models, because:

- Adsorption to suspended matter can be an important physical elimination process from water in sewage treatment plants (STPs). This in turn may mean that sewage sludge, if spread to land, is a major source of the substance in soil.
- Adsorption to suspended matter in receiving waters affects both the concentration in surface water and the concentration in sediment.
- Desorption of a substance from soil directly influences its mobility and potential to reach surface or groundwaters.

Consequently, information on adsorption/desorption is also an important factor in test strategies for assessing toxicity to sediment- or soil-dwelling organisms. For example:

Substances with a K_{oc} below 500–1,000 L/kg are generally unlikely to adsorb to sediment (SETAC, 1993). To avoid extensive testing of chemicals, a log K_{oc} (or log K_{ow}) $\geq$3 can be used as a trigger value for sediment effects assessment.

As a screening approach, both sediment and soil risks may be estimated using aquatic toxicity data using the equilibrium partitioning method. For substances with a log K_{ow} above 5 (or with a corresponding K_{oc}), however, the resulting PEC/PNEC ratio is increased by a factor of 10, to account for the potential additional accumulation via sediment/soil ingestion that may occur for certain types of invertebrate. In practice this means that the assessment conducted for the aquatic compartment will also cover the sediment compartment for chemicals with a log K_{ow} up to 5.

Strong binding behaviour to soil particles (e.g. log K_{ow} >5, log K_{oc} >4) might justify immediate long-term soil organism toxicity testing if particular sensitivity and/or persistence is anticipated. This is specifically mentioned in Column 2 of REACH Annex IX, 9.4 for effects on terrestrial organisms.

Substances that adsorb strongly to biological surfaces (e.g., gills, skin, etc.) may lead to toxic effects in higher organisms after biomagnification.

The information is also relevant for assessing environmental persistence. For example: Degradation rates in sediment and soil are also assumed to be reduced by default if a substance is highly sorptive (since it is less bioavailable to microorganisms). This may lead to consideration of soil/sediment simulation testing in some cases.

Finally, there may be practical implications for test performance: Substances that adsorb strongly to surfaces can be difficult to test in aquatic systems.

Sediment organism toxicity tests should be designed to minimise desorption from the sediment particles during the test (which would lead to an underestimation of the toxicity).

Strong binding to sediment or soil particles may affect the ability to measure test substance concentrations analytically.

Definition of adsorption/desorption

Adsorption is caused by temporary (reversible) or permanent bonding between the substance and a surface (e.g. due to Van der Waals interactions, hydrogen bonding to hydroxyl groups, ionic interactions, covalent bonding, etc.). The following definitions are taken from OECD (2001).

1) The distribution coefficient (K_d) is the ratio of equilibrium concentrations of a dissolved substance in a two-phase system consisting of a sorbent (typically soil or sewage sludge) and an aqueous phase.

$$K_d = C_s / C_{aq}$$

Where C_s = concentration of test substance in soil or sludge at equilibrium

C_{aq} = concentration of test substance in aqueous phase at equilibrium

The value may be dimensionless if the concentration in both phases is given on a weight/weight basis. Where the concentration in the aqueous phase is expressed on a weight/volume basis, then K_d has units of L/kg. K_d can vary with sorbent properties (e.g., organic carbon content, clay content, texture, ion exchange capacity, redox potential and pH) and can be concentration dependent.

2) When the equilibrium concentration in the aqueous phase (C_{aq}) is equal to one, the concentration of the test substance in the solid phase is known as the Freundlich adsorption coefficient (K_f). This is expressed in units of g/kg sorbent, and the value can vary with sorbent properties.

3) Electrostatic interactions between the substance and mineral surfaces in the solid can be difficult to assess, and so it is often assumed that all adsorption can be related to the organic matter content of the medium (i.e., the influence of the mineral matrix is ignored; this is not valid for ionic substances). When the distribution coefficient is normalised to the organic carbon content of the soil/sludge used, it is called the organic carbon-water partition coefficient (K_{oc}). This can be dimensionless or may be expressed as L/kg or g/kg (note that there is a further conversion factor if normalisation is in terms of organic *matter* rather than carbon – see OECD (2000) for details).

$$K_{oc} = K_d \times 100 / \%oc$$

Where %oc = percentage of organic carbon in the sorbent

For K_{oc} to be a true constant, the K_d from the linear portion of the isotherm should be used since it is independent of concentration.

Organic carbon contents in the environment can range from 30-40% (in sewage works) to below 1% (in sea water). The ability of the K_{oc} parameter to normalise for the behaviour of substances over such a wide range is doubtful. Nevertheless, in risk assessment, the range is assumed to be from 2% in soils to 10% for suspended matter. Over *this* range, the K_{oc} is an approximate indicator for the extent of adsorption between a substance and the sorbent (particularly for non-polar organic chemicals) and allows comparisons to be made between different chemicals. K_{oc} values can vary depending on the actual sorbent, but their variability is greatly reduced compared to K_d (or K_f) values.

4) In practice, the K_{oc} of the substance determined for a specific sorbent (such as soil) is used to estimate solid-water partition coefficients (K_p) for sewage sludge, sediment and soil (since these are rarely measured directly). This is done by multiplying the K_{oc} by the weight fraction of organic carbon (F_{oc}) in the specific compartment.

R.7.1.15.1 Information requirements on adsorption/desorption

Screening information on adsorption (and desorption) is required for substances manufactured or imported in quantities of 10 t/y or more. Depending on the results, further information (for example, a test) may be required for substances manufactured or imported in quantities of 100 t/y or more.

Column 2 of REACH Annexes VIII and IX provides two exemptions. A study does not need to be conducted if:

- the substance can be expected to have a low potential for adsorption based on its physico-chemical properties (e.g. low octanol-water partition coefficient); or
- the substance and its relevant degradation products decompose rapidly. Note that if a substance hydrolyses then it might be more appropriate to also determine the degree of adsorption of the hydrolysis products.

In practice a cut off value for log K_{ow} of 3 can be applied for adsorption potential. However, caution should be exercised in this criterion's use as substances that are water soluble and have a low octanol-water partition coefficient do not necessarily always have a low adsorption potential. A *measured* adsorption coefficient is usually needed for ionising substances, since it is important to have information on pH-dependence (cationic substances in particular generally adsorb strongly). Similarly, measured values will normally be needed for surface active substances (e.g. surfactants), because K_{ow} values (predicted or measured) are likely to be poor predictors of adsorption for these types of substance. For ionisable substances, partition coefficients should also be corrected according to the pH of the environment being assessed (see Annex 2). For complex mixtures (e.g. UVCBs) a single value of K_{oc} will not be definitive. In such cases a range of values or a representative value can be given, depending on the substance.

R.7.1.15.2 Available information on adsorption/desorption

Testing data on adsorption/desorption

The adsorption of a substance to sewage sludge, sediment and/or soil can be measured or estimated using a variety of methods, which are tabulated in Table R.7.1-33 in order of increasing complexity. The dissociation constant (if appropriate) should be known before testing. Information on vapour pressure, solubility in water and organic solvents, octanol-water partition coefficient and stability/degradability is also useful.

Table R.7.1-33 Methods for the measurement of adsorption

Method and Description	Applicability/Notes
Adsorption control within an inherent biodegradability test Estimate of the extent of adsorption to STP sludge made from the elimination level in a Zahn-Wellens inherent biodegradation test. (e.g. OECD TG 302B). 3-hour value recommended. Values beyond 24 hours not normally used. Where data are not available for adsorption up to 24 hours, data from time scales beyond this can only be used if adsorption is the only removal mechanism, with an upper limit of 7 days.	Highly adsorptive substances that are water soluble
HPLC method: OECD TG 121; EU C.19: Estimation of the Adsorption Coefficient (K_{oc}) on Soil and on Sewage Sludge using High Performance Liquid Chromatography (HPLC) (Original Guideline, adopted 22nd January 2001) Calibration with reference substances (preferably structurally related to the test substance) of known K_{oc} allows the K_{oc} of the test substance to be estimated. Test substance K_{oc} value should lie within the calibration range of the reference substances.	Measurement of log K_{oc} in the range 1.5-5.0. Validated for several chemical types, see test guideline for details. Poorly soluble and volatile substances as well as mixtures. Ionisable substances: test both ionised and unionised forms in appropriate buffer solutions where at least 10 % of the test compound will be dissociated within pH 5.5 to 7.5 May not be suitable for: substances that react with the column, solvent or other test system components; surface active substances; substances that interact in a specific way with inorganic soil components such as clay minerals; inorganic compounds; moderate to strong acids and bases.
Batch test of adsorption of substances on activated sludge (International Standard 18749) Screening method to determine the degree of adsorption of substances on activated or primary sludge in sewage treatment plants (ISO, 2004). The method does not differentiate between adsorption and other elimination methods (such as complex formation, flocculation, precipitation, sedimentation or biodegradation).	Suitable for substances that: are water soluble, or allow for stable suspensions/dispersions/emulsions, are not significantly removed by abiotic processes (e.g. stripping/foaming), do not de-flocculate activated sludge, are not readily biodegradable, and have a sufficiently sensitive analytical method.
Sediment and soil adsorption/desorption isotherm (OPPTS 835.1220) Screening method according to US-EPA guideline (OPPTS, 1996) using three soil types.	
Batch equilibrium method (OECD TG 106; EU C.18: Absorption – Desorption Using a Batch Equilibrium Method (Updated Guideline, adopted 21st January 2000) Test uses a range of actual soils and so represents a more realistic scenario than the HPLC (OECD 112) method.	Used for substances with K_{oc} values that cannot be reliably determined using other techniques (e.g. surfactants). Requires a quantitative analytical method for the substance, reliable over the range of test concentrations. For ionisable substances, soil types should cover a wide range of pH. Adjustments for poorly soluble substances given in the test guideline.

OECD TG 312: Leaching in Soil Columns (Original Guideline, adopted 13th April 2004) K_d values can be derived from column leaching studies.	Appropriate study design to estimate K_d values particularly for unstable test substances that degrade significantly during the equilibrium time of 'shake flask' sorption studies
Simulation tests and direct field measurement: including OECD guidance document no. 22 (OECD, 2000b). Monolith lysimeters can be used to study the fate and behaviour of substances in an undisturbed soil profile under outdoor conditions. They allow for monitoring of the volume of leaching/drainage water as well as the concentrations of a chemical and its transformation products. They are mainly used in pesticide studies. Field leaching studies can also be carried out where hydrodynamically isolated soil layers are analysed *in situ*. Although such studies are the most realistic, their reproducibility and representativity may be limited (e.g. due to the effects of large-scale soil structure, weather events, the soil conditions at the time of application, etc.). Since data from these methods are unlikely to be encountered for the vast majority of industrial chemicals, they are not considered further here. Further information can be found in guidance for pesticide registration.	

Published data on adsorption/desorption

Adsorption can be estimated from the octanol/water partition coefficient (K_{ow}) – or in some cases the water solubility – for organic, non-ionic substances[14].

The equation for *nonhydrophobic* substances is preferred as default. See also, 2 ECETOC reports, references in the test guidelines, Boethling ref, Dearden review & TAPIR report Appendix 9 (ECB, 2005) for further examples. See also Section R.7.1.17.3 for more information on this topic.

In the absence of measured data for the substance in question, read-across from structurally similar chemicals (analogues) may be acceptable where other parameters such as water solubility and K_{ow} have been measured and are within an acceptable limit of variation. In such cases full justification for the analogue's selection should be included, starting with a comparison of its, and the assessed substance's, molecular structure and as stated including other physico-chemical data such as log K_{ow} and water solubility.

R.7.1.15.3 Evaluation of available information on adsorption/desorption

Experimental data on adsorption/desorption

In general, partition coefficients that are measured with a suitable standard method are preferred (and they are usually essential for surfactants and ionic substances that dissociate at environmentally relevant pH). Estimated values may be sufficient for non-dissociated organic molecules. Where there is some uncertainty over the use of a particular value, it may be possible to determine the significance of the property to a specific assessment by investigating the consequence of varying the value of K_{oc} between high and low extremes – a sensitivity analysis

Non-experimental data on adsorption/desorption

Soil sorption (K_{oc}) of organic non-ionic chemicals can be estimated from their octanol-water partition coefficient (K_{ow}), as well as from other properties such as aqueous solubility. Such methods, including QSPR, are useful in the first instance to indicate the qualitative/quantitative adsorption coefficient of a substance. In some instances an estimated value may be sufficient for this endpoint. In all such cases the estimated method must be proven to be valid for the type of chemical considered (see the general guidance for use and applicability of QSPR), and if possible a

[14] K_{ow} is experimentally difficult to determine for surfactants and this parameter may not be sufficiently descriptive of surface activity or adsorption/desorption. Measured Koc/K_p values are more appropriate in such cases.

sensitivity analysis should be conducted with values generated from different models. Using a range of values in the CSA will help to highlight if the adsorption coefficient is an important factor for environmental behaviour of the substance. In general an estimated value will be sufficient if it is indicated that the adsorption coefficient will not affect the CSA, i.e. no risk is identified for the sediment/soil compartments. Estimated values are essential for substances for which experimental measurement is not feasible i.e. for *difficult substances*. Estimated values are also useful for comparing screening tests [e.g. HPLC method (OECD 121; EC C19)] against. A number of reviews of K_{oc} prediction have been published recently (Lyman 1990, Reinhard & Drefahl 1999, Doucette 2000, Delle Site 2001, Doucette 2003, Dearden 2004). That of Doucette (2000) contains a number of worked examples of the estimation of log K_{oc} values.

Sabljić *et al* (1995) correlated log K_{oc} values of 19 chemical classes with log K_{ow} values, and obtained reasonably good correlations. They found, however, that slopes and intercepts varied widely from class to class. For example, for hydrocarbons and halogenated hydrocarbons the correlation was:

$$\log K_{oc} = 0.81 \log K_{ow} + 0.10$$

$n = 81 \quad r^2 = 0.887 \quad s = 0.451$

That for anilines was:

$$\log K_{oc} = 0.62 \log K_{ow} + 0.85$$

$n = 20 \quad r^2 = 0.808 \quad s = 0.341$

where n = no of data, r^2 = correlation coefficient, s = standard error

This work was discussed in the EU TGD. A full list of the equations for the 19 chemical classes can be found in the EU Technical Guidance Document on Risk Assessment Part III, Chapter 4, Section 4.3 (pp 24 – 27).

It might be thought that such differences should mean that good correlations could not be obtained for diverse data sets. In fact, that has been shown not to be the case. Gerstl (1990) found a correlation as good as most of those of Sabljić *et al* (1995) for a large diverse data set:

$$\log K_{oc} = 0.679 \log K_{ow} + 0.094$$

$n = 419 \quad r^2 = 0.831$ s not given

Briggs (1981) found a good correlation for a large set of pesticides:

$$\log K_{om} = 0.53 \log K_{ow} + 0.64$$

$n = 105 \quad r^2 = 0.90$ s not given

Note that the soil sorption term here is K_{om}, where *om* stands for *organic matter*. The relationship between K_{oc} and K_{om} is: $\log K_{oc} = \log K_{om} + 0.2365$ (Nendza 1998).

Hence from the two correlations above an estimate of K_{oc} or K_{om} can readily be obtained. Calculated log K_{ow} values can quickly be obtained from the ChemSilico website (www.logp.com).

The Abraham descriptors have been used (1999) to model K_{oc} values of a large diverse data set:

$$\log K_{oc} = 0.74\ R - 0.31\ \Sigma\alpha^H - 2.27\ \Sigma\beta^O + 2.09\ V_X + 0.21$$

$n = 131 \quad r^2 = 0.955 \quad s = 0.245$

where R = excess molar refractivity (a measure of polarisability), $\Sigma\alpha^H$ = hydrogen bond donor ability, $\Sigma\beta^O$ = hydrogen bond acceptor ability of oxygen, and V_X = McGowan molecular volume. The descriptors are approximately auto-scaled, so that the magnitude of each coefficient is an indication of the relative contribution of each descriptor to soil sorption. Hence hydrogen bond acceptor ability and molecular size appear to be the most important factors controlling soil sorption. The Abraham descriptors can be calculated using the Absolv-2 software (www.ap-algorithms.com).

Tao *et al* (1999) used a combination of 74 fragmental constants and 24 structural factors to model soil sorption of 592 diverse organic chemicals, with a standard error of 0.366 log unit. Although this is a good prediction, fragmental constant methods are not always easy to use, and can be tedious.

Although soil sorption varies to some extent with temperature, there do not appear to be any QSPR studies concerning this. One study has been published concerning the effect of ionisation on K_{oc} values. Bintein and Devillers (1994) reported the following QSPR based on 229 data points for 53 diverse chemicals:

$$\log K_p = 0.93 \log K_{ow} + 1.09\ f_{oc} + 0.32\ CFa - 0.55\ CFb' + 0.25$$

$$n = 229 \quad r^2 = 0.933 \quad s = 0.433$$

where K_p = sorption coefficient uncorrected for organic content, f_{oc} = fraction of organic carbon in soil, CFa = correction factor for acid ionisation, and CFb' = correction factor for base ionisation.

Bearing in mind the large experimental error associated with soil sorption measurements (Nendza 1998), the standard errors given above are as good as can be hoped for.

There are three software programs that calculate log K_{oc} values. Using a test set of 100 diverse organic chemicals, Dearden (2004) compared performance of two of them, and the results are shown in Table R.7.1-34. No indication of performance is available for the other software, Pharma Algorithms ADME Boxes (www.ap-algorithms.com).

Table R.7.1-34 Software programs that calculate log K_{oc} values

Software	Website	Availability	% Predicted within +/- 0.5 Log unit	Mean absolute error [Log unit]
PCKOCWIN	www.epa.gov/oppt/exposure/pubs/episuitedl.htm	Freely downloadable	82%	0.490
Absolv-2	www.ap-algorithms.com	Purchase	70%	0.569

Remaining uncertainty on adsorption/desorption

It should be noted that even for the OECD 106 test guideline, a ring test showed that K_{oc} values below 1,000 may differ by over 100% between laboratories (ECETOC, 1998). At K_{oc} values above 1,000, the differences were over one order of magnitude. Where a range of measured values are available and the predicted value lies within this range, the predicted value might be preferred as the most representative result.

Where an estimated K_{oc} value is of questionable validity, it may be appropriate to calculate additional values using appropriate methods and take the geometric mean for risk assessment purposes. In addition, it should be noted that exposure models might not be valid for substances with a log K_{oc} above 6.

R.7.1.15.4 Conclusions on adsorption/desorption

Information on adsorption/desorption is required because it indicates where a substance will be found in the environment and how it may partition between environmental compartments. This information has a marked bearing on the ecotoxicological assessment of the substance and also its bioavailability.

Screening information on adsorption (and desorption) is required for substances manufactured or imported in quantities of 10 t/y or more. For substances manufactured or imported in quantities of 100 t/y or more, depending on the results of the screening information, further information (for example, a test) may be required.

Exceptions to screening information/testing:

- substances that are likely to have a low potential for adsorption due to high water solubility coupled with very low octanol-water partition coefficient and the absence of functional groups that would suggest the ability to adsorb.
- substances, and relevant degradation products, that decompose rapidly. However consideration should be given to collecting screening information for degradation products of substances that hydrolyse rapidly ($t_{1/2}$ <12 hours).

Read-across and/or QSPR prediction for K_{oc} are important predictive tools and should be the first method used to predict Koc if reliable measured data do not exist and the model is valid for the substance. However if these options do not given meaningful and valid information or if K_{oc} is an important factor in the CSA (i.e. risks are indicated for sediment/soil compartments based on a predicted value and log K_{ow} is >3), then an experimental value should be measured. Several screening test methods and full tests exist for K_{oc}; these are discussed in Section R.7.1.15.2.

If the substance is ionisable, knowledge of dissociation constant (pKa) should be gleaned before testing. If no measured data are available for pKa an estimate can be made (see Section R.7.1.17.3). Estimation methods for K_{oc} are not appropriate for surface active or ionisable (at environmentally-relevant pH) substances. A batch equilibrium test may need to be considered at the 10 t/y band, and would be essential at the 100 t/y band.

Information on vapour pressure, solubility in water and organic solvents, octanol-water partition coefficient and stability/degradability is also useful before beginning screening for K_{oc}.

Concluding on C&L and chemical safety assessment

The K_{oc} is not directly relevant for environmental classification or the PBT assessment. However, it is a key property for exposure assessment so the information requirement should not be waived.

For all organic substances manufactured or supplied in quantities of 10 t/y or more, the K_{oc} should be estimated using read-across or QSPR methods as a first step. If the property is likely to be a significant determinant in the calculation of risk (e.g. following a sensitivity analysis), then a test should be conducted to provide a more reliable value for substances manufactured or supplied in quantities of 100 t/y or more. In general, confirmatory testing would not be expected for non-

ionising substances with a log K_{ow} value below 3, or for substances that degrade rapidly (in which case the degradation products may be more relevant). The HPLC method may be used as a first step in testing, with the batch equilibrium method being considered only if more definitive data become necessary for the Chemical Safety Assessment. Column leaching studies might be an option under some circumstances (e.g. for unstable test substances that degrade significantly during the equilibrium time of *shake flask* sorption studies).

If estimation methods are not appropriate (e.g. because the substance is a surfactant or ionisable at environmentally-relevant pH), then a batch equilibrium test may need to be considered at the 10 t/y band, and would be essential at the 100 t/y band.

R.7.1.15.5 Integrated testing strategy (ITS) for adsorption/desorption

In depicting the ITS (Figure R.7.1-8) for adsorption coefficient, directions for four basic chemical types can be followed: substances meeting the criteria (see Annex VII of Directive 1907/2006 REACH text), surface active substances, ionisable substances, and other substances. Substances that meet the exclusion criteria would generally have a log K_{ow} value <3, and be non-ionisable and not surface active. For analogue values used as read-across, justification for the analogue must be given including its log K_{ow} and water solubility values (see Section R.7.1.15.2).

For *other* substances, it is possible to use QSPR prediction for the K_{oc} value in the absence of a literature or analogue value. However if supply exceeds 100 t/y then a screening test may be required depending on how a sensitivity analysis of predicted K_{oc} value affects the environmental fate of the substance in the chemical safety assessment.

In the case of ionisable (within the environmentally relevant range of pH 5–9) and surface active substances, QSPR prediction techniques are not usually valid. In these cases a test is required if no valid literature or analogue value is identified.

Examples and Case studies

The following examples are taken from the Existing Substances Regulation, EC 793/93.

No measured K_{oc} data were available for octabromodiphenyl ether. Values were estimated from the chemical structure and from the K_{ow} value. Measured values for adsorption coefficients were available for 2 commercial homologues, one with a lesser degree of bromine substitution (pentabromodiphenyl ether) and the other with a greater degree of bromine substitution (decabromodiphenyl ether). Interpolating a K_{oc} value for commercial octabromodiphenyl ether from the 2 indicated a K_{oc} of around 1,363,040 L/kg. This was higher than the predicted values by a factor of 10. As there were uncertainties in the log K_{ow} value used for one estimation, the interpolated value was preferred for risk assessment purposes. In summary, if a measured value is not available but a reliable and robust value developed from analogue(s) data is available (i.e. an interpolated or extrapolated value), this should first be compared to calculated values. If there is a large discrepancy between the interpolated value and calculated value(s), then the former should be used in preference along with full justification backing up the decision.

K_{oc} values were measured for three structurally related chloroalkyl phosphate esters by the HPLC method. The measured values were all somewhat higher than the predicted values based on the log K_{ow}, significantly so for the two higher molecular weight substances. The difference suggested that these two substances adsorb more strongly than might be expected on the basis of K_{ow} alone (e.g. interaction with inorganic minerals may be possible). The measured K_{oc} was used in the risk assessment, but a sensitivity analysis was also performed using a high and a low value. The risk

characterisation ratios were significantly affected, especially for soil, and so a soil/sediment adsorption study according to OECD 106 was requested for one of the substances to check this further. In summary, when uncertainty exists for K_{oc} and this uncertainty proves to affect the outcome of the risk assessment, then further testing should be considered in order to clarify K_{oc} and so refine this aspect of the risk assessment.

Tetrabromobisphenol-A is an ionisable substance, with two pK_a values in the environmentally-relevant pH range. K_{oc} values were measured in both sediment and soil, as well as estimated using the usual QSPR methods. The interpretation of the adsorptive behaviour of tetrabromobisphenol-A is complicated as it could be expected to vary with pH of the soil or sediment system. For some of the adsorption studies, the pH of the sediment was below 6, which indicates that the undissociated form would predominate. All the available K_{oc} data were therefore converted back to the sediment-water partition coefficient ($Kp_{sediment}$) values using the respective organic carbon contents and plotted against the organic carbon content of the sediment. The slope of the plot corresponds to the K_{oc} value, which was used for the risk assessment. The large intercept on the y-axis on this plot indicated that a substantial amount of the adsorption of tetrabromobisphenol-A to the sediment may not be governed by the organic carbon content (i.e. adsorption onto mineral fractions may also be important). Given the uncertainty and variation in the available data, and to take account of the possible natural variation of the adsorptive behaviour of the substance in the environment, the environmental assessment also considered a set of higher adsorption coefficient values (including the 90^{th} percentile value for the $Kp_{sediment}$) as part of the sensitivity analysis. Conclusion: for substances which are likely not to have a single representative K_{oc} value (e.g. multiply ionisable substances), the CSA should include a sensitivity analysis so that more than one value of K_{oc} can be compared.

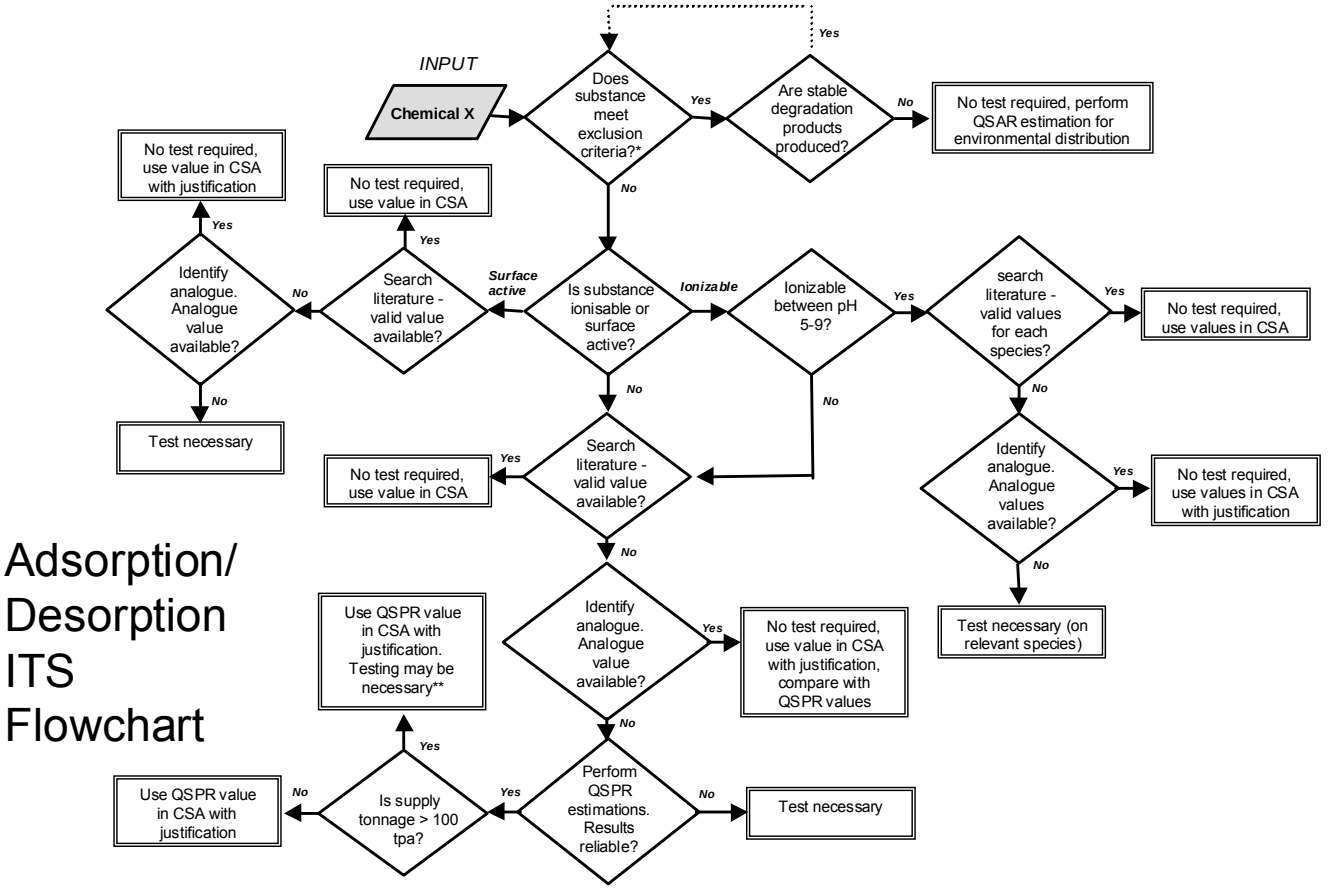
Primary alkylamines are surface-active substances with alkyl chains 8 to 20 carbon atoms in length and are predominantly protonated at environmentally relevant pHs. QSPR methods are therefore inappropriate and adsorption behaviour of ^{14}C-labelled n-octadecylamine was measured in a batch equilibrium experiment according to OECD 106. Adsorption onto soil with a 2.6% organic matter content was much higher than to one with a 6.6% organic matter content. It was assumed that ionic interactions play a more important role than hydrophobic partitioning with organic matter for this type of substance, since alkyl ammonium ions can interact with the surface of mineral particles or with negative charges of humic substances. The influence of the chain length on the sorption behaviour is therefore expected to be low, and the experimental results obtained in the test with octadecylamine were taken to be representative for the other products. In addition, it was assumed that for octadecenylamine the double bond would not influence sorption for the same reasons. Conclusion: functional aspects of the substance that may affect Koc should be considered (e.g. substances which exist predominantly as cations under environmental conditions can be expected to have very much greater K_{oc} values than calculations based on the neutral form would suggest).

Aniline has a log K_{ow} of 0.9, which suggests that adsorption will be negligible. However, experiments revealed that the substance forms covalent bonds with the organic fraction in soils and sediments. The reaction is between the amino group and aldehyde or keto groups as well as double bonds of quinoid systems that are typically found in humic substances. Some of the reaction products are hydrolysable.

Because of the specificity of the reaction, chemisorption onto sewage sludge was not expected. The adsorption of aniline onto sludge was therefore predicted by QSPR from the log K_{ow}. An experimentally determined K_{oc} value was used for sediment and soil, recognising that the term does not really apply if chemisorption occurs. Conclusion: functional aspects of the substance that may affect K_{oc} should be considered (e.g. substances which contain a nucleophile may have very much greater K_{oc} values in certain situations than calculations based on the structure/K_{ow} would suggest.

This is because they may undergo nucleophilic substitution or addition reactions with organic components of soils and sediments)

Figure R.7.1-8 Integrated Testing Strategy for adsorption/desorption



*As stated in Annex VIII 9.3.1 of 1907/2006 REACH text: "The study does not need to be conducted if:

- based on the physicochemical properties the substance can be expected to have a low potential for adsorption (e.g. the substance has a low octanol-water partition coefficient); or

- the substance and its relevant degradation products decompose rapidly."

However consider collecting screening information for degradation products of substances that hydrolyse rapidly ($t_{1/2}$ < 12 hours)
**If supply tonnage reaches 100 tpa, then testing may be necessary to quantify better the actual adsorption coefficient and how it affects the fate and behaviour of the substance, depending on the result of sensitivity analysis. In some cases this may include carrying out a full test (eg according to OECD 106) rather than an estimation (eg. according to OECD 112). SEE TEXT FOR DETAILS.

R.7.1.15.6 References on adsorption/desorption

Bintein S. and Devillers J. QSAR for organic chemical sorption in soils and sediments. *Chemosphere* (1994) 28, 1171-1188.

Briggs G.G. (1981) Theoretical and experimental relationships between soil adsorption, octanol-water partition coefficients, water solubilities, bioconcentration factors and the parachor. *J. Agric. Food Chem.* **29**, 1050-1059.

Dearden J.C. (2004) QSAR modelling of soil sorption. In Cronin M.T.D. and Livingstone D.J. (Eds.), *Predicting Chemical Toxicity and Fate*, CRC Press, Boca Raton, FL, pp. 357-371

Delle Site, A., (2001) Factors affecting sorption of organic compounds in natural sorbent/water systems and sorption coefficients for selected pollutants. *J. Phys. Chem. Ref. Data* 30, 187-439

Doucette W.J. (2000) Soil and sediment sorption coefficients. In Boethling R.S. and Mackay D. (Eds.), *Handbook of Property Estimation Methods for Chemicals: Environmental and Health Sciences.* Lewis, Boca Raton, FL, pp. 141-188

Doucette W.J. (2003) Quantitative structure-activity relationships for predicting soil/sediment sorption coefficients for organic chemicals. *Environ. Toxicol. Chem.* 22, 1771-1788

ECETOC (1998). Technical Report No. 74: QSARs in the Assessment of the Environmental Fate and Effects of Chemicals. European Centre for Ecotoxicology and Toxicology of Chemicals. Brussels.

EU C18 Adsorption – desorption using a batch equilibrium method

EU C19 Estimation of the adsorption co-efficient (K_{oc}) on soil and on sewage sludge using high performance liquid chromatography (HPLC).

Gerstl Z. Estimation of organic chemical sorption by soils. *J. Contaminant Hydrology* (1990) 6, 357-375

ISO (2004). Water quality: adsorption of substances on activated sludge – batch test using specific analytical methods. International Standards Guideline 18749. First edition February 2004.

Lyman W.J. Adsorption coefficient for soils and sediments. In Lyman W.J., Reehl W.F. and Rosenblatt D.H. (Eds.), *Handbook of Chemical Property Estimation Methods*, American Chemical Society, Washington DC, 1990, pp. 4.1-4.33.

Mueller, M. and Kordell, W. (1996). Comparison of screening methods for the estimation of adsorption coefficients on soil. Chemosphere 32(12), 2493-2504.

Nendza M. *Structure-Activity Relationships in Environmental Sciences.* Chapman & Hall, London, 1998

OECD (2000a). Adsorption – desorption using a batch equilibrium method. Organisation for Economic Co-operation and Development (OECD) Guideline for the testing of chemicals no 106.

OECD (2000b) Guidance Document No. 22: Performance of Outdoor Monolith Lysimeter Studies. Organisation for Economic Co-operation and Development (OECD), Paris.

OECD (2001). Estimation of the adsorption co-efficient (K_{oc}) on soil and on sewage sludge using high performance liquid chromatography (HPLC). Organisation for Economic Co-operation and Development (OECD) Guideline for the testing of chemicals no 121.

OECD (2002) Guidelines for Testing of Chemicals (Draft): Leaching in Soil Columns. Organisation for Economic Co-operation and Development (OECD), Paris.

OPPTS (1996). Sediment and soil adsorption/desorption isotherm. United States Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances Fate, Transport and Transformation Test Guideline 835.1220. Draft of April 1996.

Poole S.K. and Poole C.F. (1999) Chromatographic models for the sorption of neutral organic compounds by soil from air and water. *J. Chromatogr. A* 845, 381-400

Reinhard M. and Drefahl A. *Handbook for Estimating Physico-chemical Properties of Organic Compounds.* Wiley, New York, 1999.

Sabljić A., Güsten H., Verhaar H. and Hermens J. QSAR modelling of soil sorption. Improvements and systematics of log K_{oc} *vs.* log K_{ow} correlations. *Chemosphere* (1995) 31, 4489-4514

SETAC (1993). Guidance Document on Sediment Toxicity Tests and Bioassays for Freshwater and Marine Environments. From the Workshop on Sediment Toxicity Assessment at Renesse, Netherlands on 8-10 November 1993. Hill I, Mathiessen P, Heimbach F (eds). Society of Environmental Toxicology and Chemistry – Europe, Brussels.

Tao S., Piao H., Dawson R., Lu X. and Hu H. (1999) Estimation of the organic carbon normalised sorption coefficient (K_{oc}) using the fragment constant method. *Environ. Sci. Tech.* (1999) 33, 2719-2725

R.7.1.16 STABILITY IN ORGANIC SOLVENT AND DEGRADATION PRODUCTS

There are rare occasions when it is important to have information on the stability of a compound in an organic solvent, to ensure confidence in the test results. However, for many substances, the stability in organic solvents will not be critical and testing need not be conducted.

Examples of when stability in organic solvents could be important are given below:

- for certain solubility measurements (e.g. octanol-water partition coefficient)
- to check on the stability of reagent solutions, fortification standards or calibration standards
- when a test substance is dosed as a solution in an organic solvent (e.g. ecotoxicity studies)
- when a test substance is extracted from an environmental sample, plant or animal tissue or diet matrix (arising from a variety of physico-chemical property, ecotoxicity and animal toxicity studies) into an organic solvent and stored pending analytical measurement.

Definition of the stability of a substance

A study of the stability of a test compound in an organic solvent is normally undertaken for a specific time period and under certain storage conditions (e.g. temperatures experienced under normal test conditions for that solution) to confirm whether the test compound is stable under these storage conditions for the duration of the storage of the organic solvent or extract containing the test substance. Often several time periods are selected to check whether there is any particular downward trend in stability over time.

The stability of the test substance at a particular time period during the study is normally expressed as a percentage of the concentration of the test substance in the solvent extract at that time period compared with the initial starting concentration of the test substance at t = 0

$$\frac{C_t}{C_{t0}} 100$$

Where C_t = concentration of test substance in solvent extract at t_1, t_2, t_3…, t_n and C_{t0}=concentration of test substance in solvent extract at t_0.

If there is evidence of the compound breaking down under these conditions, further investigation of the identification of the subsequent degradation products may be carried out by using confirmatory spectroscopic techniques such as LC-NMR, LC-MS, LC-DAD, GC-FTIR and GC-MS. The choice of technique will depend on the particular test substance. In the case of samples treated with radio-labelled test substance, further investigation of the stability may be carried out by scintillation counting and radio-thin layer chromatography.

R.7.1.16.1 Information requirements on stability of a substance

Only if their stability in organic solvent is considered critical is this study required for substances that are manufactured or imported in quantities of ≥100 t/y. The study does not need to be conducted if the substance is inorganic.

R.7.1.16.2 Available information on stability of a substance

Testing data on stability of a substance

A number of physical, chemical and biological processes can result in a decline in the actual concentration of a test substance in an organic solvent over time. Information on the stability of a test substance in a solvent is desirable, particularly when samples are to be stored. However, there does not appear to be any generally accepted methodology for performing such stability studies.

Factors affecting the rate of degradation include rates of hydrolysis, of photolysis and of oxidation. Enzymatic degradation by these or other pathways is considered to be of minor importance at low temperatures. Storage in the dark can minimise photodegradation. Oxidation can sometimes be an important process (e.g. thio compounds) but generally the rate of oxidation of organic compounds is slow (Egli, 1982). Hydrolysis, however, is suspected to be a main route of degradation. Studies of the stability of test substances in organic solvents are designed to closely simulate storage of such samples or extracts under test conditions before analysis.

Typically one or more concentrations of the test substance in the solvent are made up and analysed immediately after preparation (i.e. t = 0). They are then stored in appropriate vessels under the required test conditions (e.g. temperature, absence of light, etc) and analysed, along with a freshly prepared solution of the test substance at the original test concentration(s), at regular intervals during the period of interest. A single vessel containing the test substance under test may be re-sampled over time, or preferably, particularly when samples are frozen, samples may be stored in individual vessels for withdrawal and testing at the appropriate time period. All samples are maintained under the same storage conditions for the required period of time.

At each time of analysis, a sample is withdrawn from storage and mixed thoroughly before taking any aliquot for analysis. The analysis is carried out using the recommended method to determine whether any significant loss of the test substance has occurred during storage. It is important to analyse freshly made standards of the test substance in the organic solvent at the same time as analysing stored samples, so that any losses that may occur of the test substance during sampling, sample treatment and analysis are taken into consideration.

Glass bottles are usually appropriate, but cannot be used in the case of deep-freezing. The use of metal bottles, e.g. aluminium, is not to be recommended (non-transparent, risk of corrosion) (German Chemists Association, 1981). The storage vessel should contain sufficient sample for either a single analysis or a series of time-dependent analyses (whichever is required).

It is important to be able to have a check on the temperature, either by visual observation of a thermometer or from a calibrated recording device, to ensure that the temperature regime has been maintained throughout the period of the stability study.

Unlabelled reference material of suitable known purity (>95% m/m) may be used where a reliable method of analysis is available. Where an analytical method is still under development or is unlikely to be sufficiently sensitive, then radio-labelled [^{14}C] compounds should be used if available. Use of radio-labelled compounds can shorten the analysis time and help facilitate identification of any degradation products, should the test substance not be stable in the organic solvent.

Recovery or spiking experiments should normally be run. The number of spiking levels or the range of concentrations tested within a project should be left to the judgement of the analyst.

Recovery rate of spiking experiments should be ±20% of the average recovery rate given in the analytical method, but not less than 50% and no more than 100%.

It is assumed that the recurrent standard deviation of the test methods applied is less than ±10%. Consequently a deviation in the measured concentrations by more than 10% in terms of the initial concentration after a given storage period of the preserved sample is to be regarded as a real change in the corresponding contents of the sample. Samples should on no account be preserved for longer than the stated time and the actual storage period should, if appropriate, be shorter depending on the accuracy requirements.

Published data on stability of a substance

Generation of stability data is normally part of the analytical method development stage. Such data are not recorded in standard text books.

R.7.1.16.3 Evaluation of available information on stability of a substance

Experimental data on stability of a substance

Stability data of substances in organic solvents are not normally reported in standard published sources of physico-chemical data. Relevant sources of basic information regarding stability and storage conditions of substances are the Hazardous Substances Data Base (HSDB) and Sax's "Dangerous Properties of Industrial Materials" (see Section R.7.1.1.2).

Non-experimental data on stability of a substance

There is no estimation programme for stability of substances, although a knowledge of the structure of the test substance might indicate reactive groups that may give rise to instability of the test substance under the storage conditions.

Remaining uncertainty on stability of a substance

As there is often some basic variability in sampling from vessels and in analysis of test substances, tests designed to assess the stability of a test substance often require an analytical method with good precision and a reasonable period of time to be able to confirm the stability (or instability) of the test substance in the organic solvent of choice.

R.7.1.16.4 Conclusions on stability of a substance

A basic knowledge of the structure of the test substance might indicate reactive groups that may give rise to instability of the test substance. Further information should be obtained by checks on the stability of standards of the test substance in organic solvents as part of routine analytical protocols to confirm whether the test substance in unstable under normal storage conditions.

Further tests may be necessary to identify storage conditions which minimise any degradation of the test substance not only in organic solvents, but also during the conducting of other tests, such as water solubility, surface tension and in the preparation of test media for ecotoxicity studies (OECD, 2000). Identification of the degradation product(s) will allow an assessment of whether they are likely to be more toxic than the parent material in subsequent ecotoxicity studies.

R.7.1.16.5 Integrated testing strategy (ITS) for stability of a substance

Since there are no non-testing methods, stability in organic solvents for chemicals will be experimentally determined

R.7.1.16.6 References on stability of a substance

Egli, H. (1982). Storage Stability of Pesticide Residues. J. Agric. Food Chem., 30, 861- 866

German Chemists Association – Working Party on "Stabilization of Samples" from the Hydrochemistry Team (1981). Preservation of Water Samples", Water Research, 15, 233-241.

OECD Series on Testing and Assessment Number 23 Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures, ENV/JM/MONO(2000)6 (http://www.olis.oecd.org/olis/2000doc.nsf/LinkTo/env-jm-mono(2000)6)

R.7.1.17 DISSOCIATION CONSTANT

Information on the dissociation constant of a chemical in water (relating to the acidity constant, pK_a) is required for substances supplied at > 100 t/y. This property is important for ionisable organic substances, since it indicates which chemical species will be present at a particular pH (e.g. in fresh or marine waters, or in the gut). The fate (e.g. water solubility, adsorption and bioconcentration potential) and toxicity of the ionised form of a substance may be markedly different from the corresponding neutral molecule.

The dissociation of a substance may affect its fate and behaviour (e.g. bioconcentration and adsorption to soil/sediment). It is also important information for the interpretation of ecotoxicity and mammalian toxicokinetics data. For the latter, dissociation constant indicates the potential for absorption from the gastrointestinal tract and dermal absorption (certain ionic compounds are not absorbed dermally). Dissociated compounds are thought not to cross biological membranes and so knowledge of pKa allows amounts of unionised compound at the various pH values of the gastrointestinal tract to be calculated.

It should be noted that water solubility measurements for regulatory purposes are usually made in distilled water with pH of 6-9, whereas pH of aquatic toxicity test media is usually 7-8. The solubility of an ionisable substance may be significantly different between the two as a result, especially if the pK_a is between 5 and 9. This is because the extent of ionisation may vary according to pH or the level of counter ions in the test medium, and relatively small changes may significantly alter the equilibrium between dissociated and non-dissociated species. The dissociated and non-dissociated species may have different water solubilities and partition coefficients, and therefore bioavailability and toxicity (which in turn may cause the expression of different toxicities in different environments).

The pK_a may also be an important factor in deciding which method or conditions should be used to determine the octanol-water partition coefficient (K_{ow}) and soil adsorption partition coefficient (K_{oc}).

Definition of dissociation constant

Dissociation is the reversible splitting of a substance into two or more chemical species, which may be ionic (OECD, 1981). The process can be represented as:

$$RX \rightleftarrows R^+ + X^-$$

The dissociation constant (K) for this process is expressed as the ratio of concentrations of the species on either side of the equation in water at equilibrium:

$$K = [R^+][X^-] / [RX]$$

Where the cation R^+ is hydrogen, the substance can be considered an acid, and so this constant becomes an *acid* dissociation constant (K_a). This is usually represented by a slightly different equilibrium equation in aqueous solution, but is not presented here for succinctness. A substance can have more than one acidic (or basic) group, and a dissociation constant can be derived for each dissociation step in a similar way.

The K_a is related to pH as follows:

$$pK_a = pH - \log_{10}\left(\frac{[X^-]}{[HX]}\right) \qquad \text{(where p is } -\log_{10})$$

In practice, this means that the pK_a is equivalent to the pH at which the ionised and non-ionised forms are present in equal concentration (i.e. the substance has undergone 50% dissociation). The percentage of the dissociated and the neutral form of the compound can be determined from the dissociation constant. For example, for an acid with a pK_a of 5.5, the pH dependency of the behaviour of the substance can be described as follows:

1% dissociated at pH 3.5;

10% dissociated at pH 4.5;

50% dissociated at pH 5.5;

90% dissociated at pH 6.5;

99% dissociated at pH 7.5.

This means that even slight changes in pH can considerably affect the form in which the substance is present in solution, especially if the pKa value is close to environmentally-relevant pH values[15].

Strong acids have low pK_a values (<3); weak acids have pK_a values in the range 3-10. Conversely, strong bases have high negative and weak bases low negative pK_a values (because base strength is expressed as the acidity of the conjugate acid[16]). An ionisable substance with a pK_a above 10 will tend not to dissociate in water under normal environmental conditions.

Salts are reaction products of acids and bases that retain their ionic character, and may be prepared from any combination of acid and base. Their solubility depends upon the properties of the crystalline form; when dissolved in water (at moderate concentration) the ionisation state depends upon pH and the pK_a of each half. Substances that contain both an acidic and a basic functional group are described as amphoteric, and can exist as zwitterions or *internal salts* where both groups are charged (examples include amino acids which have carboxylic acid and amine functionality).

[15] Fresh surface waters have pH values in the range 4-9, whereas marine environments have a stable pH of about 8. pH normally varies between 5.5 and 7.5 for agricultural soils and sewage treatment plant tanks.

[16] The term pKb was once used to express basicity so that the same scale could be used alongside acidity – care should be taken when citing older sources to check which term has been used.

R.7.1.17.1 Information requirements on dissociation constant

Information on the dissociation constant of a chemical in water (pK_a) is required for substances manufactured or supplied in quantities of 100 t/y or more; it may also be helpful at lower tonnage. Clearly, if the substance cannot dissociate due to a lack of relevant functional groups, the dissociation constant is irrelevant. However, ionisable groups might not always be obvious (e.g. in sulphonyl urea herbicides, which contain the function $-S(=O)_2NH.C(=O)NH-$, the acid group is $S(=O)_2NH$). If a substance is much more soluble in water than expected, this may be an indication that dissociation has occurred.

For substances which contain multiple ionisable functionalities, pKa values should be reported for each dissociated species likely to exist in solution. In practice such compounds will exist in equilibrium with the other possible ionic forms at a given pH.

Column 2 of the REACH Annex IX provides two further exemptions. A study does not need to be conducted if:

- the substance is hydrolytically unstable (half-life less than 12 hours) or is readily oxidisable in water; or
- it is scientifically not possible to perform the test (e.g. because the analytical method is not sensitive enough).
- For complex mixtures (e.g. UVCBs) containing ionisable components the assessment of pKa is clearly complicated. Estimation of the individual components' pKa values, if appropriate, should be considered.

R.7.1.17.2 Available information on dissociation constant

Testing data on dissociation constant

OECD test guideline 112 (Dissociation constants in water, adopted May 1981) describes three laboratory methods to determine the pK_a of a substance. The three methods are appropriate for particular types of chemical as given in Table R.7.1-35.

Table R.7.1-35 OECD test guideline 112

OECD 112 Method	Applicability/Notes
Titration Method known quantity of the substance dissolved in distilled water and titrated against a standard acid or base solution. At least ten additions of titrant required and pH of the solution measured in order to complete a titration curve (Albert and Sergeant, 1962; Nelson and Faust 1969).	method is not suitable for poorly soluble substances
Spectrophotometric Method Involves determination of the ratio of the un-dissociated molecule to the ionised species. UV/visible absorbance spectrum of the non-ionised species is obtained by dissolving the substance in a non-absorbing buffer of known pH in which the substance does not dissociate.	only applicable for substances that have considerably different UV/Vis absorbances in the ionised and un-ionised forms. suitable method for low solubility compounds
Conductometric Method Conductivity of a 0.1 molar solution of the substance is measured. Additionally, measures of the conductivity of a range of dilutions are also made (Albert and Sergeant, 1969; ASTM 1974).	ECETOC (1998) reports that the coefficient of variation using this guideline is ± 0.1 log unit, equivalent to 10% at a pK of 1.

Other methods reported by Albert and Sergeant (1969) include:

Partition coefficient method: Octanol-water partition coefficient (K_{ow}) values for the substance are obtained using a range of buffer solutions of differing pH. A single replicate at each pH is sufficient. The pK_a may be derived using the following formulae:

For acids: $$pK_a = pH - \log\left[\left(\frac{P}{P'}\right) - 1\right]$$

For bases: $$pK_a = pH + \log\left[\left(\frac{P}{P'}\right) - 1\right]$$

where P is the partition coefficient of the undissociated species, and

P' is the apparent partition coefficient (of the neutral and ionised forms combined)

Solubility in water method: The water solubility of ionisable compounds varies with pH since the ionised form has a greater affinity for water. The solubility of the un-ionised form is first measured at a pH where this species is expected to predominate. Solubility is then determined at two further pHs (0.5 units above and below the first pH value, respectively), to test whether the same solubility figure is obtained (the three results should agree within the experimental error of the analytical technique). The solubility is then measured at a pH near to where the pK_a is expected to lie. This result can be used to calculate an approximate pK_a value. The solubility is then determined at a series of pH values evenly distributed within a range of pK_a ±1. The pK_a can be derived using the following formulae:

For acids: $pK_a = pH - \log\left(\left(S'/s\right) - 1\right)$

For bases: $pK_a = pH + \log\left(\left(S'/s\right) - 1\right)$

where S is the solubility of the undissociated species, and

S' is the apparent solubility (of the neutral and ionised forms combined)

Solubilities should be determined at constant ionic strength. This method is rather laborious but may be useful in those rare cases where the:

- substance is too insoluble in water for the titration method, and
- absorbance spectra for the different species are either too similar or completely absent, and
- partition coefficient method is impractical.

Published data on dissociation constant

Many literature sources for dissociation constant exist; many reference textbook and on-line sources are listed in Table R.7.1-2. These should be searched for published, valid data. Similar chemicals (analogues) for which measured pKa data according to a reliable method may be considered for read-across. Such values should be reinforced by estimated methods for pKa (e.g. the result of a QSPR prediction; see Section R.7.1.17.3). For most ionisable chemicals supplied at greater than 100 t/y that are predicted to dissociate at environmentally relevant pHs, a test will be required for dissociation constant. In some instances it may be acceptable to read-across dissociation constant from an analogue. However if there is significant variation between the analogue read-across and predicted pKa then a test should be conducted.

R.7.1.17.3 Evaluation of available information on dissociation constant

In general, pK_a values that are measured with a suitable method are preferred. If an estimated pK_a value suggests that the substance will dissociate significantly at environmentally relevant pH, a test may be required to confirm the result.

Experimental data on dissociation constant

The analytical method used to determine the amounts of dissociated and undissociated forms present in solution should not affect the equilibrium, and should also be capable of distinguishing between the chemical species involved.

Non-experimental data on dissociation constant

Within a congeneric series of chemicals, pKa is often closely correlated with the Hammett substituent constant, and this is the basis for a number of attempts at pKa prediction. Harris and Hayes (1990) and Livingstone (2003) have reviewed the published literature in this area.

The Hammett substituent constant σ was derived from a consideration of acid dissociation constants Ka, and most non-computerised methods of calculating Ka and pKa values are based on σ values:

pKa (derivative) = pKa (parent) – ρσ

where ρ is the series constant, which is 1.0 for benzoic acids. Harris and Hayes (1990) list ρ values for other series.

Harris and Hayes (1990) give several examples of pKa calculation, for example for 4-t-butylbenzoic acid. The pKa value of benzoic acid is 4.205, the ρ value for benzoic acids is 1.0, and the σ value for 4-tert-butyl is – 0.197. Hence the pKa value of 4-tert-butylbenzoic acid is calculated as 4.205 – (– 0.197) = 4.402. This value is virtually identical to the measured value for this compound.

A number of publications have dealt with estimation of pKa values from chemical structure, but these relate mostly to specific chemical classes, e.g. amines (Nagy *et al* 1989), 4-aminoquinolines (Kaschula *et al* 2002) and imidazol-1-ylalcanoic acids (Soriano *et al* 2004). There have, however, been a few attempts to model pKa values of diverse sets of chemicals. Klopman and Fercu (1994) used their MCASE methodology to model the pKa values of a set of 2464 organic acids, and obtained good predictions; a test set of about 600 organic acids yielded a standard error of 0.5 pKa unit. Klamt *et al* (2003) employed their COSMO-RS methodology to predict pKa values of 64 organic and inorganic acids, with a standard error of 0.49 pKa unit.

There are several software programs that predict multiple pKa values of organic chemicals (Table R.7.1-36), but there are no published comparisons of their performance; ACDLabs has a claimed standard error of 0.39 pKa unit for 22 compounds, and one of 0.36 pKa unit for 26 drugs. pKalc is claimed to be accurate to within 0.25 pKa unit (Tsantili-Kakoulidou *et al* 1997), QikProp is claimed to have a mean error of 0.19 pKa unit, and SPARC is claimed to have an RMS error of 0.37 pKa unit when evaluated on 3685 compounds (Hilal & Karickhoff 1995). ADMET Predictor is claimed to have a mean error of 0.56 pKa unit for a test set of 2143 diverse chemicals.

Table R.7.1-36 Software programs that predict multiple pKa values of organic chemicals

Software	Website	Availability	Format	Batch operation
ACDLabs	www.acdlabs.com	Purchase	molfile	Yes
pKalc	www.compudrug.com	Purchase	molfile	Yes
QikProp	www.schrodinger.com	Purchase	molfile	Yes
SPARC	ibmlc2.chem.uga.edu/sparc	Free on line	SMILES	No
ADMET Predictor	www.simulationsplus.com	Purchase	SMILES & others	Yes
ADME Boxes	www.ap-algorithms.com	Purchase	Various	Yes
ASTER	www.epa.gov	Not currently available	Not known	Yes
ChemSilico	www.chemsilico.com	Purchase	Various	Yes
Pipeline Pilot	www.scitegic.com	Purchase	Various	Yes
VCCLAB	www.vcclab.org	Free on line	Various	Yes

If only a few pKa predictions are required, it is recommended that SPARC be used. Structures are inputted into SPARC in the form of a SMILES string.

If many pKa predictions are required, software programs that can operate in batch mode should be used; any of those whose performance is mentioned above should give good results. The performance of the other software programs is not known.

As measured data for pKa are required only above a supply tonnage of 100 t/y, predictive methods for pKa are very important. This is particularly the case for chemicals that may be ionised at environmentally relevant pHs but for which supply is less than 100 t/y. Their speciation in solution can have a marked effect on other physico-chemical parameters (water solubility, partition coefficient, adsorption coefficient) that affect environmental behaviour and uptake from the gastrointestinal tract in mammals and dermal absorption.

Remaining uncertainty on dissociation constant

As measured data for this endpoint are only required for substances supplied at >100 t/y, there will be uncertainty in predicted values used for substances below this tonnage threshold. In the absence of valid literature values for the substance analogue values, backed up by predicted (QSPR) values, are preferred. If no analogue value is available and the predicted value falls within the environmentally relevant pH range (5-9) then the assessor should consider carrying out a test, although there is no requirement to do so at the lower (10) tonnage level. Measured data would help to understand better the species present in solution, and how they might affect other required physico-chemical endpoints at this tonnage (e.g. soil adsorption coefficient, water solubility).Dissociation constant also indicates the potential for absorption from the gastrointestinal tract and dermal absorption (certain ionic compounds are not absorbed dermally). Dissociated compounds are thought not to cross biological membranes and so knowledge of pKa allows amounts of unionised compound at the various pH values of the gastrointestinal tract to be calculated.

R.7.1.17.4 Conclusions on dissociation constant

Knowledge of an ionisable substance's pKa is useful for substances supplied at levels above 10 tonnes per annum, but is not a testing requirement. Ideally a literature value, analogue or QSPR prediction can be obtained for such substances. For ionisable substances supplied at tonnages greater than 100 t/y, dissociation constant is a requirement. Details of how to obtain a valid value for pKa are contained in Section R.7.1.17.5.

The endpoint is important as it indicates partitioning behaviour of the substance in the environment and what chemical species are likely to be present in the environment and *in vivo* (and therefore what organisms might in reality be exposed to). The process of dissociation (depending on pH) in solution can have a huge effect on other physico-chemical properties of a substance, including water solubility and soil adsorption, and on environmental fate and behaviour parameters including Henry's Law constant.

Concluding on C&L and chemical safety assessment

Ionisation can markedly alter the toxicity of a substance, and so it is important that toxicity test data are derived under appropriate conditions. Toxicity tests should preferably be carried out at both sides of the pKa, to fully characterise the possible differences in toxicity. Since this may not be possible in every case, the role of pH in the interpretation and use of such data should at least be discussed qualitatively in the assessment.

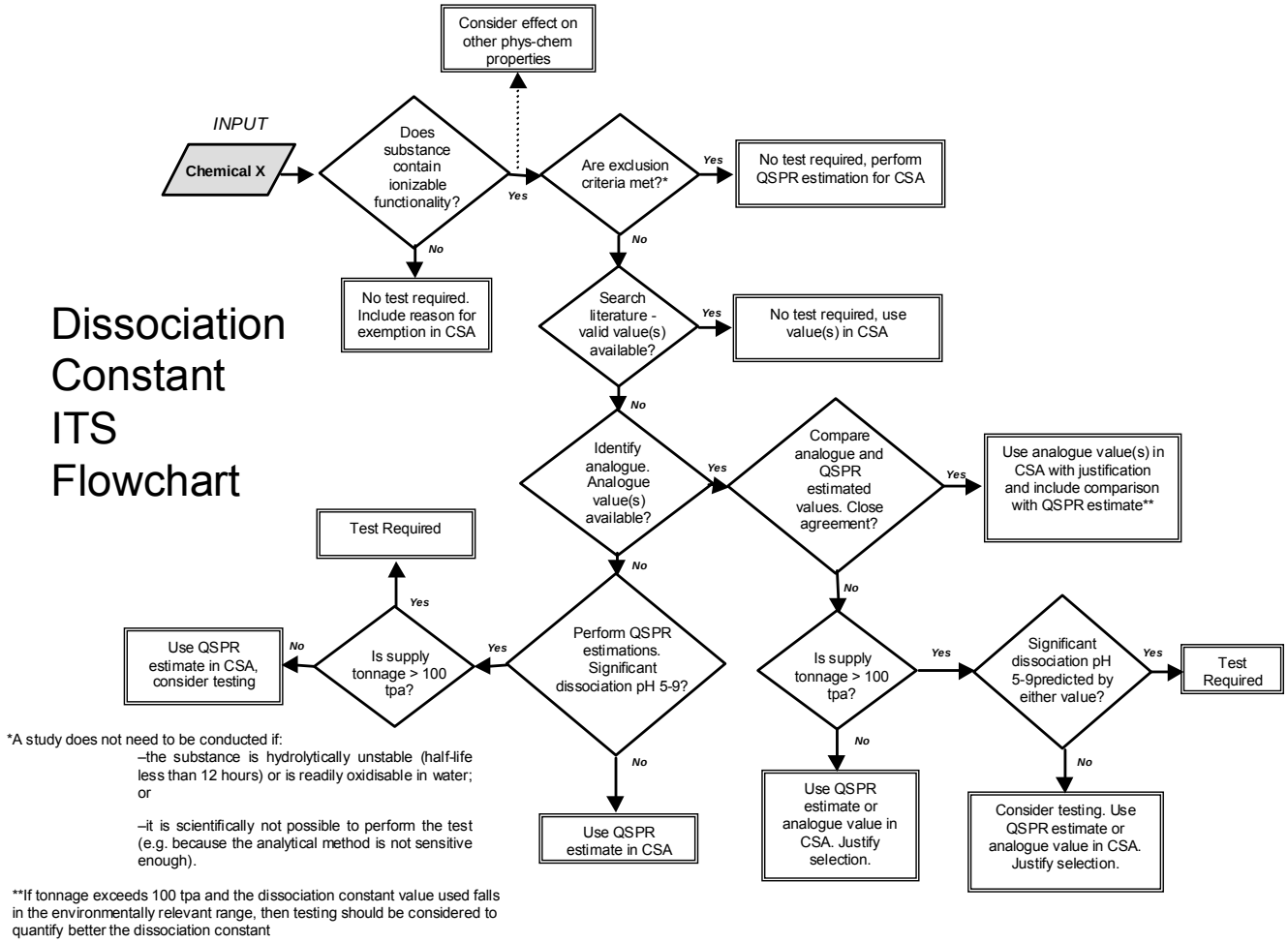
Similarly, it is important that ionisation is taken into account when assessing partitioning behaviour for the chemical safety assessment. Care should be taken that the PEC and the PNEC in the risk

characterisation represent similar conditions. PEC/PNEC comparisons should preferably be made at both sides of the pKa values, within the environmentally relevant pH-range. The higher PEC/PNEC ratio should be used in the risk characterisation, following the realistic worst-case approach. If it is not possible to carry out a quantitative analysis, the assessor should take the pH effect into account qualitatively.

R.7.1.17.5 Integrated testing strategy (ITS) for dissociation constant

The chemical structure should be examined to see if there are any functional groups that could dissociate. In general, the pK_a may be estimated using read-across or QSPR methods as a first step. If the pK_a appears to be within the environmentally-relevant pH range, a test should be conducted to confirm the result at the 100 t/y supply level (unless the substance degrades rapidly in water or there are other technical reasons that make it impossible to perform the test).

Figure R.7.1-9 Integrated testing strategy for dissociation constant



Examples and Case studies

The following examples are taken from the Existing Substances Regulation, EC 793/93.

- The pK_a value for phenol is reported to be 9.9. Alkyl substituents are electron donating, and so alkylphenols would be expected to be slightly less acidic than phenol, with a consequently higher pK_a. This suggests that the pK_a of alkylphenols is outside the environmentally relevant pH range, and that they would therefore be undissociated under normal conditions. This line of argument was accepted for nonylphenol without the need for a test as confirmation. Conclusion: careful consideration of the substance, using other available data sources (e.g. read-across), can help the assessor to construct a reasoned argument for the dissociation behaviour of the substance without the need for further testing.

- Tetrabromobisphenol-A has two acidic hydrogen atoms. The pK_a values had been reported to be 7.5 and 8.5 in the secondary literature, but no further details of the study were available. A repeat test was carried out using a method based broadly on OECD Guideline 112. However, this test only gave a single pK_a value and was considered unreliable. No QSPR methods were used to compare with the original data. However, the dissociation profile estimated from the original data was consistent with the available data on the variation of water solubility with pH (the monobasic and dibasic forms of tetrabromobisphenol-A would be expected to be of higher solubility than the undissociated form). This was used as a weight of evidence (WoE) argument to use the original data. Conclusion: careful consideration of the substance, using other available data sources (e.g. WoE approach), can help the assessor to construct a reasoned argument for the dissociation behaviour of the substance without the need for further testing.

R.7.1.17.6 References on dissociation constant

Albert, A. and Sergeant, E. P. (1962). Ionisation constants of acids and bases, Wiley Inc., New York.

Annual ASTM standards (1974). American Society for testing and materials, Philadelphia.

ECETOC (1998). Technical Report No. 74: QSARs in the Assessment of the Environmental Fate and Effects of Chemicals. European Centre for Ecotoxicology and Toxicology of Chemicals. Brussels.

Harris J.C. and Hayes M.J. Acid dissociation constant. In Lyman W.J., Reehl W.F. and Rosenblatt D.H. (Eds.), *Handbook of Chemical Property Estimation Methods.* American Chemical Society, Washington DC, 1990, pp. 6.1-6.28.

Hilal S.H. and Karickhoff S.W. A rigorous test for SPARC's chemical reactivity models: estimation of more than 4300 ionisation pKa's. *Quant. Struct.-Act. Relat.* (1995) 14, 348-355.

Kaschula C.H., Egan T.J., Hunter R., Basilico N., Parapini S., Taramelli D., Pasini E. and Monti D. (2002) Structure-activity relationships in 4-aminoquinoline antiplasmodials. The role of the group at the 7-position. *J. Med. Chem.* (2002) 45, 3531-3539.

Klamt A., Eckert F., Diedenhofen M. and Beck M.E. (2003) First principles calculations of aqueous pK(a) values for organic and inorganic acids using COSMO-RS reveal an inconsistency in the slope of the pK(a) scale. *J. Phys. Chem. A* 107, 9380-9386.

Klopman G. and Fercu D. (1994) Application of the multiple computer automated structure evaluation methodology to a quantitative structure-activity relationship study of acidity. *J. Comput. Chem.* (1994) 15, 1041-1050.

Livingstone D.J. Theoretical property prediction. *Current Topics in Med. Chem.* (2003) 3, 1171-1192.

Nagy P., Novak K. and Szasz G. (1989) Theoretical calculations on the basicity of amines. 1. The use of molecular electrostatic potential for pKa prediction. *J. Mol. Struct. – Theochem* 60, 257-270.

Nelson, N. H., and Faust, S. D. (1969). Acidic dissociation constants of selected aquatic herbicides, *Environmental Science and Technology*, 3 II: 1186-1188.

OECD (1981). Dissociation constants in water (titration method – spectrophotometric method – conductometric method). Organisation for Economic Co-operation and Development (OECD) Guideline for the testing of chemicals no 112.

Soriano E., Cerdan S. and Ballesteros P. (2004) Computational determination of pK(a) values. A comparison of different theoretical approaches and a novel procedure. *J. Mol. Struct. – Theochem* (2004) 684, 121-128.

Tsantili-Kakoulidou A., Panderi I., Csizmadia F. and Darvas F. Prediction of distribution coefficient from structure 2. Validation of PrologD, and expert system. *J. Pharm. Sci.* (1997) 86, 1173-1179^

R.7.1.18 VISCOSITY

The viscosity of fluids is environmentally relevant owing to the penetration of fluids into the soil and the harmful effect on the groundwater which may thus possibly be caused. From the point of view of the problem, surface tension, as well as questions of wettability, miscibility or solubility play a part in addition to viscosity, so that it is usually not sufficient to consider viscosity alone.

There is no environmentally relevant limit in the direction of low viscosities. The lower the viscosity the more easily a fluid seeps into the soil. The lowest dynamic viscosity of liquids occurring at room temperature is approximately 0.2 mPa s, that is to say 1/5 of the viscosity of water at 20ºC. A limit in the direction of high viscosities cannot be precisely quoted. Dynamic viscosities above approximately 10^7 mPa s are so high that penetration into the soil is no longer a problem.

In the case of substances which have a yield value (pastes, ointments), the substance may still not penetrate into the soil; although the dynamic viscosity may be low after the yield value has been exceeded. If the substance is soluble in water or can be emulsified, environmental damage may occur despite the existence of a flow limit.

Certain liquid substances and preparations may present an aspiration hazard in humans because of their low viscosity. Aspiration (in the context of pneumonitis) is the entry of liquid or solid materials into the trachea and lower respiratory tract. A variety of hydrocarbons (e.g. paint thinners, kerosene, gasolines, turpentine, lamp oil, motor oil, etc.) have been implicated in human poisoning incidents. Primary alcohols and ketones have been shown to pose an aspiration hazard only in animal studies.

Definition of viscosity

Viscosity is the property of a fluid substance of absorbing a stress during deformation which depends on the rate of the deformation. Similarly, the stress can be regarded as the cause which brings about a deformation rate.

The shear stress τ and the shear rate D are related by the equation

$$\tau = \eta D$$

η is defined as the dynamic viscosity.

For Newtonian liquids, the viscosity is constant at all shear rates and depends only on the variable pressure and temperature. For non-Newtonian liquids, the viscosity will vary with shear rate.

If the viscosity is measured with capillary viscometers without applied pressure, the measured quantity obtained is the ratio of dynamic viscosity to density, the so-called kinematic viscosity v. The SI unit of dynamic viscosity is the Pascal second, Pa s. For practical use a submultiple is more convenient; 1mPa s = 10^{-3} Pa s (one centipoise [cP] in the obsolete cgs-system).

The SI unit of kinematic viscosity is the square metre per second, m^2/s. The normal sub-unit derived from this is the square millimetre per second, $mm^2/s = 10^{-6}$ m^2/s (1 mm^2/s = 1 centistoke [cSt] in the obsolete cgs-system).

The classification criteria for aspiration hazard refer to kinematic viscosity. The following provides the conversion between dynamic and kinematic viscosity :

$$\text{Dynamic viscosity (mPa s)} = \text{Kinematic viscosity } (mm^2/s) \times \text{Density } (g/cm^3)$$

R.7.1.18.1 Information requirements on viscosity

It is noted that two physical properties are key to determination of aspiration hazard potential. Low viscosity and low surface tension (<33 mN/cm) determine the potential of a substance to constitute an aspiration hazard to the lung. Low viscosity leads to flow and low surface tension leads to spread of a liquid through the respiratory tract. In addition low solubility appears to be correlated with the ability of substances to penetrate the lungs (Gerarde and Ahlstrom, 1966). The characteristic of low solubility points to non-polar organic compounds.

R.7.1.18.2 Available information on viscosity

It is noted that two physical properties are key to determination of aspiration hazard potential. Low viscosity and low surface tension (<33 mN/cm) determine the potential of a substance to constitute an aspiration hazard to the lung. Low viscosity leads to flow and low surface tension leads to spread of a liquid through the respiratory tract. In addition low solubility appears to be correlated with the ability of substances to penetrate the lungs (Gerarde and Ahlstrom, 1966). The characteristic of low solubility points to non-polar organic compounds.

Testing data on viscosity

Viscosity measurements are carried out predominantly according to three measurement principles:

- The flow under gravity though a capillary (Capillary viscometer or flow cap)
- Shearing of the fluid between concentric cylinders, coneplate and parallel plate (rotational viscometer)
- Dynamic viscosity can be measured by movement of a ball in a vertical or inclined liquid-filled cylindrical tube (e.g. a rolling ball viscometer by Höppler, drawing ball viscometer, etc.)

With the Höppler viscometer the density must be known in order to calculate the dynamic viscosity.

The five methods listed are appropriate in principle for the investigation of Newtonian fluids. The measurement of non-Newtonian fluids is possible only with the rotational viscometer. For non-Newtonian fluids the results obtained are preferred in the form of flow curves, which must be interpreted, assuming the validity of various laws of flow.

The various methods of determining viscosity of liquids, as outlined in OECD TG 114 “viscosity of liquids”, are compared in Table R.7.1-37.

Table R.7.1-37 Methods of determining viscosity of liquids

Method of measurement	Viscosity Dynamic (mPa s)	Viscosity Kinematic (mm2/s)	Measuring range (mPa s or mm2/s)	Standardisation	Temperature constancy required (ºC)
Capillary viscometer		X	0.5 to 10^5	ISO 3104	± 0.1
Flow cup		X	8 to 700	ISO 3105	± 0.5
Rotational viscometer	X		10 to 10^9	ISO 3218.2	± 0.2
Rolling ball viscometer	X		0.5 to 10^5	No international standards, see DIN 53105	± 0.1
Drawing ball viscometer	X		0.5 to 10^7	No international standards, see DIN 52007 part 2	± 0.1

Reference substances need not be employed in all cases when investigating a new substance. They are provided primarily so that calibration of the method may be performed from time to time and to offer the chance to compare the results when another method is applied. I.U.P.A.C have recommended suitable reference substances of mineral oils, polyisobutenes and isobutylenes, issued by National laboratories (I.U.P.A.C, 1976).

Published data on viscosity

Viscosity data are not available in commonly used environmental handbooks nor on various environmental databases, such as HSDB (http://www.toxnet.nlm. nih.gov/cgi-bin/sis/htmlgen?HSDB), which provides summaries of chemical and physical properties of substances.

R.7.1.18.3 Evaluation of available information for viscosity

Experimental data on viscosity

The majority of methods described are based on both international and national standards. Appropriate ISO-Standards describing the concerned methods are cited. Measurements are carried out according to the specifications in the respective standards. Each determination of viscosity must be accompanied by the temperature at which the measurement was made. The determination should preferably be made at a temperature of 20ºC and one other temperature approximately 20ºC higher (temperature control limits are shown in the Table above). At least two determinations should be made at each temperature.

Non experimental data on viscosity

Liquid viscosity (η_L) can be regarded as a measure of the force needed to overcome the mutual attraction of molecules so that they can be displaced relative to each other (Grain 1990). The prediction of liquid viscosity has been reviewed by Grain (1990) and Reinhard and Drefahl (1999).

The method of van Velzen *et al* (1972) is based on the following equation:

$$\log \eta_L = B_3(1/T - 1/T_0)$$

where B_3 = a class-dependent constant, and T_0 (in K) is the temperature at which the viscosity is 1 centipoise (cp). Grain (1990) gives details of how to calculate B_3.

Grain's method (Grain 1990) allows the calculation of viscosity at different temperatures, given the viscosity at the boiling point T_b (K):

$$\ln \eta_L = \ln \eta_{Lb} + B_4(1/T - 1/T_b)$$

Values of η_{Lb} are: alcohols and amines (aliphatic and aromatic) 0.45; all other organic liquids 0.2. The calculation of B_4 is given by Grain (1990).

Skubla (1985) developed a group contribution scheme for calculating the viscosity for various homologous series:

$$\log \eta_L = a_0 + a_1 P_{vap}$$

where a_0 and a_1 are derived from group contributions, and P_{vap} is vapour pressure.

For liquid hydrocarbons, halogenated hydrocarbons, and O-containing compounds Joback and Reid (1987; reviewed also in Reinhard and Drefahl, 1999) proposed the following relationship based upon group contributions:

$$\eta_L(Nm^{-2}s) = M \exp\left[\frac{\sum n_i(\Delta\eta_A)_i - 597.82}{T} + \sum n_i(\Delta\eta_B)_i - 11.202\right]$$

Where M is the molecular mass in g/mol and T is the temperature in K. The viscosity is computed by using two terms of group contributions, denoted by A and B.

$(\Delta\eta_A)_i$ and $(\Delta\eta_B)_i$ are the contributions for the *i*th group type. The summations are carried out on all groups and n_i is the number of times the group occurs in the molecule.

Kauffman and Jurs (2001) used their ADAPT software to develop a QSPR for liquid viscosity, based on viscosity values for a number of common organic solvents. Using a neural network approach, they obtained RMS errors of 0.147 mPa.s for the training set of 159 chemicals, and 0.122 mPa.s for a test set of 19 chemicals.

The software programs ChemProp, Molecular Modeling Pro, PREDICT and ProPred predict liquid viscosity. PREDICT is reported to yield errors of 2 – 20%, depending on the calculation method used. The performances of the others are not known.

It is recommended that the method of van Velzen *et al* (1972) or of Grain (1990) be used for the estimation of liquid viscosity.

Remaining uncertainty on viscosity

The methods are capable of greater precision than is likely to be required for environmental assessment.

R.7.1.18.4 Conclusions on viscosity

Viscosity is relevant only to liquids, therefore for many substances this determination is not required. This would form the basis of a suitable justification for non-testing. For those substances that are tested, the results can be used to assign a suitable aspiration hazard class.

For organic substances, experimentally derived viscosity values, or values which are evaluated in reviews and assigned *recommended values*, are preferred over other determinations of viscosity. Validated quantitative structure activity relationships (QSARs) for viscosity may be used as a first step. However viscosity measurements are relatively quick and straightforward to undertake and it is recommended that experimental measurements are made to confirm predicted values for those chemicals that are close to the classification criteria.

Concluding on C&L and chemical safety assessment

Liquid substances and preparations may be classified R65 (Harmful: may cause lung damage if swallowed) by presenting an aspiration hazard in humans because of their low viscosity:

(a) for substances and preparations containing aliphatic, alicyclic and aromatic hydrocarbons in a total concentration equal or greater than 10% and having either:

- a flow time of <30 seconds in a 3 mm ISO cup according to ISO 2431 (April 1996/July 1999 edition) relating to 'Paints and varnishes – Determination of flow time by use of flow cups'
- a kinematic viscosity measured by a calibrated glass capillary viscometer in accordance with ISO 3104/3105 of less than 7 x 10^{-6} m^2/sec. at 40ºC (ISO 3104, 1994 edition, relating to 'Petroleum products – Transparent and opaque – Determination of kinematic viscosity and calculation of dynamic viscosity ': ISI 3105, 1994 edition, relating to 'Glass capillary kinematic viscometers – Specifications and operating instructions'), or
- a kinematic viscosity derived from measurements of rotational viscometry in accordance with ISO 3219 of less than 7 x 10^{-6} m^2/sec. at 40ºC (ISO 3219, 1993 edition, relating to 'Plastics – Polymers/resins in the liquid state or as emulsions or dispersions – Determination of viscosity using a rotational viscometer with defined shear rate').

Note that substances and preparations meeting these criteria need not be classified if they have a mean surface tension greater than 33 mN/m at 25ºC as measured by the du Nouy tensiometer or by the test methods shown in Annex V, Part A.5;

(b) for substances and preparations, based on practical experience in humans

Under the proposed GHS classification scheme, there are two hazard categories for aspiration toxicity. Category 1 is for chemicals known to cause human aspiration toxicity hazards or to be regarded as if they caused human aspiration toxicity hazard, where the viscosity criterion is if it is a hydrocarbon and has a kinematic viscosity of 20.5 mm^2/s or less, measured at 40ºC.

Examples of substances included in Category I are certain hydrocarbons, turpentine and pine oil. Category 2 is for chemicals known to cause concern owing to the presumption that they cause human aspiration toxicity hazard, where the viscosity criterion is a kinematic viscosity of 14 mm^2/s or less, measured at 40ºC. Some authorities would consider that n-primary alcohols with a composition of at least 3 carbon atoms but no more than 13, as well as isobutyl alcohol and ketones with a composition of no more than 13 carbon atoms, should be included in this category.

R.7.1.18.5 Integrated testing strategy (ITS) for viscosity

The tiered approach to testing (Section R.7.1.1.4) in conjunction with the choice of an appropriate test method represents an Intelligent Testing Strategy for this endpoint.

EXAMPLES AND CASE STUDIES ON VISCOSITY

A limited review of current risk assessments of chemical products (e.g. alkanes, ketones) has not provided any useful examples or case studies of products where viscosity measurements have been linked to the R65 classification.

R.7.1.18.6 References on viscosity

Gerarde, H.W. and Ahlstrom, D.B. (1966). The aspiration hazard and toxicity of a homologous series of alcohols. Archives of Environmental Health, 13:457-461.

Grain C.F. Liquid viscosity. In Lyman W.J., Reehl W.F. and Rosenblatt D.H. (Eds.), *Handbook of Chemical Property Estimation Methods*. American Chemical Society, Washington DC, 1990, pp. 22.1-22.20.

Kauffman G.W. and Jurs P.C. Prediction of surface tension, viscosity, and thermal conductivity for common organic solvents using quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* (2001) 41, 408-418.

Joback, K. G. and Reid, R.C., (1987) Estimation of Pure-Component Properties from group-contribution. *Chem. Eng. Commun.* 57, 233-243.

OECD Guidelines for Testing of Chemicals – Method 114 "Viscosity of Liquids"

Reinhard M. and Drefahl A. *Handbook for Estimating Physico-chemical Properties of Organic Compounds.* Wiley, New York, 1999.

Skubla P. Prediction of viscosity of organic liquids. *Coll. Czech. Chem. Commun.* (1985) 50, 1907-1916.

van Velzen D., Lopes Cardozo R. and Lagenkamp H. A liquid viscosity-temperature-chemical constitution relation for organic compounds. *Ind. Eng. Chem. Fundam.* (1972) 11, 20-25.

Appendices 1-4 to Section R.7.1

Appendix R.7.1-1 Comments on thermodynamic consistency of physico-chemical properties

The behaviour of chemicals released in the environment is described by the partitioning coefficients that quantify their distribution between the pure liquid phase, the gas phase, and the dissolved phases in water and *n*-octanol. When performing model calculations, internally consistent substance data have to be used. Beyer *et al*. (Beyer *et al*., 2002) suggested a procedure that enables the computation of a consistent set of data while taking advantage of all measured properties and exploiting the fundamental thermodynamic constraints that relates the investigated properties (Cole and Mackay, 2000). One typical example is Henry's law coefficient, which can be estimated as the ratio of saturation vapour pressure and water solubility. Beyer *et al*. (Beyer *et al*., 2002) published visual basic code for the computational procedures that proposed which is freely downloadable from http://www.usf.uos.de/ projects/elpos/.

In the same article Beyer et co-authors highlighted the importance of temperature for chemical fate models and the role of supercooled liquid properties when computing solubilities. Indeed, it is important that all solubilities apply to the same state and, generally, the liquid state is preferred because it better represents the conditions of chemicals in solution at low environmental concentrations and liquid or supercooled liquid properties are used in quantitative structure–property relationships (Beyer *et al*., 2002).

The adjustment methodology proposed by Beyer *et al*. was successfully used by Wania and Dugani (Wania and Dugani, 2003) in order to eliminate thermodynamic inconsistencies during the selection of physico-chemical properties for polybrominated diphenyl ethers.

Another comprehensive work where data were collected evaluated and rated in order to develop a consistent data set for polychlorinated biphenyls was carried out by Li *et al*. (Li *et al*., 2003).

References for Appendix R.7.1-1

Beyer, A., Wania, F., Gouin, T., Mackay, D., Matthies, M. (2002) Selecting internally consistent physico-chemical properties of organic compounds. *Environ. Toxicol. Chem.* 21(5), 941-953

Cole, J.G., Mackay, D. (2000) Correlating environmental partitioning properties of organic compounds: the three solubility approach. *Environ. Toxicol. Chem.* 19(2), 265-270

Wania, F., Dugani, C.B. (2003) Assessing the long range transport potential of polybrominated diphenyl ethers: a comparison of four multimedia models. *Environ. Toxicol. Chem.* 22(6), 1252-1261

Li N, Wania F, Lei YD, Daly GL. (2003) A comprehensive and critical compilation, evaluation and selection of physical chemical property data for selected polychlorinated biphenyls. J Phys Chem Ref Data 32:1535-1590.

Appendix R.7.1-2 pH correction of partition coefficients for ionisable substances

Partition coefficients used in environmental exposure assessment (log K_{ow}, Henry's law constant, adsorption/desorption coefficients) should be corrected according to the pH of the environment being assessed, to take only the undissociated fraction of the compound into account. This can be done using the following correction factor:

$$CORR = \frac{1}{1 + 10^{A(pH - pKa)}}$$

where: A = 1 for acids, -1 for bases

pH = pH-value of the environment

pKa = acid dissociation constant

This correction can only be used for partitioning coefficients that refer to the unionised form of the substance. This means that for estimated partitioning coefficients, water solubility and K_{ow} need to be determined for the neutral form. The choice of relevant pH values to be used in the calculation should be based on the pKa of the compound of concern and any relevant knowledge of the actual toxic form of the substance. For experimentally determined partition coefficients the need for correction should be assessed on a case-by-case basis, depending on the pH in the test. Because the pH of seawater (approximately 8) tends to be more constant than that of freshwater, the procedure to correct partition coefficients for ionisable substances may also be considered sufficiently reliable for marine conditions.

These principles apply also to the fate of the substance in sewage treatment plant. However, since the STP is a well buffered environment, a default pH of 7 can be used in the calculations. The role of pH in the experimental determination of the bioconcentration should also be acknowledged.

Appendix R.7.1-3 Temperature correction

Extract from current guidance document (environment chapter):

If experimentally determined physico-chemical data have been obtained at a temperature which for the substance under consideration would significantly change when extrapolated to the relevant temperature of the exposure models employed (e.g. 12°C in the regional model) then such an extrapolation should be considered. In most cases this will not be necessary.

However, the vapour pressure may for some substances change considerably according to the temperature even within a temperature range of only 10°C. In this case a general temperature correction should be applied according to the following equation:

$$VP(TEMP_{env}) = VP(TEMP_{test}) \cdot e^{\left(\frac{H_{0vapor}}{R} \cdot \left(\frac{1}{TEMP_{test}} - \frac{1}{TEMP_{env}} \right) \right)} \qquad (2)$$

Explanation of symbols:

$VP(TEMP_{env})$	vapour pressure at the environmental temperature	[Pa]	
$VP(TEMP_{test})$	vapour pressure as give in the data set	[Pa]	data set
$TEMP_{env}$	environmental temperature (scale-dependent)	[K]	
$TEMP_{test}$	temperature of the measured experimental VP	[K]	
$H_{0vapour}$	enthalpy of vaporisation	[J/mol]	$5 \cdot 10^4$
R	gas constant	$[Pa \cdot m^3/(mol \cdot K)]$	8.314

Care must be taken when the melting point is within the extrapolated temperature range. The vapour pressure of the solid phase is always lower than the extrapolated vapour pressure of the liquid phase. Extrapolation will therefore tend to overestimate the vapour pressure. There is no general solution to this problem.

The same approach can be followed for correcting the water solubility:

$$SOL(TEMP_{env}) = SOL(TEMP_{test}) \cdot e^{\left(\frac{H_{0solut.}}{R} \cdot \left(\frac{1}{TEMP_{test}} - \frac{1}{TEMP_{env}} \right) \right)} \qquad (3)$$

Explanation of symbols:

$SOL(TEMP_{env})$	solubility at the environmental temperature	[Pa]	
$SOL(TEMP_{test})$	solubility as give in the data set	[Pa]	data set
$TEMP_{env}$	environmental temperature (scale-dependent)	[K]	
$TEMP_{test}$	temperature of the measured experimental SOL	[K]	
H_{0solut}	enthalpy of solution	[J/mol]	$1 \cdot 10^4$
R	gas constant	$[Pa \cdot m^3/(mol \cdot K)]$	8.314

Appendix R.7.1-4 Henry's law constant and evaporation rate

Content of Appendix 7.1-4

Henry's law constant (H) is one of the most important factors in determining the environmental fate of chemicals. This physical law states that the mass of gas dissolved by a given volume of solvent is proportional to the pressure of the gas with which it is in equilibrium. The relative constant quantifies the partitioning of chemicals between the aqueous phase and the gas phase such as rivers, lakes and seas with respect to the atmosphere (gas phase). Indeed, this constant is a fundamental input for fugacity models that estimate the multimedia partitioning of chemicals (Mackay, 1991).

For many chemicals, volatilisation can be an extremely important removal process, with half lives as low as several hours. Henry's law constants can give qualitative indications of the importance of volatilisation. For chemicals with H values less than 0.01 Pa m^3/mole, the chemical is less volatile than water and as water evaporates the concentration of the chemical in the aqueous phase will increase; for chemicals with H values around 100 Pa m^3/mole, volatilisation will be rapid. However, the degree of volatilisation of substances from the aquatic environment is highly dependent on the environmental parameters for the specific water bodies in question, such as the depth and the gas exchange coefficient (i.e. wind speed and water flow rate). Henry's law constant cannot be used for evaluation of the removal of a chemical from the water phase without considering these factors.

To estimate volatilisation (as the half life of a compound in a river), a model outlined by Lyman et al. (1982) can be used. This model assumes a river of 1 metre depth, flowing at 1 m/sec with a wind velocity of 3 m/sec at 20ºC and requires only the Henry's law constant and the molecular weight of the chemical for input.

Definition

Different measurement units can be used for the definition of H (Staudinger and Roberts, 1996). It is conventional to define H in terms of gas concentrations in atmospheres (or Pa) and liquid concentrations in mol/m^3, thus the most typical units are atm m3/mole (or equivalent 100 Pa m^3/mole). However, it is often easier to work with dimensionless H by converting gas concentrations from atmospheres to mol/m^3. A dimensionless H is usually found in the engineering literature:

$$H_{cc} = \frac{C_{i,G}}{C_{i,L}} \quad \text{(Eq. X.1)}$$

$$H_{y,X} = \frac{Y_i}{x_i} \quad \text{(Eq. X.2)}$$

Where C_i = compound concentration of compound i on a volume (e.g. g/m^3 or mol/m^3) and G and L stand for Gas (air) and Liquid (water)

Another common definition which can be found in the physical sciences is :

$$H_{px} = \frac{y_i P_t}{x_i} \quad \text{(Eq. X.3)}$$

Where H_{px} = HLC [atm], Pt = total atmospheric pressure [atm] and x_i, y_i = mol fraction of compound i in aqueous solution and in the air phase at equilibrium [mol/mol].

Hcc (i.e., the ratio of mass or molar concentrations) is the preferred form for environmental engineering applications.

Experimental determination of Henry's law constant

The experimental approaches can be classified into two major groups: dynamic equilibration approach (often referred to as the *gas purge*) and the static equilibration approach. The following table briefly summarises the reviewing work done by Staudinger and Roberts (1996).

Table R.7.1-38 Experimental approaches for the determination of HLC

Approach	Average RSDs/Notes
Dynamic approach	
Batch air stripping (bubble column) HLC values are determined by measuring the rate of loss the compound of interest from water by isothermally stripping with a gas (typically air) in a suitable bubble column apparatus.	Average RSDs determined from different literature sources ranged from 2.8 to 21
Concurrent flow (wetted wall column) Values are determined based on the use of a wetted wall (desorption) column. The wetted wall column equilibrates an organic solute between a thin film of water and a concurrent flow of gas. Compound-laden water is introduced into the wetted wall column where it comes in contact with a compound-free gas stream flowing concurrently. HLC. The knowledge of flow rates and compound masses present in the separated phase streams enables the direct calculation of HLC.	Average RSDs determined from different literature sources ranged from 19 to 52 Preliminary work must be performed to ensure that phase equilibrium is reached.
Static approach	
Single equilibration A known mass of compound is introduced into an air-tight vessel with a known volume of water and air. When the equilibrium is attained the compound concentration is determined in one or both phases.	Average RSDs determined from different literature sources ranged from 2.8 to 30
Multiple Equilibration A liquid sample containing a known quantity of solute is allowed to equilibrate with a known volume of solute-free air. The air is the expelled and a new equilibration with the same amount of solute-free air is started. This process can be repeated until the number of equilibrations exhausts the mass of solute remaining in the system.	RSDs ranged from 0.7 to 3.5 This method is applicable for compounds with $0.1 \leq HLC \leq 2$ The experimental error is reduced with a larger number of equilibrations.
EPICS Technique HLC is determined by measuring the gas headspace concentration ratios from pairs of sealed bottles. Relative rather than absolute air-phase concentrations are required.	Average RSDs determined from different literature sources ranged from 2.9 to 19
Variable Headspac The method is based upon the measurement of the relative equilibrium air-phase concentration (gas chromatography peak areas) from aliquots of the same solution in multiple containers having different headspace-to-liquid volume ratios.	Average RSDs determined from different literature sources ranged from 0.5 to 7.9

A data-analysis of reviewed experimental studies for HLC can be found in Staudinger and Roberts (1996). HLC values can also be found in one or more of the following references: Sander (1999), CRC Handbook of Chemistry and Physics (2000), the NIST Chemistry WebBook (1998), and the : "The Handbook of Environmental Data on Organic Chemicals" (Verschueren K, 2001).

Main factors affecting Henry's Law Constant values

Staudinger and Roberts (1996) thoroughly explain all the factors affecting HLC values and report equations that quantify the effect of temperature and pH. Based on their work, in a majority of cases, temperature is the main parameter affecting HLC values for natural waters with moderate contamination (1 mg/ml or less). Other conditions that have influence on HLC values are listed in Table R.7.1-39 (Staudinger and Roberts, 1996):

Table R.7.1-39 Conditions that have influence on HLC values

pH	Important for compound classes that dissociate to a significant extent in water because only nondissociated species undergo air-water exchange. For most natural waters (6 < pH < 8) the apparent H will be significantly less than the intrinsic H.
Compound Hydration	Important for aldehydes, which hydrate nearly completely in water, resulting in H apparent being several orders of magnitude lower than the intrinsic constant.
Compound concentration/ Complex mixtures effects	If a solution cannot be regarded as diluted (e.g. concentration approaching 10.0 mg/ml) H apparent will be lower than H values determined at lower concentrations.
Dissolved salts	If the ionic strength of a solution is high (e.g. seawater) the apparent H will be higher than the H determined in pure water.
Suspended solids /Dissolved Organic Matter (DOM)	If a compound is easily adsorbed (e.g. pesticides) the apparent H will be higher than the H determined in pure water.
Surfactants	Compounds with high Kow are expected to have an effect on H by lowering its value. Recorded effects increase in direct proportion with Kow

It is worth noting that because of the complex nature of the water matrix the net effect of a possible combination of the parameters listed above may be more than the simple sum of individual effects (Staudinger and Roberts, 1996).

QSPR prediction of Henry's law constant

The prediction of HLC has been reviewed by Schwarzenbach *et al* (1993), Reinhard and Drefahl (1999), Mackay *et al* (2000) and Dearden and Schüürmann (2003). The most important approaches are:

- Ratio of water solubility (c_w) to vapour pressure (vp);
- Estimation using connectivity indices;
- Estimation using group and bond contribution methods.

The first method for estimating HLC is not strictly a QSAR method as it uses the water solubility (c_w) and vapour pressure (vp). It is not a highly accurate method, but neither is the measurement of HLC, especially for chemicals with very high or very low H values. vp/c_w can be converted to the dimensionless form of HLC (ratio of concentrations in air and water, c_a/c_w) or K_{aw} by the following equation, which is valid for 25°C:

$c_a/c_w = 40.874\ vp/c_w$

Since both water solubility and vapour pressure can be calculated by QSAR methods, then this approach might in some circumstances be a QSAR based method. The method is limited to substances of low water solubility (< 1.0 mol/l). If QSAR calculated values are used for vp and/or cw, then the respective uncertainties must be considered. For miscible compounds or the compounds with water solubility > 1 mol.l^{-1} the vp/ c_w method is not valid.

The second method is based on a combination of connectivity indices and calculated polarisability (Nirmalakhandan and Speece, 1988). A relatively narrow range of chemical types was used to develop the model, so it is not widely applicable. Moreover, Schüürmann and Rothenbacher [1992] found it to have poor predictive power.

Most prediction methods for *H* use a group or bond contribution approach, although some have used physico-chemical properties [Dearden et al 2000]. The group and bond contribution methods were first used by Hine and Mookerjee [1974], who obtained, for a set of 263 diverse simple organic chemicals, a standard deviation of 0.41 log unit for the group contribution method and one of 0.42 for the bond contribution method. Cabani et al. [1981] claimed an improvement in the group contribution method over that of Hine and Mookerjee, whilst Meylan and Howard [1991] extended the bond contribution method and obtained, for a set of 345 diverse chemicals, a standard error of 0.34 log unit. Their method, together with a group contribution method, is incorporated in the HENRYWIN module of the Episuite software.

Russell et al [1992] used their ADAPT software to develop a 5-descriptor model of log K_{aw} for a relatively small but diverse data-set:

$$\log K_{aw} = -0.547\ N\text{HEAVY} + 0.0402\ \text{WPSA} + 0.0360\ \text{RNCS} + 10.1\ \text{QHET} - 215\ \text{QRELSQ} + 0.73$$

$n = 63 \quad R^2 = 0.956 \quad s = 0.375$

where *N*HEAVY = number of heavy atoms, WPSA = (total solvent-accessible surface area) x (sum of surface areas of positively charged atoms), RNCS = (charge on most negative atom) x (surface area of most negative atom)/(sum of charges on negatively charged atoms), QHET = (total charge on heteroatoms)/(number of heteroatoms), and QRELSQ = square of (total charge on heteroatoms)/(number of atoms). Note that the ADAPT descriptors are available in the Pharma Algorithms ADME Boxes software (www.ap-algorithms.com).

The Ostwald solubility coefficient *L* (the reciprocal of K_{aw}) of a very diverse data-set of chemicals was modelled by Abraham et al [1994]:

$$\log L = 0.577\ R + 2.549\ \pi + 3.813\ \Sigma\alpha + 4.841\ \Sigma\beta - 0.869\ V_X + 0.994$$

$n = 408 \quad R^2 = 0.996 \quad s = 0.151$

where *R* = excess molar refractivity (a measure of polarisability), π = a polarity/polarisability term, α and β = hydrogen bond donor and acceptor abilities respectively, and V_X = the McGowan characteristic volume (see next section on prediction of relative density of liquids). The Abraham descriptors are approximately auto-scaled, so that the magnitudes of the coefficients in the above equation indicate the relative contributions of each term. It is clear that hydrogen bonding is the most important factor controlling water-air distribution; the greater magnitude of the $\Sigma\beta$ term probably reflects the strong hydrogen bond donor ability of water. Molecular size, represented by V_X, appears to play only a minor role in determining air-water partitioning. It may be noted that the very high correlation coefficient and low standard error of the equation suggest possible overfitting;

no external validation of the equation was provided. The Abraham descriptors are available in the Absolv-2 software (www.ap-algorithms.com).

Katritzky et al [1996] used their CODESSA software (www.semichem.com) to model the data-set of Abraham et al [1994]:

$\log L = 42.37\ \text{HDCA(2)} + 0.65\ [N(\text{O}) + N(\text{N})] - 0.16\ \Delta\text{E} + 0.12\ \text{PCWT} + 0.82\ N_R$

$+ 2.65$

$n = 406 \quad R^2 = 0.942 \quad s = 0.52$

where HDCA(2) = hydrogen bond donor ability, $N(\text{O}) + N(\text{N})$ = a linear combination of the number of oxygen and nitrogen atoms, ΔE = HOMO-LUMO energy difference, PCWT = most negative partial charge-weighted topological electronic index, and N_R = number of rings. It may be noted that the standard error of 0.52 log unit is more realistic than is that of 0.151 reported by Abraham et al [1994].

Katritzky et al [1998] used predicted vapour pressure and aqueous solubility to calculate Henry's law constant according to equation 20 for 411 diverse chemicals. The table giving their results was inadvertently omitted in their paper, but they reported a standard error of 0.63 log unit, which is not very much greater than that found (0.52 log unit) in their correlation shown in equation 23 above.

Very recently QSPRs have been developed by Modarresi et al [2007] using a very large (940-compound) diverse data set. Using genetic algorithm selection of descriptors, they obtained a 10-descriptor QSPR with a root mean square error of 0.571 log unit.

There are seven software programs that calculate Henry's law constant, namely Episuite, Absolv-2, ADME Boxes, ASTER, ChemProp, ProPred and SPARC. The performances of the last five are not known.

Dearden and Schüürmann [2003] tested a number of methods for prediction of log *H*, using a large, diverse test set of 700 chemicals. Only one of the methods, the bond contribution method in the HENRYWIN module of the Episuite software, allowed prediction of log *H* for all 700 chemicals, with a mean absolute error of prediction of 0.63 log unit.

It is recommended that the HENRYWIN module of the Episuite software be used for the prediction of Henry's law constant.

Evaporation rate

Evaporation rates generally have an inverse relationship to boiling points, i.e. the higher the boiling point, the lower the rate of evaporation. Knowledge of the evaporation rate of spills of volatile liquids can be useful in several respects. If it is known that a spill of a high vapour pressure liquid will evaporate completely in a short period of time, it may be preferable to isolate the area and avoid any intervention or clean-up. The evaporation rate also controls the atmospheric concentration of the vapour and hence the threat of explosion or fire. Data on the volatility properties of the liquid, its temperature, the wind speed, and the spill dimensions are used to calculate the evaporation rate and hence the fraction evaporated at any time.

The chemical's tendency to partition into the atmosphere is controlled by the vapour pressure, which is essentially the maximum vapour pressure that a pure chemical can exert in the atmosphere. This can be viewed as a kind of *solubility* of the chemical in the atmosphere. Using the ideal gas

law (PV=nRT), the vapour pressure P (units Pa) can be converted into a solubility (mol/$m^{3)}$, where the gas constant R is 8.314 Pa.m^3/mol.K and T is absolute temperature (K).

Conversion from vapour pressure into concentration in air under ambient temperature:

% volume = vapour pressure (Pa)/101,325 x 100

or ppm = vapour pressure (Pa)/101,325 x 1,000,000

Since the molar volume is the same for all ideal gases (equal volumes of all gases under the same conditions of temperature and pressure contain the same number of molecule) ppm ≡ volume (i.e. ml/m^3). To convert to weight per unit volume:

X ppm = X x MW/24.041 mg/m^3, 1 mg/m^3 = 24.041/MW ppm

In the formulation of paints and related products, solvents are chosen based on their evaporation characteristics appropriate to the application technique and the curing temperature. To a large extent the evaporation rate of a solvent determines where and how it can be used. In determining the evaporation rate of solvents, n-butyl acetate is used as the standard and is assigned an evaporation rate value of 1. Other solvents are assigned evaporation rate values that indicate how fast they evaporate in relation to n-butyl acetate. For instance, a solvent that evaporates three times as fast as n-butyl acetate would be assigned a value of 3, whereas a solvent that evaporates half as fast as n-butyl acetate would be assigned a value of 0.5.

The rate of evaporation is determined using ASTM D3539-87. A known volume of liquid is spread on a known area of filter paper that is suspended from a sensitive balance in a cabinet. Dry air or nitrogen at 25ºC is passed through the cabinet at a known rate. The loss of weight is determined and plotted against time.

References for Appendix R.7.1-4

Abraham M.H., Andonian-Haftvan J., Whiting G.S., Leo A. and Taft R.S. Hydrogen bonding. Part 34. The factors that influence the solubility of gases and vapours in water at 298 K, and a new method for its determination. *J. Chem. Soc. Perkin Trans. 2*, (1994) 1777-1791.

ASTM D 3539-87 (Reapproved 2004) "Standard Test Methods for Evaporation Rates of Volatile Liquids by Shell Thin-Film Evapourometer".

Cabani S., Gianni P., Mollica V. and Lepori L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Solut. Chem.* (1981) **10**, 563-595.

CRC (2000) CRC Handbook of Chemistry and Physics 81st Edition. Editor in Chief, D. Lide. CRC Press

Dearden, J.C., Cronin, M.T.D., Ahmed, S.A. and Sharra, J.A. QSPR prediction of Henry's law constant: improved correlation with new parameters. In Gundertofte, K. and Jørgensen, F.S. (Eds), *Molecular Modelling and Prediction of Bioactivity*. Kluwer Academic/Plenum Publishers, New York, 2000, pp. 273-274.

Dearden J.C. and Schüürmann G. Quantitative structure-property relationships for predicting Henry's law constant from molecular structure. *Environ. Toxicol. Chem.* (2003) 22, 1755-1770.

Hine J. and Mookerjee P.K. The intrinsic hydrophilic character of organic compounds. Correlations in terms of structural contributions. *J. Org. Chem.* (1974) **40**, 292-298.

Katritzky A.R., Mu L. and Karelson M. A QSPR study of the solubility of gases and vapors in water. *J. Chem. Inf. Comput. Sci.* (1996) **36**, 1162-1168.

Katritzky A.R., Wang Y., Sild S., Tamm T. and Karelson M. QSPR studies on vapour pressure, aqueous solubility, and the prediction of air-water partition coefficients. *J. Chem. Inf. Comput. Sci.* (1998) **38**, 720-725.

Kim, B.R., Kalis, E.M., DeWulf, T., Andrews, K.M. (2000) Water Environ. Res. 72, 65-74.

Lyman, W.J et al. *Handbook of Chemical Property Estimation Methods*. McGraw-Hill, New York, 1982.

MackayD. (1991) Multimedia environmental models. Lewish Publishers. Chelsea, MI

Mackay D., Shiu W.Y. and Ma K.C. Henry's law constant. In Boethling R.S. and Mackay D. (Eds.), *Handbook of Property Estimation Methods for Chemicals.* Lewis, Boca Raton, FL, 2000, pp. 69-87.

Meylan W.M. and Howard P.H. Bond contribution method for estimating Henry's law constants. *Environ. Toxicol. Chem.* (1991) **10**, 1283-1293.

W.G. Mallard and P.J. Linstrom, Eds., NIST Chemistry WebBook, NIST Standard Reference Database Number 69. November 1998. National Institute of Standards and Technology, Gaithersburg, MD, 20899

Modarresi H., Modarress H. and Dearden J.C. QSPR model of Henry's law constant for a diverse set of organic chemicals based on genetic algorithm-radial basis function network approach. *Chemosphere* (2007) **66**, 2067-2076.

Nirmalakhandan NN and Speece (1988). QSAR model for predicting Henry's Law Constant. Environ. Sci. Technol. **22**, 1349-1357.

Reinhard M. and Drefahl A. *Handbook for Estimating Physico-chemical Properties of Organic Compounds.* Wiley, New York, 1999.

Russell C.J., Dixon S.L. and Jurs P.C. Computer-assisted study of the relationship between molecular structure and Henry's law constant. *Anal. Chem.* (1992) **64**, 1350-1355.

Sander, R.. (1999) Compilation of Henry's Law Constants for Inorganic and Organic

Species of Potential Importance in Environmental Chemistry, Version 3

(http://www.mpch-mainz.mpg.de/~sander/res/henry.html)

Schwarzenbach R.P., Gschwend P.M. and Imboden D.M. *Environmental Organic Chemistry*. Wiley, New York, 1993.

Schüürmann G. and Rothenbacher C. Evaluation of estimation methods for the air-water partition coefficient. *Fresenius Environ. Bull.* (1992) **1**, 10-15.

Staudinger, J. and Roberts, P.V. (1996) A critical review of Henry's Law Constants for Environmental Applications. *Crit. Rev. Environ. Sci. Technol*. 26(3):205-297

Verschueren K (2001) Handbook of Environmental Data on Organic Chemicals 4th Edition (2 volume set). John Wiley & Sons Inc, New York, USA

R.7.2 Skin- and eye irritation/corrosion and respiratory irritation

R.7.2.1 Introduction

Irrespective of whether a substance can become systemically available, changes at the site of first contact (skin, eye, mucous membrane/ gastro-intestinal tract, or mucous membrane/ respiratory tract) can be caused. These changes are considered local effects. A distinction in local effects can be made between those observed after single and those after repeated exposure. In this guidance document, the focus will be on local effects after single ocular or dermal exposure. However, wherever possible, use should also be made of existing repeated dose data as far as they may contain valuable information for the purpose of assessing and classifying effects after single ocular or dermal exposure.

Substances causing local effects after single exposure can be further distinguished as irritant or corrosive substances, depending on the severity, reversibility or irreversibility of the effects observed. *Corrosive substances* are those which may destroy living tissues with which they come into contact. *Irritant substances* are non-corrosive substances which, through immediate contact with the tissue under consideration may cause inflammation. These tissues are in the present context skin, eye (cornea and conjunctiva) and mucous epithelia such as the respiratory tract. Criteria for classification of irritant and corrosive substances are given in Annex VI to Directive 67/548/EEC[17].

Substances may also cause irritant effects only after repeated exposure, for example organic solvents. This type of chemicals may have defatting properties (Ad-hoc Working group on Defatting substances, 1997). Chemicals that have a similar mechanism need to be considered for labelling with the risk phrase 'repeated exposure may cause skin dryness or cracking'.

Information on the mechanism underlying corrosion and irritation from skin, eye and respiratory tract are given in

Appendix R.7.2-1 Mechanisms of local toxicities: skin corrosion/irritation, eye and respiratory irritation

R.7.2.1.1 Definitions of skin- and eye irritation/corrosion/respiratory irritation

Dermal irritation: Defined in OECD TG 404/EU B.4 as *"…the production of reversible damage of the skin following the application of a test substance for up to 4 hours"*.

Dermal irritation after repeated exposure: Substances which may cause skin dryness, flaking or cracking upon repeated exposure but which can not be considered a skin irritant.

Dermal corrosion: Defined in OECD TG 404/EU B.4 as *"…the production of irreversible damage to skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test substance for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discolouration due to blanching of the skin, complete areas of alopecia, and scars…."*.

[17] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

Eye irritation: Defined in OECD TG 405/EU B.5 as *"…the production of changes in the eye following application of a test substance to the anterior surface of the eye, which are fully reversible within 21 days of application"*.

Eye corrosion: Defined in OECD TG 405/EU B.5 as *"…the production of tissue damage in the eye, or serious physical decay of vision, following application of a test substance to the anterior surface of the eye, which is not fully reversible within 21 days of application"*.

Respiratory tract irritation: There is no EU or OECD TG for respiratory irritation and testing for respiratory irritation is not required under REACH. Respiratory irritation is often used to describe either or both of two different toxicological effects, *sensory irritation* and *local cytotoxic effects*.

Risk phrases and hazard codes to be considered within the EU classification and labelling information system (EC, 2001)**:**

a) Corrosion

■ Corrosive: Risk phrase "R34", "Causes burns". Hazard code: "C"
Full thickness destruction of the skin occurs as a result of up to 4 hours exposure.

■ Corrosive: Risk phrase R35, "Causes severe burns". Hazard code: "C"
Full thickness destruction of the skin occurs as a result of up to 3 minutes exposure.

b) Irritation

■ Irritant: Risk phrase "R38", "Irritating to skin". Hazard code: "Xi",

■ Irritant: Risk phrase "R66", "Repeated exposure may cause skin dryness or cracking". Hazard code: "Xi" **a**

■ Irritant: Risk phrase "R36", "Irritating to eyes". Hazard code: "Xi"

■ Irritant: Risk phrase: "R41", "Risk of serious damage to eyes". Hazard code: "Xi"

■ Irritant: Risk phrase: "R37", "Irritating to respiratory system". Hazard code: "Xi"

Note that cytotoxic irritation of the respiratory tract, if observed in repeated dose studies at critical concentrations and if composed of a clearly nectrotic character, has been classified according to the criteria for R48

Information on the Globally Harmonised System (GHS) for the classification and labelling of chemicals can be found at (UN/ECE, 2003):

http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html

Note that dermal/respiratory irritation following repeated exposure are not discussed in the present context, since this report focuses on acute effects after single exposure. However, data from repeated exposure studies may be useful in certain cases (e.g. if the substance was identified as a corrosive or strong irritant after the first application or for deriving quantitative information). Nevertheless, for the sake of completeness, both the definition of dermal irritation after repeated exposure as well as the related risk phrase (R66) are given here. More guidance on local effects after repeated exposure can be found in Section R.7.5 on repeated dose toxicity.

R.7.2.1.2 Objective of the guidance on skin- and eye irritation/corrosion/respiratory irritation

The general objectives are:

a. to establish whether information from physical/chemical data, from non-testing methods (grouping, QSARs and expert systems), from *in vitro* studies, from animal studies or human experience provide evidence that the substance is, or is likely to be, corrosive.
b. to establish whether information from physical/chemical data, from non-testing methods (grouping, QSARs and expert systems), from *in vitro* studies, from animal studies or human experience provide evidence of significant skin, eye or respiratory irritation.
c. to establish the time of onset and the extent and severity of the responses and information on reversibility.
d. to gather, in the process of hazard identification, any quantitative data on dose-response relationships that might allow the derivation of DNELs essential for a complete risk assessment.
If a risk assessment is necessary, both the severity of the identified hazard (in so far at it can be judged from the test data) and the probability of the occurrence of an acute corrosive or irritant response in humans must be assessed based on the likelihood of any exposure to the substance and in relation to the route, pattern and extent of the expected exposure.

Please note that there are currently no standard tests and no OECD TG available for respiratory irritation and there is no testing requirement for respiratory irritation under REACH. Consequently respiratory irritation is not included in the testing strategies suggested in this report.

Nevertheless, account should be taken of any existing and available data that provide evidence of the respiratory irritation potential of a substance. Moreover, the data on local dermal or ocular corrosion/irritation might contain information that is relevant for the respiratory endpoint and this should be considered accordingly. It is for instance a reasonable precaution to assume that corrosive (and severely irritating) substances would also cause respiratory irritation when vaporised or in form of aerosol, though formal classification with R37 is not justified in this case. Furthermore, information from cases where symptoms have been described associated with occupational exposures can be used on a case-by-case basis to characterise the respiratory irritation potency of a substance. Information from acute and repeated dose inhalation toxicity studies may also be considered sufficient to show that the substance causes respiratory irritation at a specific concentration level or range. The data need to be carefully evaluated with regard to the exposure conditions (sufficient documentation required). Possible confounding factors should be taken into account.

R.7.2.2 Information requirements on skin/eye irritation/corrosion

The information requirement for irritation and corrosion that shall be submitted for registration and evaluation purposes is specified in REACH Annexes VI to XI. According to Annex VI, the registrant should gather and evaluate all available information before considering further testing. These include physico-chemical properties, (Q)SAR, grouping, *in vitro* data, animal studies, and human data. Furthermore, information on exposure, use and risk management measures should also be collected and evaluated.

If these data are inadequate for hazard and risk assessment, further testing should be carried out in accordance with the requirement in REACH Annexes VII (≥1tpa) and VIII (≥10tpa).

Information requirements for quantities of ≥1 tpa (Annex VII)

If new testing data are necessary, these must be derived from *in vitro* methods only. Annex VII does not foresee *in vivo* testing for irritancy or corrosivity.

The standard information required (column 1) at this tonnage level for skin corrosion/irritation can be satisfied by following four steps: (1) assessment of the available human and animal data, (2) assessment of the acid or alkaline reserve, (3) *in vitro* skin corrosivity study, (4) an *in vitro* skin irritation study.

Column 2 lists specific adaptations that specify when step 3 or 4 do not have to be conducted. These are:

1. when the available information already indicates that the criteria are met for classification as corrosive to the skin or irritating to eyes.
2. the substance is flammable in air at room temperature (Please note that this rule should actually read: "the substance is spontaneously flammable in air at room temperature").
3. the substance is classified as very toxic in contract with skin.
4. an acute toxicity study by the dermal route does not indicate skin irritation up to the limit dose level (2000 mg / kg body weight).

The standard information required (column 1) at this tonnage level for eye irritation can be satisfied by following four steps: (1) assessment of the available human and animal data, (2) assessment of the acid or alkaline reserve, (3) *in vitro* eye irritation study.

Column 2 lists specific adaptation that specify when step 3 *in vitro* eye irritation testing is not necessary. These are:

1. when the available information already indicates that the criteria are met for classification as corrosive to the skin or irritating to eyes.
2. the substance is flammable in air at room temperature (Please note that this rule should actually read: "the substance is spontaneously flammable in air at room temperature").

Information requirement for quantities of ≥10 tpa (Annex VIII)

For substances manufactured or imported in quantities of ≥10 tpa *in vivo* testing is required to meet the standard information requirements of Annex VIII column 1. Column 2 lists specific rules that allow deviating from the standard testing regime. More importantly, the standard testing regime of Annex VII and VIII can be adapted by the rules laid down in Annex XI, e.g. allowing to avoid unnecessary animal testing as required in Annex VIII (see Section R.7.2.4.1 for possible alternatives). For detailed information, see the REACH legislative text.

In summary these rules for adapting the standard testing are for:

a) skin irritation:

- the substance is classified as corrosive to the skin or as a skin irritant or
- the substance is a strong acid (pH < 2) or base (pH > 11.5) or
- the substance is flammable in air at room temperature (Please note that this rule should actually read: "the substance is spontaneously flammable in air at room temperature") or

- the substance is classified as very toxic in contact with skin or
- or an acute toxicity study by the dermal route does not indicate skin irritation up to the limit dose level (2000 mg/kg body weight).

b) eye irritation:

- the substance is classified as irritating to eyes with risk of serious damage to eye or
- the substance is classified as corrosive to the skin and provided that the registrant classified the substance as eye irritant or
- the substance is a strong acid (pH<2,0) or base (pH > 11,5) or
- the substance is flammable in air at room temperature (Please note that this rule should actually read: "the substance is spontaneously flammable in air at room temperature").

Guidance on the application of these rules is given in the integrated testing strategies described in Section R.7.2.6.

R.7.2.3 Information and its sources on irritation/corrosion

R.7.2.3.1 Non-human data on irritation/corrosion

Non-testing data on irritation/corrosion

Physico-chemical properties

Information of relevance to irritation/corrosion can be inferred from basic physico-chemical characteristics of a substance (extreme pH). Substances with *extreme* pH values will be inevitably skin corrosives or severe eye irritants:

IF pH≤2 or pH≥11.5, THEN predict to be corrosive to skin and severely irritating to eyes. See also Section R.7.2.4.1)

Grouping, (Q)SARs and expert systems

Non-testing methods can be divided into three categories: 1) grouping approaches (read-across, SARs and categories), 2) QSARs, and 3) expert systems, generally incorporating multiple (Q)SARs, expert rules and data. These methods can be used for the assessment of skin and eye irritation and corrosion, if they provide relevant and reliable data for the chemical of interest. Generally this means that the use of non-testing methods should be justified by means of detailed descriptions. In the case of QSARs and expert systems, the justification is provided by means of a QSAR Model Reporting Format (QMRF). In this guidance document, it is not possible to provide QMRFs for all existing models. However, QMRFs for potentially useful models are available from the JRC QSAR Model Database, which will be accessible via the website (http://qsardb.jrc.it). More detailed guidance on QSAR models, their use and reporting formats, including the QMRF, is provided in Section R.6.1.

In the case of skin irritation and corrosion, many of the models have a mechanistic basis, which provides additional information on the relevance of the model.

SAR and read-across on skin irritation and corrosion:

SARs and read-across are treated together because the existence of a SAR (structural alert or set of fragments) provides one means of justifying read-across.
The occurrence of structural analogues that exhibit corrosion (or irritation) potential can be used to predict the effect in the substance of interest and derogate from further assessment, as indicated in the OECD testing strategy for skin irritation/corrosion (OECD, 2001). Negative data from structural analogues may also be used to make predictions in certain cases, provided that there are no other substructures in the substance that are thought likely to cause the effect. Structural alerts are generally considered to reflect some kind of chemical or biochemical reactivity that underlies the toxicological effect.

The non-reactive chemicals, which lack alerts for reactivity, will normally not exhibit irritant or corrosive effects. However, irritant effects such as irritant contact dermatitis can occur in the case of exposure to organic solvents, which have defatting properties. Chemicals that have a similar mechanism need to be considered for 'Repeated exposure may cause skin dryness or cracking' (R66) (Ad-hoc Working group on Defatting Substances, 1997).

An example of a simple SAR is the use of the hydroperoxide group as an alert for corrosivity, which is mechanistically based on the fact that hydroperoxides are both acidic and oxidisers. Another SAR is the peroxide group (R1-C-O-O- R2), based on the fact that peroxides are oxidising agents. These SARs are mentioned in the Classification and Labelling guide (EC, 2004). The validity of these models, however, is not given there. Rorije et al. (2007) showed that 75 and 60% of the hydroperoxides and peroxides are classified for corrosivity and irritancy, respectively.

A variety of SARs for predicting the presence of irritation or corrosion have been described by Hulzebos et al. (2001, 2003, 2005), and others have been incorporated into the BfR rulebase and the SICRET tool (Walker et al., 2005, see Appendix R.7.2-2 QSARs and expert systems for skin irritation and corrosion).

Read-across has been used to a limited extent in the New Chemicals notification procedure for the classification of skin irritants (Hoffmann et al., 2005). As of May 2006, one substance has been classified as R38 by read-across from an analogue, and seven substances have been unclassified for R38 on the basis of read-across from analogues that were not found to meet the classification criteria for skin irritation (Thomas Cole, ECB, personal communication).

QSARs and expert systems on skin irritation and corrosion:

QSARs and expert systems for skin irritation and corrosion have been described in several reviews (Hulzebos et al., 2001, 2003, 2005; Patlewicz et al., 2003; Gallegos Saliner et al., 2006). A few examples are presented in Appendix R.7.2-2 QSARs and expert systems for skin irritation and corrosion, including literature-based QSAR models, commercial models, and expert systems.

Most of the QSARs reported in the literature have been developed from small data sets of specific groups of compounds, although in some cases more diverse and larger datasets were also examined. In general, it has been suggested that basic physico-chemical parameters such as acidity, basicity, hydrophobicity, and molecular size as well as electrophilic reactivity, are useful to predict the toxic potential of homologous chemicals. In contrast, models intended to predict the toxic potential of heterogeneous groups of chemicals emphasise the commonality of structural features.

Commercial models are coded in the form of expert systems, which are computer programs that guide hazard assessment by predicting toxicity endpoints of certain chemical structures based on the available information. Expert systems can be based on an automated rule-induction system (e.g.

TOPKAT, HazardExpert and MultiCASE), or on a knowledge-based system (e.g. DEREK for Windows, the BfR-DSS, and SICRET). More details on commercial expert systems are reported in Appendix R.7.2-2.

Not all of the models were developed with EU regulatory purposes in mind, so it is important to assess in each case whether the endpoint or effect being predicted corresponds with the regulatory endpoint of interest. In principle, such models could be redeveloped (re-parameterised) by using updated or alternative datasets, and used instead of the published models. The BfR model (also reported in Appendix R.7.2-2) has been developed to predict EU regulatory endpoints, and it has been recently validated (Rorije & Hulzebos, 2005 and Gallegos Saliner et al., 2007).

Use of (Q)SAR models for skin corrosion:

In the case of classification models for skin corrosion, where it is not indicated in the supporting documentation whether the predicted classification should be R34 or R35, it is recommended to treat the prediction as equivalent to R35 (severe corrosive). Very few models are available (see Gallegos Saliner et al., 2006 for review). Available models tend to focus on defined chemical classes (e.g. acids, bases, phenols) and might be useful as an alternative to *in vitro* testing for such chemicals.

SARs and read-across for eye irritation and corrosion:

The occurrence of structural analogues that exhibit corrosion (or irritation) potential can also be used to predict the effect in the substance of interest and derogate from further assessment. Negative data from structural analogues may also be used to make predictions in certain cases, provided that there are no other substructures present that are thought likely to cause the effect.

Read-across has been used in the New Chemicals notification procedure for the classification of eye irritants. An example is provided by the classification as R36 of Neodol HS, a branched alcohol ethoxy sulphate, by read-across from structurally related anionic surfactants. The adequacy of the read-across was justified in multiple ways:

i) by comparing the *in vitro* results of Neodol HS with that of SLS in the Cytosensor Microphysiometer Test. Since SLS is classified as R36 and used as the positive control in this assay, and since the test result showed that Neodol has a lower eye irritancy than SLS, it was argued that Neodol HS should also be (conservatively) classified as R36;

ii) by referring to the Critical Micelle Concentration (CMC). Below this concentration, the surfactant is in the monomer form, which has irritant properties, whereas above the CMC, the surfactant form micelles, which are less irritant. Thus, the higher the CMC, the greater the proportion of monomers present, and the more likely the surfactant will be an irritant. Neodol HC was shown to have a lower CMC than similar chemicals classified as R36;

iii) by referring to the fact that alkyl ethoxy sulphates, such as Neodol HC, tend to be weaker eye irritants than alkyl sulphates and sulphonates, and that alkyl sulphates and sulphonates with similar chain lengths to Neodol HC are classified as R36.

This illustrates the use of *in vitro* data to support read-across by comparing the *in vitro* effect of the chemical of interest with that of a suitable benchmark chemical.

QSARs and expert systems for eye irritation and corrosion:

An extensive review of the current state-of-the-art has been published by the ECB (Gallegos Saliner et al. 2006). In Appendix R.7.2-3 QSARs and expert systems for eye irritation and corrosion some

examples are given to illustrate currently available models and the techniques that have been used to develop them. These models include literature-based QSAR models, commercial models, and expert systems.

From the scientific literature, it appears that more emphasis has been placed on the QSAR modelling of ocular irritation compared with dermal irritation. Examples of models based on classical regression and classification techniques, together with more innovative approaches, are collected in Appendix R.7.2-3.

The most widely used commercial expert systems for assessing eye irritation are the same as those used for assessing skin irritation and corrosion. Details on automated rule-induction systems (e.g. TOPKAT and MultiCASE), and on knowledge-based systems (e.g. DEREK for Windows, and the BfR-DSS) are reported in Appendix R.7.2-3.

Not all of the models were developed with EU regulatory purposes in mind, so it is important to assess in each case whether the endpoint or effect being predicted corresponds with the regulatory endpoint of interest. In the case of the more transparent, literature-based models, the examples could be more useful in terms of illustrating the feasibility of developing a model by using defined descriptors and by applying a defined statistical approach to a suitable dataset. If alternative or extended datasets are available, such models could be redeveloped (re-parameterised) and used instead of the published models. The BfR model for the prediction of eye irritation has been developed to predict EU regulatory endpoints, and it has been recently validated (Tsakovska et al., 2005 and Tsakovska et al., 2007).

Use of (Q)SAR models for eye irritation/corrosion:

In the case of classification models for eye irritation, the classification criteria used in the model develop should be compared with the EU classification criteria, to assess the relevance of the model. Where it is not indicated in the supporting literature whether the predicted classification should be R36 or R41, the risk phrase chosen should be supported with expert judgment.

Table R.7.2-1 Overview of available (Q)SARs for skin and eye irritation/corrosion

and the availability of QSAR model reporting formats (QMRFs), in which the application of the OECD principles for QSARs is illustrated

Category of model or source	Reference or name of the model	Type of model	Applicability domain	Draft QMRF* developed
Literature models	Barratt, 1995	Statistical model	Acids, Bases , Phenols and pKa,	no
	Berner *et al*.,1988, 1990a, 1990b	Mathematical model	pKa related acids	no
	Nangia *et al*., 1996	Mathematical model	pKa related for bases	no
	Barratt, 1996b	Statistical model	Electrophiles	no
	Smith *et al*. 2000 a,b	Statistical model	Esters	no
	Barratt, 1996b	Statistical model	Neutral organics	no

	Gerner *et al*., 2004, 2005; Walker *et al*., 2004	Rule-based model	New Chemicals Database, organic chemicals	yes
Computerised models	TOPKAT commercial	Mathematical model using connectivity descriptors	Organic chemicals	yes
	DerekfW, commercial	Expert system using structural alerts	Organic chemicals and some metals	yes
	MultiCASE, commercial	Mathematical model using fragments	Organic chemicals	no
	Hazard expert, commercial	Organic chemicals using structural alerts	Organic chemicals	no
	BfR rulebase, free, available in-house at BfR	Rule-based model	New Chemicals Database, organic chemicals	yes
Review papers	Hulzebos *et al*., 2001, 2003, 2005	N.A.	N.A.	N.A.
	Patlewicz *et al*., 2003	N.A..	N.A.	N.A.
	Gallegos Saliner *et al* (2006)	N.A..	N.A.	N.A.

*) QMRF: (Q)SAR model reporting format see Section R.6.1 (available at http://qsardb.jrc.it).

**) see Annex II and III for more information on these models

Testing data for irritation/corrosion

The internationally accepted testing methods for skin irritation and eye irritation are described in OECD TGs. Those regarding skin effects can be found in TGs 404, 430, 431 and 435 (EU B.4, B.40, B.40bis), those for the endpoint eye in TG 405 (EU 5). The testing strategies developed (see Section R.7.2.6) emphasise the need to evaluate all available information (including physico-chemical properties) before attempting any *in vivo* testing. They both employ screening elements designed to avoid, as far as possible, *in vivo* testing of corrosive substances and to limit *in vivo* testing of severely irritating substances. In particular, it is recommended to test *in vitro* for skin corrosion (method B.40) before any attempts to assess skin or eye irritation/corrosion by animal testing and when no other information is available. There is no method for respiratory irritation in Annex V of Directive 67/548/EC[18]

In vitro data

***In vitro* tests for skin corrosivity:**

[18] All the test methods previously included in Annex V to Directive 67/548/EEC will be incorporated in a new Test Methods (TM) Regulation that is currently (February 2008) under adoption. The TM Regulation will be adapted to technical progress whenever a new test method has been developed, scientifically validated and accepted for regulatory use by the National Coordinators of the Member states.

Accepted *in vitro* tests for skin corrosivity are listed in Annex V of Directive 67/548/EC[18] and as OECD TG (EU, 2000; OECD 2004ab; OECD 2006).
These are the following (see also Table R.7.2-2):

i. The Transcutaneous Electrical Resistance (TER using rat skin) test (OECD TG 430/EU B.40)

ii. Human Skin Model tests (OECD TG 431/EU B.40 bis)

iii. The *in vitro* Membrane Barrier test method (not yet included as EU B.40 method; OECD TG 435)

For acceptable use in OECD TG 431/EU B.40 bis , human skin models need to satisfy the conditions for general and functional models given in the guideline. Models currently accepted as valid are EPISKIN™ and EpiDerm™ (Fentem *et al*., 1998; Liebsch *et al*., 2000; ECVAM, 1998; ECVAM, 2000); SkinEthic™ has undergone testing for this purpose (Kandárová *et al*., 2006) and been endorsed by ESAC as a method able to distinguish between corrosive and non-corrosive chemicals within the context of OECD TG 431/EU B.40 bis.

The TER assay and the human skin model assays do not allow the sub-categorisation of corrosive substances as permitted in the GHS.

The *in vitro* Membrane Barrier Test Method for Skin Corrosion (commercially available as Corrositex®) is recognised to enable identification of corrosive substances and mixtures and allowing the sub-categorisation of corrosive substances as required under the GHS. However, a limitation of the test is that many non-corrosive substances and preparations and some corrosive substances and preparations do not qualify for testing (i.e., test substances and preparations not causing a colour change in the Chemical Detection System; aqueous substances with a pH in the range of 4.5 to 8.5 often do not qualify for testing). Both ECVAM and ICCVAM have therefore concluded that this test may only be used for determining the corrosivity/ non-corrosivity of a specific categories of substances, *e.g.,* organic and inorganic acids, acid derivatives, and bases (ECVAM, 2001; ICCVAM, 2002). The test is accepted for testing purposes related to the transportation of chemicals of these specific classes by the US Department of Transport (US DOT, 2002).

***In vitro* tests for skin irritation:**

After prevalidation (Fentem et al., 2001) and extensive optimisation (Zuang *et al*., 2002; Cotovió *et al.*, 2005; Kandárová *et al*., 2005), two human skin assays EPISKIN™ (EPISKIN SNC, France) and EpiDerm™ (MatTek Inc., USA), have undergone a formal ECVAM validation (2003-2006) and are currenly undergoing ESAC peer review.
Irritant substances are identified in the human skin assays by:

i. their ability to induce a decrease in cell viability (measured by the MTT test) below defined threshold levels.

ii. their ability to release inflammatory mediators (Interleukin 1-α) when the cell viabiliy is above the defined threshold levels.

If the ESAC peer-review concludes that the test(s) are scientifically valid they will be forwarded to the EU and OECD for regulatory acceptance. At the time of writing this report it is expected that the EPISKIN text will be validated and endorsed as a full replacement of the *in vivo* test. In this case the test should be used for Annex VII and for Annex VIII under provisions laid down in Annex

XI 1.4 to avoid *in vivo* testing for skin irritation under the standard testing regime in compliance with Article 25 of the REACH legislation.

The Skin integrity function test (SIFT; Heylings *et al*., 2003), which uses the electrical resistance of mouse skin and transepidermal water loss across mouse skin as endpoints, was discontinued after phase 1 of the validation study.

The validation trial was designed to test these assays against current EU irritant/non irritant classifications. A post-hoc assessment using GHS classifications was subsequently undertaken (see Section R.7.2.4).

***In vitro* tests for eye irritation:**

At present there are no validated or OECD adopted *in vitro* tests for eye irritation. Within the EU, the 64th Competent Authority (CA) meeting November 2002, agreed that where there are positive results in the non-validated *in vitro* tests below, a substance can be considered a severe eye irritant (R41) and can be labelled accordingly (negative results require further testing *in vivo;* EC, 2006a):

i. isolated rabbit eye (IRE) test

ii. isolated chicken eye (ICE) test

iii. bovine corneal opacity & permeability (BCOP) test

iv. hen's egg test – chorio-allantoic membrane (HET-CAM) test.The above tests are currently undergoing evaluation by ICCVAM (with ECVAM collaboration) as to their validation status for the identification of severe eye irritants (ICCVAM, 2006).

There are two human corneal epithelium models available commercially, EpiOcular™ (MatTek Inc.) and SkinEthic™ HCE (SkinEthic, France) which have undergone assessment in industry-organised trials from which pre-validation and validation data have been submitted to ECVAM for evaluation. ECVAM is also taking the lead in the evaluation of promising cell cytotoxicity/cell function-based *in vitro* methods (e.g., red blood cell haemolysis, neutral red release, fluorescein leakage and silicon microphysiometer). These assays have previously undergone validation studies which were not successful, but they may be currently used as screening tests within companies and may be considered suitable for particular chemical domains following evaluation of supporting data.

The above tests are mainly concerned with modelling the immediate effects of chemicals on the cornea**.** *In vivo* eye irritation endpoints which are not covered by the above-mentioned optimised protocols are the following:

i. Persistence/reversibility of effects

ii. Effects on conjunctivae or other eye tissue

iii. Mechanical irritation produced by solid materials

Integrated Testing Strategies combining the different tests according to their applicability domain and capacity to classify in the different ranges of irritation will be developed, once the individual tests will be completely evaluated (Scott et al., in preparation).

Table R.7.2-2 Validation status, regulatory acceptance, relevant guidelines

area of concern	Test	validation status, reg. acceptance, use, limitations	OECD guideline	Dir 67/548/EEC	ECVAM-Invittox Nr.
skin corrosion					
	TER (1)	Validated	TG 430	Part of annex V	115
	EpiDerm	Validated	TG 431	Part of annex V	119
	EPISKIN	Validated	TG 431	Part of annex V	118
	SkinEthic	Validated	N.A.	N.A.	No protocol
	Corrositex	Validated	TG 435	Not yet	116
skin irritation					
	EpiDerm	Validated	not yet	Not yet	No protocol
	EPISKIN	Validated	not yet	Not yet	No protocol
	SIFT (2)	Only prevalidation so far. Applicability domain limited.	N.A.	N.A.	No protocol
eye irritation					
	IRE (3)	Pending, but regulatory acceptance for severe irritants *	N.A.	N.A.	85
	ICE (4)	Pending, but regulatory acceptance for severe irritants*	N.A.	N.A.	80
	BCOP (5)	Pending, but regulatory acceptance for severe irritants*	N.A.	N.A.	98, 124
	HET-CAM (6)	Pending, but regulatory acceptance for severe irritants*	N.A.	N.A.	47, 96
	RBC (7)	Pending. Used by industry for screening purposes.	N.A.	N.A.	37, 99
	FL (8)	Pending. Used by industry for screening purposes.	N.A.	N.A.	71, 82, 120
	NRR (9)	Pending. Used by industry for screening purposes.	N.A.	N.A.	54

	CMP / SMP (10)	Pending. Used by industry for screening purposes.	N.A.	N.A.	97, 102
	EpiOcular TM	Pending. Used by industry for screening purposes.	N.A.	N.A.	No protocol
	SkinEthic TM	Pending. Used by industry for screening purposes.	N.A.	N.A.	No protocol

*) see: EC 2004.

1) TER = Transcutaneous Electrical Resistance. 2) SIFT = Skin Integrity Function Test in Mouse. 3) IRE = Isolated Rabbit Eye. 4) ICE = Isolated Chicken Eye. 5) BCOP = Bovine Corneal Opacity and Permeability. 6) HET-CAM = Hen's Egg Test on Chorioallantoic Membrane. 7) RBC = Red Blood Cell Haemolysis Test. 8) FL = Fluorescein Leakage. 9) NRR = Neutral Red Release. 10) CMP / SMP = Cytosensor Microphysiometer / Silicon Microphysiometer

status within Dir 67/548/EEC and availability of invittox protocols of relevant tests in the field of skin corrosion, skin irritation and eye irritation.

Animal data

Skin and eye irritation:

Annex VI of the Dangerous Substances Directive (Directive 67/548/EEC) defines both skin and eye irritation as a local toxic effect, and, as such, an assessment of irritation is normally part of the acute testing phase of a toxicity programme and it is an early requirement of all regulatory programmes. As a consequence, testing for irritation has, historically, used animal models and a variety of test methodologies depending upon, for example, the laboratory undertaking the test, the era and intended application.
Current approaches for irritation testing are covered by:

i. OECD TG 404, Acute Dermal Irritation/Corrosion (adopted 12 May 1981; most recently updated 24 April 2002);

ii. Commission Directive 2004/73/EC, Method B4, Acute Toxicity: Dermal Irritation/Corrosion.

iii. OECD TG 405, Acute Eye Irritation/Corrosion (adopted 12 May 1981; most recently updated 24 April 2002)

iv. Commission Directive 2004/73/EC, EU B.5, Acute Toxicity: Eye Irritation/Corrosion.

The guidelines for skin and eye irritation testing require a tiered approach, using one animal (the rabbit is the preferred species) initially, which in the absence of severe effects is followed by a further two animals (a total of three animals).

Both OECD and EU methods use the scoring system developed by Draize (1944). The EU criteria for classification are based on the mean tissue scores obtained over the first 24-72 hour period after exposure and on the reversibility or irreversibility of the effects observed. Currently for both eye and skin, *irritants* (labels R36 and R38, respectively) cause significant inflammation of the eye (conjunctiva redness/oedema, cornea and/or iris) and/or skin (erythema and/oedema) but these effects are transient i.e. the affected sites are repaired within the observation period of the test. A *severe eye irritant* causes considerable damage to the cornea and/or iris and is labelled with R41. The criteria for R41 include persistence of effects (any score), irreversible staining of the eye and/or criteria for the degree of severity. Guidance on how industry interprets eye irritation data in the light of EU classification and labelling is summarised in a publication by ECETOC (1997).

A corrosive substance causes full thickness destruction of the skin tissue and is classified as *corrosive* and assigned a label (R34 or R35) depending upon the exposure time (3 min and 4 hours, respectively).

For existing substances, the use of methods other than those specified in Annex V of Directive 67/548/EC [19], or corresponding OECD methods, such as LVET (Griffith et al., 1980) may be accepted on a case-by-case basis.

[19] All the test methods previously included in Annex V to Directive 67/548/EEC will be incorporated in a new Test Methods (TM) Regulation that is currently (February 2008) under adoption. The TM Regulation will be adapted to technical progress whenever a new test method has been developed, scientifically validated and accepted for regulatory use by the National Coordinators of the Member states.

In addition to the OECD guidelines and Commission Directives mentioned above, further animal data may be available from:

i. Acute dermal toxicity test (OECD TG 402/EU B.3)

ii. Skin sensitization (OECD TGs 406 and 429/EU B.6 and B.42)

See Section R.7.2.6 for comments on how to use information from these tests in an Integrated Testing Strategy for skin and eye irritation/corrosion.
Data on chemosensory effects obtained in the Alarie test for respiratory irritation (Alarie, 1973, Arts et *al.,* 2006) may be useful as supportive evidence for human eye irritation after exposure to airborne chemicals (e.g. vapours).

Respiratory irritation:

There are currently no OECD adopted test guidelines that deal specifically with respiratory tract irritation. The type of information from animal studies that could inform on the respiratory irritation potential of the chemical concerned are the Alarie assay (information on sensory irritation, Alarie, 2000; ASTM, 2004) and single or repeated inhalation exposure studies (information on (histo)pathological changes).

In rodents, sensory irritation leads to a concentration-dependent reduction in the respiratory rate (breath-holding) mediated via the trigeminal nerve reflex; this reflex effect on respiration can be measured experimentally as the RD_{50} value in the Alarie assay.

Single inhalation exposure studies may provide information on nasal irritation such as rhinitis, whereas histopathological examination of respiratory tract tissues of animals repeatedly exposed by inhalation (28-day and 90-day inhalation studies) may provide information on inflammatory/cytotoxic effects such as hyperemia, edema, inflammation or mucosal thickening.

Data from bronchoalveolar lavage may give additional information on the inflammatory response.

R.7.2.3.2 Human data for irritation/corrosion

Existing human data include historical data that should be taken into account when evaluating intrinsic hazards of chemicals. *New* testing in humans for hazard identification purposes is not acceptable for ethical reasons.

Existing data can be obtained from case reports, poison information centres, medical clinics, and occupational experience or from epidemiological studies. Their quality and relevance for hazard assessment should be critically reviewed. However, in general human data can be used to determine a corrosive or irritating potential of a substance. Good quality and relevant human data have precedence over other data. However, lack of positive findings in humans does not necessarily overrule good quality animal data that are positive.

Specifically with regard to respiratory irritation, there is a view in the occupational health literature that sensory irritation may be a more sensitive effect than overt tissue-damaging irritation, given that its biological function is to serve as an immediate warning against substances inhaled during a short period of time which could damage the airways, and that it triggers physiological reflexes that limit inhalation volumes and protect the airways. However, there is a lack of documented evidence to indicate that this is a generic position that would necessarily apply to all inhaled irritants.

R.7.2.4 Evaluation of available information on irritation/corrosion

R.7.2.4.1 Non-human data on irritation/corrosion

Non-testing data on irritation/corrosion (skin and eye)

Physico-chemical properties

According to the current EU and OECD guidelines, substances should not be tested in animals for irritation/corrosion if they can be predicted to be corrosive from their physico-chemical properties. In particular, substances exhibiting strong acidity (pH ≤2) or alkalinity (pH ≥11.5) in solution are predicted to be corrosive, and should not be tested. However, no conclusion can be made regarding corrosivity when the pH has an intermediate value (when 2< pH <11.5).

Physico-chemical properties for skin corrosion/irritation:

Chemicals that have other pH values will need to be considered further for their potential for skin and eye irritation/corrosion.

The following decision rule can be used in a tiered testing strategy:

IF pH ≤ 2 or pH ≥ 11.5 THEN assume the chemical to be corrosive (R34 or R35).

This model is included in OECD testing strategy for skin irritation and corrosion (OECD, 2001). Several studies have investigated and confirmed the usefulness of pH as a predictor of corrosion (Worth & Cronin, 2001) and as an element in tiered testing strategies (Worth, 2004).

However, where extreme pH is the only basis of classification as corrosive, it may also be important to take into consideration the acid/alkaline reserve, a measure of the buffering capacity of a chemical substance (Young *et al*., 1988; Botham *et al*., 1998; Young & How, 1994), as mentioned in the OECD test guideline 404. However, the buffering capacity should not be used alone to exonerate from classification as corrosive. Indeed, when the Acid/Alkaline reserve suggests that the substance might be non-corrosive, further in vitro testing should be considered.

Physico-chemical properties for eye irritation:

A chemical known or predicted to be corrosive to the skin is automatically considered to be severely irritating to the eye (R41). However, no conclusion can be made regarding eye irritation potential when the pH has an intermediate value (when 2< pH <11.5). Thus, the following decision rule may be used in a tiered testing strategy:

IF pH ≤2 or pH ≥11.5 THEN consider the chemical for classification as a severe eye irritant.

To predict the eye irritation potential of non-corrosive chemicals, the distribution of pH values for irritants and non-irritants in a data set of 165 chemicals has been analysed (Worth, 2000). The irritants spanned a wide range of pH values from 0 to about 12, whereas the non-irritants spanned a much narrower range from about 3 to 9. Using the cut off values generated by classification tree analysis, the following model was formulated:

IF pH <3.2 or if pH >8.6, then consider the chemical for eye irritation classification; otherwise make no prediction.

According to the way the model was developed, *irritant* can either be R41 or R36. Further information and/or reasoning is needed to conclude on the risk phrases. The more severe classification (R41) should be assumed if no further information is available.

This model had a sensitivity of 53% (and therefore a false negative rate of 47%), a specificity of 97% (and therefore a false positive rate of 3%), and a concordance of 76%. A QSAR Model Reporting Format (QMRF) has been developed (see Section R.6.1 and JRC QSAR Model Database: http://qsardb.jrc.it).

Based on these statistics, this model is not recommended for the stand-alone discrimination between eye irritants and non-irritants. However, could be used in the context of a tiered testing strategy to identify eye irritants (due to its very low false positive rate) but not non-irritants (due to its relatively high false negative rate).

Grouping, (Q)SARs and expert systems

Guidance has been developed by the ECB (Worth et al., 2005) on how to apply (Q)SARs for regulatory use. Guidance on how to assess the validity and suitability of (Q)SAR models and adequacy of their predictions is given in Section R.6.1, and guidance on the use of read-across/category approaches is given in Section R.6.2.

First the model should be descripted in accordance with OECD principles on (Q)SARs (OECD, 2004c), and documented by means of a QMRF. Interpretation of the model is additionally needed. For example a model based on the logarithm of the octanol/water partition coefficient (K_{ow}) might indicate how the log K_{ow} should be derived, measured, calculated, with which program, whether ionised chemicals can be used as well. For more complicated parameters e.g. the quantum descriptors HOMO (Highest Occupied Molecular Orbital energy) and LUMO (Lowest Unoccupied Molecular Orbital energy) this is even more crucial as the calculation outcome depend on the configuration state of the molecule. The performance parameters for the model (i.e. correlation coefficient, sensitivity/specificity, etc.) have to be reported. When the predictivity of a model is assessed, it should be assessed whether the test set is within the applicability domain of the model. The guidance given by the authors/builders of the model should be a starting point.

The second step is to evaluate the prediction of a specific chemical. The OECD principles on (Q)SARs can be used again. One of the most important principle is the chemical's fit in the applicability domain (i.e. is the submitted chemical similar to the training set and does information exist on the predictivity) The outcome of the prediction should be assessed and documented in the form of a QPRF.

The third and last step of the evaluation explicitly needs to meet regulatory requirements. In this last evaluation the (Q)SAR prediction is weighed against the possible mechanism of skin irritation and corrosion. It has to be compared with the effects that can be observed in the *in vivo* test, to see whether all skin irritation/corrosion pathways are covered. In this last step, the hazard of defatting properties has to be assessed as well. (Q)SAR models have to be evaluated in considering the possible mechanism and how this would relate to EU hazard classification.[20]

The mechanism of irritation and corrosion has toxicodynamic and toxicokinetic parameters. Models that solely predict irritation and corrosionon toxicodynamics properties such as acidity or basicity, electrophilicity, other reactivity, surfactant activity, solving membranes, have to be additionally

[20] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

evaluated for their toxicokinetic parameters. These parameters can be physical chemical parameters or others and indicate the potential to cross the skin (stratum corneum) and be active in the living tissue underneath the stratum corneum. Also models that solely predict (the absence of) activity, irritation and corrosion, e.g. by physical chemical properties that illustrate the toxicokinetic behaviour of chemicals, have to be evaluated for their activity (toxicodynamics).

For example, the BfR physico-chemical rulebase predicts the absence of skin and eye irritation. Evaluations of the BfR rulebases for the prediction of no skin irritation/corrosion (Rorije and Hulzebos, 2005; Gallegos Saliner *et al*, 2007) and for the prediction of no eye irritation (Tsakovska *et al*, 2005) have been carried out independently. However, when the absence of irritation cannot be excluded, further information on the structure of the chemical is needed to predict presence of irritation/corrosion.

The absence of skin and eye irritation and corrosion is well predicted with the BfR rulebase and therefore these rules can be applied.

There is no other model yet available which sufficiently describes the absence of effects. Neutral organics are expected not to be irritants, however their defatting potential should be discussed. The definition of a neutral organic is a chemical which do not have potential reaction centres, even after skin metabolism. The absence of reactivity needs to be described in sufficient detail or be substantiated with other information.

The presence of effects has been well established using the pH cut off values for high acidity and basicity and can be applied. Structural alerts for the presence of effects can be used, however further characterisation of the effect needs to be described in sufficient detail or be substantiated with other information. For instance, the BfR structural rulebases for the prediction of skin irritation/corrosion (Rorije *et al*., 2007 and Gallegos Saliner *et al*. 2007) and for the prediction of eye irritation (Tsakovska *et al*., 2007) have been recently validated.

Testing data on irritation/corrosion (skin and eye)

In vitro data

There are OECD adopted guidelines for tests (see Section R.7.2.3) under which substances can be classified as corrosive. A negative result in these tests should be supported by a *Weight of Evidence* determination using other existing information, e.g. pH, (Q)SAR, human and/or animal data. These tests do not provide information on skin irritation and, therefore, further information is required to evaluate the skin irritation potential of non-corrosives. If a substance is shown to be non-corrosive in an *in vitro* test, unless this is confirmed by other data, an *in vivo* test needs to be conducted at the appropriate tonnage level.

Annex VII of REACH requires information from *in vitro* tests for skin and eye irritation, not from animal tests.

In order to accept an *in vitro* skin or eye irritation test under Annex VII or VIII, it is of great importance that a proper quality assessment of any such reports should be done.

As a consequence of the general rules in Annex XI, data from the following types of tests may be accepted as described below.

Skin irritation

Validated Tests:

The human skin model tests, EpiDerm™ and EPISKIN™ have undergone formal validation by ECVAM. The predictive capacity (expressed as sensitivity and specificity by comparison of in vitro data with animal data from the Draize skin irritation test carried out according to OECD TG 404) of the EPISKIN skin irritation test (SIT) using two endpoints (cytotoxicity (MTT test) and Interleukin 1-alpha release) was 90.7% (sensitivity) and 78.8% (specificity). Thus the test is considered scientifically valid for the prediction of irritant and non-irritant chemicals for Annex VII, and also Annex VIII according to the rules laid down in Annex XI.

The sensitivity and specificity of the EpiDerm SIT using one endpoint (cytotoxicity by MTT reduction measurement) was 60,1% and 88,8%, respectively. In its current form, the test is suitable for the identification of irritant chemicals as it has a low false positive rate, but not for the identification of non-irritant chemicals because of a high false negative rate. A positive result from the assay could thus be used for classification as irritant at Annex VII and VIII levels, but since negative data would however need to be supported by additional data, the EPISKIN™ test (SIT; see above) is the preferred method.

The methods were primarily validated against the EU classification scheme (irritants vs non-irritants; R38 vs no-label). A post-hoc evaluation of the EPISKIN assay performance against the GHS classification showed that the assay is not able to discriminate the GHS mild irritants from the GHS non-irritants and GHS irritants.

According to the Proposal for a EU Regulation on Classification and Labelling of Substances and Mixtures based on the GHS, GHS non-irritants and GHS mild irritants will become EU non-irritants (EC, 2006b). Considering this proposal for the new EU classification system and based on the results of the skin irritation validation study, the assay discriminated the irritants (GHS irritants) from the non-irritants (GHS mild and GHS non-irritants) with a sensitivity of 100% and a specificity of 64.4%. Note, that the final published EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS) should be taken into account (see Section R.7 Introduction).

Non-validated tests:

Positive data from the following tests may be accepted and used under Annexes VII and VIII (exploiting the possibilities provided by Annex XI Section 1.4). Negative results would however need to be supported by other data (see skin ITS box 9c).

The Skin integrity function test (SIFT) has completed a pre-validation followed by an optimisation phase, but more work is required for full validation. The test in its current state has a specific applicability domain (surfactants). In contrast, the pig's ear test and PREDISKIN™ assay only underwent a prevalidation study.

Eye irritation:

Positive outcome from four *in vitro* assays, the BCOP, ICE, IRE and HET-CAM are accepted by the EU to classify severe eye irritants under Annex VII and Annex VIII using the adaptations of the standard testing regime specified in Annex XI. They have undergone a formal retrospective evaluation, and their scientific validity has been the subject of a statement by ESAC (2007).

For the lower ranges of irritancy no assay is currently accepted by regulators but the following assays exist: Two reconstituted human tissue models, the EpiOcular™ and SkinEthic™ HCE models, have undergone corporate validation (EpiOcular: Blazka et al, 1999, 2000, 2003) and prevalidation trials (SkinEthic: van Goethem et al., 2006) respectively. The results are undergoing a

formal evaluation. Positive data from these may be accepted under Annex VII and VIII (see adaptation rules in Annex XI) if there is sufficient background information on the performance of the assay.

Four cytotoxicity and cell/tissue function based assays such as the Red Blood Cell haemolysis test, the Neutral Red Release assay, the Fluorescein Leakage test and the Silicon/Cytosensor Microphysiometer assay are currently undergoing a retrospective evaluation. Companies may use several of these for internal purposes and resultant positive data may be appropriate for Annex VII and VIII (see adaptation rules in Annex XI).

Quality Aspects:

In such a quality assessment that will lay the basis for later possible *Weight of Evidence* considerations, see Sections R.4.4 and R.5.2.1.2 for aspects that need to be taken into account in such a WoE.

Animal data

Well-reported studies particularly if conducted in accordance with principles of GLP, can be used to identify substances which would be considered to be, or not to be, corrosive or irritant to the skin or eye. There may be a number of skin or eye irritation studies already available for an existing substance, none of which are fully equivalent to a EU test method such as those in the Annex V to Directive 67/548/EEC. If the results from such a batch of studies are consistent, they may, together, provide sufficient information on the skin and/or eye irritation potential of the substance.

If the results from a variety of studies are unclear, based on the criteria given below for evaluation of the data, the registrant will need to decide which of the studies are most reliable, relevant for the endpoint in question and will be adequate for classification purposes.

Particular attention should be given to the persistence of irritating effects, even those which do not lead to classification. Effects such as erythema, oedema, fissuring, scaling, desquamation, hyperplasia and opacity which do not reverse within the test period may indicate that a substance will cause persistent damage to the human skin and eye.

Data from studies other than skin or eye irritation studies (e.g. other toxicological studies on the substance in which local responses of skin, eye mucous membranes and/or respiratory system have been reported) may provide useful information though they may not be well reported in relation to, for example, the basic requirements for information on skin and eye irritation. However, information from studies in animals on mucous membrane and/or respiratory system irritation can be very useful for risk assessment provided the irritation is clearly substance-induced, and particularly if it can be related to exposure levels.

Quality Aspects

Data from *existing* irritation studies in animals must be taken into account before further testing is considered. A quality assessment of any such reports should be done using, for example, the system developed by Klimisch et al. (1997), as described in Section R.4.2, and a judgement will need to be made as to whether any further testing is required. Some examples to note are:

i. Was the animal species the rabbit or was it another such as rat or mouse? Rat and mouse, as species, are not as sensitive as the rabbit for irritation testing.

ii. How many animals were used? Current methodology requires 3 but 6 was frequently used in the past.

iii. How many dose levels were used? If dilutions were included, what solvent was used (as this may have influenced absorption)? Which dose volume was used?

iv. For skin, which exposure period was used? Single or repeated exposure?

v. The method used to apply the chemical substance to the skin should be noted i.e. whether occluded or semi-occluded, whether the application site was washed after treatment.

vi. Check the observation period used post exposure. Shorter periods than in the current guideline may be adequate for non-irritants but may require a more severe classification for irritants when the observation period is too short to measure full recovery.

vii. For eye irritation, was initial pain noted after instillation of the test substance, was the substance washed out of the eye, was fluorescent staining used?

viii. For eye irritation, how was the test material applied into the eye?

Irritation scores from old reports, reports produced for regulatory submission in the USA or in publications may be expressed as a Primary Irritation Score. Without the original data it is not always possible to convert these scores accurately into the scoring system used in the EU. For extremes i.e. where there is either no irritation or severe irritation, it may not be necessary to look further, but average irritation scores pose a problem and judgement may be required to avoid repeat testing.

Observations such as the above can all be used to assess whether the existing animal test report available can be used reliably to predict the irritation potential of a substance, thus avoiding further testing.

Specific considerations for eye irritation

A refinement of the classical Draize test is the rabbit low volume eye test (LVET). The test protocol deviates from OECD TG 405 in that in the LVET, 10uL is directly applied onto the cornea. The grading scale and the data interpretation in the LVET is exactly the same as those used in OECD TG 405. The validity of the LVET is currently under review of ECVAM for the detergent and cleaning preparations applicability domain. Anatomical and physiological considerations for rabbit and human eyes indicate that a dose volume of 10uL is appropriate (A.I.S.E. 2006): the tear volume in both rabbit and man is approximately the same (~ 7-8uL), and after blinking, the volume capacity in the human eye is ~10uL. These considerations suggest that the LVET is also potentially a suitable test to demonstrate toxicological effects on man of potential eye hazards of substances. The LVET has been used in industry safety evaluation of single chemicals (Griffith et al, 1980) and detergent and cleaning preparations (Freeberg *et al*, 1984; Freeberg *et al.* 1986a,b; Cormier *et al.,* 1995; Roggeband et al, 2000), and has shown to be a very good predictor of the effects on man. It still overpredicts, but much less so than the classical Draize test of OECD TG 405.

In summary, available data from the LVET on substances and preparations should be considered and must be carefully evaluated. For the classification of substances however it must be taken into account that the test up to now has a limited applicability domain (detergent and cleaning products). Consequently, positive LVET data (be it R36 or R41) are a trigger for the appropriate classification for eye irritancy, but negative data from LVET as a *stand alone method* (in the absence of any other information) are not conclusive for *no classification*.

Specific considerations for respiratory irritation

All data available should be evaluated to estimate a substance potential to induce respiratory tract irritation. Sources of information could be:

Human data:

- Experience from occupational exposure
- Published data on volunteers (objective measurements, psychophysical methods, and subjective reporting)
- Other data (e.g. from nasal lavage)

Animal data:

- Alarie assay
- Data from other inhalation studies (acute, repeated exposure):
    1. Clinical symptoms of dyspnoea or breathing difficulties,
    2. Histomorphology of the respiratory tract,
    3. Lavage examination (nasal, bronchoalveolar)

Data indicating the cytotoxic type of respiratory irritation, which were mainly gained from histopathological examinations of tissues, are considered in the DNEL derivation for the acute toxic effects or for the repeated dose toxic effects (Section R.8.2.1 and Appendix R.8-8).

With respect to the sensory irritation response, the evidence from all sources has to be considered for the quantitative risk assessment procedure.

Although the Alarie test for various reasons has never become an OECD TG, results of the Alarie assay can be used for hazard identification of sensory irritation as the Alarie test detects the potential of a substance to stimulate the trigeminal nerve. Like in acute inhalation toxicity testing, results from Alarie tests may show high inter-laboratory variability. Therefore, the use of Alarie data for deriving quantitative information for instance to establish short-term DNELs for irritation should be done with caution (i.e. taking into account the actual breathing pattern, whether a response plateau is being reached; see the review by Bos *et al*, 1992). In that review it was shown that data of the Alarie test could not be used to establish TLV values for lifetime exposure. It can be expected that a substance that is capable to stimulate the trigeminal nerve in mice will also have this potential in humans. However, because the human response at an exposure concentration equal to the RD_{50} cannot quantitatively be determined and because responses in the Alarie-test of less than 10-12% are considered to be within the expected normal variation (Boylstein, 1996; Doty *et al*, 2004; ASTM, 2004), use of the Alarie-bioassay in a quantitative risk assessment, if any, is suggested to start from an RD10 rather than from an RD_{50}.

Although anatomical differences in rodents and humans do exist (f.i. rodents are obligate nose breathers and humans not), sensory irritation will be present in both but the location and the type of effect may differ, i.e. in rodents a decrease in breathing frequency may be observed whereas in humans this may result in coughing.

Sensory irritation does not necessarily lead to tissue damage. Effects characterising overt tissue damage are covered by inhalation studies for acute or repeated exposure toxicity. In this sense the Alarie assay is not designed to predict such pathological changes (Bos *et al*, 2002). If available from

other studies with the inhalation route (acute and repeated exposure) the characterisation of histomorphological lesions at the respiratory tract could be used as supplemental information.

Although both the Alarie test and for instance human nasal pungency threshold determinations are aimed to test for sensory irritation, correlation of the results of the Alarie test with such human data is difficult as the first is looking at rather strong effects upon exposure for at least 20 min (a 50% decrease in breathing frequency may be experienced by humans as unbearable) whereas human data are based on, for instance, very short exposure durations (sniffing for a few seconds). The results of a study by Cometto-Muniz et al. (1994) indicated that RD_{50} values in animals are not easily comparable with 'nasal pungency thresholds' in humans (see also Bos *et al*, 2002).

R.7.2.4.2 Human data for irritation/corrosion

Human data for skin corrosion, skin irritation and eye irritation

Well-documented *existing* human data of different sources can often provide very useful information on skin and/or respiratory irritation, sometimes for a range of exposure levels. Often the only useful information on respiratory irritation is obtained from human experience (occupational settings). The usefulness of all human data on irritation will depend on the extent to which the effect, and its magnitude, can be reliably attributed to the substance of interest. Experience has shown that it is difficult to obtain useful data on substance-induced eye irritation, but data may be available on human ocular responses to certain types of preparations (e.g. Freeberg *et al*, 1986a).

The quality and relevance of existing human data for hazard assessment should be critically reviewed. For example, in occupational studies with mixed exposure it is important that the substance causing the irritation or corrosion has been accurately identified. There may also be a significant level of uncertainty in human data due to poor reporting and lack of specific information on exposure.

Examples of how existing human data can be used in hazard classification for irritancy are provided in a recent ECETOC monograph (ECETOC, 2002).

Human data on local skin effects may be obtained from existing data on single or repeated exposure. The exposure could be of accidental nature or prolonged, for example in occupational settings. The exposure is usually difficult to quantify. When looking at the effects, corrosivity is characterised by destruction of skin tissue, namely visible necrosis through the epidermis and into the dermis. Corrosive reactions are typified by ulcers, bleeding and bloody scabs. After recovery the skin will be discoloured due to blanching of the skin, complete areas of alopecia and scars (see Chapter 3.2 of GHS), i.e. corrosivity is an irreversible damage. With this characterisation it should be possible to discern corrosive properties in humans. However, to distinguish between "Causes severe burns", R35, and "Causes burns", R34, (3 minutes' and 4 hours' exposure in rabbits, respectively) may not be so obvious in practice. A clear case for R35 classification would be an accidental splash which gave rise to necrosis of the skin. In cases where it is obvious that a prolonged exposure is needed (not to be mixed with delayed effects) before necrosis occurs, R34 seems more reasonable. If the distinction between R35 and R34 is not clearly apparent then the more stringent classification should be chosen. Discrimination between corrosives and skin irritants in rabbits is made on the effects caused after 4 hours' exposure. Irritants to the skin cause a significant inflammation which is reversible.

Severe eye irritants (R41) give more severe corneal opacity and iritis than eye irritants (R36). R41 compounds induce considerable tissue damage which can result in serious physical decay of vision. The effects normally do not reverse within 21 days (relates to animals); see Chapter 3.3 of the GHS.

In contrast, the effects of R36 compounds are reversible within 21 days. In humans, a sight control by a physician would reveal a decay of vision. If it is not transient but persistent it implies classification with R41. If the discrimination between R41 and R36 is not obvious, then R41 should be chosen.

Human data for respiratory irritation

Consideration should be given to real-life human observational experience, if this is properly collected and documented (Arts *et al*, 2006), e.g. data from well-designed workplace surveys, worker health monitoring programmes. For substances with an array of industrial uses and with abundant human evidence, the symptoms of respiratory irritation can sometimes be associated with certain concentrations of the irritants in the workplace air and might thus allow derivation of DNELs. However, the exposure details need to be well documented and due consideration should be given to possible confounding factors.

Data on sensory irritation of the airways may be available from volunteer studies including objective measurements of respiratory tract irritation such as electrophysiological responses, data from lateralization threshold testing, biomarkers of inflammation in nasal or bronchoalveolar lavage fluids. Including anosmics as subjects could exclude odor as a bias.

R.7.2.4.3 Exposure considerations for irritation/corrosion

Exposure-based waiving from testing is not applicable to the endpoints of skin corrosion, skin and eye irritation. Exposure-based waiving from testing as specified in Annex XI (3) applies to Sections 8.6 and 8.7 of Annex VIII, Annex IX and Annex X according to the REACH text.

R.7.2.4.4 Remaining uncertainty on irritation/corrosion

Usually it is possible unequivocally to identify (or accept) a substance as being corrosive, whatever type of study provides the information.

There may be a significant level of uncertainty in human data on irritant effects (because of poor reporting, lack of specific information on exposure, subjective or anecdotal reporting of effects, small numbers of subjects, etc.).

Data from studies in animals according to internationally accepted test methods will usually give very good information on the skin or eye irritancy of a substance in the test species, and, in general, it is assumed that substances which are irritant in Annex V studies in animals will be skin and/or eye irritants in humans, and those which are not irritant in Annex V studies will not be irritant in humans. Good data, often clearly related to exposure levels, can be obtained on respiratory and mucous membrane irritation, from well-designed and well-reported inhalation studies in animals. However, inconsistent results from a number of similar studies increases the uncertainty in deriving data from animal studies.

The data obtained from *in vitro* studies may include many dose levels and replicates: when such a study has a well-defined mechanistic basis and indicates that a substance is expected to be irritating, this may suffice for defined hazard identification purposes.

R.7.2.5 Conclusions for irritation/corrosion

R.7.2.5.1 Concluding on suitability for Classification and Labelling

In order to conclude on C&L, all the available information needs to be taken into account, and consideration should be given to both Annex VI of the Directive 67/548/EEC[21] and the various remarks (as they relate to classification and labelling) made throughout this guidance document.

R.7.2.5.2 Concluding on suitability for Chemical Safety Assessment

A dose-response assessment is difficult to make for irritation and corrosion simply because up to the present time most data have been produced with undiluted chemicals in accordance with test guidelines and traditional practice (which continues today). From a risk characterisation perspective it is therefore advisable to use the outcome of the classification procedure, i.e. a substance that is classified is assumed to be sufficiently characterised. However, a complete risk assessment requires both hazard, as well as dose-response data. Consequently, if the latter are available, they must be taken into account (see flowchart 1). For instance,dose-response information might be available from sub-acute dermal, repeated dose dermal and inhalation toxicity studies as well as from human experience.

Guidance on the possibilities for derivation of DNELs for skin and eye irritation/corrosion and respiratory irritation is given in Appendix R.8-9.

However, with specific regard to respiratory irritation, special attention needs to be given to as to whether extrapolation of the dose-response assessments from animal tests to the human situation is possible (see Section R.7.2.4.2).

R.7.2.5.3 Information not adequate

A *Weight of Evidence* approach comparing available adequate information with the tonnage-triggered information requirements by REACH may result in the conclusion that the requirements are not fulfilled. In order to proceed in further information gathering the following testing strategies can be adopted (see Section R.7.2.6).

R.7.2.6 Integrated Testing Strategy (ITS) for irritation/corrosion

R.7.2.6.1 Objective / General principles

For substances with no or very few data, the following sequential test strategy is recommended for developing adequate and scientifically sound data for assessment/evaluation and classification of the corrosive and irritating properties of substances. For existing substances with insufficient data, this strategy can also be used to decide which additional data, beside those available, are needed.

[21] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS);,

The objective of the testing strategies is to give guidance on a stepwise approach to hazard identification with regard to skin and eye irritation/corrosion. A principle of the strategy is that the results of one study are evaluated before another study is initiated. The strategy seeks to ensure that the data requirements are met in the most efficient and humane manner so that animal usage and costs are minimised.

Some guidance for testing is provided by the specific rules for adaptation from standard information requirements, as described in column 2 of Annexes VII-X, together with some general rules for adaptation from standard information requirements in Annex XI.

R.7.2.6.2 Testing strategy for irritation/corrosion

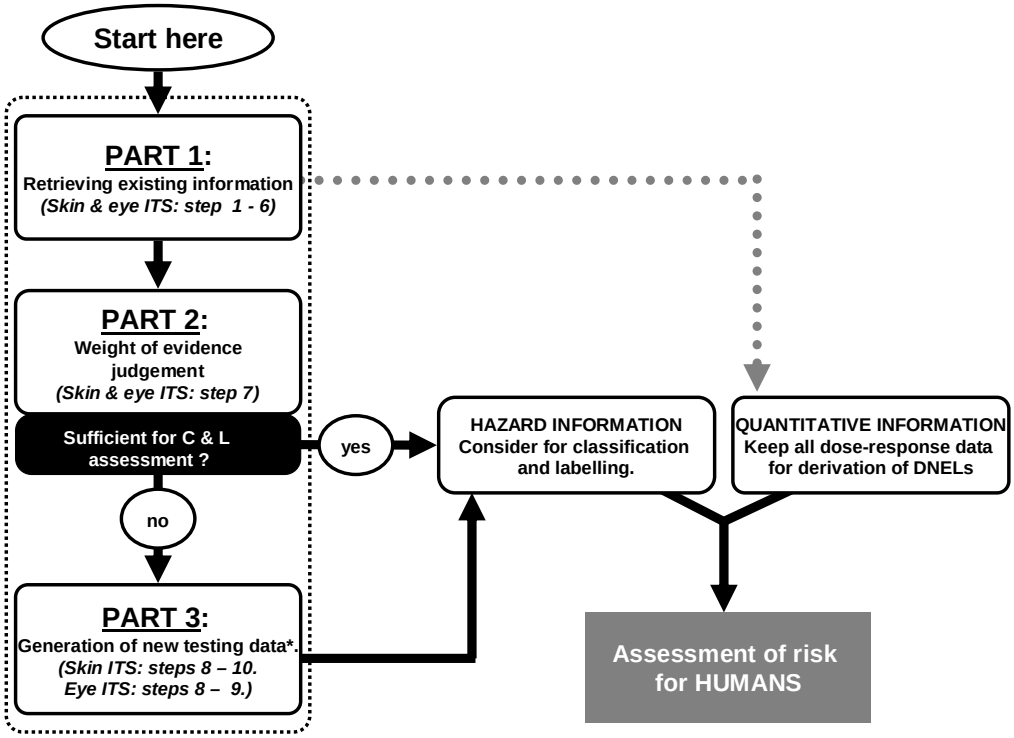
Risk assessment of the irritating potential of a substance is normally made in a qualitative way provided the substance has been classified as being irritant or corrosive to skin. Existing test guidelines do not contain dose-response assessment, so that a quantitative analysis will often not be possible. Therefore, hazard identification and appropriate classification is the key determinant in the information strategy below. As a consequence, the use of *Assessment Factors* is of limited use in order to take into account uncertainty of data. However, the registrant is encouraged to keep and use all quantitative data that might be encountered in the process of retrieving hazard information in the context of the present ITS and to perform a complete risk assessment, comprising hazard as well as quantitative information.

It is recommended that the information strategy is followed to step 6 (Figure R.7.2-1 & Figure R.7.2-2) in all cases and thereafter the weight of the evidence (WoE) analysis is performed. Clearly, not all steps will necessarily be accompanied by data, but it is important, that all potential data sources are explored prior to starting the WoE analysis. Note that before the WoE analysis in step 7, no new *in vitro* or *in vivo* tests should be conducted: Instead the assessment should be solely based on existing data. Furthermore, prior to perform any new *in vivo* test, the use of *in vitro* methods should be fully exploited (see Article 25 of REACH) by using the general rules of Annex XI which allow to adapt the standard testing regime set out in Annexes VIII to X.

If the substance is not classified for skin irritation/corrosion, no risk assessment for this endpoint is performed, regardless of the exposure. Please note that there is no option for exposure-based waiving for this endpoint in the REACH regulation.

The following flow chart (Figure R.7.2-1)gives an overview of the overall strategy for defining a testing strategy for irritation and corrosion.

Figure R.7.2-1 Overview of the Integrated Testing Strategy for irritation/corrosion



*Generation of new testing data according to Annex VII to VIII and with due observation of the rules for adaptation of the standard testing regime laid down in Annex XI.

The ITS presented here comprises three sequential parts (see flow chart below): Part 1 (in light grey) is about retrieving existing information (step 1 to 6), part 2 (in dark grey) represents a *Weight of Evidence* (*WoE*) analysis and judgement (step 7) and part 3 (white background) is about the generation of new information by testing (step 8 to 10).
In the information retrieval part, existing and available information from the literature and databases is gathered and considered in a stepwise process. At the end of this part all information collected is analysed using a WoE approach (step 7). It is therefore necessary to run through all steps before arriving at step 7. This means that in cases of "yes, consider to classifiy…", the registrant should nevertheless proceed to the next step. However, the ITS may be exited in the sole exception if the substance is spontaneously flammable at room temperature in contact with air or water (box 1a). In this case, testing is not required.

In the information generation part, new information on the irritation potential of substances is created by means of *in vitro* or, as a last resort (see Article 25 of the REACh legislation), *in vivo* testing. Therefore, before concluding the WoE analysis in step 7, new *in vivo* tests should not be conducted.

Figure R.7.2-2 Integrated testing strategy (ITS) for assessing the skin corrosion and skin irritation potential of substances

Step	Information	Conclusion
Existing data on physico-chemical properties		
1a	Is the substance spontaneously flammable) in contact with air(pyrophoric) or water at room temperature? ⟶ ↓ NO ↓	YES: No testing required. No need to proceed.
1b	Is the substance an organic hydro peroxide or an organic peroxide? ⟶ ↓ NO ↓	YES: Consider to classify as ■ corrosive (R34; "causes burns") if the substance is a hydro peroxide or ■ irritating as R38 ("Irritating to skin") if the substance is a peroxide. OR Provide evidence for the contrary Proceed to next step
1c	Is the pH of the substance lower than 2 or higher than 11.5? [a] ⟶ ↓ NO ↓	YES: Consider to classify as corrosive. Where classification is based upon consideration of pH alone (see step 7!), R35 should be applied. Proceed to next step
1d	Are there other physical or chemical properties that indicate that the substance is irritating/corrosive? ⟶ ↓ NO ↓	YES: Use this information for WoE analysis (step 7). Proceed to next step
Existing human data		
2	Are there adequate existing human data [b] which provide evidence that the substance is an irritant or corrosive ⟶ ↓ NO ↓	YES: Consider to classify accordingly. Proceed to next step

Existing animal data from irritation/corrosivity studies		
3	Are there data from existing studies *on irritation and corrosion* in laboratory animals, which provide sound conclusive evidence that the substance is a corrosive, irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (either R35 or R34 or R38 or no classification). Proceed to next step
Existing data from general toxicity studies via the dermal route and from sensitization studies		
4a	Is the substance acutely toxic (LD_{50}≤400 mg/kg bw) or very toxic (LD_{50}≤50 mg/kg bw) via the dermal route? [c] → ↓ NO ↓	YES: The substance will be classified for its acute dermal toxicity. Proceed to next step
4b	Has the substance proven to be a corrosive, irritant or non-irritant in a suitable acute dermal toxicity test? [d] → ↓ NO ↓	YES: If test conditions are consistent with OECD 404, consider to classify accordingly (R35 or R34 or R38 or no classification). Proceed to next step
4c	Has the substance proven to be a corrosive or an irritant in sensitisation studies or after repeated exposure? [e] → ↓ NO ↓	YES: This information cannot be used for considering a concrete classification conclusion but must be used exclusively within the integrated WoE judgement. Proceed to next step
Existing (Q)SAR data and read-across		
5a	Are there structurally related substances (suitable "read-across" or grouping), which are classified as corrosive (R34, R35) on the skin, or do suitable QSAR methods indicate corrosion potential of the substance? [f] → ↓ NO ↓	YES: Consider to classify as R35 Proceed to next step
5b	Are there structurally related substances (suitable "read-across" or grouping), which are classified as irritant on the skin (R38), or do suitable (Q)SAR methods indicate irritating potential of	YES: Consider to classify as R38. Proceed to next step

	the substance?[f] → ↓ NO ↓	
Existing in vitro data		
6a	Has the substance demonstrated corrosive properties in an OECD adopted *in vitro* test? → ↓ NO[g] ↓	YES: Consider to classify as corrosive. If discrimination between R34 and R35 is not possible, R35 must be chosen. Proceed to next step
6b	Are there acceptable data from a validated *in vitro* test (adopted by OECD or not), which provide evidence that the substance is an irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (R38 or no classification). Proceed to next step
6c	Are there data from a non-validated *in vitro* test, which provide sound conclusive evidence that the substance is an irritant[h]? → ↓ NO ↓	Yes: Consider to classify as R38, Proceed to next step
Weight of evidence analysis		
7	Taking all existing and relevant data (steps 1-6) into account, is there sufficient information to make a decision of whether classification/labelling is necessary, and – if so – how to classify and label? → ↓ NO ↓	YES: Classify accordingly (R35 or R34 or R38 or no classification)
New in vitro/ex vivo tests for corrosivity (Annex VII)		
8	Does the substance demonstrate corrosive properties in an OECD adopted *in vitro* or *ex vivo* tests for skin corrosion? → ↓ NO[g]	YES: Classify R34 or R35. If discrimination between R34 and R35 is not possible, R35 must be chosen.

	↓	
New in vitro/ex vivo tests for irritation (Annex VII)		
9a	Does the substance demonstrate irritating or non-irritating properties in validated *in vitro* tests (adopted by OECD or not) for skin irritation? ⟶ ↓ NO ↓	YES: Classify accordingly
9b	Does the substance demonstrate irritating properties in a non-validated *in vitro* test for skin irritation[h]? ⟶ ↓ NO ↓	YES: Classify accordingly.
New in vivo test for irritation (Annex VIII)[i]		
10	Does the substance demonstrate irritancy in an OECD adopted *in vivo* test? ⟶ ↓ NO ↓ No classification	YES: Classify accordingly.

Notes to the information scheme skin irritation/corrosion

a) Note that if the buffering capacity suggests that the substance may not be corrosive, further data are needed to confirm this.

b) data from case reports, occupational experience, poison information centres or from clinical studies.

c) if the substance is acutely toxic (LD_{50}≤400 mg/kg bw) or very toxic (LD_{50}≤50 mg/kg bw) via the dermal route further testing for irritation/corrosion would result in severe suffering or death of the animal. Thus, further testing is not required and sufficient labelling (warning) is provided by the risk phrases: "R24: toxic in contact with the skin" or "R27: very toxic in contact with the skin" and the symbol with T or T^+, shown below. Please note, that although the derogation regarding acute toxicity (LD_{50}≤400 mg/kg bw) is not a specific rule for adaptation from column 1 in REACH, it is considered here to be scientific common sense.

d) Has the substance proven to be either an irritant or a corrosive in an acute dermal toxicity test carried out with rabbits with the undiluted test substance (liquids) or with a suitable suspension (solids)? In case of signs of skin corrosion: classify as R35. In all other cases: calculate or estimate the amount of test substance per cm^2 and compare this to the test substance concentration of 80 µl or 80 mg/cm^2 employed in the OECD TG 404 for dermal irritation/corrosion test with rabbits. If in the same range and adequate scoring of skin effects is provided: classify or not as R38. In case conclusive negative data was obtained in rabbits, stop. If not in the same range and inadequate scoring of skin effects: use for WoE analysis and proceed.

In case the test was performed in other species, which may be less sensitive, evaluation must be made with caution. Usually, the rat is the preferred species for toxicity studies within the EU. The limit dose level of 2000 mg/kg bw of a solid is normally applied as a 50% suspension in a dose volume of 4 ml/kg bw onto a skin surface area of ca 5x5 cm.

Assuming a mean body weight of 250 grams, a dose of 1 ml of the suspension will be applied to an area of 25 cm^2, i.e 20 mg test substance per cm^2. In case of an undiluted liquid, 0,5 ml is applied to 25 cm^2, i.e. 20 µl/cm^2. Considering the fact that the rat skin is less sensitive compared to rabbit skin, much lower exposures are employed and, in general, the scoring of dermal effects is performed less accurate, the results of dermal toxicity testing in rats will not be adequate for classification with respect to skin irritation. Only in case of evidence of skin corrosivity in the rat dermal toxicity test, the test substance can be classified as R35. All other data should be used for WoE.

e) Regarding data from skin sensitiation studies, the skin of guinea pigs is less sensitive than the skin of rats which is less sensitive than the skin of rabbits. Only in case of evidence of skin corrosivity in the sensitization test (Maximization or Buhler) with the neat material or dilutions of solids in water, physiological saline or vegetable oil, the test substance should be classified as R35. However, care should be exercised when interpreting findings from guinea pig studies, particularly from maximisation protocols, as intradermal injection with adjuvant readily causes necrosis. All other data should be used for WoE only. Information on irritating properties from skin sensitisation tests cannot be used to conclude a specific classification regarding acute skin irritation but may be used in a WoE analysis. In general, irritation data from the Local Lymph Node Assay are not usable. The test substance is applied to the dorsum of the ear by open topical application, and specific vehicles for enhancement of skin penetration are used.

f) Conclusion on no classification can be made if the in silico model has been shown to predict adequately the absence of the classified effect and also fulfils the requirements of Annex XI.

g) No classification for corrosivity if a negative result can be supported by a WoE determination using other existing information, e.g. pH, SAR, human and/or animal data (according to OECD TG 430 and 431/EU B.40 and B.40 bis). If not corrosive, the irritating potential needs to be determined, proceed.

h) Conclusion on no classification can only be made if it has been concluded in the evaluation process that the test allows the identification of non-irritants and the data are used in a WoE approach following Annex XI 1.2.

i) In the light of a recently finished ECVAM validation trial, the *in vivo* test might be avoided in the near future by using the EPISKIN *in vitro* model. At the time of writing this report, the model has not finally be endorsed by ESAC, but it is likely that it will be recommended as a stand-alone replacement method for the animal test. *In vivo* testing as specified in Annex VIII for the appropriate tonnages might therefore be avoided using the tool of Annex XI 1.4 *in vitro* methods, that allows adaptation of the standard testing regime using suitable and, for the case of negative identification, validated *in vitro* tests.

The ITS for eye irritation is completely analogous in structure to that of skin corrosion, irritation. The ITS consists of an information retrieval part (steps 0a to 6 in light grey) and a part on the generation of new information by testing (step 8 to 9, no background colour). These two parts are separated by a WoE analysis and judgement (step 7 in dark grey).

In the information retrieval part, existing and available information from the literature and databases is gathered and considered in a stepwise process. At the end of this part all information collected is analysed using a WoE approach (step 7). It is therefore necessary to run through all steps before arriving at step 7. This means that in cases of "yes, consider to classify…", one should nevertheless proceed to the next step ("Proceed to next step"). An exception is a "yes" in one or all of the following boxex: 0a, 1a or 1c: if the substance is classified as a skin corrosive or its pH is < 2 and > 11.5 (taking the buffer capacity into due consideration), the process of information retrieval can stop at this point, since the substance's eye irritation potential is implicit in this classification. If the substance is spontaneously flammable at room temperature in contact with air (pyrophoric) or water, testing is not required.

In the information generation part (steps 8 to 9), new information on the irritation potential of substances is created by means of *in vitro* or, as a last resort (see article 25 of the REACH legislation), *in vivo* testing. Therefore, before concluding the WoE analysis in step 7, new *in vivo* tests should not be conducted.

Figure R.7.2-3 Integrated testing strategy (ITS) for assessing the eye irritation potential of substances.

Step	Information	Conclusion
Conclusion of the information strategy on skin irritation/corrosion		
0a	Is the substance classified as a skin corrosive? → ↓ NO ↓	YES: when assigned R34 or R35, the risk of severe damage to eyes is considered implicit. No need to proceed.
Existing data on physico-chemical properties		
1a	Is the substance spontaneously flammable in contact with air (pyrophoric) or water at room temperature? → ↓ NO ↓	YES: no testing required No need to proceed
1b	Is the substance an organic hydro peroxide or an organic peroxide? → ↓ NO ↓	YES: Consider to classify for ■ corrosivity (hydro-peroxide) using R34 ("causes burns'), thus implicitly also for severe ocular irritancy (R41 "risk of serious damage to eyes") or ■ for irritation (peroxide) using R36 ("irritating to eyes"). Proceed to next step
1c	Is the pH of the substance lower than 2 or higher than 11.5? [a] → ↓ NO ↓	YES: when assigned R35, the risk of severe damage to eyes is considered implicit. No need to proceed
1d	Are there other physical or chemical properties that indicate that the substance is irritating to the eye [b]? → ↓ NO ↓	YES: Use this information for WoE analysis (step 7). Proceed to next step
Existing human data		
2	Are there adequate existing human data [c] which provide evidence that the substance is irritating to the eye? →	YES: Consider to classify (R41 or R36), or use for WoE analysis (step 7).

	↓ NO ↓	Proceed to next step
Existing animal data from eye irritation studies		
3	Are there data from existing studies *on eye irritation* in laboratory animals, which provide sound conclusive evidence that the substance is an eye irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (R41 or R36 or no classification). Proceed to next step
Existing data on acute dermal toxicity		
4	Is the substance acutely toxic (LD_{50}≤400 mg/kg bw) or very toxic (LD_{50}≤50 mg/kg bw) via the dermal route?[d] → ↓ NO ↓	YES: The substance will be classified for its acute dermal toxicity. Proceed to next step
Existing (Q)SAR data and read-across		
5	Are there structurally related substances (suitable "read-across" or grouping), which are classified as irritating to the eye, or do valid QSAR methods indicate eye irritation of the substance?[e] → ↓ NO ↓	YES: Consider to classify accordingly (R41 or R36). If discrimination between R41 and R36 is not possible, R41 must be chosen. Proceed to next step
Existing in vitro data		
6a	Are there data from a validated *in vitro* test (adopted by OECD or not), which provide evidence that the substance is an eye irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (R36, R41 or no classification). If discrimination between R41 and R36 is not possible, R41 must be chosen. Proceed to next step
6b	Are there acceptable data from a non-validated *in vitro* test, which provide evidence that the substance is an irritant to the eye[f]? → ↓ NO	YES: Consider to classify R41, Proceed to next step

	↓	
Weight of evidence analysis		
7	Taking all existing and relevant data (steps 1-6) into account, is there sufficient information to make a decision of whether classification/labelling is necessary, and – if so – how to classify and label? ⟶ ↓ NO ↓	YES: Classify for accordingly (R36, R41 or no classification).
New in vitro/ex vivo tests for eye irritation (Annex VII)		
8a	Does the substance demonstrate irritating or non-irritating properties in validated *in vitro* or *ex vivo* tests (adopted by OECD or not) for eye irritation? ⟶ ↓ NO ↓	YES: Classify accordingly (R36, R41 or no classification). If discrimination between R41 and R36 is not possible, R41 must be chosen.
8b	Does the substance demonstrate severe irritating properties in acceptable non-validated *in vitro* or *ex vivo* tests for eye irritation (at present only IRE, ICE, BCOP and HET-CAM)[f]? ⟶ ↓ NO ↓	YES: Classify R41
New in vivo test for eye irritation (Annex VIII)		
9	Does the substance demonstrate irritancy in an OECD adopted *in vivo* test? ⟶ ↓ NO ↓ No classification	YES: Classify accordingly.

Notes to the information scheme eye irritation

[a] Note that if the buffering capacity suggests the substance be non-corrosive, further data are needed to confirm this.

[b] If pH < 3.2 or pH > 8.6, the substance is very likely to be an eye irritant.

[c] Data from case reports, occupational experience, poison information centres or from clinical studies.

[d] If the substance is acutely toxic (LD_{50}≤400 mg/kg bw) or very toxic (LD_{50}≤50 mg/kg bw) via the dermal route further testing for eye irritation would result in severe suffering or death of the animal. Thus, further testing is not required and

sufficient labelling (warning) is provided by the risk phrases: “R24: toxic in contact with the skin” or “R27: very toxic in contact with the skin” and the symbol with T or T^+, shown below.

[e] Conclusion on no classification can be made if the model has been shown to adequately predict the absence of the classified effect and if it fulfils the requirements of Annex XI.

[f] Conclusion on no classification can only be made if it has been concluded in the evaluation process that the test allows the identification of non-irritants and the data are used in a WoE approach following Annex XI 1.2.

Appendices 1-3 to Section R.7.2

Appendix R.7.2-1 Mechanisms of local toxicities: skin corrosion/irritation, eye and respiratory irritation

Content of Appendix 7.2-1

Mechanisms of skin corrosion and irritation

Mechanisms of eye irritation

Mechanisms of respiratory irritation

Mechanisms of skin corrosion and irritation

Clinically, different types of irritant contact dermatitis (ICD) exist, and have been classified on the basis of differences in morphology and mode of onset, as: acute irritant dermatitis (primary irritation); irritant reaction; delayed, acute irritant contact dermatitis; cumulative irritant dermatitis; traumatic irritant dermatitis, pustular and acneiform irritant dermatitis; non-erythematuous irritant dermatitis; and subjective irritation (Lammintausta & Maibach, 1990).

Two different pathogenetic pathways may be involved in ICD. Acute ICD is characterised by an inflammatory reaction which mimics allergic contact dermatitis, with the release of inflammatory mediators and cytokines. Chronic ICD, on the other hand, is characterised by disturbed barrier function, associated with an increased epidermal turnover which leads clinically to lichenification (Berardesca and Distante, 1994).

The clinically relevant elements of skin irritation are a disturbance of the desquamation process, resulting in scaling or hyperkeratosis (chronic effects), i.e. epidermal events, and an inflammatory response with vasodilation and redness in combination with extravasation of water, which may be observed as papules, vesicles and/or bullae and oedema (acute effects), i.e. events essentially taking place in the dermis (Serup, 1995). The onset of irritation takes place at the stratum corneum level and later in the dermis, whereas early events in sensitisation occur in the dermis. Variations in the skin reactions are dependent on the degree of injury induced, as well as on the effects of an irritant substance on different cell populations. For example, pigmentary alterations are due to effects on melanocytes, whereas ulcerations are due to extensive keratinocyte necrosis (skin corrosion). The release of cytokines and mediators can be initiated by a number of cells, including living keratinocytes and those of the stratum corneum, which thus modulate inflammation and repair (Sondergard *et al*., 1974; Hawk *et al*., 1983; Barker *et al*., 1991; Baadsgaard and Wang, 1991; Hunziker *et al*., 1992; Berardesca *&* Distante, 1994).

The physico-chemical properties, concentration, volume and contact time of the irritant give rise to variations in the skin response. Furthermore, inter-individual differences exist, based on age, gender, race, skin colour and history of any previous skin disease. In the same individual, reactivity differs according to differences in skin thickness and skin sensitivity to irritation of the different body regions. Finally, a greater sensitivity to some irritants (DMSO, propylene glycol, SLS and soap) has been reported during winter, because of the reduced hydration state of the skin (Frosch and Pilz, 1995). Although clinically different types of irritant reactions can be observed, they are all based on cellular and biochemical mechanisms which induce the irritant response. It is not yet possible to conclude whether the observed clinical differences are actually due to differences in biochemical mechanisms, and further investigations are needed.

According to Barratt (1995) and further elaborated by Walker et al. (2004), for organic chemicals, the mechanisms leading to skin irritation are normally described by a two-stage process where a chemical first has to penetrate the *stratum corneum* and then trigger a biological response in deeper epidermal or dermal layers.

For strong inorganic acids and bases, no *stratum corneum* penetration is needed because they erode the *stratum corneum*. According to the Technical Guidance Document (TGD) supporting Commission Directive 93/67/EEC on risk assessment for new notified and existing substances (EC, 2003), the percutaneous absorption of acrylates, quaternary ammonium ions, heterocyclic ammonium ions and sulphonium salts is slow, since these chemicals are binding to macromolecules in skin. As a result of binding, corrosion can occur as the *stratum corneum* is eroded. Reactivity can be caused by electrophiles and/or pro-electrophiles. Electrophiles contain atoms, such as N, O or halogens attached to a C-atom, which makes that specific C-atom positively charged and therefore reactive with electron-rich regions of peptides and proteins. This causes irritation via covalent binding to the skin.

At this time, the following mechanisms are proposed for inducing skin irritation or skin corrosion by affecting the structure and function of the *stratum corneum* :

1. Mechanisms of skin irritation:

   - Reaction with skin proteins and interference with lipids in the *stratum corneum* by surface-active agents (denaturation of proteins, disruption of plasma membrane lipids)

   - Dissolving of plasma membrane lipids and thus defatting and disintegration of skin by low molecular weight organic chemicals.

2. Mechanisms of skin corrosion:

   - Erosion of the *stratum corneum* by most inorganic acids and bases and by strong organic acids with pH <2.0 and bases with pH >11.5 and

   - Binding to skin components in the *stratum corneum* by cationic surfactants and percutaneous absorption of acrylates, quaternary ammonium ions, heterocyclic ammonium ions and sulphonium salts.

3. Mechanisms that may lead to both skin irritation and corrosion:

   - Penetration of the *stratum corneum* by anionic or non-surfactant organic chemicals with sufficient hydrophobic and hydrophilic properties and

   - Elicitation of a inflammatory and/or cytotoxic response in the epidermis or dermis. The severity of these responses may determine whether irritation or corrosion occurs.

Mechanisms of eye irritation

Eye injury can be caused by many insults. These can be physical such puncture by sharp objects. Eye injury can be caused by chemicals such as systemic drugs that can enter into the eye through the blood stream (examples are Cyclosporine, Vaccines, Intravenous immunoglobulines, Intravenous streptokinase). Various degrees of eye injury can also be caused by direct (topical) contact with chemicals or chemical mixtures such as acids, alkalis, solvents or surfactants. These materials may contact the eye intentionally e.g. through the use of eye drops, medications, products

intended for use around the eyes but also unintentionally e.g. accidental spills and splashes of consumer products or accidental exposures in the workplace.

In general, chemicals or chemical mixtures which contact the eye directly may cause local effects on the frontal tissues and substructures of the eye e.g. cornea, conjunctiva, iris, lachrymal system and eye lids. There are several modes of action by which topical chemicals and chemical mixtures cause eye injury (see Table R.7.2-3).

Table R.7.2-3 Categories of irritant chemicals and their typical mode of action in eye irritation.

Chemical/chemical mixtures	Mode of Action
Inert chemicals	May cause effect due to large size. Protrusions may cause direct puncture of the eye
Acids	May react directly with eye proteins and cause coagulation or precipitation resulting in relatively localised injury
Bases (Alkalis)	May actively dissolve cell membranes. May penetrate to the deeper layers of the eye tissue
Solvents	May dissolve lipids in plasma membranes of epithelial and underlying cells resulting in loss of the cells affected and, as a result, tissue degradation, that might be – depending on the repair mechanisms (cell proliferation, tissue restoration) transient.
Lachrymators	May stimulate the sensory nerve endings in the corneal epithelium causing an increase in tearing.

The degree of eye injury is usually dependent on the characteristics (chemical category/class) and concentration of the chemical or chemical mixture. Acids and alkalis usually cause immediate irritation to the eyes. Other substances may cause eye injuries that start as mild but progress to be more severe at a later period.

Upon exposure of the ocular surface to eye irritants, inflammation of the conjunctiva can be induced. This includes dilation of the blood vessels causing redness, increased effusion of water causing swelling (oedema/chemosis) and an increase in the secretion of mucous leading to an increase in discharge. Visual acuity can be impaired. Irritants may also produce an increase in tear production and changes to the tear film integrity such as increased wetness. Iritis may result from direct irritation or become a secondary reaction to the corneal injury. Once the iris is inflamed, infiltration of fluids can follow which affects the ability to adjust the size of the pupil and decreases the reaction to light leading to decreased visual acuity. Due to the richness of nerves in the iris, irritation also causes subjective symptoms such as itching, burning and stinging.

Eye injury can be reversible or irreversible depending on the degree of damage and degree of repair. Damage to the corneal epithelium alone can repair quickly, often with no permanent eye damage. The cornea may still repair fairly well if the damage goes beyond the basement membrane into the superficial part of the stroma but the repair process may take days or even weeks to occur. Once the damage extends significantly into the stroma, corneal ulceration can occur due to the subsequent series of inflammatory processes. If damage extends to and beyond the endothelium, corneal perforation may occur which is irreversible and may cause permanent loss of vision. Eye injury can

cause different degrees of functional loss e.g. increase of tear production, opacification of the cornea, oedema and so decrease visual acuity.

The body has its own defence mechanisms e.g. sensing the pain, stinging and burning and the eyelids will blink to avoid full exposure to the chemical. Increased tear production and blinking of the eyes with the help of the drainage apparatus help to dilute or clear the causative agent. Such defence mechanisms are highly developed in man with rapid blinking and profuse tear production resulting from exposure of the eye to a foreign material that is irritating. It is well reported in the literature that species differences occur in the rate of blinking and tear production mechanism that can influence how effectively foreign materials are removed from the eye.

Mechanisms of respiratory irritation

The term "respiratory tract irritation" is often used to indicate either or both of two different toxicological effects. These are i) cytotoxic effects in the affected tissue, and ii) sensory irritation. The first type of irritation is comparable to deremal and eye irritation.

Cytotoxic irritant effects are characterised by inflammation (increased blood flow (hyperemia), local infiltration with white blood cells, swelling, oedema) and there may also be haemorrhage, and eventual necrosis and other pathological changes. The effects are in principle reversible.

Chronic irritation can lead to repeated episodes of cell proliferation in the affected tissues, and this may increase the risk of tumor development. The nature of effects depends on the chemical compound and its primarily targeted region, the severity of effects depends on the concentration and duration of exposure. In general, repeated exposure studies in animals tend to focus on observing (histo)pathological evidence for tissue damage rather than for sensory irritant effects. In case overt tissue damage (mucosal erosion and ulceration) occurs, a non-specific cytotoxic action at the site of contact along the respiration route can be assumed. Depending on the concentration and duration of exposure a severity gradient of lesions from anterior to posterior regions can be observed (in contrast to effects in certain mucosa types depending on the metabolic activation of the test substance) and, depending on the severity and the extent of the lesions, adjacent submucosal tissues can also be affected (e.g., by cartilage destruction). Such lesions are not fully reversible due to scar formation or replacement of the original mucosa, or may induce other serious health effects as marked bleeding or persistent airway obstruction.

"Sensory irritation" refers to the local and central reflex interaction of a substance with the autonomic nerve receptors, which are widely distributed in the mucosal tissues of the eyes and upper respiratory tract. Compound or compound-group specific target sites of sensory irritation generating different responses can be identified: a) nasal (and eye) irritation, i.e. interaction with the trigeminal nerve, b) pharyngeal irritation, i.e. interaction with the glossopharyngeal nerve, and c) larynx and lower respiratory tract, i.e. interaction with the vagus nerve.

Sensory irritation leads to unpleasant sensations such as pain, burning, pungency, and tingling. The severity depends on the airborne concentration of the irritant rather than on the duration of exposure. Sensory irritation is a receptor-mediated effect, and usually occurs almost immediately upon exposure to the inhaled irritant. It leads to reflex involuntary responses such as sneezing, lacrimation, rhinorrhea, coughing, vasodilatation of blood vessels in the nasal passages, and changes in the rate and depth of respiration. In humans, protective behavioural responses such as covering the nose and mouth can also occur. Sensory irritation is distinct from odor sensation, which is mediated via different nerve pathways (olfactory). However, there is evidence that odor perception and other cognitive influences can affect the perception of sensory irritation in humans.

In rodents, sensory irritation leads to a reflex reduction in the respiratory rate (breath-holding); this reflex effect on respiration can be measured experimentally (determination of the RD_{50} value in the Alarie assay) although results may vary considerably depending on the species and strain of rodents, on the exposure duration (time should be long enough to induce changes), and results also show inter-laboratory variability. Investigations into the correlation of the results of the Alarie test with human data are difficult since the parameters examined in humans and mice are different and adequate human data to determine a human equivalent to the RD_{50} is not available at the moment. The results of a study by Cometto-Muniz et al. (1994) indicate that RD_{50} values in animals are not easily comparable with 'nasal pungency thresholds' in humans.

As indicated, human data are mostly based on subjective experiences and need to be carefully controlled in order to prevent confounding by odour perception (Dalton, 2003; Doty *et al*., 2004). Validated questionnaires have been developed for the investigation of sensory irritation responses in human volunteers. During recent years, emphasis was given to develop a spectrum of objective measurements (see review by Arts *et al*., 2006).

There is a view in the occupational health literature that sensory irritation may be a more sensitive effect than overt tissue-damaging irritation (which is a non-receptor mediated unspecific mode inducing cell death at the site of contact). Sensory irritation-related effects are fully reversible given that its biological function is to serve as a warning against inhaled substances that could damage the airways, and that it triggers physiological reflexes that limit inhalation volumes and protect the airways. However, there is a lack of documented evidence to indicate that this is a generic position that would necessarily apply to all inhaled irritants. It should be noted that no clear relationship between the RD_{50} value and the onset of histologically observable lesions in animals has been observed.Appendix R.7.2-2 - QSARs and expert systems for skin irritation and corrosion

Appendix R.7.2-2 QSARs and expert systems for skin irritation and corrosion

Content of Appendix 7.2-2

Literature-based QSAR models

Commercial models

BfR decision support system

SICRET

Literature-based QSAR models

In the open scientific literature, (Q)SARs have been based on continuous (e.g. Primary Irritation Indices) or categorical (e.g. EU classifications) measures of skin irritation.

For defined classes of chemicals, categorical QSARs have been reported for discriminating between corrosives and non-corrosives (Barratt 1996a, 1996b), and between skin irritants and non-irritants (Smith *et al.*, 2000a; Smith *et al.*, 2000b). These studies did not actually provide a transparent algorithm for classifying chemicals, so they are of limited value for regulatory use. However, they illustrate the feasibility of developing such models, so it should be possible for a QSAR specialist to redevelop the models in such a way that an algorithm is clearly defined.

A linear discriminant model for distinguishing between irritant and non-irritant liquid esters in human volunteers was reported by Smith *et al*. (2000a). As mentioned above the exact algorithm is not clear. In addition the primary irritation index for human irritation may need translation when these scores are considered for classification. However, the results could be informative for future model development for esters, since they indicate that irritant esters can be distinguished from non-irritants on the basis of a limited number of physico-chemical parameters.

For defined classes of chemicals, continuous QSARs for predicting the Primary Irritation Index (PII) have also been published (Barratt 1996b; Hayashi *et al.*, 1999; Kodithala *et al.*, 2002). For example, the application of stepwise regression analysis to a set of 52 neutral and electrophilic organic chemicals produced the following model:

$$PII = 1.047 \log P - 0.244\ MV + 0.888\ DM + 0.353$$

$N=52,\ r^2 = 0.422,\ r_{cv}^2 = 0.201,\ s=1.376,\ F=11.70$

This equation indicates that the PII has a positive dependence on log P (logarithm of the octanol-water partition coefficient) and DM (dipole moment), and a negative dependence on MV (molecular volume). This model has a low goodness-of-fit (r^2) and a poor predictivity (as reflected by r_{cv}^2), so is not recommended for regulatory use. Nevertheless, the model does reveal three potentially useful descriptors for the development of new models for PII prediction. More research is needed into the development of models for predicting PII and it should be considered whether the information generated could be used in the setting of DNELs.

Some limited evidence indicates that the reactive effects of acids and bases can be predicted by using the acid/base dissociation constant (pKa), which can itself be predicted by using commercially available software products, such as the SPARC program. Evidence for the usefulness of pKa as a predictor of skin irritation for acids has been provided by Berner *et al.* (1988, 1990a, 1990b), whereas evidence for the usefulness of pKa as a predictor of skin irritation for bases has been provided by Nangia *et al*. (1996). Barratt also used pKa for predicting the effects of acids and bases (Barratt, 1995). These studies did not address the question of how to use pKa where there are multiple functional groups in the chemical of interest, and therefore multiple ionization constants. Based on current knowledge, no clear recommendations can be made about how to use pKa information.

Commercial models

TOPKAT, which is commercialised by Accelrys (http://www.accelrys.com/products/topkat), incorporates models to discriminate severe irritants from non-severe irritants, as well as mild/moderate irritants from non-irritants. These models are based on work by Enslein *et al*. (1987), but due to a lack of documentation, it is not clear whether the current version of the software encodes the models that were originally published. A QMRF for the TOPKAT skin irritation model is provided as an appendix. The algorithm of the TOPKAT is not transparent. The model predicts a probability of a weak/mild/moderate and severe irritation. It states that probabilities <0.3 and >0.7 give sufficient certainty of the prediction. The model gives the sensitivity and specificity values of the specific classes such as acyclic etc, which are mostly around or above 90%. It also shows similar structures from the TOPKAT perspective including the experimental result. The TOPKAT predictions of weak/mild/moderate and severe irritation need to be translated to consider them for classification. The models indicate whether the prediction is in the applicability domain of the model. Due to the limitations of the model (lack of transparency for the algorithm, no external validation, no mechanistic reasoning), it cannot be used as stand alone method. The TOPKAT prediction should be supported with mechanistic reasoning, using other models or expert judgment.

There is a rulebase for irritation in **Derek for Windows** (Sanderson & Earnshaw, 1991; Combes & Rodford, 2004), which is developed and regularly updated by LHASA Ltd (http://www.chem.leeds.ac.uk). To predict toxicity, the program checks whether any alerts within the query structure match previously characterised toxicophores (substructure with potential toxic effect) in the knowledge base. The reasoning engine then assesses the likelihood of a structure being toxic, and a message indicating the nature of the toxicological hazard is provided together with relevant literature references. There are nine levels of confidence: certain, probable, plausible, equivocal, doubted, improbable, impossible, open, contradicted. The DerekfW8.0 rulebase has 25 structural alerts for the prediction of skin irritancy/corrosion; four alerts are specific to eye irritancy, and some combined for the respiratory irritation and gastrointestinal tract, but none is specific to skin irritancy or corrosivity. If DerekfW does not make a prediction of irritancy or corrosivity, it cannot be concluded that there is no effect – it could mean that none of known alerts was found to be present in the chemical of interest or it was outside the applicability domain of that specific alert. The DerekfW model is transparent in its algorithm, when the model is fired showing the structural alert and its limitations. The alert is supported with literature references and sometimes with example chemicals, although this is not sufficient to consider them validated. The example chemicals support the mechanistic reasoning. The DerekfW model can be used for positive identification of skin irritation. The confidence levels have to be translated to consider them for classification. Due to the limitations (lack of validation) it cannot be used as stand alone method, though the mechanistic reasoning provides supporting information. The DerekfW model cannot be

used to predict non-irritation/corrosion as the model only contains alerts that detect the presence of irritation/corrosion.

HazardExpert is a rule-based software tool developed and commericalised by CompuDrug Chemistry Ltd. (http://www.compudrug.com) for predicting the toxicity of organic compounds in humans and in animals (Smithing & Darvas 1992). HazardExpert uses a fragment-based approach to predict toxicokinetic effects and various human health effects, including membrane irritation. Since this endpoint is not clearly defined in HazardExpert, it is recommended not to use it directly for the assessment of skin or eye irritation. However, it could be used as supplementary information in a *Weight of Evidence* approach for positive prediction.

The Multiple Computer Automated Structure Evaluation (**MultiCASE**) program, developed by MultiCASE Inc. (http://www.multicase.com), is an automated rule induction tool that automatically identifies molecular fragments likely to be relevant to the activity of molecules (Klopman, 1992; Klopman *et al.*, 1993). It also provides an indication of the importance of these fragments in relation to the potency of the molecules containing them. MultiCASE can be used to predict various human health endpoints, including eye irritation (Klopman *et al.*, 1993; Rosenkranz *et al.*, 1998). However, it is not clear how to relate the MultiCASE scoring system to Draize scores or regulatory classifications. In principle, the MultiCASE model can be used for positive and negative indications of skin irritation. The structural alert is provided as well as information on its internal validation. The MultiCASE model also indicates whether it is in the applicability domain of the model. The MultiCASE predictions of weak/mild/moderate and severe irritation need to be translated to consider them for classification. Due to limitations (lack of external validation and mechanistic reasoning) the model cannot be used as a stand alone method. The prediction should be supported with mechanistic reasoning using other models or expert judgment.

The Danish EPA has developed an in-house MultiCASE model for predicting severe versus mild skin irritation based on 800 test results taken from RTECS (Registry of Toxic Effects of Chemical Substances), the HSDB (Hazardous Substances Data Bank) and the official list of EU-classified substances (Annex I of Directive 67/548/EEC). It is not clear how the RTECS and HSDB classification criteria for irritation comply with the EU criteria. Due to limitations in the information for assessing the reliability of the prediction, these predictions are difficult to use in the regulatory context.

BfR decision support system

A decision support system (DSS) developed by the German Federal Institute for Risk Assessment (BfR) uses physico-chemical exclusion rules to predict the absence of skin irritation/corrosion potential in combination with structural inclusion rules (SARs) to predict the presence of such potential (Gerner *et al.*, 2004; Walker *et al.*, 2004). The exclusion rules are based on physico-chemical properties such as molecular weight, aqueous solubility, and log K_{ow}, whereas the inclusion rules are based on substructural molecular features. The physico-chemical rules implicitly take into account bioavailability (skin penetration) whereas the structural rules take reactivity into account. The physico-chemical and structural rulebases are designed to predict the EU risk phrases for skin irritation (R38) and skin corrosion (R34 and R35). Further details are given in QSAR Reporting Format for the BfR skin and eye irritation rulebases (http://qsardb.jrc.it).

The exclusion rules have the following general form:

IF (physico-chemical property) A THEN predict the absence of toxic effect B

Example: IF Log K_{ow} < -3.1 THEN the chemical does not need to be considered for classification

The structural inclusion rules take the following general form:

IF (substructure) A THEN predict the occurrence of toxic effect B

Example: IF *Chlorosilane* THEN the chemical needs to be considered for "corrosive" classification

The performance of the BfR physico-chemical rulebase for predicting the absence of skin effects has been validated by the RIVM (Rorije & Hulzebos, 2005), whereas the structural rulebase for predicting the occurrence of skin effects has been validated by the ECB (Gallegos Saliner *et al.*, 2007). The endpoint is EU classification, the algorithms and domain of applicability are transparent, the rules and alerts are independently validated by ECB and RIVM (Gallegos Saliner *et al.*, 2007, Rorije & Hulzebos, 2005). Though the rules are empirically derived, a mechanism of action can be deduced. For chemicals in the applicability domain of the rulebase, the rules may be used on their own to predict the presence or absence of hazard. Thus, the resulting predictions can be used as the basis for classification. It should be determined, on a case-by-case basis, whether the predictions for a given chemical provide a sufficient basis for classification, or whether additional information is needed in a weight-of-evidence approach.

SICRET

The so-called "Skin Irritation Corrosion Rules Estimation Tool" (SICRET), has been developed by Walker *et al.* (2005) to estimate whether chemicals are likely to cause skin irritation or skin corrosion SICRET is not actually a computer-based tool but a tiered approach based on the use of physico-chemical property limits, structural alerts and *in vitro* tests to classify chemicals that cause skin irritation or skin corrosion. The physico-chemical rules and alerts include those in the BfR rulebases as well as some additional rules and alerts published by Hulzebos *et al.* (2001, 2003, 2005).

Appendix R.7.2-3 QSARs and expert systems for eye irritation and corrosion

Content of Appendix 7.2-3

Literature-based QSAR models

Commercial models

BfR decision support system

Literature-based QSAR models

In the open scientific literature, (Q)SARs have been based on continuous (e.g. molar eye scores) or categorical (e.g. EU classifications) measures of eye irritation. Examples of mathematical (continuous) models have been published models by Sugai *et al.* (1991) & Cronin *et al.* (1994), whereas examples of categorical models have been published by Sugai *et al.* (1990) and by Barratt (1997).

Regression models based on solvatochromic parameters can be used for predicting the degree of eye irritation, as illustrated by Abraham and coworkers (Abraham, 1994; Abraham *et al.*, 1998). The mecahnistic basis of these models is that a substance is transferred from a pure organic liquid to an organic solvent phase consisting of the tear film and cell membranes on the surface of the eye. The more soluble the organic liquid in the initial phase, the greater the degree of irritation is. These models are worthy of further characterisation. However, for routine regulatory use, information on a number of so-called Abraham descriptors would also need to be made available.

Neural network approaches can also be used to model eye irritation (e.g. Patlewicz *et al.*, 2000). At present, however, many of these models lack the transparency, especially in the algorithm. However if the training sets are provided as well as validation information they could possibly be used in a *Weight of Evidence* approach. Mechanistic reasoning should also be provided.

An approach called Membrane-Interaction QSAR analysis, developed by Kulkarni *et al.*(2001), provides a means of incorporating molecular dynamic simulations to generate membrane–solute interaction properties. The development and application of models based on molecular simulations requires the use of specialised expertise and software. They could be used to increase understanding of the mechanisms of eye irritation.

A classification approach called Embedded Cluster Modelling (ECM) provides a means of generating *elliptic models* in two or more dimensions (Worth & Cronin, 2000), so that irritants can be transparently identified as those chemicals located within the boundaries of the ellipse. The statistical significance of these "embedded clusters' can be verified by cluster significance analysis (CSA), as illustrated for an eye irritation dataset by (Cronin, 1996).

Applying the methods of ECM and CSA, the following model, applicable to undiluted organic liquids, was developed by Worth & Cronin (2000):

Classify an undiluted, organic liquid as an eye irritant if:

$$(\log P\text{-}1.07)^2 / 2.06^2 + (dV1 + 0.98)^2 / 0.99^2 \leq 1$$

This model was based on 73 diverse organic chemicals, using two descriptors: LogP (which accounts for diffusion) and a size-independent molecular connectivity index (dV1, which accounts for the degree of branching and cyclicity). The sensitivity, specificity and concordance of the model were 73%, 78% and 75%, respectively, whereas the positive and negative predictivities were 77% and 74% respectively. The model is an explicit algorithm with a defined applicability domain and predicts EU classifications directly.

The different methods were applied to a dataset of 119 organic liquids classified as I or NI according to EU classification criteria. The classification models (CMs) were developed by applying linear discriminant analysis (LDA), binary logistic regression (BLR), and classification tree (CT) analyses, using a single predictor variable (molecular weight), and assigning equal probabilities for the two classes (I/NI). The cut off values below which a chemical should be predicted to be irritating to the eye were 121, 77, and 137 g/mol, in the LDA, BLR, and CT classification models, respectively (Table R.7.2-4) (Worth & Cronin, 2003).

Table R.7.2-4 Classification results of the different models of eye irritancy

CM ($p<0.01$)	Cut off value	Sensitivity	Specificity	Accuracy
Linear Discriminant Analysis (LDA)	if MW ≤ 121 g/mol, then predict I; otherwise, predict NI	73	62	65
Binary Logistic Regression (BLR)	if MW ≤ 77 g/mol, then predict I; otherwise, predict NI	27	93	76
Classification Tree (CT)	if MW ≤ 137 g/mol, then predict I; otherwise, predict NI	97	49	61

All of these models are simple to apply and are associated with a transparent algorithm. The statistics illustrate the inevitable trade-offs that result from the selection of different cut off values. Thus, the BLR model does not identify many irritants (only 27%), but is does so with a high degree of confidence (i.e. low false positive rate of 7%). Conversely, the CT does not identify many of the non-irritants (49%), but it has a low false negative rate of 3%). Thus, the combined use of the BLR and CT models could be useful for distinguishing between eye irritants and non-irritants.

Commercial models

The **TOPKAT** software includes models for eye irritation based on structural fragments. These models were originally developed by Enslein *et al.* (1988), but the algorithms are not well defined in the TOPKAT documentation The TOPKAT algorithm is not transparent. The model predicts a probability of a weak/mild/moderate and severe irritation. It states that probabilities <0.3 and >0.7 give sufficient certainty of the prediction. The model gives the sensitivity and specificity values of the specific classes such as acyclic, which are mostly around or above 90%. It also shows similar structures from the TOPKAT perspective including the experimental result. The TOPKAT predictions weak/mild/moderate and severe irritation need to be translated to consider them for classification. The models indicate whether the prediction is in the applicability domain of the model. Due to the limitations of the model (lack of transparency for the algorithm, no external validation, no mechanistic reasoning), it cannot be used as stand alone method. The TOPKAT prediction should be underlined with a mechanistic reasoning, using other models or expert judgment.

There is a rulebase for irritation in **Derek for Windows** (Sanderson & Earnshaw, 1991; Combes & Rodford, 2004), which is developed and regularly updated by LHASA Ltd (http://www.chem.leeds.ac.uk). See for a general outline the skin irritation section on (Q)SARs. The DerekfW8.0 rulebase has four four alerts are specific to eye irritancy. If DerekfW does not make a prediction of irritancy or corrosivity, it cannot be concluded that there is no effect – it could mean that none of known alerts was found to be present in the chemical of interest or it was outside the applicability domain of that specific alert. The DerekfW model is transparent in its algorithm, when the model is fired showing the structural alert and its limitations. The alert is underlined with literature references and sometimes with example chemicals, which is not sufficient to consider them internally validated. The example chemicals underline the mechanistic reasoning. The DerekfW model can be used for positive identification of skin irritation. The confidence levels have to be translated to consider them for classification. Due to the limitations (lack of internal and external validation) it cannot be used as stand alone method, though the mechanistic reasoning possibly provides sufficient information. The DerekfW model cannot be used to predict for non-irritation/corrosion as the model only contains alerts that detect the presence of irritation/corrosion.

The fragment-based **MultiCASE** approach has been used to model eye irritation (Klopman *et al.*, 1993; Enslein *et al.*, 1988; Rosenkranz *et al.*, 1998; Klopman (1998). The publications on these models do not define the algorithms. In principle, the MultiCASE model can be used for positive and negative indication for eye irritation. The structural alert is provided as well as the internal validation. The MultiCASE model also indicates whether it is in the applicability domain of the model. The MultiCASE predictions of weak/mild/moderate and severe irritation need to be translated to consider them for classification. Due to limitations (lack of external validation and mechanistic reasoning) the model cannot be used as a stand alone method. The prediction should be underlined with mechanistic reasoning using other models or expert judgment.

BfR decision support system

The decision support system (DSS) developed by the German Federal Institute for Risk Assessment (BfR) uses physico-chemical exclusion rules to predict the absence of eye irritation/corrosion potential in combination with structural inclusion rules (SARs) to predict the presence of such potential (Gerner *et al.*, 2005). These rules are used analogously to those described in the skin irritation and corrosion section above. The physico-chemical and structural rulebases are designed to predict the EU risk phrases for eye irritation (R36) and severe eye irritation/corrosion (R41). Independent validation exercises by the ECB support the performance of the physico-chemical rulebase for predicting the absence of eye effects (Tsakovska *et al.*, 2005), as well as the performance of the structural rulebase for predicting the occurrence of eye effects (Tsakovska *et al*., 2007)

R.7.2.7 Useful links

- JRC QSAR Model Database: http://qsardb.jrc.it
- ECVAM page: http://ecvam.jrc.it/index.htm
- ECVAM database service on alternative methods to animal experimentation (DB-ALM):
-

R.7.2.8 References on skin and eye irritation/corrosion and respiratory irritation

Ad-hoc Working group on Defatting substances (1997). ECBI/22/96-Add. 10

A.I.S.E. (2006) Low Volume Eye Test (LVET) – A.I.S.E. Submission to ECVAM – January 2006

Abraham, M.H. (1994). Scales of solute hydrogen-bonding: their construction and application to physico-chemical and biochemical processes. *Chemical Society Reviews* **22**, 73-83.

Abraham, M.H., Kumarsingh, R., Cometto-Muñiz, J.E., Cain W.S. (1998). A Quantitative Structure-Activity Relationship (QSAR) for a Draize Eye Irritation Database. *Toxicology in Vitro* **12**, 201-207.

Alarie Y (1973). Sensory irritation in the upper airways by airborne chemicals. Toxicol Appl Pharmacol **24**: 279-297.

Alarie, Y. (2000). Computerized animal bioassay to evaluate the effects of airborne chemicals on the respiratory tract. In: Indoor Air Quality Handbook (J. D. Spengler, J. M. Samet, and J. F. McCarthy, Eds.), p. 24.1-24.25. McGraw-Hill, New York.

Arts JHE, de Heer C., and Woutersen R (2006). Local effects in the respitory tract: relevance of subsjectively measured irritation for setting occupational exposure limits. Int Arch Occup Environ Health **79**: 283-298.

ASTM, 2004. Standard test method for estimating sensory irritancy of airborne chemicals; Designation 981-04. In: *Annual Book of ASTM Standards*. Philadelphia: American Society for Testing and Materials.

Baadsgaard, O. & Wang, T. (1991) Immune regulation in allergic and irritant skin reactions. *International Journal of Dermatology* 30**, 161-172.**

Barker, J.N., Mitra, R.S., Griffiths, C.E.M., Dixit, V.M. & Nickoloff, B.J. (1991) Keratinocytes as initiators of inflammation. ***Lancet*** 337**, 211-214.**

Barratt, M.D. (1995) Quantitative structure activity relationships for skin corrosivity of organic acids, bases and phenols. *Toxicology Letters* **75**, 169-176.

Barratt, M.D. (1996a.) Quantitative Structure-Activity Relationships for Skin Corrosivity of Organic Acids, Bases and Phenols: Principal Components and Neural Network Analysis of Extended Sets. *Toxicology in Vitro* **10**, 85-94.

Barratt , M.D. (1996b). Quantitative Structure-Activity Relationships for Skin Irritation and Corrosivity of Neutral and Electrophilic Organic Chemicals. *Toxicology in Vitro* **10**, 247-256.

Barratt , M.D. (1997). QSARs for the Eye Irritation Potential of Neutral Organic Chemicals. *Toxicology in Vitro* **11**, 1-8.

Berardesca, E. & Distante, F. (1994) The modulation of skin irritation. *Contact Dermatitis* **31**, 281-287.

Berner, B., Wilson, D.R., Guy, R.H., Mazzenga, G.C., Clarke, F.H., Maibach, H.I. (1988). The relationship of pKa and acute skin irritation in man. *Pharmaceutical Research* **5**, 660-663.

Berner, B., Wilson, D.R., Steffens, R.J., Mazzenga, G.C., Hinz, R., Guy, R.H., Maibach, H.I. (1990a). The relationship between pKa and acute skin irritation for a series of basic penetrants in man. *Fundamental and Applied Pharmacology* **15**, 760–766.

Berner, B., Wilson, D.R., Guy, R.H., Mazzenga, G.C., Clarke, F.H., Maibach, H.I. (1990b). Relationship of pKa and acute skin irritation in humans. *Journal of Toxicology-Cutaneous and Ocular Toxicology* **9**, 481–492.

Blazka, M.E. et al. (1999) Colgate-Palmolive's programme to validate the EpiOcular TM human tissue construct model (OCL-200). ATLA, 27:331

Blazka, M.E. et al. (2000) Colgate-Palmolive's program – validate the EpiOcular human tissue construct model (OCL200). Toxicological Sciences, 54(1) 188, Abstract #884.

Blazka, M.E. et al. (2003) Evaluating the ocular irritation potential of 54 tst articles using the epiocular human tissue construct model (OCL-200). Society of Toxicology Meeting, Poster #1070.

Bos P.M, Zwart A, Reuzel P.G, Bragt P.C (1991). Evaluation of the sensory irritation test for the assessment of occupational health risk. Crit Rev Toxicol. 21(6):423-50. Review.

Bos PM, Busschers M., Arts J.H. (2002) Evaluation of the sensory irritation test (Alarie test) for the assessment of respiratory tract irritation. J Occup Environ Med. Oct;44(10):968-76.

Botham, P.A., L.K. Earl, J.H. Fentem, R. Roguet and J.J.M. van de Sandt (1998). Alternative methods for skin irritation testing: the current status. ECVAM Skin Irritation Task Force Report 1. *Alternatives To Laboratory Animals* **26**, 195-211.

Boylstein, L., Luo, J., Stock, M.F., Alarie, Y. (1996). An attempt to define a just detectable effect for airborne chemicals in the respiratory tract in mice. *Arch. Toxicol*. **70:**567-578.

Combes, R.D. & Rodford, R. (2004). The Use of Expert Systems for Toxicity Prediction - Illustrated With Reference to the DEREK Program. In *Predicting Toxicity and Fate*. Cronin, M.T.D. & Livingstone, D.J. Eds. CRC Press, pp. 193-204.

Cometto-Muniz J.E. & Cain W.S. (1994) Am Ind Hyg Assoc J. 1994 Sep;55(9):811-7. Sensory reactions of nasal pungency and odor to volatile organic compounds: the alkylbenzenes.

Cormier EM, Hunter JE, Billhimer W, May J and Farage MA (1995) Use of clinical and consumer eye irritation data to evaluate the low-volume eye test. J Toxicol Cut Ocul Toxicol 14, 197-205

Cotović, J., Grandidier, M.H., Portes, P., Roguet, R. & Rubinstenn, G. (2005). The in vitro skin irritation of chemicals: optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process. *ATLA* **33**, 329-349.

Cronin, M.T.D., Basketter, D.A., York, M. (1994). A quantitative structure- activity relationship (QSAR) investigation of a Draize eye irritation database. *Toxicology in Vitro* **8**, 21-28.

Cronin, M.T.D. (1996). The use of cluster significance analysis to identify asymmetric QSAR datasets in toxicology. An example with eye irritation data. *SAR and QSAR in Environmental Research* **5**, 167-175.

Dalton, P. (2003). Upper airway irritation, odor perception and health risk due to airborne chemicals. Toxicol. Lett. 140-141:239-248

Doty, R.L., Cometto-Muñiz, J.E., Jalowayski, A.A., Dalton, P., Kendal-Reed, M., Hodgson, M. 2004. Assessment of upper respiratory tract and ocular irritative effects of volatile chemicals in humans. *Crit. Rev. Toxicol.* **34:**85-142.

Draize, J.H., Woodard, G. & Calvery, H.O. (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. The Journal of Pharmacology and Experimental Therapeutics **82**, 377-390.

EC (2001). Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 29th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances *Official Journal L 225, 21/08/2001 P. 1 – 333,* Office for Official Publications of the European Communities, Luxembourg, 2004.

EC (2003): Technical Guidance Document 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.

EC (2006a) Manual of Decisions for Implementation of the 6th and 7th Amendments to Directive 67/548/EEC on Dangerous Substances. Updated version of July 2004 (EUR 22311). 190pp. Ispra, Italy: European ChemicalsBureau, European Commission JRC. Website http://ecb.jrc.it/classification-labelling.

EC (2006b) Proposal for a Regulation of the European Parliament and of the Council on Classification and Labelling of Substances and Mixtures based on the Globally Harmonised System (31 October 2006) agreed revision 061031, Volume I.

ECETOC (1995) Technical report No.66, Skin irritation and corrosion: reference chemicals data bank.

ECETOC (1997) ECETOC Document No.37, EC Classification of Eye Irritancy.

ECETOC (2002) ECETOC Monograph No.32, Use of Human Data in Hazard Classification for Irritation and Sensitisation.

ECVAM (1998). Statement on the scientific validity of the EPISKIN™ test (an *in vitro* test for skin corrosivity). *ATLA* **26**, 277-280.

ECVAM (2000). Statement on the application of the Epiderm™ human skin model for skin corrosivity testing. *ATLA* **28**, 365-366.

ECVAM (2001). Statement on the application of the CORROSITEX® assay for skin corrosivity testing. *ATLA* **29**, 96-97.

Enslein, K, Borgstedt, H.H., Blake, B.W., Hart, J.B. (1987). Prediction of rabbit skin irritation severity by structure activity relationships. *In Vitro Toxicology* **1**, 129-147.

Enslein, K, Blake, B.W., Tuzzeo, T.M., Borgstedt, H.H., Hart, J.B., Salem, H. (1988). Estimation of rabbit eye irritation scores by structure activity equations. *In vitro Toxicology* **2**, 1-14.

Epstein E. (1974) Letter: Patch test reactions. *Arch Dermatol.* Aug;110(2):299-300.

EU (2000). Annex I to Commission Directive 2000/33/EC adapting to technical progress for the 27th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal of the European Union* **L136**, 91-97.

Fentem, J.H., D. Briggs, C. Chesné, G.R. Elliott, J.W. Harbell, J.R. Heylings, P. Portes, R. Roguet, J.J.M. van de Sandt and P.A. Botham (2001). A prevalidation study on *in vitro* tests for acute skin irritation: results and evaluation by the management Team. *Toxicology in Vitro* 15, 57-93.

Fentem, J. H., Archer, G. E. B., Balls, M., Botham, P. A., Curren, R. D., Earl, L. K., Esdaile, D. J., Holzhütter, H. G. & Liebsch, M. (1998). The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the Management Team. *Toxicology in vitro* **12**, 483-524.

Freeberg FE, Griffith JF, Bruce RD and Bay PHS (1984) Correlation of animal test methods with human experience for household products. *J Toxicol Cut Ocul Toxicol* 1(3), 53-64

Freeberg FE, Hooker DT and Griffith JF (1986a) Correlation of animal eye test data with human experience for household products: an update. *J Toxicol Cut Ocul Toxicol* 5(2), 115-123

Freeberg FE, Nixon GA, Reer PJ, Weaver JE, Bruce RD, Griffith JF and Sanders III, LW (1986b) Human and rabbit eye responses to chemical insult. *Fund Appl Toxicol* 7, 626-634

Frosch, P.J. & Pilz, B. (1995) Irritant patch test techniques. In *Handbook of Non-invasive Methods and the Skin*, pp. 587-591 (Eds. Serup, J. & Jemec, G.B.E.) CRC Press, Boca Raton.

Gallegos Saliner, A., Tsakovska, I., Pavan, M., Patlewicz, G., Worth AP. (2007). Evaluation of SARs for the prediction of skin irritation/corrosion potential - structural inclusion rules in the BfR decision support system. *SAR & QSAR in Environmental Research* 18. *in press*.

Gallegos Saliner, A., Patlewicz, G., Worth, A.P. (2006). Review of literature-based models for skin and eye irritation and corrosion. JRC report EUR 22320 EN. European Chemicals Bureau, Ispra, Italy. Accesible from: http://ecb.jrc.it .

Gerner I., Schlegel K., Walker J.D., Hulzebos, E. (2004). Use of physico-chemical property limits to develop rules for identifying chemical substances with no skin irritation or corrosion potential. *QSAR and Combinatorial Science* **23**, 726-733.

Gerner, I., Liebsch, M., Spielmann, H. (2005). Assessment of the Eye Irritating Properties of Chemicals by Applying Alternatives to the Draize Rabbit Eye Test: The Use of QSARs and *In Vitro* Tests for the Classification of Eye Irritation. *Alternatives to Laboratory Animals* **33**, 215-237.

Grant, W.M. (1974) Toxicology of the eye. 2nd ed. Charles C. Thomas, Springfield, Illinois.

Griffith JF, Nixon GA, Bruce RD, Reer PJ and Bannan EA (1980) Dose-response studies with chemical irritants in the albino rabbit eye as a basis for selecting optimum testing conditions for predicting hazard to the human eye. Toxicol Appl Pharmacol 55, 501-513.

Hawk, J.L.M., Black, A.K., Jaenicke, K.F., Barr, R.M., Soter, N.A., Mallett, A.I., Gilchrest, B.A., Hensby, C.N., Parrish, J.A. & Greaves, M.W. (1983) Increased concentrations of arachidonic acid, prostaglandins E2, D2, and 6-oxo-F1 alpha, and histamine in human skin following UVA irradiation. *Journal of Investigative Dermatology* 80, 496-499.

Hayashi, M., Nakamura, Y., Higashi, K., Kato, H., Kishida, F., Kaneko, H. (1999). A Quantitative Structure-Activity Relationship Study of the Skin Irritation Potential of Phenols. *Toxicology in Vitro* **13**, 915-922.

Heylings, J.R., Diot, S., Esdaile, D.J., Fasano, W.J.; Manning, L.A. & Owen, H.M. (2003) A prevalidation study on the in vitro skin irritation function test (SIFT) for prediction of acute skin irritation in vivo: results and evaluation of ECVAM Phase III. *Toxicol In Vitro*. 2, 23-38.

Hoffmann, S. Cole, T., Hartung, T. (2005). Skin irritation: prevalence, variability, and regulatory classification of existing in vivo data from industrial chemicals. *Regulatory Toxicology and Pharmacology* **41**, 159-166.

Hulzebos, E.M., Janssen, P.A.H., Maślankiewicz, L., Meijerink, M.C.M., Muller, J.J.A., Pelgrom, S.M.G., Verdam, L., Vermeire, T.G. (2001). The application of structure-activity relationships in human hazard assessment: a first approach. RIVM report 601515 008, Bilthoven, The Netherlands.

Hulzebos, E.M., Maslankiewicz, L., Walker, J.D. (2003) Verification of literature derived SARs for skin irritation and corrosion. *QSAR & Combinatorial Science* **22**, 351-363.

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

Hunziker, T., Brand, C.U., Kapp, A., Waelti, E.R. & Braathen, L.R. (1992) Increased levels of inflammatory cytokines in human skin lymph derived from sodium lauryl sulphate-induced contact dermatitis. *British Journal of Dermatology* 127, 254-257.

ICCVAM (2002) OECD Guideline for the Testing of Chemicals. Proposal for a New Guideline. In vitro Membrane Barrier Test System for Skin Corrosion. 16pp. NC, USA: ICCVAM, NIEHS.

ICCVAM: Current Status of In Vitro Test Methods for Detecting Ocular Corrosives and Severe Irritants Background Review Documents (BRDs) March 2006. http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_brd_ice.htm

Kandárová H., Liebsch, M., Spielmann, H.,Genschow, E., Schmidt, E., Traue, D., Guesr, R., Whittingham,A., Warren, N., Gamer, A.O., Remmele, M., Kaufmann, T., Wittmer, E., De Wever, B. & Rosdy, M. (2006) Assessment of the human epidermis model SkinEthic RHE for in vitro skin corrosion testing of chemicals according to new OECD TG 431. Toxicology in vitro 20, 547-559.

Kandárová, H.; Liebsch, M., Gerner, I.,Schmidt, E., Genschow,E., Traue, D and Spielmann, H. (2005) The EpiDerm test protocol for the upcoming ECVAM validation study on in vitro skin irritation tests—an assessment of the performance of the optimised test. *ATLA* **33**, 351-367.

Kligman AM, Epstein W. (1975) Updating the maximization test for identifying contact allergens.

Contact Dermatitis. Aug;1(4), 231-9.

Klimisch HJ, Andreae M, Tillmann U (1997). Systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharm 25, 1-5.

Klopman, G. (1992) MULTICASE.1. A Hierarchical Computer Automated Structure Evaluation Program. *Quantitative Structure-Activity Relationships* **11**, 176-184.

Klopman, G., Ptchelintsev, D., Frierson, M., Pennisi, S., Renskers, K., Dickens, M. (1993). Multiple computer automated structure evaluation methodology as an alternative to *in vitro* eye irritation testing. *Alternatives to Laboratory Animals* **21**, 14-27.

Klopman, G. (1998) The MultiCASE Program II. Baseline Activity Identification Algorithm (BAIA). *Journal of Chemical Information and Computer Sciences* **38**, 78-81.

Kodithala, K., Hopfinger, A.J., Thompson, E.D., Robinson, M.K. (2002). Prediction of skin irritation from organic chemicals using membrane-interaction QSAR Analysis. *Toxicological Sciences* **66**, 336-346.

Kulkarni, A., Hopfinger, A.J., Osborne, R., Bruner, L.H., Thompson, E.D. (2001). Prediction of Eye Irritation from Organic Chemicals Using Membrane-Interaction QSAR Analysis. *Toxicological Sciences* **59**, 335-345.

Lammintausta, K.H. & Maibach, H.I. (1990) Contact dermatitis due to irritation. In *Occupational Skin Disease*, p. 1 (Ed. Adams, R.M.) 2nd edition, W.B. Saunders, Philadelphia.

Liebsch, M., Traue, D., Barrabas, C., Spielmann, H., Uphill, P., Wilkins, S., Wiemann, C., Kaufmann, T., Remmele, M. & Holzhütter, H. G. (2000). The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. *ATLA* **28**, 371-401.

Nangia, A., Anderson, P.H., Berner, B., Maibach, H.I. (1996). High dissociation constants (pKa) of baasic permeants are associated with in vivo skin irritation in man. *Contact Dermatitis* **34**, 237-242.

OECD (2001) OECD Attachment to the Test Guideline 404: A Sequential Testing Strategy for Skin Irritation and Corrosion. OECD, Paris, France.

OECD (2004a). *OECD Guidelines for the Testing of Chemicals No. 430: In vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER).* Paris, France: Organisation for Economic Cooperation and Development.

OECD (2004b) *OECD Guidelines for the Testing of Chemicals No. 431. In vitro Skin Corrosion: Human Skin Model Test.* Paris, France: Organisation for Economic Cooperation and Development.

OECD (2004c) 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 pp.
[http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html]

OECD (2006) *OECD Guidelines for the Testing of Chemicals No. 435:In vitro Membrane Barrier Test Method for Skin Corrosion.*

Patlewicz, G.Y., Rodford, R.A., Ellis, G., Barratt, M.D. (2000). A QSAR Model for the Eye Irritation of Cationic Surfactants. *Toxicology in Vitro* **14**, 79-84.

Patlewicz, G., Rodford, R., Walker, J.D. (2003). Quantitative structure-activity relationships for predicting skin and eye irritation. *Environmental Toxicology and Chemistry* **22**, 1862-1869.

Roggeband R, York M, Pericoi M and Braun W (2000) Eye irritation responses in rabbit and man after single applications of equal volumes of undiluted model liquid detergent products. Food Chem Toxicol, 38, 727-734

Rorije, E.& Hulzebos, E. (2005). Evaluation of (Q)SARs for the prediction of Skin Irritation/Corrosion Potential. Physico-chemical exclusion rules. Final report for ECB contract IHCP.B430206. European Chemicals Bureau, Joint Research Centre, Ispra, Italy. Accesible from: http://ecb.jrc.it .

Rorije, E., Herzler, M., Hulzebos, E., Gerner, I., Hakkert, B., Walker, J.D. (2007). Evaluation of structural alerts for skin irritation or corrosion. Alerts for unspecific activity. *SAR & QSAR in Environmental Research*, submitted.

Rosenkranz, H.S., Zhang, Y.P., Klopman, G. (1998). The Development and Characterisation of a Structure-activity Relationship Model of the Draize Eye Irritation Test. *Alternatives to Laboratory Animals* **26**, 779-809.

Sanderson, D.M. & Earnshaw, C.G. (1991) Computer Prediction of Possible Toxic Action from Chemical Structure; The DEREK System. *Human & Experimental Toxicology* **10**, 261-273.

Serup, J. (1995) The spectrum of irritancy and application of bioengineering techniques. In *Irritant Dermatitis. New Clinical and Experimental Aspects. Current Problems in Dermatology*, vol.23, pp.131-143 (Eds. Elsner, P. & Maibach, H.I.) Karger, Basel.

Smith, J.S., Macina, O.T., Sussman, N.B., Luster, M.I., Karol, M.H. (2000a). A robust structure-activity relationship (SAR) model for esters that cause skin irritation in humans. *Toxicological Sciences* **55**, 215-222.

Smith, J.S., Macina, O.T., Sussman, N.B., Karol, M.H., Maibach, H.I. (2000b). Experimental validation of a structure-activity relationship model of skin irritation by esters. *Quantitative Structure-Activity Relationships* **19**, 467–474.

Smithing, M.P.& Darvas, F. (1992) HazardExpert: An Expert System for Predicting Chemical Toxicity. In *Food Safety Assessment*. Finlay, S.F., Robinson, S.F., Armstrong, D.J. Eds. American Chemical Society, Washington, DC, pp. 191-200

Sondergard, J., Graevers, M.W. & Jorgenson, H.P. (1974) Recovery of prostaglandins in human primary irritant dermatitis. *Archives of Dermatology* 110**,** 556-557.

Sugai, S., Murata, K., Kitagaki, T., Tomita, I. (1990) Studies on eye irritation caused by chemicals in rabbits—1. A quantitative structure-activity relationships approach to primary eye irritation of chemicals in rabbits. *The Journal of Toxicological Sciences* **15**, 245-262.

Sugai, S., Murata, K., Kitagaki, T., Tomita, I. (1991). Studies on eye irritation caused by chemicals in rabbits—II. Structure-activity relationships and in vitro approach to primary eye irritation of salicylates in rabbits. *Journal of Toxicological Sciences* **16**, 111-130.

Tsakovska, I., Netzeva, T., Worth AP. (2005). Evaluation of (Q)SARs for the prediction of Eye Irritation/Corrosion Potential -physico-chemical exclusion rules. JRC Report EUR 21897 EN, 42pp. European Chemicals Bureau, Ispra, Italy. Accesible from: http://ecb.jrc.it.

Tsakovska, I., Gallegos Saliner, A., Netzeva, M., Pavan, M., Worth AP. (2007). Evaluation of SARs for the prediction of eye irritation/corrosion potential - structural inclusion rules in the BfR decision support system. *SAR & QSAR in Environmental Research* 18, *in press*.

United Nations-Economic Commission for Europe (UN/ECE) (2003). Globally Harmonised System of Classification and Labelling of Chemicals (GHS). New York and Geneva: UN.

U.S. DOT (2002) Exemption DOT-E-10904 (fifth revision). September 20, 2002. Washington, D.C., U.S. DOT. Available at: [http://hazmat.dot.gov/exemptions/E10904.pdf]

Van Goethem F, Adriaens E, Alepee N, Straube F, De Wever B, Cappadoro M, Catoire S, Hansen E, Wolf A, Vanparys P. (2006) Prevalidation of a new in vitro reconstituted human cornea model to assess the eye irritating potential of chemicals. Toxicol In Vitro, 20(1):1-17

Walker, J.D., Gerner, I., Hulzebos, E., Schlegel, K. (2004) (Q)SARs for Predicting Skin Irritation and Corrosion: Mechanisms, Transparency and Applicability of Predictions. *QSAR & Combinatorial Science* **23**, 721-725.

Walker, J.D., Gerner, I., Hulzebos, E., Schlegel, K. (2005) The Skin Irritation Corrosion Rules Estimation Tool (SICRET). *QSAR & Combinatorial Science* **24**, 378-384.

Worth, A.P. (2000). The integrated use of physico-chemical and in vitro data for predicting chemical toxicity. PhD thesis. Liverpool John Moores University, Liverpool, UK.

Worth, A.P.& Cronin, M.T.D. (2000). Embedded cluster modelling: a novel quantitative structure-activity relationhip for generating elliptic models of biological activity. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation*. Balls M, van Zeller A-M, Halder ME, eds. Elsevier Science, Amsterdam, The Netherlands, pp. 479-491.

Worth, A.P.& Cronin, M.T.D. (2001). The use of pH measurements to predict the potential of chemicals to cause acute dermal and ocular toxicity. *Toxicology* **169**, 119-131.

Worth, A.P. & Cronin, M.T.D. (2003) The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. *Journal of Molecular Structure (Theochem)* **622**, 97-111.

Worth, A.P. (2004). The tiered approach to toxicity assessment based on the integrated use of alternative (non-animal) tests. In *Predicting Chemical Toxicity and Fate*. Cronin MTD, Livingstone DJ, eds. CRC Press, Boca Raton, FL, USA, pp. 389-410.

Worth, A.P., Bassan, A., Gallegos, A., Netzeva, T.I., 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.

Young, J.R., How, M.J., Walker, A.P., Worth W.M.H. (1988). Classification as corrosive or irritant to skin of preparations containing acidic or alkaline substances, without testing on animals, *Toxicology In Vitro* **2**, 19-26.

Young J.R. and How M.J. (1994) Product Classification as Corrosive or Irritant by measuring pH and Acid/Alkali Reserve. Altrenative Methods in Toxicology, Vol 10, 23-27. Published by Mary Ann Liebert, Inc.

Zuang, V., M. Balls, P.A. Botham, A. Coquette, E. Corsini, R.D. Curren, G.R. Elliott, J.H. Fentem, J.R. Heylings, M. Liebsch, J. Medina, R. Roguet, J. van de de Sandt, C. Wiemann and A. P. Worth (2002). Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation. ECVAM skin irritation task force report 2. *Alternatives To Laboratory Animals* 30, 109-129.

R.7.3 Skin and respiratory sensitisation

R.7.3.1 Introduction

A number of diseases are recognised as being, or presumed to be, allergic in nature. These include asthma, rhinitis, conjunctivitis, allergic contact dermatitis, urticaria and food allergies (the latter is not discussed in this document). In this Section, the endpoints discussed are those traditionally associated with occupational and consumer exposure to chemicals (proteins are not discussed in this document). Photosensitisation is potentially important but its mechanism of action is poorly understood, and it is not discussed in this document.

R.7.3.1.1 Definition of skin and respiratory sensitisation

A sensitizer is an agent that is able to cause an allergic response in susceptible individuals. The consequence of this is that following subsequent exposure via the skin the characteristic adverse health effects of allergic contact dermatitis or atopic dermatitis may be provoked. After inhalation exposure, adverse health effects include asthma (and related respiratory symptoms such as rhinitis) or extrinsic allergic alveolitis.

Respiratory hypersensitivity is a term that is used to describe asthma and other related respiratory conditions, irrespective of the mechanism (immunological or non-immunological) by which they are caused. In contrast, dermal allergy is based on an immunological mechanism.

It is perhaps helpful to attempt to define the term chemical respiratory hypersensitivity. One approach taken by the UK Health and Safety Executive was to describe the induction phase as the process of rendering the airways unusually sensitive (hypersensitive) such that following subsequent inhalation exposure an asthmatic reaction might be elicited associated with classical symptoms of airway narrowing, chest-tightening and bronchial restriction (HSE, 1997). Other approaches to definition of relevant terms are available elsewhere. For instance, various definitions are provided for specific sensitising agents in the workplace – all of which imply a mechanism whereby hypersensitivity of the respiratory tract is induced as the result of workplace exposure – and that this may result later in the development of occupational asthma (Bernstein et al., 1993). Lists of chemicals cited here, by the HSE, and elsewhere, as causes of respiratory sensitisation and occupational asthma are very similar, and in some instances identical (Chan-Yeung et al., 1993). Among the chemicals populating these lists are: diisocyanates, acid anhydrides, certain platinum salts, some reactive dyes, cyanuric chloride, and plicatic acid (from Western Red Cedar).

When directly considering human data in this document, the clinical diagnostic terms asthma, rhinitis and extrinsic allergic alveolitis have been retained.

These definitions are reflected in the criteria for the classification of skin and respiratory sensitizers, which provide a useful tool against which the hazardous properties of a substance can be judged. These criteria are given in the 22nd Adaptation to Technical Progress to Directive 67/548/EEC [Directive 96/54/EC, Official Journal L248; pp 227-229]; Annex VI has been recast in the 28th Adaptation to Technical Progress (ATP) (Directive 2001/59, Official Journal L225; pp 1- 333).

R.7.3.1.2 Objective of the guidance on skin and respiratory sensitisation

The general objectives are to determine:

- whether there are (Q)SAR data, existing *in vitro* or *in vivo* data, or human evidence indicating that the agent has skin or respiratory sensitisation potential
- whether the agent has skin sensitisation potential based on new tests according to the strategy as presented in this document.

Therefore, in the sections on skin sensitisation and respiratory sensitisation firstly an overview of types of data is given that may provide information on sensitisation, followed by guidance on the process of judging the available data in terms of adequacy, completeness and remaining uncertainty. In Section R.7.3.7 guidance is given on application of the data to reach a conclusion on suitability for classification and labelling and possibly potency. Finally in Section R.7.3.8 an integrated testing strategy (ITS) for skin sensitisation and an integrated evaluation strategy (IES) for respiratory sensitisation is presented.

R.7.3.1.3 Mechanisms of immunologically-mediated hypersensitivity

Among the key steps required for a chemical to induce sensitisation via skin contact are gaining access to the viable epidermis, protein binding, metabolic activation (if required), internalization and processing by Langerhans cells (LC), transport of antigen by LC to draining lymph nodes, and presentation to and recognition by T lymphocytes. For chemicals that sensitise via the respiratory tract, the relevant mechanisms are believed to be essentially similar, although gaining access to the respiratory epithelium may be somewhat easier than at skin surfaces due to the lack of a stratum corneum. Moreover, because the lining of the respiratory tract, the professional antigen presenting cells, and regulatory mechanisms in the respiratory tract differ from those in the skin, they all may have an impact on the type of immune response evoked. Although the site of induction of an adaptive immune response to a chemical allergen may be influenced by local conditions and local immunoregulatory mechanisms, the fact remains that the inherent properties of the chemical itself play a major role in determining whether an immune responses is induced and the qualitative characteristics of that response.

Although it is sometimes assumed that immune responses induced following encounter with antigen in or on the skin are often of selective Th_1-type, this is not necessarily the case. It is clear that cutaneous immune responses can be of either Th_1- or Th_2-type according to the nature of the antigen.

In the respiratory tract, chemical respiratory allergens appear to preferentially elicit Th_2-immune responses (Maestrelli et al., 1997); observations that are consistent with experimental experience in mice (Dearman et al., 2002; Herrick et al., 2003; Farraj et al., 2004), and possibly also rats (Arts et al., 1998). Th2 type immune responses are characterised by the production of cytokines such as IL4 and IL5 and by the production of IgE antibodies. However, the mechanisms through which chemicals are able to induce sensitisation of the respiratory tract are not fully understood and there remains controversy about the roles played by IgE antibody-mediated mechanisms, and whether IgE represents a mandatory universal requirement for the induction by chemicals of allergic sensitisation of the respiratory tract. The area is complicated because although for all chemical respiratory allergens there are patients who display serum IgE antibodies of the appropriate specificity, in other instances (and particularly with respect to the diisocyanates) there are symptomatic subjects in whom it is not possible to detect IgE antibody. There are two, non-mutually exclusive, possibilities. The first is that IgE does play a central role but that for one or

more of various reasons it is not being detected accurately in the serum of patients with occupational asthma. The second is that allergic sensitisation of the respiratory tract by chemicals can be effected through IgE antibody-independent immunological mechanisms (Kimber et al., 2002 and 2005). These may also include Th_1-type immune responses. In this context it has been reported, for instance, that inhalation challenge of sensitised rodents with contact allergens may elicit respiratory allergic reactions (Garssen et al., 1991; Garcia et al., 1992; Buckley et al., 1994; Zwart et al., 1994; Satoh et al., 1995; Arts et al., 1998). This comes as no surprise because it is clear that contact sensitisation is systemic in nature and that there is no reason to suppose that encounter of sensitised animals with the relevant contact allergen at respiratory epithelial surfaces will not cause an adverse immunologic reaction. However, it is important to note that in reality only a very few precedents for the elicitation of pulmonary reactions by skin sensitising chemicals in humans have been observed, and in practice it may not represent a significant health issue.

In addition, there is a growing body of evidence that effective sensitisation of the respiratory tract by chemicals defined as respiratory allergens (such as for instance the acid anhydrides, diisocyanates and others) can and does occur in response to dermal contact (reviewed by Kimber et al., 2002). There are also experimental animal data and human evidence for sensitisation by inhalation and skin effects following dermal challenge (Kimber et al., 2002, Baur et al., 1984, Ebino et al., 2001, Stadler et al., 1984). Therefore, it is not necessarily the case that chemicals that cause allergic dermal reactions require sensitisation via the skin, or that chemicals that cause allergic airway reactions require sensitisation via the respiratory tract.

R.7.3.2 Information requirements for skin and respiratory sensitisation

The information requirements for sensitisation are described in REACH Annexes VI to XI, where the information that shall be submitted for registration purposes is specified.

Column 1 of Annex VII clearly informs on the standard information requirement for skin sensitisation data for substances produced or imported in quantities of ≥1 t/y.

The assessment of skin sensitisation shall comprise the following consecutive steps:

1. an assessment of the available human, animal and alternative data,

2. In vivo testing

Column 2 of Annex VII lists specific rules according to which the required standard information may be omitted, replaced by other information, or adapted in another way. If the conditions are met under which column 2 of this Annex allows adaptations, the fact and the reasons for each adaptation should be clearly indicated in the registration. For skin sensitisation column 2 reads:

Step 2 does not need to be conducted if:

- *the available information indicates that the substance should be classified for skin sensitisation or corrosivity; or*
- *the substance is a strong acid (pH<2.0) or base (pH>11.5); or*
- *the substance is flammable in air at room temperature.*

The Murine Local Lymph Node Assay (LLNA) is the first-choice method for in vivo testing. Only in exceptional circumstances should another test be used. Justification for the use of another test

shall be provided. This means that in certain cases other in vivo methods may be conducted. In such cases convincing scientific justification for the use of another test shall be provided.

No information requirements are present for respiratory sensitisation. Respiratory sensitizers are indicated for harmonised classification and labelling in REACH Article 115, and respiratory sensitisation is mentioned in Annex I and XV which deal with respectively chemical safety report and preparation of these dossiers.

In addition to these specific rules, the required standard information set may be adapted according to the general rules contained in Annex XI. In this case as well, the fact and the reasons for each adaptation should be clearly indicated in the registration.

General requirements for generation of information on intrinsic properties of substances are given in REACH Article 13 which states that this information may be generated by means other than tests, provided the conditions specified in Annex XI are met.

R.7.3.3 Information for skin sensitisation and its sources

R.7.3.3.1 Non-human data for skin sensitisation

Non-testing data for skin sensitisation

Non-testing methods for skin sensitisation cover a breadth of different approaches namely read-across/chemical categories, chemistry considerations and (Q)SARs. Read-across/chemical categories are described in Sections R.6.1 and R.6.2.

A compendium of available (Q)SARs is not in existence at the present time, work is being carried out by ECB to develop an inventory of evaluated (Q)SARs which will populate the (Q)SAR Application Toolbox, a larger project currently led by the OECD. The JRC QSAR Model Database is being designed to help a user determine the validity and applicability of a model for a specific chemical and purpose. This is relevant to the assessment of adequacy. The OECD principles (described on Website http://www.oecd.org/document/23) will help to characterise the validity of a given model. Preliminary practical guidance on their interpretation has been developed (Worth et al., 2005). Evaluated (Q)SARs will be documented in (Q)SAR Reporting Formats (see Section R.6.1.9). More generic information on evaluating QSARs, their predictions and reporting formats is provided in Section R.6.1.6.

Exploring the reaction chemistry of compounds forms the basis of most read-across justifications and many of the available skin sensitisation (Q)SARs. The skin sensitisation potential of a chemical is related to its ability to react with skin proteins to form covalently linked conjugates and recognition of these by the immune system. In the vast majority of cases, this is dependent on electrophilic reactivity of the skin sensitizer or a derivative produced (usually by oxidation) *in vivo* or abiotically (Barratt et al., 1997). There are various types of electrophile-nucleophile reactions in skin sensitisation, perhaps the most frequently encountered are: Michael-type reactions; S_N2 reactions; S_NAr reactions; acylation reactions and Schiff-base formation. These chemical reaction mechanisms can serve as a means of describing the domain of applicability (the scope) of a (Q)SAR or form the basis for grouping chemicals into chemical categories. Recent work in this area has been described in (Aptula et al., 2005, Aptula and Roberts 2006, Roberts et al., 2007).

There are relatively few (Q)SARs for skin sensitisation reported in the peer reviewed literature. Available models include local and global (Q)SARs as well as expert systems.

Local (Q)SAR models

The majority of local models available have been developed for direct-acting electrophiles using the relative alkylation index (RAI) approach. This is a mathematical model derived by Roberts and Williams (1982). It is based on the concept that the degree of sensitisation produced at induction, and the magnitude of the sensitisation response at challenge, depends on the degree of covalent binding (haptenation; alkylation) to carrier protein occurring at induction and challenge. The RAI is an index of the relative degree of carrier protein haptenation and was derived from differential equations modelling competition between the carrier haptenation reaction in a hydrophobic environment and removal of the sensitizer through partitioning into polar lymphatic fluid. In its most general form the RAI is expressed as:

$$RAI = \log D + a \log k + b \log P \quad (1)$$

Thus the degree of haptenation increases with increasing dose D of sensitizer, with increasing reactivity (as quantified by the rate constant or relative rate constant k for the reaction of the sensitizer with a model nucleophile) and with increasing hydrophobicity (as quantified by log P, P being the octanol/water partition coefficient). This RAI model has been used to evaluate a wide range of different datasets of skin sensitising chemicals. Examples include sulfonate esters (Roberst and Basketter 2000), sulfones (Roberts and Williams 1982), primary alkyl bromides (Basketter et al., 1992), acrylates (Roberts 1987), aldehydes and diketones (Patlewicz et al., 2001, Patlewicz et al., 2002, Patlewicz et al., 2004, Roberts et al., 1999, Roberts and Patlewicz 2002, Patlewicz et al., 2003).

This approach has been shown to be mechanistically robust but the breadth of available models so far is still somewhat limited. These types of models assume a reasonable appreciation of chemistry.

The covalent hypothesis has served and continues to be the most promising way of developing mechanistically based robust QSARs. These are local in that their scope is characterised by a mechanistic reactivity domain as outlined in Aptula et al., 2005, Aptula and Roberts 2006, Roberts et al., 2007. An example of this type of mechanistic model has been recently published (Roberts et al., 2006). In the RAI model, log$k$, has been typically modelled by experimental rate constants, substituents' constants or molecular orbital parameters. More effort is needed to encode reactivity into descriptors, this could be achieved through the systematic generation of *in vitro* reactivity data as outlined in (Aptula and Roberts 2006, Aptula et al., 2006b, Schultz et al., 2006, Gerberick et al., 2004) and in the next section.

Global statistical models

Global Statistical models usually involve the development of empirical QSARs by application of statistical methods to sets of biological data and structural descriptors.

These are perceived to have the advantage of being able to make predictions for a wider range of chemicals. In some cases, the scope/domain of these models are well described, in most other cases a degree of judgement is required in determining whether the training set of the model is relevant for the chemical of interest. Criticism often levied at these types of models is that they lack mechanistic interpretability. The descriptors might appear to lack physical meaning or are difficult to interpret from a chemistry perspective. The sorts of descriptors used may encode chemical reactivity/electrophilicity e.g. LUMO (the energy of the lowest molecular orbital) and partitioning effects e.g. Log P, but more commonplace is that a large number of descriptors are calculated that encode structural, topological and/or geometrical information. A number have been reported in the recent literature, examples include those developed using LLNA data (Devillers 2000, Estrada et

al., 2003, Fedorowicz et al., 2005, Fedorowicz et al., 2005, Li et al., 2005, Miller et al., 2005, Ren et al., 2006, Li et al., 2007).

Expert systems

There are several commercial (Q)SAR models for skin sensitisation available. Examples include TOPKAT, CASE, Derek for Windows and TIMES.

Statistical Models

TOPKAT (current version 6.2) marketed by Accelrys Inc (San Diego, USA) comprises two suites of models; one for aromatics (excluding chemicals with 1 benzene ring) and the other for aliphatics and chemicals with 1 benzene ring. The first set of models discriminate between non-sensitizers and sensitizers, a probability is calculated for the submitted chemical structure. If the probability is greater than or equal to 0.7, the chemical is predicted to be a sensitizer, a non-sensitizer would have a probability of less or equal to 0.30. The second set of models resolve the potency: weak/moderate vs. strong where a probability of 0.7 or more indicates a strong sensitizer and a probability below 0.30 indicates a weak or moderate sensitizer. Probability values between 0.30 and 0.70 are referred to as indeterminate. An optimum prediction space algorithm ensures that predictions are only made for chemicals within the model applicability domain (Enslein et al., 1997, http://www.accelrys.com/products/topkat/).

CASE methodology and all its variants were developed by Klopman and Rosenkranz. There are a multitude of models for a variety of endpoints and hardware platforms. The CASE approach uses a probability assessment to determine whether a structural fragment is associated with toxicity (Cronin et al., 2003). The MCASE models that have been developed for skin sensitisation are described further in primary articles (Gealy et al., 1996, Graham et al., 1996, Johnson et al., 1997). There are two sensitisation modules available for purchase from MultiCase Inc (Ohio, USA) (http://www.multicase.com/products/prod0911.htm). In addition the (Q)SAR estimates for one MCASE skin sensitisation model are included in the Danish Environmental Protection Agency (EPA) (Q)SAR database which is currently hosted on the European Chemicals Bureau (ECB) website http://ecb.jrc.it/QSAR/ .

Knowledge based systems

Derek for Windows (DfW) 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. It is marketed and developed by LHASA Ltd (Leeds, UK) a not-for-profit company and educational charity (http://www.lhasalimited.org/index.php).

Within DfW (version 9), there are 361 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 alongside associated literature references, comments and examples. The skin sensitisation knowledge base in DfW was initially developed in collaboration with Unilever in 1993 using its historical database of guinea pig maximisation test (GPMT) data for 294 chemicals and contained approximately forty alerts (Barratt et al., 1994). Since that time, the knowledge base has undergone extensive improvements as more data have become available (Payne and Walsh 1994). The current version (version 9) contains seventy alerts for skin sensitisation and the closely-related endpoint of photoallergenicity (Barratt et al., 2000, Langton et al., 2006).

Hybrids

TIssue MEtabolism Simulator (TIMES) software has been developed to integrate a skin metabolism simulator with 3D-QSARs for evaluating reactivity of chemicals in order to predict

their skin sensitisation potency (Dimitrov et al., 2005, Dimitrov et al., 2005). The simulator contains 236 hierarchically ordered spontaneous and enzyme controlled reactions. Covalent interactions of chemicals/metabolites with skin proteins are described by 47 alerting groups. 3D-QSARs (COREPA) are applied for some of these alerting groups.

Clearly there are a breadth of different (Q)SARs and expert systems available for the estimation of skin sensitisation hazard. The approaches are quite varied and each has been developed on different sets of *in vivo* data (principally GPMT and LLNA). Whilst efforts have been made to characterise a number of the literature based models in terms of the OECD principles for QSAR validation (see Roberts et al., 2007 as an example), further work is still required for some of the commercial systems (ECETOC 2003). In addition, in many cases these models have been demonstrated to be reasonable for predicting skin sensitizers correctly but are limited in predicting non-sensitizers correctly (Roberts et al., 2007, ECETOC 2003). For this reason, careful interpretation of model predictions needs to be considered in light of other information e.g. analogue read-across (other similar chemicals with respect to their mechanistic domain).

Further work should explore encoding more knowledge/rules for non-reactive chemicals as well as those chemicals likely to undergo chemical or metabolic transformation.

Consideration of which model(s) to apply will be dependent on the specific chemical of interest, the underlying training set data and the applicability domain. These issues are described more fully in Section R.6.1. An example is illustrated here; if the chemical falls into a chemistry reactivity domain that is well characterised, then a local (Q)SAR model developed for this domain (such as those previously described) will give rise to the most robust prediction of skin sensitisation. Where the mechanism is not understood or not known *a priori* one or more of the expert systems such as TOPKAT, Derek for Windows or the others already described will be best placed to provide an estimate. These systems whilst not wholly transparent do provide a reasonable amount of supporting information to enable the robustness of a prediction to be evaluated. This is discussed in more detail in Section R.7.3.4.1.

Testing data for skin sensitisation

In vitro data

At present, no officially adopted EU-OECD *in vitro* tests for skin sensitisation exist. However, several systems are in the course of development (Eskes et al., 2005), based on an improved understanding of the biochemical and immunological mechanisms underlying the process (Worth et al., 2002). Currently, *in vitro* assays to detect the sensitising properties of a chemical are under development for the following areas:

- **Epidermal bioavailability**: skin penetration is a prerequisite for skin sensitisation. Information about the skin penetration properties can help to evaluate the potential of a chemical to be identified as a skin sensitizer (ECVAM, 2007).

- **Chemical reactivity**: since the majority of chemical allergens is electrophilic and reacts with nucleophilic amino acids, peptide reactivity assays can give an indication of skin sensitisation potency or potential to form a complete antigen (Gerberick et al., 2004, Aptula et al., 2006b).

- **Cell-based assays**: the knowledge that changes occur in epidermal Langerhans cells as a result of exposure to chemical allergens (e.g. the expression of surface markers and/or cytokines release) and that Langerhans cells can be replaced by blood derived dendritic-like cells or cell lines have been applied to design *in vitro* alternative tests

(Kimber et al., 2001, Tuschl et al., 2000, Casati et al., 2005, Ryan et al., 2005, Sakaguchi et al., 2006, Aeby et al., 2004, Azam et al., 2006, Python et al., 2007). These systems have been shown to selectively express various mediators and/or markers of activation following exposure to chemical sensitizers and attempts to develop robust assays have started. Beside Langerhans cells, keratinocytes play a prominent role in the sensitisation process (Corsini et al., 1998, van Och et al., 2005, Vandebriel et al., 2005). In addition to chemical processing, LC activation requires the binding of cytokines produced by keratinocytes as a result of initial chemical exposure. Moreover the assessment of keratinocytes cytokine expression as a function of the ability of chemicals to induce cutaneous sensitisation is also the object of several investigations (Aiba et al., 2000, Herouet et al., 2000). Keratinocytes have been tested both in primary cultures, in co-culture with dendritic cells and as reconstituted epidermis (Casati et al., 2005, Kubilus et al., 1986, Coquette et al., 2003). The use of reconstituted skin models for the assessment of contact allergens is under investigation.

Owing to the complexity of the mechanisms of skin sensitisation, a single test will probably not be able to replace the currently required animal procedures. Efforts are still needed to identify the most relevant endpoints in the optimisation of existing tests. However, a combination of several *in vitro* tests, covering the relevant mechanistic steps of skin sensitisation, into a test battery could possibly lead to replacement of *in vivo* tests (Eskes et al., 2005). How the outputs from these tests could be combined is not as yet determined, although a general strategy has been presented (Jowsey et al., 2006). Until that date, *in vitro* tests may be used as supportive evidence in combination with other types of data for the identification of allergens (see Section R.7.3.8.3 for an ITS based on a WoE approach).

Animal data

Guideline-compliant tests

For new *in vivo* testing of skin sensitisation potential, the murine local lymph node assay (LLNA) is the REACH Annex VII-endorsed method. This assay has been validated internationally and has been shown to have clear animal welfare benefits and scientific advantages compared with the guinea pig tests described below. The LLNA is designed to detect the potential of chemicals to induce sensitisation as a function of lymphocyte proliferative responses induced in regional lymph nodes. This method is described in OECD TG 429/EU B.42.

Two further animal test methods for skin sensitisation are described in OECD TG 406/EU B.6: the guinea pig maximisation test (GPMT) and the Buehler test. The GPMT is an adjuvant-type test in which the acquisition of sensitisation is potentiated by the use of Freund's Complete Adjuvant (FCA) and in which both intradermal and topical exposure are used during the induction phase. The Buehler test is a non-adjuvant method involving for the induction phase topical application only.

Both the GPMT and the Buehler test are able to detect chemicals with moderate to strong sensitisation potential, as well as those with relatively weak sensitisation potential. In such methods activity is measured as a function of challenge-induced dermal hypersensitivity reactions elicited in test animals compared with controls. Since the LLNA is the preferred method for new *in vivo* testing, the use of the standard guinea pig tests to obtain new data on skin sensitisation potential will be acceptable only in exceptional circumstances and will require scientific justification. However, existing data of good quality deriving from such tests will be acceptable and will, if providing clear results, preclude the need for further *in vivo* testing.

ECETOC Monograph 29 (2000) contains a useful discussion of these tests.

Non-guideline compliant tests and refinements to the standard assays

Existing data may be available from tests that do not have an OECD guideline, for example:

i. other guinea pig skin sensitisation test methods (such as the Draize test, optimisation test, split adjuvant test, open epicutaneous test);

ii. additional tests (such as the mouse ear swelling test);

Information may also be available from other endpoints, for example, repeated dose dermal studies that show effects indicative of an allergic response, such as persistent erythema and/or oedema.

For new testing, refinements to the existing guideline methods may also be possible. In such cases, care should be taken to ensure that any modifications or deviations from standard methodologies are scientifically justified. For example, it might be feasible to conduct a reduced version of the LLNA (rLLNA) in which assessments are made on the basis of results from a vehicle control and a single (highest) concentration of the test substance (Eskes et al., 2005). In such cases, it is recommended that expert advice be sought before commencing the tests.

R.7.3.3.2 Human data on skin sensitisation

Human data on cutaneous (allergic contact dermatitis and urticarial) reactions may come from a variety of sources:

- consumer experience and comments, preferably followed up by professionals (e.g. diagnostic patch tests)
- diagnostic clinical studies (e.g. patch tests, repeated open application tests)
- records of workers' experience, accidents, and exposure studies including medical surveillance
- case reports in the general scientific and medical literature
- consumer tests (monitoring by questionnaire and/or medical surveillance)
- epidemiological studies
- human experimental studies such as the human repeat insult patch test (Stotts, 1980) and the human maximisation test (Kligman, 1966), although it should be noted that *new* experimental testing for hazard identification in humans, including HRIPT and HMT, is not acceptable for ethical reasons.

R.7.3.4 Evaluation of available information on skin sensitisation

For both steps of the effects assessment, i.e. hazard identification and dose (concentration)-response (effect) assessment, it is very important to evaluate the data with regard to their adequacy and completeness. The evaluation of adequacy shall address the reliability and relevance of the data. The completeness of the data refers to the conclusion on the comparison between the available adequate information and the information that is required under the REACH proposal for the applicable tonnage level of the substance. Such a conclusion relies on WoE approaches, mentioned in REACH Annex XI Section 1.2, which categorise available information based on the methods used: *guideline tests*, *non-guideline tests*, and other types of information which may justify

adaptation of the standard testing regime. Such a WoE approach also includes an evaluation of the available data as a whole, i.e. both *over* or *across* endpoints: i.e. for a sensitive evaluation of sensitisation effects, it is necessary to efficiently integrate the information gathered for sensitisation with that obtained from the study of skin and eye irritation (and acute dermal toxicity).

This approach provides a basis to decide whether further information is needed on endpoints for which specific data appear inadequate or not available, or whether the requirements are fulfilled.

For this specific endpoint some additional remarks are made on the adequacy of the various types of data that may be available.

R.7.3.4.1 Non-human data on skin sensitisation

Non-testing data on skin sensitisation

The evaluation and assessment of a chemical using (Q)SARs is dependent on both the chemical of interest and the (Q)SAR model(s) used to make a prediction. Here we attempt to provide some specific advice for skin sensitisation. More general advice on (Q)SARs including evaluation of OECD principles is described in Section R.6.1.3).

One of the first steps to consider is what information already exists on chemicals *similar* to the one of interest. Chemical similarity is a widely used concept in toxicology, and is based on the hypothesis that similar compounds have similar biological activities. This forms the underlying basis for developing (Q)SARs. In the case of skin sensitisation, the most robust means of comparing two or more chemicals is through an evaluation of their likely chemical reactivity. Recent work in this area has been investigating means of encoding reactivity for the different mechanistic domains in form of rules (Aptula and Roberts 2006, Aptula et al., 2006). (Note: This approach might involve the systematic generation of *in vitro* reactivity data for these different mechanistic domains. (see Aptula et al., 2006 as an example) .If the chemical reactivity is not known, or can not be determined through experimentation then a pragmatic means of identifying similar chemicals can be done through a substructural/analogue search.

There are a number of available computational tools and databases that facilitate the search and retrieval of similar analogues. Some like Leadscope (http://www.leadscope.com) are commercial, others like Chemfinder (www.chemfinder.com), ChemID (http://chem.sis.nlm.nih.gov/chemidplus/) or DssTox (http://www.epa.gov/nheerl/dsstox/) are freely available to use on the internet.

Some of the available search engines are linked to databases (through hyperlinks and indexes) whereas other facilities such as DssTox provide a repository of available QSAR datasets which can be downloaded for subsequent use in appropriate QSAR /database software tools.

Many of currently available tools containing public data have focussed on endpoints such as carcinogenicity, mutagenicity or acute toxicity. This means that an additional search is needed to identify skin sensitisation data. Much of the available skin sensitisation experimental data resides in peer reviewed publications. Cronin and Basketter (1994) published the results of over 270 *in vivo* skin sensitisation tests (mainly from the guinea pig maximisation test). All data were obtained in the same laboratory and represent one of the few occasions when large amounts of information from corporate databases was released into the open literature. A larger database of animal and human studies for 1034 compounds is described by Graham et al. (1996), the MCASE database. A comparatively large number of data have been published for the local lymph node assay, examples include publications by Ashby et al. (1995) and Gerberick et al (2005).

These publications are invaluable to identify analogues with associated skin sensitisation test data.

The second step involves an assessment of the similarity of the analogues identified. Considerations will include whether:

- the same endpoint is considered
- there are any additional functional groups or additional substituents that might influence the reactivity and sensitising behaviour (applicability domain considerations)
- the physico-chemical parameters similar (e.g. LogP, applicability domain considerations)
- there are impurities that influence the sensitisation profile
- the likely chemical mechanism is the same

These considerations may help identify an available local (Q)SAR for that chemical class/mechanistic group.

If an appropriate local model can not be identified then a third step of evaluating a chemical using one of the available global models/expert systems is merited.

Here a prediction needs to be evaluated in the context of the likely chemistry and the available *like* chemicals available within the training set. i.e. is the compound of interest within the scope of the model and are similar chemicals in the training set of the model well predicted. This type of information provides additional weight to whether the estimate derived is meaningful and relevant. For global models available in the literature, the training sets and the algorithm(s) are usually available to allow such comparisons to be made.

For expert systems such as Derek for Windows, TOPKAT etc, the training sets and to an extent the algorithms or descriptors used are often kept latent within the software. Some supporting information is provided on the robustness and relevance for a given prediction. For example, within DfW it is possible to see representative example chemicals and explanations of the mechanistic basis for the SAR developed. Within TOPKAT, it is possible to obtain an assessment of whether the chemicals falls within the applicability domain of the model (both with respect to the fragment and descriptor space), whether it is an example chemical in the database as well as perform a similarity assessment to identify analogues. Similar functionalities and features are present in many of the other commercial expert systems available.

Although the main factors driving skin sensitisation (and therefore the (Q)SARs) is the underlying premise of the electrophilicity of a chemical, other factors such as hydrophobicity encoded in the octanol/water partition coefficient (log P) may also be considered as playing a role in the modifying the sensitisation response observed. Within DfW, an assessment of the likely skin penetration ability is made using the algorithm by Potts and Guy. This relates the Kp value to log P and MW (Potts and Guy 1992). It is then possible to rationalise the output in terms of bands of penetration potential. Some have been described in (Howes et al., 1996).

Specific model and prediction information can be described in more detail in reporting formats ((Q)SAR Reporting Format). This summarises the pertinent information to consider for given model when evaluating an estimate as well as the estimate itself. More details are provided in Section R.6.1.

Other information such as results in other assays such as the Ames test (a common feature of genotoxic substances is that they can bind covalently to DNA and cause direct DNA damage) or

aquatic toxicity tests may provide supporting information about the electrophilicity of the chemical of interest and hence its likely sensitisation ability. Some of this work is still at an early stage but correlations have been explored between mutagens and sensitizers (Wolfreys and Basketter 2004) and between aquatic toxicants and sensitizers (Aptula et al., 2006).

Testing data on skin sensitisation

In vitro data

Even though a number of *in vitro* methods are currently under development, none of these methods has yet undergone a formal validation process. According to Annex XI, *in vitro* data obtained with non-validated methods can only be used in a WoE approach. If such data are considered for the evaluation, expert judgement is needed to assess their reliability. In particular, attention should be paid to the level of optimisation of the method that should meet at least the ECVAM criteria for entering pre-validation (Curren et al., 1995), including evidence of the reproducibility of the method, its mechanistic relevance and predictive capacity (Balls et al., 1995, Hartung et al., 2004, Worth et al., 2001).

In vitro assays only cover a (specific) part of the process of sensitisation that occurs *in vivo*, therefore it is unlikely that a single method will be able to substitute for the animal test.

Animal data

Well reported studies using internationally acceptable protocols, particularly if conducted in accordance with the principles of GLP, can be used for hazard identification. Other studies (see Section R.7.3.3.1 and below), not fully equivalent to OECD test protocols, can, in some circumstances, provide useful information. Particular attention should be paid to the quality of these tests and the use of appropriate positive and negative controls. The specificity and sensitivity of all animal tests should be monitored through the inclusion of appropriate positive and negative controls. In this context, positive controls are the 6-monthly sensitivity checks with an appropriate positive control substance, and negative controls are the vehicle-treated control animals included as part of each test.

Guideline-compliant tests

For the conduct and interpretation of the LLNA the following points should be considered:

i. the vehicle in which the test material and controls have been applied;

ii. the concentrations of test material that have been used;

iii. any evidence for local or systemic toxicity, or skin inflammation resulting from application of the test material;

iv. whether the data are consistent with a biological dose response;

v. the submitting laboratory should be able to demonstrate its competency to conduct the LLNA.

OECD TG 429/EU B.42 provides guidance on the recommended vehicles, number of animals per group, concentrations of test chemical to be applied and substances to be used as a positive control. A preliminary study or evaluation of existing acute toxicity/dermal irritation data is normally conducted to determine the highest concentration of test substance that is soluble in the vehicle but does not cause unacceptable local or systemic toxicity. The submission of historical control data will demonstrate the ability of the test laboratory to produce consistent responses. Based on the use

of radioactive labelling, chemicals that result in a stimulation index (SI) of ≥3 at one or more test concentrations are considered to be positive for skin sensitisation. Both positive and negative responses in the LLNA conducted as described in OECD TG 429/EU B.42 meet the data requirements for classification of a substance as a skin sensitizer: no further testing is required.

Alternative vehicles to those listed in OECD TG 429/EU B.42 may be used in the LLNA if sufficient scientific justification is provided. OECD TG 429/EU B.42 also states that endpoints other than radioactive labelling may be used to assess proliferation, on condition that justification and scientific support, which will include full citations and a description of the methodology, are provided.

The guinea pig test methods described in OECD TG 406/EU B.6, the GPMT (Magnusson et al., 1969, Schlede et al., 1995) and the Buehler, can also be used for hazard identification. Recommendations on conducting and analysing these methods are provided by Steiling et al., 2001. Particular attention should be paid to the quality of these tests with consideration given to the following points:

i. numbers of test and control guinea pigs;

ii. number or percentage of test and control animals displaying skin reactions;

iii. whether skin irritation was observed at the induction phase;

iv. whether the maximal non-irritating concentration was used at the challenge phase;

v. the choice of an appropriate vehicle (ideally, one that solubilises or gives a stable suspension or emulsion of the test material, is free of allergenic potential, is non-irritating, enhances delivery across the stratum corneum, and is relevant to the usage conditions of the test material, although it is recognised that it will not always be possible to meet all these conditions);

vi. whether there are signs of systemic toxicity (a sighting study should be performed to determine an appropriate induction dose that causes irritation but not systemic toxicity);

vii. staining of the skin by the test material that may obscure any skin reactions (other procedures, such as chemical depilation of the reaction site, histopathological examination or the measurement of skin fold thickness may be carried out in such cases);

viii. results of rechallenge treatments if performed;

ix. checking of strain sensitivity at regular intervals by using an appropriate control substance (as specified in OECD guidelines and EU Test Methods). Currently (2007), the recommended interval is 6 months.

The investigation of doubtful reactions in guinea pig tests, particularly those associated with evidence of skin irritation following first challenge, may benefit from rechallenge of the test animals. In cases where reactions may have been masked by staining of the skin, other reliable procedures may be used to assist with interpretation; where such methods are used, the submitting laboratory should provide evidence of their value.

Non-guideline compliant tests and refinements to the standard assays

The submitted dossier should include scientific justification for conducting any new test that is a modification or deviation from guideline methods. In such cases, it would be advisable to seek appropriate expert advice on the suitability of the assay before testing is begun.

For hazard identification, it may be possible to use a reduced LLNA (rLLNA) (Kimber et al., 2006) which reduces the use of animals by requiring only a single (high) dose group (≥10%) and a concurrent negative control group. A preliminary study or evaluation of existing acute toxicity/dermal irritation data is normally conducted to determine the highest concentration of test substance that is soluble in the vehicle, but that does not cause unacceptable local or systemic toxicity. As with the full LLNA, although a concurrent positive control group is not required, registrants would be required to submit historical positive control data supportive of their competence. The rLLNA should be used only in appropriate circumstances:

i. where hazard identification is the primary objective and

ii. where potency data are not required

As in the standard (OECD guideline-compliant) LLNA, group sizes should comprise four or five animals. A positive result in a rLLNA will suffice in circumstances where risk assessment and/or risk management is NOT required. Registrants should be aware that the rLLNA is less scientifically rigorous than the standard LLNA, with an associated increased level of uncertainty.

Historically, guinea pig studies that are not fully equivalent to OECD test protocols have been conducted and can provide useful hazard information. These studies include, but are not limited to, the following: Draize test, optimisation test, split adjuvant test, open epicutaneous test and the cumulative contact enhancement test. In the case of positive results the substance may be considered as a potential skin sensitizer. If, taking into account the above quality criteria, especially the positive and negative control data, there is a clear negative result, i.e. no animals displaying any signs of sensitisation reactions, then no further animal testing is required. Where there is a low level of response, the quality of the study is questionable, or where unacceptably low concentrations of the test material have been used for induction and/or challenge, further testing may be required.

R.7.3.4.2 Human data on skin sensitisation

When reliable and relevant human data are available, they can be useful for hazard identification and even preferable over animal data. However, lack of positive findings in humans does not necessarily overrule positive and good quality animal data.

Well conducted human studies can provide very valuable information on skin sensitisation. However, in some instances (due to lack of information on exposure, a small number of subjects, concomitant exposure to other substances, local or regional differences in patient referral etc) there may be a significant level of uncertainty associated with human data. Moreover, diagnostic tests are carried out to see if an individual is sensitised to a specific agent, and not to determine whether the agent can cause sensitisation.

For evaluation purposes, existing human experience data for skin sensitisation should contain sufficient information about:

- the test protocol used (study design, controls)
- the substance or preparation studied (should be the main, and ideally, the only substance or preparation present which may possess the hazard under investigation)
- the extent of exposure (magnitude, frequency and duration)
- the frequency of effects (versus number of persons exposed)

- the persistence or absence of health effects (objective description and evaluation)
- the presence of confounding factors (e.g. pre-existing dermal health effects, medication; presence of other skin sensitizers)
- the relevance with respect to the group size, statistics, documentation
- the *healthy worker* effect

- Evidence of skin sensitising activity derived from diagnostic testing may reflect the induction of skin sensitisation to that substance or cross-reaction with a chemically very similar substance. In both situations, the normal conclusion would be that this provides positive evidence of the skin sensitising activity of the chemical used in the diagnostic test.

Human experimental studies on skin sensitisation are not normally conducted and are generally discouraged. Where human data are available, then quality criteria and ethical considerations are presented in ECETOC monograph no 32.

Ultimately, where a very large number of individuals (e.g.10^5) have frequent (daily) skin exposure for at least two years and there is an active system in place to pick up complaints and adverse reaction reports (including via dermatology clinics), and where no or only a very few isolated cases of allergic contact dermatitis are observed then the substance is unlikely to be a significant skin sensitizer. However, information from other sources should also be considered in making a judgement on the substance's ability to induce skin sensitisation.

It is emphasised that testing with human volunteers is strongly discouraged, but when there are good quality data already available they should be used as appropriate in well justified cases.

R.7.3.5 Information and its sources on respiratory sensitisation

R.7.3.5.1 Non-human data on respiratory sensitisation

Non-testing data on respiratory sensitisation

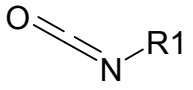
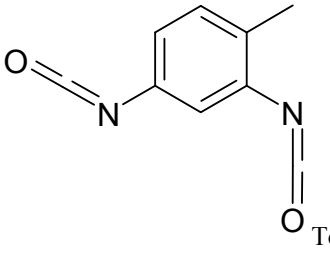
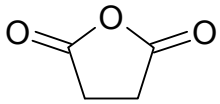
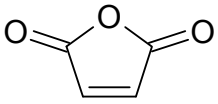
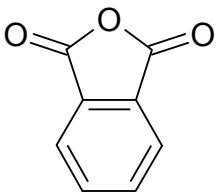
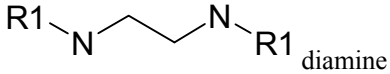
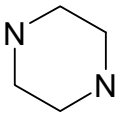
Attempts to model respiratory sensitisation have been hampered by a lack of a predictive test protocol for assessing chemical respiratory sensitisation. (Q)SAR models are available but these have largely been based on data for chemicals reported to cause respiratory hypersensitivity in humans. Examples of some structural alerts are shown in Table R.7.3-1.

Agius et al (1991) made qualitative observations concerning the chemical structure of chemicals causing occupational asthma. This work drew attention to the large proportion of chemical asthmagens with at least two reactive groups, e.g., ethylene diamine and toluene diisocyanate. The earlier work was followed up by a simple statistical analysis of the occurrence of structural fragments associated with activity, with similar conclusions (Agius et al, 1994 and 2000).

The MCASE group has developed three models for respiratory hypersensitivity (Karol et al., 1996, Graham et al., 1997, Cunningham et al., 2005). The Danish (Q)SAR Database has an in-house model for respiratory hypersensitivity for which estimates can be extracted from the on-line database (available at http://ecb.jrc.it/QSAR). Derek for Windows contains several alerts derived from a set of respiratory sensitisers/asthmogens (Payne et al., 1995).

Whilst the available structural alerts (SAR) are transparent and easily to apply (Aigus et al., 1991, 1994 and 2000, Payne et al., 1995), it should be stressed that these are derived on the basis of chemical asthmagens not specifically chemical respiratory allergens. A need therefore remains to develop new (Q)SARs as and when a robust predictive test method becomes available.

Table R.7.3-1 Examples of structural alerts for respiratory sensitisation

Structural Alert Description	Examples of structures
 isocyanate	 Toluene-2,4-diisocyanate
 cyclic anhydride	 maleic anhydride  trimellitic anhydride
 diamine	 piperazine

Testing data for respiratory sensitisation

In vitro data

No *in vitro* tests specific for respiratory sensitisation are available yet, owing to the complexity of the mechanisms of the sensitisation process.

Efforts are still needed to identify the most relevant endpoints in the optimisation of existing tests. However, a combination of several *in vitro* tests, covering the relevant mechanistic steps of respiratory sensitisation, into a test battery could eventually lead to replacement of the *in vivo* tests.

Animal data

At present, although a number of test protocols has been published to detect respiratory allergenicity of low molecular weight compounds, none of these are validated nor are these widely accepted. One approach that might be of some value in characterising the likely respiratory sensitising activity of chemicals is application of the LLNA, or of other tests for measuring skin sensitisation potential. Although the LLNA was developed and validated for the identification of contact allergens, there is evidence that chemical respiratory allergens will also elicit positive responses in this assay (Kimber, 1995). That is, chemicals known to cause respiratory allergy and occupational asthma have been shown to test positive in the LLNA. Among such chemicals are acid anhydrides (such as trimellitic anhydride and phthalic anhydride), diisocyanates (including

diphenylmethane diisocyanate and hexamethylene diisocyanate) and certain reactive dyes. In fact, the view currently is that most, if not all, chemical respiratory allergens are able to elicit positive responses in the LLNA, or in other tests for skin sensitisation, such as the M&K (guinea pig maximisation) test. This is true even of those chemical respiratory allergens, such as phthalic anhydride, for instance, that are implicated virtually exclusively with the induction of chemical respiratory allergy and have rarely, if ever, been shown to cause allergic contact dermatitis. Against this background and in combination with other data it might be possible to conclude in a WoE assessment that chemicals that (at an appropriate test concentration and test conditions, i.e. skin penetration should have occurred) are negative in the LLNA, as well as being considered as not being skin sensitizers, can also be regarded as lacking the potential to cause allergic sensitisation of the respiratory tract.

One approach that has been proposed for the identification of chemicals that have the potential to cause allergic sensitisation of the respiratory tract is one in which activity is measured as a function of the profiles of cytokines produced by draining lymph node cells in mice exposed more chronically (over a 2 week period) to the test chemical (Dearman et al., 2002). This method is predicated on an understanding that allergic sensitisation of the respiratory tract is favoured by selective Th_2-type immune responses and that in many instances chemical respiratory allergy and occupational asthma are associated with IgE antibody. Using this approach chemical respiratory allergens are identified as a function of their ability to stimulate in mice the selective development of preferential Th2-type immune responses associated with a predominance of type 2 cytokine secretion by draining lymph node cells (Dearman et al., 2002 and 2003). Specifically, chemical contact allergens promote Th1 responses characterised by an enhanced production of IFN-gamma, whereas chemical respiratory allergens promote Th_2 responses characterised by enhanced production of IL-4, IL-5 and IL-13. Many variables other than the compound itself, such as concentration used to induce sensitisation, duration of the sensitisation period, and presence or absence of mitogens to reveal differences in cytokine expression have all been noted to have impact on the outcome (Van Och et al., 2002). There are general guidelines now available for the conduct of the method (Dearman et al., 2003), however, this method has not yet been formally validated nor is it widely accepted.

Another, relatively simple approach may serve the purpose to specifically predict sensitisation of the respiratory tract: i.e. increases in total serum IgE antibodies after induction. This method is based on statistically significant increases in total serum IgE (see review by Arts and Kuper, 2007).

Methods that use both an induction and an inhalation elicitation or challenge phase and which include different parameters such as total and/or specific IgE antibody determinations, lung function testing, tests for a specific hyperreactivity (e.g. methacholine challenges), bronchoalveolar lavage measurements, and histopathological examination of the entire respiratory tract, may provide (additional) information on the potential of chemicals to cause respiratory sensitisation. These methods usually use high IgE-responding animal strains; to test for Th1-mediated responses low IgE-responding strains should typically be used. Several of these models have been reviewed recently (Arts and Kuper, 2007).

There are currently no predictive methods to identify chemicals that induce asthma through non-immunological mechanisms, however, when performing challenge tests including non-sensitised but challenged controls information can be obtained on non-immunological effects of these chemicals.

R.7.3.5.2 Human data on respiratory sensitisation

Human data on respiratory reactions (asthma, rhinitis, alveolitis) may come from a variety of sources:

- consumer experience and comments, preferably followed up by professionals (e.g. bronchial provocation tests, skin prick tests and measurements of specific IgE serum levels)
- records of workers' experience, accidents, and exposure studies including medical surveillance
- case reports in the general scientific and medical literature
- consumer tests (monitoring by questionnaire and/or medical surveillance)
- epidemiological studies

R.7.3.6 Evaluation of available information for respiratory sensitisation

R.7.3.6.1 Non-human data for respiratory sensitisation

Non-testing data for respiratory sensitisation

Given the lack of available (Q)SARs for respiratory sensitisation, it is not possible to provide any additional guidance.

Testing data for respiratory sensitisation

In vitro data

Presently (March 2007) there are no *in vitro* tests available to assess respiratory sensitisation. If such a method were to become available then it would need to be assessed for its relevance and reliability (Hartung et al., 2004).

Animal data

Although the LLNA does not represent a method for the specific identification of chemical respiratory allergens, there is evidence that chemical respiratory allergens will also elicit positive responses in this assay (Kimber, 1995). The interpretation is therefore that a chemical which fails to induce a positive response in the LLNA (at an appropriate test concentration) most probably lacks the potential for respiratory allergy. Conversely, it cannot be wholly excluded that a chemical that induces a positive response in the LLNA, might sensitise the respiratory tract upon inhalation or via dermal exposure. Any potential hazard for respiratory sensitisation could only be positively identified by further testing, although such testing is neither validated nor widely accepted.

One further approach to the identification of chemicals that have the potential to induce allergic sensitisation of the respiratory tract is *cytokine fingerprinting* (Dearman et al., 2002; see Section R.7.3.5.1). This method is predicated on an understanding that allergic sensitisation of the respiratory tract is favoured by selective Th_2-type immune responses and that in many instances chemical respiratory allergy and occupational asthma are associated with IgE antibody.

In addition, there are other approaches that have been proposed and these have been reviewed recently (Arts and Kuper, 2007) - although again it is important to emphasise that there are

currently available no fully evaluated or validated animal models for the predictive identification of chemical respiratory allergens.

As indicated previously, some chemicals may have the potential to induce pulmonary reactions via Th1-type immune responses. Studies with typical skin allergens such as DNCB, DNFB and picryl chloride (trinitrochlorobenzene) in BALB/c mice, guinea pigs or Wistar rats have shown the potential of these chemicals to induce allergic reactions in the lungs that are independent of IgE (Garssen et al., 1991; Garcia et al., 1992; Buckley et al., 1994; Zwart et al., 1994; Satoh et al., 1995; and see for a review Arts and Kuper, 2007). Sensitisation and challenge with DNCB resulted in laryngitis in low IgE-responding Wistar rats (Arts et al., 1998). [In addition, cellular immune responses to these sensitizers were shown to be associated with hyperreactivity of the airways to non-specific stimuli (Garssen et al., 1991).] For these reasons, it might be the case that people who are sensitised via the skin might suffer adverse pulmonary reactions if they were to inhale sufficient amounts of the contact allergen to which they were sensitised. As indicated previously, very few precedents for the elicitation of pulmonary reactions by skin sensitising chemicals in humans have been observed. In practice it appears not to represent a health issue.

R.7.3.6.2 Human data for respiratory sensitisation

Although human studies may provide some information on respiratory hypersensitivity, the data are frequently limited and subject to the same constraints as human skin sensitisation data.

For evaluation purposes, existing human experience data for respiratory sensitisation should contain sufficient information about:

- the test protocol used (study design, controls)
- the substance or preparation studied (should be the main, and ideally, the only substance or preparation present which may possess the hazard under investigation)
- the extent of exposure (magnitude, frequency and duration)
- the frequency of effects (versus number of persons exposed)
- the persistence or absence of health effects (objective description and evaluation)
- the presence of confounding factors (e.g. pre-existing respiratory health effects, medication; presence of other respiratory sensitizers)
- the relevance with respect to the group size, statistics, documentation
- the *healthy worker* effect

Evidence of respiratory sensitising activity derived from diagnostic testing may reflect the induction of respiratory sensitisation to that substance or cross-reaction with a chemically very similar substance. In both situations, the normal conclusion would be that this provides positive evidence for the respiratory sensitising activity of the chemical used in the diagnostic test.

For respiratory sensitisation, no clinical test protocols for experimental studies exist but tests may have been conducted for diagnostic purposes, e.g. bronchial provocation test. The test should meet the above general criteria, e.g. be conducted according to a relevant design including appropriate controls, address confounding factors such as medication, smoking or exposure to other substances, etc. Furthermore, the differentiation between the symptoms of respiratory irritancy and allergy can

be very difficult. Thus, expert judgment is required to determine the usefulness of such data for the evaluation on a case-by-case basis.

Although predictive models are under validation, there is as yet no internationally recognized animal method for identification of respiratory sensitisation. Thus human data are usually evidence for hazard identification.

Where there is evidence that significant occupational inhalation exposure to a chemical has not resulted in the development of respiratory allergy, or related symptoms, then it may be possible to draw the conclusion that the chemical lacks the potential for sensitisation of the respiratory tract. Thus, for instance, where there is evidence that a large cohort of subjects have had opportunity for regular inhalation exposure to a chemical for a sustained period of time in the absence of respiratory symptoms, or related health complaints, then this will provide reassurance regarding the absence of a respiratory sensitisation hazard.

R.7.3.7 Conclusions on skin and respiratory sensitisation

The preceding paragraphs on skin and respiratory sensitisation are summarised in the separately provided summary tables. However, it is emphasised that the complete guidance text should be read in order to gain a correct and complete view of the described area.

R.7.3.7.1 Remaining uncertainty on sensitisation

Reliable data can be generated on skin sensitisation from well designed and well conducted studies in animals. The use of adjuvant in the GPMT may lower the threshold for irritation and so lead to false positive reactions, which can therefore complicate interpretation (running a pre-test with FCA treated animals can provide helpful information). In international trials, the LLNA has been shown to be reliable, but like the guinea pig tests is dependent on the vehicle used, and it can occasionally give false positive results with irritants. Careful consideration should be given to circumstances where exposure may be sub-optimal due to difficulties in achieving a good solution and/or a solution of sufficient concentration. In some circumstances inconsistent results from guinea pig studies, or between guinea pig and LLNA studies, might increase the uncertainty of making a correct interpretation. Finally, for existing human data consideration must be given to whether inter-individual variability is such that it is not scientifically sound to generalize from a limited test panel.

When considering whether or not a substance is a respiratory sensitizer, observations of idiosyncratic reactions in only a few individuals with hyper-reactive airways are not sufficient to indicate the need for classification.

Major uncertainties remain in our understanding of the factors that determine whether or not a substance is an allergen, and if so, what makes it a skin or a respiratory sensitizer.

R.7.3.7.2 Concluding on suitability for Classification and Labelling

REACH demands that all available information for a chemical is gathered and any lack of information is reported.

Skin sensitizers

Standard information required for skin sensitisation is described in Annex VII of REACH, i.e. for any substance manufactured or imported in quantity of 1 ton or more.

A substance can be classified as *skin sensitizer* following the flow chart for integrated testing strategy (ITS) reported in Figure R.7.3-1 in Section R.7.3.8.3.

According to Directive 67/548/EEC[22], labelling for skin sensitisation is with symbol Xi, the indication of danger *irritant* and the risk phrase R43 (R43: May cause sensitisation by skin contact).

Respiratory sensitizers

In REACH, respiratory sensitizers are indicated for harmonised classification and labelling and regulated in Annex I of Directive 67/548/EEC. Annex XV in REACH lays down general principles for preparing dossiers to propose and justify harmonised classification and labelling of CMRs (carcinogenic, mutagenic, toxic for reproduction) and respiratory sensitizers.

Potential hazard for respiratory sensitisation cannot be easily addressed, as validated testing methods are currently not available. A probable hazard for respiratory sensitisation should be mentioned in the Safety Data Sheet.

Although no testing strategy is available, a substance could be classified as *respiratory sensitizer* by following the flow chart for integrated evaluation strategy (IES) reported in Section R.7.3.8.3 which is based on existing evidence.

According to Directive 67/548/EEC, labelling for *respiratory sensitizers* is with symbol Xn, the indication of danger *harmful* and the risk phrase R42 (R42: May cause sensitisation by inhalation). Concluding on suitability for chemical safety assessment: dose response assessment and potency

There is evidence that for both skin sensitisation and respiratory hypersensitivity dose-response relationships exist (although these are frequently less well defined in the case of respiratory hypersensitivity). The dose of agent required to induce sensitisation in a previously naïve subject or animal is usually greater than that required to elicit a reaction in a previously sensitised subject or animal; therefore the dose-response relationship for the two phases will differ. Little or nothing is known about dose-response relationships in the development of respiratory hypersensitivity by non-immunological mechanisms.

It is frequently difficult to obtain dose-response information from either existing human or guinea pig data where only a single concentration of the test material has been examined. With human data, exposure measurements may not have been taken at the same time as the disease was evaluated, adding to the difficulty of determining a dose response.

Dose-response data however, can be generated from local lymph node assays or, in exceptional cases, using specially designed guinea pig test methods. Such types of data can give data on induction and elicitation thresholds in these models, but it must be remembered these cannot be translated directly to human thresholds.

[22] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS). See section R.7 Introduction

Measurement of potency

Appropriate dose-response data can provide important information on the potency of the material being tested. This can facilitate the development of more accurate risk assessments. This section refers to potency in the induction phase of sensitisation.

Neither the standard LLNA nor the GPMT/Buehler is specifically designed to evaluate the skin sensitising potency of test compounds, instead they are used to identify sensitisation potential for classification purposes. However, all could be used for some estimate of potency. The relative potency of compounds may be indicated by the percentage of positive animals in the guinea pig studies in relation to the concentrations tested. Likewise, in the LLNA, the EC3 value (the dose estimated to cause a 3-fold increase in local lymph node proliferative activity) can be used as a measure of relative potency (ECETOC, 2000). Often linear interpolation of a critical effect dose from the EC3 is proposed (ECETOC), but more advanced statistical approaches basing conclusions on the characteristic of the dose response curve and variability of the results is also used (Basketter et al., 1999, van Och et al., 2000). The dose-response data generated by the LLNA makes this test more informative than guinea pig assays for the assessment of skin sensitising potency. EC3 data correlate well with human skin sensitisation induction thresholds derived from historical predictive testing (Schneider et al., 2004; Griem, 2003; Basketter et al., 2005b). Accordingly, there are proposals for how this information may be used in a regulatory sense (Basketter et al., 2005b) and for risk assessment.

Derivation of a DNEL

Potency information, such as the LLNA EC3 value, can be utilised for the derivation of no-effect levels, that is – in this instance - the threshold required for the induction of skin sensitisation. It should be noted that thresholds for skin sensitisation should be expressed in terms of dose per unit area. As mentioned above, the EC3 value correlates well with thresholds observed in previously published human predictive test data and with clinical experience (reviewed in Basketter et al., 2007a). The EC3 value can then be extrapolated by the application of assessment factors (reflecting e.g. intra and inter-individual variability and vehicle matrix effects) to derive no-effect levels (expressed in $\mu g/cm^2$of skin) for use of specific skin sensitizers in defined exposure situations (Gerberick et al., 2001; Felter et al., 2002 and 2003; Basketter et al., 2006). The approach is commonly referred to as quantitative risk assessment (QRA) and has been deployed, with considerable effect, to identify safe exposure levels for a range of skin sensitising chemicals (Zachariae et al., 2003; Basketter et al., 2003). Most recently, this has been reported extensively for fragrance and preservative sensitizers (Api et al., 2007; Basketter et al., 2007b).

Guidance on how to use the potency information for qualitative assessment (see also Section E.3.4.2) and how to derive a DNEL as a second step in the safety assessment of sensitizers is given in Appendix R.8-10.

R.7.3.7.3 Additional considerations

Chemical allergy is commonly designated as being associated with skin sensitisation (allergic contact dermatitis), or with sensitisation of the respiratory tract (asthma and rhinitis). In view of this it is sometimes assumed that allergic sensitisation of the respiratory tract will result only from inhalation exposure to the causative chemical, and that skin sensitisation necessarily results only from dermal exposure. This is misleading, and it is important for the purposes of risk management to acknowledge that sensitisation may be acquired by other routes of exposure. Since adaptive immune responses are essentially systemic in nature, sensitisation of skin surfaces may theoretically

develop from encounter with contact allergens via routes of exposure other than dermal contact (although in practice this appears to be uncommon). Similarly, there is evidence from both experimental and human studies which indicate that effective sensitisation of the respiratory tract can result from dermal contact with a chemical respiratory allergen. Thus, in this case, it appears that the quality of immune response necessary for acquisition of sensitisation of the respiratory tract can be skin contact with chemical respiratory allergens (Kimber et al., 2002). Such considerations have important implications for risk management. Thus, for instance, there is a growing view that effective prevention of respiratory sensitisation requires protection of both skin and respiratory tracts. This includes the cautious use of known contact allergens in products to which consumers are (or may be) exposed via inhalation, such as sprays. The generic advice is that appropriate strategies to minimise the risk of sensitisation to chemical allergens will require consideration of providing protection of all relevant routes of exposure.

R.7.3.7.4 Information not adequate

A WoE approach, comparing available adequate information with the tonnage-triggered information requirements by REACH, may result in the conclusion that the requirements are not fulfilled. In order to proceed in further information gathering the testing strategy given in the next Section R.7.3.8 can be adopted.

R.7.3.8 Integrated testing strategy (ITS) for sensitisation

R.7.3.8.1 Objective / General principles

Ensure that the objective of this testing strategy is to give guidance on a stepwise approach to hazard identification with regard to the endpoint; a key principle of the strategy is that the results of one study are evaluated before another is initiated. The strategy should seek to ensure that the data requirements are met in the most efficient and humane manner so that animal usage and costs are minimised.

R.7.3.8.2 Preliminary considerations

The guidance given in Sections R.7.3.2 to R.7.3.4 above will have enabled the identification of the data gaps that need to be filled in to meet the requirements of REACH as defined in Annexes VI to XI. Careful consideration of existing toxicological data, exposure characteristics and current risk management procedures is recommended to ascertain whether the fundamental objectives of the ITS (see above) have already been met. Give guidance on other factors that might mitigate data requirements for the endpoint of interest e.g. possession of other toxic properties, characteristics that make testing technically not possible.

R.7.3.8.3 Testing strategies for sensitisation

Develop a testing strategy for the endpoint that takes account of existing data on toxicity, exposure characteristics as well as the specific rules for adaptation from standard information requirements, as described in column 2 of Annexes VII-X, together with some general rules for adaptation from standard information requirements in Annex XI.

Figure R.7.3-1 Integrated testing strategy for skin sensitisation

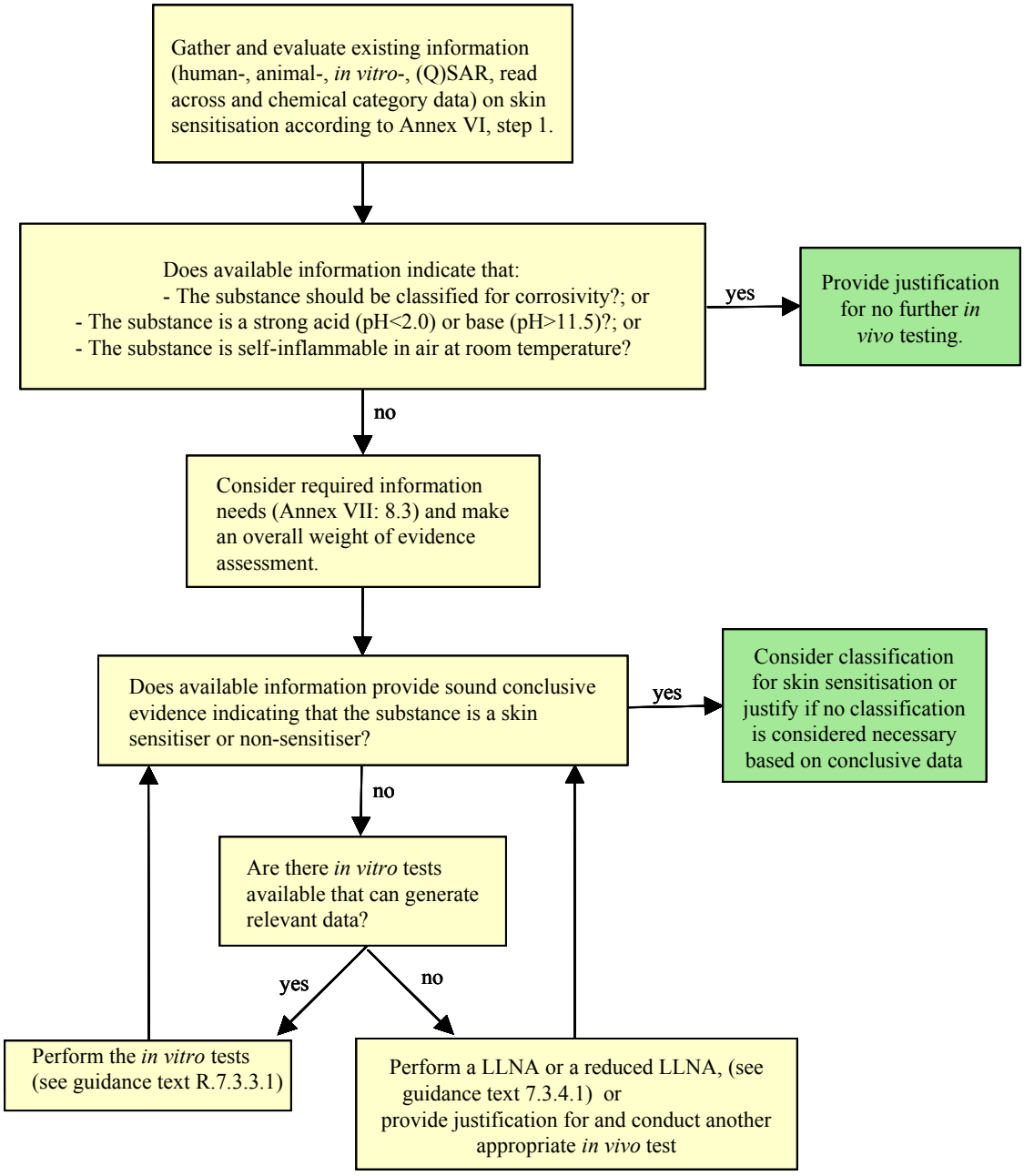
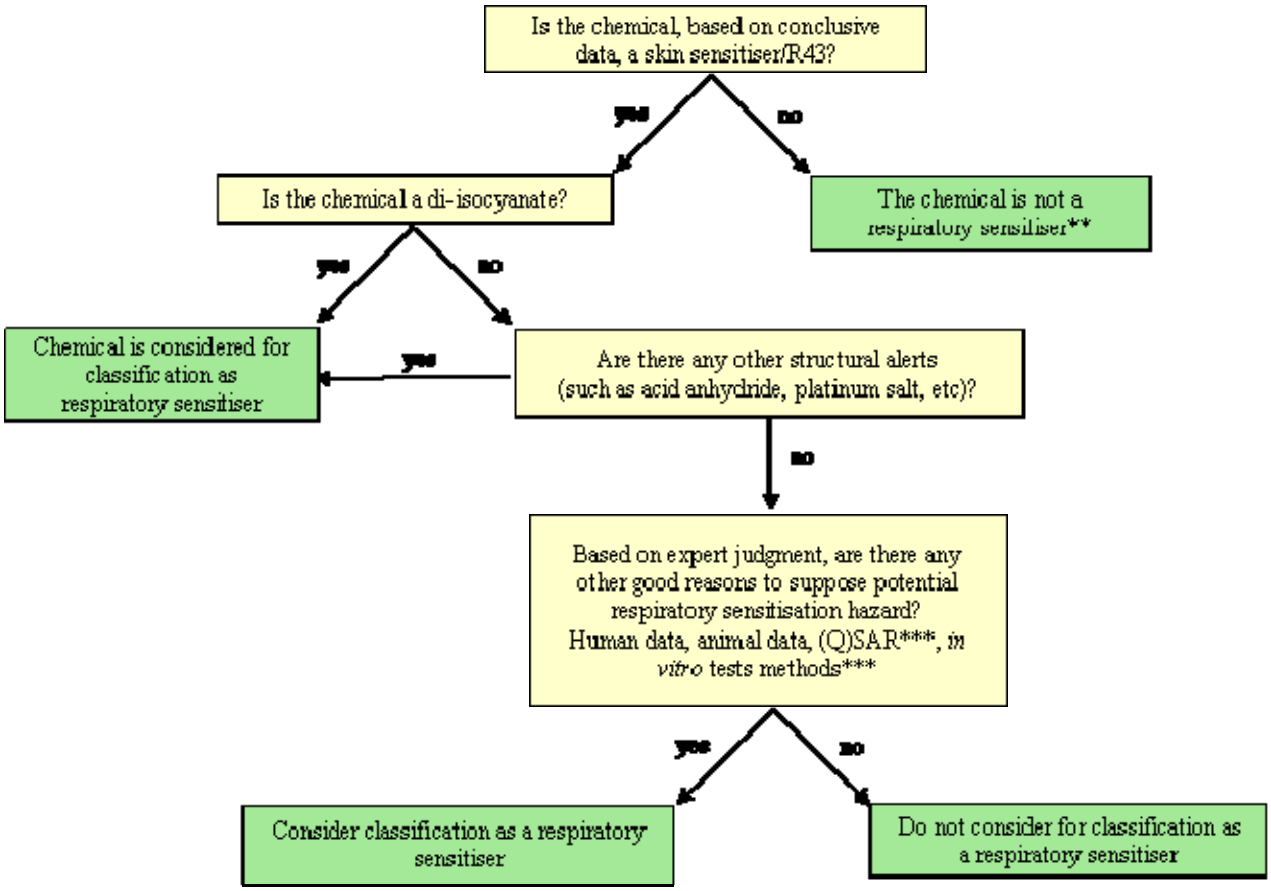


Figure R.7.3-2 Integrated evaluating strategy for respiratory sensitisation data*



* In contrast to tests for skin sensitisation, the performance of tests for respiratory sensitisation is currently not required under REACH. Therefore the present IES scheme depicts a strategy for evaluating existing data.
** This does not discount the possibility that the chemical may induce respiratory hypersensitivity through non-immunological mechanisms. Chemicals that act through such mechanisms are usually identified on the basis of evidence from human exposure.
*** not yet available

R.7.3.9 References on skin and respiratory sensitisation

Accelrys Inc., TopKat User Guide Version 6.02, San Diego: Accelrys Inc., 2004.

Aeby P, Wyss C, Beck H, Griem P, Scheffler, H and Goebel, C. 2004. Characterisation of the sensitizing potential of chemicals by in vitro analysis of dendritic cell activation and skin penetration. Journal of Investigative Dermatolog. 122, 1154-1164.

Agius RM, Nee J, McGovern B, Robertson A. 1991. Structure activity hypotheses in occupational asthma caused by low molecular weight substances. Ann Occup Hyg 35:129–137.

Agius RM, Elton RA, Sawyer L, Taylor P. 1994. Occupational asthma and the chemical properties of low molecular weight organic substances. Occup Med-State of the Art Rev 44:34–36.

Agius RM. 2000. Why are some low molecular weight agents asthmagenic? Occup Med-State of the Art Rev 50:369–384.

Aiba S, Manome H, Yoshino Y, Tagami H. 2000. In vitro treatment of human transforming growth factor-beta1-treated monocyte-derived dendritic cells with haptens can induce the phenotypic and functional changes similar to epidermal Langerhans cells in the initiation phase of allergic contact sensitivity reaction. Immunology 101, 68-75

Api AM, Basketter DA, Cadby PA, Cano M-F, Ellis G, Gerberick G F, Griem P, McNamee PM, Ryan CA, Safford B. Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients. Regulatory Toxicology and Pharmacology, submitted.

Aptula AO, Patlewicz G, Roberts DW. 2005. Skin Sensitisation: Reaction Mechanistic Applicability Domains for Structure-Activity Relationships. Chemical Research in Toxicology. 18: 1420-1426.

Aptula AO, Roberts DW. 2006a. Mechanistic applicability domains for nonanimal-based prediction of toxicological endpoints: general principles and application to reactive toxicity. Chemical Research in Toxicology 19(8):1097-105.

Aptula AO, Patlewicz G, Roberts DW, Schultz TW. 2006b. Non-enzymatic glutathione reactivity and in vitro toxicity: A non-animal approach to skin sensitisation, Tox in Vitro 20, 239-247

Arts JHE., Kuper CF, Spoor SM, Bloksma N. 1998. Airway morphology and function of rats following dermal sensitization and respiratory challenge with low molecular weight chemicals. Toxicol.Appl. Pharmacol. 152, 66-76

Arts JHE, Bloksma N, Leusink-Muis A, Kuper CF. 2003. Respiratory allergy and pulmonary irritation to trimellitic anhydride in Brown Norway rats. Toxicol. Appl. Pharmacol, 187, 38-49

Arts JHE, Kuper CF. 2007. Animal models to test respiratory allergy of low molecular weight chemicals: a guidance. Methods 41, 61-71

Ashby J, Basketter DA, Paton D, Kimber I. 1995. Structure activity relationships in skin sensitisation using the murine local lymph node assay. Toxicology 103: 177-194.

Azam P, Peiffer JL, Chamousset D, Tissier MH, Bonnet PA, Vian L, Fabre I and Ourlin JC. 2006. The cytokine-dependent MUTZ-3 cell line as an in vitro model for the screening of contact sensitizers. Toxicology and Applied Pharmacology 212, 14-23.

Balls M. et al. 1995. Practical aspects of the validation of Toxicity Test Procedures. ATLA 23:129-147

Barratt MD, Basketter DA, Chamberlain M, Admans GD, Langowski JJ. 1994. An expert system rulebase for identifying contact allergens. Toxicol in Vitro. 8: 1053-1060.

Barratt MD, Basketter DA, Roberts DW. 1997. Quantitative structure activity relationships. In Lepoittevin J-P, Basketter DA, Dooms-Goossens A, Karlberg A-T, eds, The Molecular Basis of Allergic Contact Dermatitis. Springer-Verlag Heidelberg. pp. 129-154.

Barratt MD, Langowski JJ. 1999. Validation and subsequent development of the Derek skin sensitisation rulebase by analysis of the BgVV list of contact allergens. J Chem Inf Comput Sci. 39: 294-298.

Barratt MD, Castell JV, Miranda MA, Langowski JJ. 2000. Development of an expert system rulebase for the prospective identification of photoallergens. J Photochem Photobiol B. 58: 54-61.

Basketter DA, Roberts DW, Cronin M Scholes EW. 1992. The value of the local lymph node assay in quantitative structure-activity investigations. Contact Dermatitis. 27: 137-142

Basketter DA, Lea LJ, Dickens A, Briggs D, Pate I, Dearman RJ, Kimber I. 1999. A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. J Appl Tox 19, 261-266.

Basketter DA, Angelini G, Ingber A, Kern P and Menné T. 2003. Nickel, chromium and cobalt in consumer products: revisiting safe levels in the new millennium. Contact Dermatitis, 49, 1 - 7.

Basketter DA, Clapp C, Jefferies D, Safford RJ, Ryan CA, Gerberick GF, Dearman RJ and Kimber I. 2005a. Predictive identification of human skin sensitisation thresholds. Contact Dermatitis, 53, 260 - 267.

Basketter DA, Andersen KE, Lidén C, van Loveren H, Boman A, Kimber I, Alanko K and Berggren E. 2005b. Evaluation of the skin sensitising potency of chemicals using existing methods and considerations of relevance for elicitation. Contact Dermatitis, 52: 39 - 43.

Basketter DA and Kimber I. 2006. Predictive test for irritants and allergens and their use in quantitative risk assessment. In "Contact Dermatitis", 4th Edition. Eds, Frosch PJ, Menné T and Lepoittevin J-P, Springer Verlag, Heidelberg, pp 179 – 188.

Basketter DA, Gerberick GF and Kimber I. 2007a. The local lymph node assay EC3 value: status of validation. Contact Dermatitis, submitted.

Basketter DA, Clapp CJ, Safford BJ, Jowsey IR, McNamee PM, Ryan CA, Gerberick GF. 2007b. Preservatives and skin sensitisation quantitative risk assessment: risk benefit considerations. Dermatitis, subitted.

Baur X, Dewair M, and Fruhmann G. 1984. Detection of immunologically sensitized isocyanate workers by RAST and intercutaneous skin tests. J. Allergy Clin. Immunol., 73, 610-618.

Bentley AM, Maestrelli P, Saetta M, Fabbri LM, Robinson DS, Bradley BL, Jeffery PK, Durham SR, Kay AB. 1992. Activated T-lymphocytes and eosinophils in the bronchial mucosa in isocyanate-induced asthma. J. Allergy Clin. Immunol. 89, 821-829.

Bernstein IL, Bernstein DI, Chan-Yeung M, Malo J-L. 1993. Definition and classification of asthma. In: Bernstein, I.L., Chan-Yeung, M., Malo, J-L. and Bernstein D.I. (eds). Asthma in the Workplace. Marcel Dekker, New York, 1-4.

Buckley TL, Nijkamp FP. 1994. Mucosal exudation associated with a pulmonary delayed-type hypersensitivity reaction in the mouse. Role for the tachykinins. J Immunol. 153, 4169-4178

Casati et al. 2005. Dendritic cells as a tool for the predictive identification of skin sensitisation hazard. ATLA 33, 47-62

Chan-Yeung M and Malo J-L. 1993. Compendium 1: table of major inducers of occupational asthma. In: Bernstein, I.L., Chan-Yeung, M., Malo, J-L. and Bernstein, D.I., editors. Asthma in the Workplace. Marcel Dekker, New York, pp 595-623.

Corsini E, Primavera A, Marinovich M, Galli CL. 1998. Selective Induction of cell associated interleukin-1 in murine keratinocytes by chemical allergens, Toxicology 129, 193-200

Coquette A., Berna N, Vandenbosch A, Rosdy M, De Wever B, Poumay Y. 2003. Analysis of interleukin-1α (IL-1α) and interleukin-8 (IL-8) expression and release in in vitro reconstructed human epidermis for the prediction of in vivo skin irritation and/or sensitisation. Toxicol. In Vitro 17, 311-321.

Cronin MT, Basketter DA.1994. Multivariate QSAR analysis of a skin sensitisation database. SAR QSAR Environ Res. 2(3):159-79.

Cronin MTD, Jaworska JS, Walker JD, Comber MHI, Watts CD, Worth AP. 2003. Use of QSARs in International Decision-Making Frameworks to Predict Health Effects in Chemical Substances. Environmental Health Perspectives. 111(10): 1391-1401.

Cunningham AR, Cunningham SL, Consoer DM, Moss ST, Karol MH. 2005. Development of an information-intensive structure-activity relationship model and its application to human respiratory chemical sensitizers. SAR QSAR Environ Res. 16(3):273-85.

Curren, Southee, Spielmann, Liebsch, Fentem, Balls. 1995. The Role of Prevalidation in the Development, Validation and Acceptance of Alternative Methods. ECVAM Prevalidation Task Force Report 1. ATLA 23, 211-217

Dearman RJ, Warbrick EV, Skinner R, Kimber I. 2002. Cytokine fingerprinting of chemical allergens: species comparisons and statistical analyses. Food Chem. Toxic. 40, 1881-1892

Dearman RJ, Betts CJ, Humphreys N, Flanagan BF, Gilmour NJ, Basketter DA, Kimber I. 2003. Chemical allergy: considerations for the practical application of cytokine profiling. Toxicol. Sci. 71, 137-145

Devillers J. 2000. A Neural Network SAR Model for Allergic Contact Dermatitis Toxicology Methods 10(3): 181-193.

Dimitrov SD, Low LK, Patlewicz GY, Kern PS, Dimitrova GD, Comber MH, Phillips RD, Niemela J, Bailey PT, Mekenyan OG. 2005. Skin sensitisation: modeling based on skin metabolism simulation and formation of protein conjugates. Int J Toxicol. 24(4):189-204.

Ebino K, Ueda H, Kawakatsu H, Shutoh Y, Kosaka T, Nagayoshi H, Lemus R, and Karol MH. 2001. Isolated airway exposure to toluene diisocyanate results in skin sensitization. Toxicol. Lett., 121, 79-85.

ECETOC Technical Report 89, 2003: (Q)SARs: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications

Enslein K, Gombar VK, Blake BW, Maibach HI, Hostynek JJ, Sigman CC, Bagheri D. 1997. A quantitative structure-toxicity relationships model for the dermal sensitisation guinea pig maximization assay. Food Chem.Toxicol. 35: 1091-1098.

Eskes C, Zuang V eds. 2005. Alternative (Non Animal) Methods for Cosmetic Testing: Current Status and future Prospects. ATLA Vol 33 Supplement 1, 83-103.

Estrada E, Patlewicz G, Chamberlain M, Basketter D, Larbey S. 2003. Computer-aided knowledge generation for understanding skin sensitisation mechanisms: The TOPS-MODE approach. Chem. Res. Toxicol. 16: 1226-1235.

Farraj AK, Harkema JR and Kaminski NE. 2004. Allergic rhinitis induced by intranasal sensitisation and challenge with trimellitic anhydride but not with dinitrochlorobenzene or oxazolone in A/J mice. Toxicol. Sci. 79: 315-325.

Fedorowicz A, Zheng L, Singh H, Demchuk E. 2004. Structure-Activity Models for Contact Sensitisation. Int. J. Mol. Sci. 5: 56-66.

Fedorowicz A, Singh H, Soderholm S, Demchuk E. 2005. Structure-activity models for contact sensitisation. Chem Res Toxicol 18, 954-969

Felter SP, Robinson MK, Basketter DA and Gerberick GF. 2002. A review of the scientific basis for default uncertainty factors for use in quantitative risk assessment of the induction of allergic contact dermatitis. Contact Dermatitis, 47, 257-266.

Felter SP, Ryan CA, Basketter DA and Gerberick GF. 2003. Application of the risk assessment paradigm to the induction of allergic contact dermatitis. Regulatory Toxicol Pharmacol, 37, 1 - 10.

Garcia H, Salter-Cid L, Stein-Streilein J. 1992. Persistent interleukin-2 activity and molecular evidence for expression of lymphotoxin in the hapten-immune model for pulmonary interstitial fibrosis. Am. J. Respir. Cell. Mol. Biol. 6, 22-28.

Garssen J, Nijkamp FP, Vliet van der H, Loveren van H. 1991. T-cell mediated induction of airway hyperreactivity in mice. Am. Rev. Resp. Dis. 144, 931-938

Gealy R, Graham C, Sussman NB, Macina OT, Rosenkranz HS, Karol MH. 1996. Evaluating clinical case report data for SAR modeling of allergic contact dermatitis. Hum Exp Toxicol. 15(6): 489-93.

Gerberick GF, Robinson MK, Felter S, White I and Basketter DA. 2001. Understanding fragrance allergy using an exposure-based risk assessment approach. Contact Dermatitis, 45, 333-340.

Gerberick et al., Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. 2004. Development of a peptide reactivity assay for screening contact allergens. Toxicological Sciences 81, 332-343.

Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz GY, Basketter DA. 2005. Compilation of historical local lymph node data for evaluation of skin sensitisation alternative methods. Dermatitis 16(4):157-202.

Gerner I, Barratt MD, Zinke S, Schlegel K, Schlede E. 2004. Development and prevalidation of a list of structure-activity relationship rules to be used in expert systems for prediction of the skin-sensitising properties of chemicals. Altern Lab Anim. Nov;32(5):487-509.

Graham C, Gealy R, Macina OT, Karol MH, Rosenkranz HS. 1996. QSAR for allergic contact dermatitis. Quant Struct-Act Relat. 15: 224-229.

Graham C, Rosenkranz HS, Karol MH. 1997. Structure-activity model of chemicals that cause human respiratory sensitisation. Regul Toxicol Pharmacol 26:296–306.

Griem P, Goebel C, and Scheffler H. 2003. Proposal for a risk assessment methodology for skin sensitisation potency data. Reg. Tox. Pharmacol., 38: 269-290.

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Janusch Roi A, Prieto P, Sabbioni E, Scott L, Worth A, Zuang V. 2004. A modular approach to the ECVAM Principles on Test Validity ATLA 32: 467-472

Health and Safety Executive. 1997. Critical assessments of the evidence for agents implicated in occupational asthma. HSE Books, Sudbury, Suffolk, UK

Herouet C, Cottin M, LeClaire J, Enk A. and Rousset F. 2000. Contact sensitizers specifically increase MHC class II expression on murine immature dendritic cells. In Vitro Molecular Toxicology 13, 113-123

Herrick CA, Das J, Xu L, Wisnewski AV, Redlich CA, Bottomly K. 2003. Differential roles for CD4 and CD8 T cells after diisocyanate sensitisation: genetic control of TH2-induced lung inflammation. J. Allergy Clin. Immunol. 111, 1087-1094.

Howes et al 1996. Methods for assessing percutaneous absorption ATLA 24: 81-106

Jowsey IR, Basketter DA, Westmoreland C, Kimber I. 2006. A future approach to measuring relative skin sensitising potency: a proposal. Journal of Applied Toxicology 26(4):341-50.

Johnson R, Macina OT, Graham C, Rosenkranz HS, Cass GR, Karol MH. 1997. Prioritizing testing or organic compounds detected as gas phase air pollutants: Structure-activity study for human contact allergens. Environmental Health Perspectives 105(9) 986-992.

Karol MH, Graham C, Gealy R, Macina OT, Sussman N, Rosenkranz HS. 1996. Structure-activity relationships and computer-assisted analysis of respiratory sensitisation potential. Toxicol Lett. 86:187-91

Kimber I. 1995. Contact and respiratory sensitization by chemical allergens: uneasy relationhips. Am. J. Contact Derm. 6, 34-39

Kimber I, Pichowski JS, Betts CJ, Cumberbatch M, Basketter D, Dearman R. 2001. Alternative Approaches to the Identification and Characterisaton of Chemical Allergens. Toxicology in vitro 15, 307-312

Kimber I and Dearman RJ. 2002. Chemical respiratory allergy: role of IgE antibody and relevance of route of exposure. Toxicology 181-182, 311-315.

Kimber I, Dearman RJ, Gerberick GF, Roggeband R and Basketter DA. 2003. Designation of substances as skin sensitising chemicals: a commentary. Human and Experimental Toxicology 22: 439-443.

Kimber I. and Dearman RJ. 2005. What makes a chemical a respiratory sensitizer. Current Opinion Allergy Clin. Immunol. 5, 119-124.

Kimber I, Dearman RJ, Betts CJ, Gerberick GF, Ryan CA, Kern PS, Patlewicz GY, Basketter DA. 2006. The local lymph node assay and skin sensitisation: a cut-down screen to reduce animal requirements? Contact Dermatitis 54: 181-5.

Kligman AM 1966. The identification of contact allergens by human assay. III. The maximization test: a procedure for screening and rating contact sensitizers. J. Invest. Derm. 47: 393-409.

Kubilus J, Cannon C, Neal P, Sennot H, Klausner M. 1986. Response of the EpiDerm Skin model to topically applied irritants and allergens. In vitro Toxicology 9, 157-166).

Li S, Fedorowicz A, Singh H, Soderholm SC. 2005. Application of the Random Forest method in Studies of Local Lymph Node Assay Based skin sensitisation data. J Chem Inf Model 45 952-964.

Li Y, Tseng YJ, Pan D, Liu J, Kern PS, Gerberick GF, Hopfinger AJ. 2007. 4D-Figerprint Categorical QSAR Models for skin sensitisation based on the classification of Local Lymph Node assay measures. Chem Res Tox 20(1):114-128.

Maestrelli P, Occari P, Turato G, Papiris SA, Di Stefano A, Mapp CE, Milani GF, Fabbri LM, Saetta M. 1997. Expression of interleukin (IL)-4 and IL-5 proteins in asthma induced by toluene isocyanate. Clin. Exp Allergy 27, 1292-1298.

Magnusson B. and Kligman AM. 1969. The identification of contact allergens by animal assay. The guinea pig maximisation test. J Invest Dermatol 52: 268-276.

Miller MD, Yourtee DM, Glaros AG, Chappelow CC, Eick JD, Holder AJ. 2005. Quantum Mechanical Structure-Activity Relationship Analyses for Skin Sensitisation. J. Chem. Inf. Model. 45: 924-929.

Van Och FM, Slob W, de Jong WH, Vandebriel RJ, van Loveren H. 2000. A quantitative method for assessing the sensitizing potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of the uncertainty margins. Toxicology 20;146(1):49-59.

Van Och FMM, Loveren van H, Jong de WH, Vandebriel RJ. 2002. Cytokine production induced by low-molecular-weight chemicals as a function of the stimulation index in a modified local lymph node assay: an approach to discriminate contact sensitizers from respiratory sensitizers. Toxicol. Appl.Pharmacol. 184, 46-56

Van Och FMM, Van Loveren H, Van Wolfswinkel JC, Machielsen AJC, Vandebriel RJ. 2005. Assessement of allergenic activity of low molecular weight compounds base don IL-1a and IL-18 production by a murine keratinocyte cell line. Toxicol 210, 95-109.

O'Neill R. 1995 Asthma at work. Causes, effects and what to do about them. Trades Union. Congres Sheffield Occupational Health Project Co-op Ltd., Sheffield, UK, p. 133.

Patlewicz G, Basketter DA, Smith CK,Hotchkiss SA, Roberts DW. 2001. Skin-sensitisation structure-activity relationships for aldehydes. Contact Dermatitis 44: 6 331-336.

Patlewicz G, Wright ZM, Basketter DA, Pease CK, Lepoittevin J-P, Arnau EG. 2002. Structure-activity relationships for selected fragrance allergens. Contact Dermatitis 47(4): 219-226.

Patlewicz G, Roberts DW, Walker JD. 2003. QSARs for the skin sensitisation potential of aldehydes and related compounds. QSAR Comb Sci. 22: 196-203.

Patlewicz G, Basketter DA, Pease CK, Wilson K, Wright ZM, Roberts DW, Bernard G, Gimenez Arnau E, Lepoittevin J-P. 2004. Further evaluation of quantitative structure-activity relationship models for the prediction of the skin sensitisation potency of selected fragrance allergens. Contact Dermatitis. 50: 91-97.

Payne MP, Walsh PT. Structure-activity relationships for skin sensitisation potential: development of structural alerts for use in knowledge-based toxicity prediction systems. J Chem Inf Comput Sci, 1994, 34, 154-161

Payne MP, Walsh PT. 1995. Structure-activity relationships for respiratory sensitisation. Proceedings, British Toxicology Society Meeting York, Yorkshire, UK, March/April 1995.

Potts RO, Guy RH. 1992. Predicting skin permeability. Pharmaceutical Research. 9(5): 663-669

Python, F., Goebel, C., Aeby, P. Assessment of the U937 cell line for the detection of contact allergens Toxicology and Applied Pharmacology (2007), doi: 10.1016/j.taap.2006.12.026 (in press).

Ren Y; Liu H; Xue C; Yao X; Liu M; Fan, B. 2006. Classification study of skin sensitizers based on support vector machine and linear discriminant analysis. Analytica Chimica Acta 572(2), 272-282.

Roberts, DW, Williams DL. 1982. The derivation of quantitative correlations between skin sensitisation and physico-chemical parameters for alkylating agents and their application to experimental data for sultones. J. Theor. Biol. 99: 807-825

Roberts DW. 1987. Structure-activity relationships for skin sensitisation potential of diacrylates and dimethacrylates. Contact Dermatitis. 17: 281-289.

Roberts DW, York M, Basketter DA. 1999. Structure-activity relationships in the murine local lymph node assay for skin sensitisation: alpha, beta-diketones. Contact Dermatitis 41(1):14-7.

Roberts, DW, Basketter DA. 2000. Quantitative structure-activity relationships: sulfonate esters in the local lymph node assay. Contact Dermatitis. 42(3):154-61.

Roberts DW, Patlewicz G. 2002. Mechanism based structure-activity relationships for skin sensitisation – the carbonyl group domain. SAR QSAR Env Res. 13(1): 145-152.

Roberts DW, Aptula AO, Patlewicz GY. 2006. Mechanistic Applicability Domains for Non-Animal Based Toxicological Endpoints. QSAR Analysis of the Schiff Base Applicability Domain for Skin Sensitisation. Chem Res Toxicol, 19(9):1228-33.

Roberts DW, Aptula AO, Cronin MTD, Hulzebos E, Patlewicz G. Global (Q)SARs for skin sensitization - assessment against OECD principles, SAR QSAR Environ. Res. 2007: in press.

Rosenkranz HS, Klopman G, Zhang YP, Graham C, Karol MH. 1999. Relationship between allergic contact dermatitis and electrophilicity. Environmental Health Perspectives 107: (2) 129-132

Ryan CA, Gerberick GF, Gildea LA, Hulette BC, Betts CJ, Cumberbatch M, Dearman RJ, Kimber I. 2005. Interactions of contact allergens with dendritic cells: opportunities and challenges for the development of novel approaches to hazard assessment. Toxicological Sciences 88, 4-11

Sakaguchi H, Ashikaga T, Miyazawa M. Yoshida Y, Yoneyama K, Hirota M, Itagaki H, Toyoda H, Suzuki H. 2006. Development of an in vitro skin sensitisation test using human cell lines; human Cell Activation Test (h-CLAT), I and II. Tox in Vitro 20, 767-773, 774-784.

Satoh T, Kramarik JA, Tollerud DJ, Karol MH. 1995. A murine model for assessing the respiratory hypersensitivity potential of chemical allergens. Toxicol. Lett. 78, 57-66.

Schultz TW, Carlson RE, Cronin MTD, Hermens JLM, Johnson R, O'Brien PJ, Roberts DW, Siraki A, Wallace KB, Veith GD. 2006. SAR and QSAR in Environmental Research 17(4) 413-428. A conceptual framework for predicting the toxicity of reactive chemicals" modelling soft electrophilicity.

Schlede E. and Eppler R. 1995. Testing for skin sensitisation according to the notification procedure for new chemicals: the Magnusson and Kligman test. Contact Dermatitis 32: 1-4.

Schneider K, and Akkan, Z. 2004. Quantitative relationship between the local lymph node assay and human skin sensitisation assays. Reg. Tox. Pharmacol. 39: 245-255.

Stadler J and Karol MH. 1984. Experimental delayed hypersensitivity following inhalation of dicyclohexylmethane-4,4'-diisocyanate: A concentration-response relationship. Toxicol. Appl. Pharmacol., 74, 244-249.

Steiling W, Basketter DA, Berthold K, Butler M, Garrigue J-L, Kimber I, Lea L, Newsome C, Roggeband R, Stropp G, Waterman S, Wieman C. 2001. Skin sensitisation testing - new perspectives and recommendations. Food and Chemical Toxicology 39: 293-301

Stotts J. 1980. Planning, conduct and interpretation of human predictive sensitisation patch tests. In: Current concepts in cutaneous toxicity. Academic Press, p. 41-53.

Tuschl H, Kovac R, Weber E. 2000. The Expression of Surface Markers on Dendritic Cells as Indicators for the Sensitising Potential of Chemicals. Toxicology in vitro, 14, 541-549

Vandebriel RJ, Van Och FMM, Van Loveren H. 2005. In vitro assessment of sensitizing activity of low molecular weight compounds. Toxicol Appl Pharmacol 207, S142-1

Website: http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.htm

Wolfreys A, Basketter DA. 2004. Mutagens and Sensitizers – an unequal relationship? Journal of Toxicology Cutaneous and Ocular Toxicology 23(3): 197-205.

Worth A.P. and Balls M. 2001. The importance of the prediction model in the development and validation of alternative tests. ATLA 29, 135-143

A. Worth, M Balls eds 2002. Alternative (Non-Animal) Methods for Chemical Testing: Current Status and Future Prospects. ATLA Vol. 30 Supplement 1, 49-53

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. ECB Report EUR 21866 EN Ispra, Italy: European Commission, Joint Research Centre.

Zachariae C, Rastogi S, Devantier C, Menne T and Johansen JD. 2003. Methyldibromo glutaronitrile: clinical experience and exposure-based risk assessment. Contact Dermatitis 48: 150-154.

Zinke S, Gerner I, Schlede E. Evaluation of a rule base for identifying contact allergens by using a regulatory database: Comparison of data on chemicals notified in the European Union with "structural alerts" used in the DEREK expert system. Altern Lab Anim. 2002 May-Jun;30(3):285-98.

Zwart A., Arts JHE, Kuper CF. 1994. Wave propagation: a new parameter in the description of mechanical airway impedance. Eur.Respir. Rev. 4, 203-209.

R.7.4 Acute toxicity

R.7.4.1 Introduction

Assessment of the acute toxic potential of a chemical is necessary to determine the adverse health effects that might occur following accidental or deliberate short-term exposure. The nature and severity of the acute toxic effects are dependent upon various factors, such as the mechanism of toxicity and bioavailability of the chemical, the route and duration of exposure and the total amount of chemical to which the person or animal is exposed.

R.7.4.1.1 Definition of acute toxicity

The term *acute toxicity* is used to describe the adverse effects, which may result from a single exposure (i.e. a single exposure or multiple exposures within 24 hours) to a substance. In the context of this guidance, exposure relates to the oral, dermal or inhalation routes. The adverse effects can be seen as clinical signs of toxicity (for animals, refer to OECD Guideline Document 19, 2000), abnormal body weight changes, and/or pathological changes in organs and tissues, which in some cases may result in death. In addition to acute systemic effects, some substances may have the potential to cause local irritation or corrosion of the gastro-intestinal tract, skin or respiratory tract following a single exposure. Acute irritant or corrosive effects due to the direct action of the chemical on the exposed tissue are not specifically covered by this document, although their occurrence may contribute to the acute toxicity of the chemical and must be reported. The endpoints of skin and eye irritation/corrosion and respiratory irritation are addressed in Section R.7.2.

At the cellular level acute toxicity can be related to three main types of toxic effect, (i) general basal cytotoxicity (ii) selective cytotoxicity and (iii) cell-specific function toxicity. Acute toxicity may also result from chemicals interfering with extracellular processes (ECVAM workshop report 16, 1996). Toxicity to the whole organism also depends on the degree of dependence of the whole organism on the specific function affected.

R.7.4.1.2 Objective of the guidance on acute toxicity

A chemical substance may induce systemic and/or local effects. This document is concerned with assessment of systemic effects following acute exposure.

Generally the objectives are to establish:

- whether a single exposure (or multiple exposures within 24 hours) to the substance of interest could be associated with adverse effects on human health; and/or
- what types of toxic effects are induced, their time of onset, duration and severity (all to be related to dose); and/or
- the dose-response relationships to determine the LD_{50}, the LC_{50}, the discriminating dose, or the acute toxic class; and/or
- when possible, the slope of the dose-response curve; and/or
- when possible, whether there are marked sex differences in response to the substance; and

- what information enables the classification and labelling of the substance for acute toxicity

The indices of LD_{50} and LC_{50} are statistically-derived values relating to the dose that is expected to cause death in 50% of treated animals in a given period; these values do not provide information on all aspects of acute toxicity. Indeed, information on lethality is not an essential requirement for the classification decision or risk assessment. Other parameters and observations and their type of dose response may yield valuable information. The potential to avoid acute toxicity testing should be carefully exploited by application of read-across or other non-testing means. Furthermore, there is an overriding obligation to minimize the use of animals in any assessment of acute toxicity.

For risk assessment, further considerations on the nature and reversibility of the toxic effects are necessary.

R.7.4.2 Information requirements for acute toxicity

The standard information requirements for acute toxicity under the REACH Regulations are as follows:

Annex VII (≥1 t/y): acute toxicity via the oral route of exposure is required;

Column 2 of Annex VII details specific rules for adaptation of these information requirements, notably allowing for the waiving of acute oral toxicity testing if the substance is corrosive to the skin or if a study on acute toxicity by the inhalation route is available.

Annex VIII -X (≥ 10 t/y): acute toxicity via the oral and dermal or inhalation route of exposure.

Column 2 of Annex VIII details specific rules for adaptation, notably requiring information on at least one other route of exposure depending on the nature of the substance and the likely route of human exposure (for details see Annex VIII Section 8.5); as for Annex VII, allowance is made for the waiving of acute oral toxicity testing if the substance is corrosive to the skin.

If there is any reason (alert from existing data) for a concern of acute toxicity at non-corrosive levels, one could point out needs to address this.

R.7.4.3 Information and its sources on acute toxicity

Information on acute toxicity, as detailed below, can be obtained from a variety of sources including unpublished studies, data bases and publications such as books, scientific journals, criteria documents, monographs and other publications (see Chapter R.3 for further general guidance).

R.7.4.3.1 Non-human data on acute toxicity

Non-testing data on acute toxicity

Non-testing data can be provided by the following approaches: a) structure-activity relationships (SARs) and quantitative structure-activity relationships (QSARs), collectively called (Q)SARs; b) expert systems incorporating (Q)SARs and/or expert rules; and c) grouping methods (read-across and categories. These approaches can be used to assess acute toxicity if they provide relevant and reliable (adequate) data for the chemical of interest. Guidance on how to assess the relevance and reliability of non-testing data is provided in the general guidance on (Q)SARs in Section R.6.1 and on grouping approaches in Section R.6.2. Non-testing methods should be documented according to

the appropriate reporting formats (see Sections R.6.1.9 and R.6.2.6). In the case of (Q)SARs and expert systems, a detailed description of available models is provided in the JRC QSAR Model Database (http://qsardb.jrc.it/).

Compared with some endpoints, there are relatively few (Q)SAR models and expert systems capable of predicting acute toxicity. Available approaches have been reviewed in the literature (Cronin *et al*., 1995,2003; Lessigiarska *et al*., 2005; Tsakovska et al., 2006). On the basis of these reviews, the following conclusions can be made: a) the relatively small number of models for *in vivo* toxicity is related to the nature of the endpoint – acute toxicity measurements are usually related to whole body phenomena and are therefore very complex. The complexity of the mechanisms involved leads to difficulties in the QSAR modelling process; b) most QSAR models identify hydrophobicity as a parameter of high importance for the modelled toxicity. In addition, many models indicate the role of the electronic and steric effects; c) most literature-based models are restricted to single classes of chemicals, such as phenols, alcohols, anilines. Models based on more heterogeneous data sets are those incorporated in the expert systems.

In the sections below some examples are given in order to illustrate the potential possibility for applying the (Q)SAR approaches for the acute toxicity endpoint for predictive purposes or to investigate the mechanisms of toxicity.

(Q)SAR models

QSARs on inhalation toxicity

Some simple regression models have been developed for predicting the inhalational toxicity of volatile substances, and these can be used reliably within their domains of applicability. Typically, parameters such as vapour pressure (VP) and boiling point (BP) have been found to be useful predictors of the acute toxic effect (e.g. LC_{50} value). These models are based on the assumption that toxicity occurs by the non-specific mechanism of narcosis, and that the LC_{50} data are based on tests in which a steady-state concentration has been reached in the blood. These models are suitable only for systemic acting volatile compounds.

For example, acute (non-lethal) neurotoxicity data for the neurotropic effects of some common solvents on both rats (whole-body exposures for 4h) and mice (whole-body exposures for 2h), taken from Frantik et al (1996), were subjected to QSAR analysis by Cronin (1996). Stepwise regression analysis of the 4-hr toxicity data causing the 30% depression in response ($log1/ECR_{30}$) in rats gave the following equation:

$$log1/ECR_{30} = 0.361\ ClogP - 0.117\ {}^{0}\chi - 1.76$$

$$n = 37\ \ R^2 = 0.817 \qquad s = 0.280 \qquad F = 35.2$$

This relationship demonstrates a partial dependence of neurotoxicity with the octanol-water partition coefficient, logP. The negative correlation with the zero-order molecular connectivity ${}^{0}\chi$ is thought to be an indication that the membrane permeability of blood-brain barrier is reduced for large molecules.

Stepwise regression for mouse neurotoxicity gave the following equation:

$$log1/ECM_{30} = 0.212\ ClogP + 0.00767\ BP - 0.176\ {}^{0}\chi - 2.03$$

$$n = 39\ \ R^2 = 0.811 \qquad s = 0.271 \qquad F = 22.4$$

in which BP is the boiling point of the substance (BP is inversely related to vapour pressure).

The application of principal components analysis (PCA), to separate compounds of high neurotoxicity from those of low neurotoxicity, suggested that in addition to partitioning through a membrane (determined by logP and molecular size), aqueous solubility and volatility are also important factors governing neurotoxicity (Cronin, 1996). Metabolism to more toxic compounds is suggested as a possible cause of compounds appearing as outliers in the QSARs.

QSARs for predicting LD_{50}

There are references in the literature to a few models for predicting LD_{50}, generally for small sets of compounds. For example, Hansch & Kurup (2003) developed the following QSAR to predict the toxicity of barbiturates (LD_{50}) in for female white mice, using toxicity data from Cope and Hancock (1939):

$$\log 1/LD_{50} = -1.44 \log P + 0.16 NVE - 8.70$$

$$n = 11 \quad R^2 = 0.924 \quad s = 0.077 \quad R^2_{cv} = 0.879$$

where NVE is the number of valence electrons (used as a measure of polarisability).

QSARs for predicting human toxicity

The same descriptors were used to predict the LD_{100} of miscellaneous drugs to humans, using toxicity data from King (1985):

$$\log 1/C = 0.61 \log P + 0.017 NVE + 1.44$$

$$n = 36 \quad R^2 = 0.850 \quad s = 0.438 \quad R^2cv = 0.817$$

QSARs for predicting *in vitro* effects

A number of QSAR models for predicting *in vitro* effects are cited in the literature (reviewed in Tsakovskaet al., 2006), but these are not directly relevant to the assessment of acute toxicity for regulatory purposes. In general, these models have been developed to investigate the mechanisms of cytotoxic action, and they outline the role of hydrophobicity as well electronic descriptors, including electrotopological state descriptors (Lessigiarska *et al*., 2006), bond dissociation energies (Selassie *et al*., 1999), and dissociation constants (Moridani *et al*., 2003). While these models are not directly relevant to the assessment of acute toxicity, the fact that reliable QSARs can be developed for the *in vitro* cytotoxicity of defined groups of chemicals indicates that the approach of modelling *in vitro* data should be further explored with a view to integrating such QSARs into the ITS for acute toxicity. For example, a battery of QSARs could be developed for predicting the *in vitro* data of a validated *in vitro* test, and then used to supplement or replace *in vivo* testing.

Expert systems

For heterogeneous groups of compounds, expert systems are available in which rule bases express generalised relationships between chemical structure and toxicity. In knowledge-based experts systems (see also Section R.6.1), such as HazardExpert, such rules are derived from human expert opinion. In statistically based expert systems, such as TOPKAT and MultiCASE, statistical methods were used to derive (Q)SAR models (see also Section R.6.1).

HazardExpert

HazardExpert is a module of Pallas software developed by CompuDrug Limited (http://www.compudrug.com). The program works by searching the query structure for known toxicophores, which are stored in the "Toxic Fragments Knowledge Base" and which include

substructures exerting both positive and negative modulator effects. Once a toxicophore has been identified, this triggers estimates for a number of toxicity endpoints, including neurotoxicity. The default knowledge base of the system is based on a US-EPA report (Brink and Walker, 1987) and scientific information collected by CompuDrug Limited. This program can be linked to MetabolExpert, another module of the Pallas software, to predict the toxicity of the parent compound and its metabolites. Information on the validity of the model is not available. Investigations on the validity and applicability of HazardExpert are needed before recommendations can be made about its regulatory use.

TOPKAT

The TOPKAT software package employs cross-validated quantitative structure-toxicity relationship (QSTR) models for assessing various measures of toxicity (http://accelrys.com/products/discovery-studio/toxicology/). The Rat Oral LD_{50} module of TOPKAT includes 19 QSAR regression models for different chemical classes. The models are based on a number of structural, topological and electrophysiological indices, and they make predictions of the oral acute median lethal dose in the rat (LD_{50}).

The TOPKAT rat oral LD_{50} models are based on experimental data from the RTECS. Since RTECS lists the most toxic value when multiple values exist, the TOPKAT model tends to overestimate the toxicity of query structures.

The Rat Inhalation LC_{50} module of TOPKAT contains five submodels related to different chemical classes.

TOPKAT models, including the models for acute oral toxicity, have been used by Danish EPA to evaluate the dangerous properties of around 47 000 organic substances on the EINECS list [17]. An external evaluation of this model using 1840 chemicals not contained in the TOPKAT database gave poor results (R^2 = 0.31). However, 86% of estimations fall within a factor of 10 from test results (DK EPA study).

The Danish EPA concluded that the TOPKAT model is sufficient to give an indication of the least strict classification for acute toxicity, Xn; R22. An Internet version of the Danish QSAR database is accessible from the ECB website (http://qsardb.jrc.it).

MultiCASE

The MultiCASE software (http://www.multicase.com) contains an acute toxicity module, which consists of a rat LD_{50} model based on 7920 compounds from compilations by FDA, NTP and WHO data. Information on the validity of the model is not available. Investigations on the validity and applicability of MultiCASE are needed before recommendations can be made about its regulatory use.

Testing data on acute toxicity

In vitro data

There are currently no *in vitro* tests that have been officially adopted by the EU or OECD for assessment of acute toxicity.

However, a number of *in vitro* tests for acute toxicity are undergoing a validation process:

- Two *In vitro* basal cytotoxicity assays for predicting starting doses for *in vivo* oral toxicity tests and lethal concentrations in man have undergone peer review by ICCVAM, namely the BALB/c 3T3 NRU & normal human keratinocyte (NHK) NRU assays (http://iccvam.niehs.nih.gov/methods/acutetox/inv_nru_brd.htm).

- Two *in vitro* tests pre-validated: TER and PCP in 2 renal cell lines (test battery). The loss of monolayer integrity is often an early indicator of nephrotoxicity in intact renal epithelia *in vitro* and reflects loss of renal function *in vivo*. Trans-epithelial resistance (TER), coupled with enhanced paracellular permeability (PCP), is a good measure of this integrity. (Duff *et al*., 2002). These tests should be used in a WoE approach as alerts or correctors in respect to the basal cytotoxicity assays. Their contribution is under evaluation in A-Cute-Tox (see below).

- A ECVAM validated test, the CFU-GM, to predict anticancer agents induced myelotoxicity in humans, is now under evaluation to widen its applicability domain to chemicals' induced toxicity (http://ecvam-dbalm.jrc.cec.eu.int/). If sufficiently validated and suited to the purpose of assessment of acute toxicity, this could be included in a WoE.

The integrated project A-Cute-Tox (A 5-year 6th FP project initiated in 2005) is addressing the possible replacement of the acute oral systemic toxicity tests (http://www.acutetox.org/). Particular attention should be given in the future to results of the project.

Animal data

Data may be available, particularly for phase-in substances, from a wide variety of animal studies, which give different amounts of direct or indirect information on the acute toxicity of a substance; e.g.:

- OECD TG 420 (EU B.1 bis) Acute oral toxicity – Fixed dose procedure
- OECD TG 423 (EU B.1tris) Acute oral toxicity – Acute toxic class method
- OECD TG 425 Acute oral toxicity – Up-and-down procedure
- OECD 401 (EU B.1) Acute Oral Toxicity (method deleted from the OECD Guidelines for testing of chemicals and from Annex V to Directive 67/548/EEC; see below)
- OECD TG 402 (EU B.3) Acute dermal toxicity
- OECD TG 403 (EU B.2) Acute inhalation toxicity
- Draft OECD TG 433 "Acute Inhalation Toxicity, Fixed Dose Procedure";
- Draft OECD TG 436 "Acute Inhalation Toxicity, Acute Toxic Class Method";
- Draft OECD TG 434 "Acute Dermal Toxicity, Fixed Dose Procedure";
- ICH compliant studies;
- Mechanistic and toxicokinetic studies;
- Studies in non-rodent species.

Traditionally, acute toxicity tests on animals have used mortality as the main observational endpoint, usually in order to determine LD_{50} or LC_{50} values. These values were regarded as key information for hazard assessment and supportive information for risk assessment. However, derivation of a precise LD_{50} or LC_{50} value is no longer considered essential. Indeed, some of the current standard acute toxicity test guidelines, such as the fixed dose procedures (OECD 420, EU

B.1 bis and draft OECD 433), use signs of non-lethal toxicity and have animal welfare advantages over the other guidelines.
Existing OECD TG 401 (EU B.1) data would normally be acceptable but testing using this deleted method must no longer be performed.
In addition to current regulatory methods, acute toxicity data on animals may be obtained by conducting a literature search and reviewing all available published and unpublished toxicological or general data, and the official/existing acute toxicological reference values. For more extensive general guidance see Section R.3.1.
Utilising all the available information from sources such as those above, a *Weight of Evidence* approach should be taken to maximise use of existing data and minimise the commissioning of new testing.
When several data are available, a hierarchal strategy should be used to focus on the most relevant.

R.7.4.3.2 Human data on acute toxicity

Acute toxicity data on humans may be available from:

- Epidemiological data identifying hazardous properties and dose-response relationships;
- Routine data collection, poisons data, adverse event notification schemes, coroner's report;
- Biological monitoring/personal sampling;
- Human kinetic studies – observational clinical studies;
- Published and unpublished industry studies;
- National poisoning centres.

The main obstacles to the use of human data are their limited availability and often limited information on levels of exposure (ECETOC, 2004).

R.7.4.3.3 Exposure considerations for acute toxicity

With regard to acute toxicity, exposure considerations are detailed in column 2 in Annex VIII, but not in Annex XI. If there is only one demonstrated route of exposure, this route must be addressed. Where the potential for human exposure exists, the most likely route, or routes, of exposure should be determined so that the potential for acute toxicity by these routes can be assessed. Determination of the most likely route of exposure will have to take into account not only how the substance is manufactured and handled, including engineering controls that are in place to limit exposure, but also the physico-chemical properties of the substance, for instance, whether the substance is a solid or liquid, the particle size and proportion of respirable and inhalable particles, vapour pressure and log P.

R.7.4.4 Evaluation of available information on acute toxicity

The detailed generic guidance provided in Chapter R.4 on the process of judging and ranking the available data for its adequacy (reliability and relevance), completeness and remaining uncertainty is relevant to information on acute toxicity.

R.7.4.4.1 Non-human data on acute toxicity

Non-testing data on acute toxicity

Physico-chemical properties[23]

It may be possible to infer from the physico-chemical characteristics of a substance whether it is likely to be corrosive or absorbed following exposure by a particular route and, produce acute toxic effects. Physico-chemical properties may be important in the case of the inhalation route (vapour pressure, MMAD, log K_{ow}), determining the technical feasibility of the testing and acting upon the distribution in the airways in particular for *local-acting substances*. Indeed, some physico-chemical properties of the substance or mixture could be the basis for waiving testing. In particular, it should be considered for low volatility substances, which are defined as having vapour pressures $<1 \times 10^{-5}$ kPa (7.5×10^{-5} mmHg) for indoor uses, and $<1 \times 10^{-4}$ kPa (7.5×10^{-4} mmHg) for outdoor uses. Furthermore, inhalable particles are capable of entering the respiratory tract via the nose and/or mouth, and are generally smaller than 100 μm in diameter. Particles larger than 100 μm are less likely to be inhalable. In that way, particular attention should be driven on results of aerosol particle size determination.

In particular, for substances in powder form, particle size of the material decisively influences the deposition behaviour in the respiratory tract and potential toxic effects. Particle size considerations (determined by e.g. granulometry testing, OECD 110) can be useful for:

- selecting a representative sample for acute inhalation toxicity testing
- assessing the respirable and inhalable fractions, preferably based on aerodynamic particle size
- justifying derogations from testing, for instance, when read-cross (or chemical grouping approach) data can be associated with results from particle size distribution analyses (see Section R.6.2)

Physico-chemical properties are also important for determination of the potential of exposure through the skin, for example, log K_{ow}, molecular weight and volume, molar refraction, degree of hydrogen bonding, melting point (Hostýnek, 1998).

Read-across to structurally or mechanistically similar substances (SAR)

Generic guidance on the application of grouping approaches is provided in Section R.6.2.

(Q)SAR

Several (Q)SAR systems are available that can be used to make predictions about, for example, dermal penetration or metabolic pathways (see cross-cutting QSAR guidance for list of models). However, these systems have not been extensively validated against appropriate experimental data and it has not been yet verified if the results genuinely reflect the situation *in vivo*. That is why the modelled data can be used for hazard identification and risk assessment purposes only as part of a WoE approach.

The complexity of the acute toxicity endpoint (possibility of multiple mechanisms) is one of the reasons for limited availability and predictivity of QSAR models. In the absence of complete validation information, available models could be used as a part of the WoE approach for hazard

[23] Refer also to Tables R.12-1 to R.12-6 in Section R.7.12

identification and risk assessment purposes after precise evaluation of the information derived from the model.

Evaluation of the validity of the method

An evaluation of model validity according to OECD principles should be available, as described in Section R.6.1.

Evaluation of the reliability of the individual prediction

The reliability of individual (Q)SAR predictions should be evaluated, as described in Section R.6.1.

The evaluations of model validity and estimate reliability should be documented according to the appropriate reporting formats, as described in Section R.6.1.
In the case of grouping approaches, adequacy should be assessed and documented according to guidance described in Section R.6.2.

Testing data on acute toxicity

In vitro data

The *in vitro* tests that are currently available provide supplementary information, which may be used to determine starting doses for *in vivo* studies, assist evaluation of data from animal studies, especially in identification of species differences, or to increase understanding of the toxicological mechanism of action of the substance. They cannot be used to replace testing in animals completely, although this may be possible in the future.

The outcome of the EU-US (ECVAM-ICCVAM) validation study on the Use of In Vitro Basal Cytotoxicity Test Methods For Estimating Starting Doses For Acute Oral Systemic Toxicity (http://iccvam.niehs.nih.gov/methods/acutetox/inv_nru_brd.htm) was that the Peer Review Panel agreed that the applicable validation criteria have been adequately addressed for using these *in vitro* test methods in a WoE approach to determine the starting dose for acute oral *in vivo* toxicity protocols. Moreover, on the basis of a preliminary analysis of data, there is the indication that the cytotoxicity tests might be useful in predicting low toxicity substances (LD_{50}≥ 2g/kg body weight) and that they might therefore be used to filter these out in the future. This application needs to be validated with a wider range of compounds.

In vitro data may be useful for predicting acute toxicity in humans providing that the domain of applicability for the test method is appropriate for the class of chemical under evaluation and a range of test concentrations have been investigated that permit calculation of an IC_{50} (inhibitory concentration 50%) value. Indeed, on the basis of a preliminary comparison of data, there is the indication that the results of *in vitro* cytotoxicity tests may be more predictive of acute oral toxicity in humans than rat or mouse data. This aspect needs to be further investigated.

Generic guidance is given in Chapter R.4 for judging the applicability and validity of the outcome of various study methods, assessing the quality of the conduct of a study (including how to establish whether the substance falls within the applicability domain of the method and the validation status for the given domain) and aspects such as vehicle, number of duplicates, exposure/ incubation time, GLP-compliance or comparable quality description.

Animal data

Acute toxicity tests on animals have primarily used mortality as the main observational endpoint, usually in order to determine LD_{50} or LC_{50} values, although some of the current standard protocols,

such as the fixed dose procedure (OECD TG 420, EU B.1 bis), use evident signs of toxicity in place of mortality. In many cases, there will be little information on the cause of death or mechanism underlying the toxicity, and only limited information on pathological changes in specific tissues or clinical signs, such as behavioural or activity changes.

Many acute toxicity studies on chemicals of low toxicity are performed as limit tests. For more harmful chemicals choice of optimum starting dose will minimize use of animals. When multiple dose levels are assessed, characterisation of the dose-response relationship may be possible and signs of toxicity identified at lower dose levels may be useful in estimating LOAELs or NOAELs for acute toxicity. For local acting substances, mortality after inhalation may occur due to tissue damage in the respiratory tract. In these cases, the severity of local effects may be related to the dose or concentration level and therefore, it might be possible to identify a LOAEL or NOAEL. For systemic toxicity, there could be some evidence of target organ toxicity (pathological findings have to be documented) or signs of toxicity based on clinical observations.

Whichever approach is used in determining acute toxicity critical information needs to be derived from the data to be used in risk assessment. It is important to identify those dose levels which produce signs of toxicity, the relationship of the severity of these with dose and the level at which toxicity is not observed (i.e. the acute NOAEL).

In addition to current available OECD or EU test methods (see Section R.7.4.3), alternative *in vivo* test methods for assessment of acute dermal and inhalation toxicity are in the process for adoption and use for regulatory purposes. Whichever test is used to evaluate acute toxicity on animals, the evaluation of studies takes into account the reliability based on the approach of Klimisch *et al*. (1997) (standardised methods, GLP, detailed description of the publication), the relevance, and the adequacy of the data for the purposes of evaluating the given hazard from acute exposure (for more guidance see Section R.4.2). The best studies are those that give a precise description of the nature and reversibility of the toxic effect, the number of subjects, gender, the number of animals affected by the observed effects and the exposure conditions (atmosphere generation for inhalation, duration and concentration or dose). The relevance of the data should be determined in describing the lethal or non-lethal endpoint being measured or estimated.

In addition, when several studies results are available for one substance, the most relevant one should be selected; data from others studies that have been evaluated should be considered as supportive data for the full evaluation of the substance.

The classification criteria for acute inhalation toxicity relate to a 4-hour experimental exposure period. If data for a 4-hour period are not available then extrapolation of the results to 4 hours are often achieved using Haber's Law (C.t = k). However, there are limits to the validity of such extrapolations, and it is recommended that the Haber's Law approach should not be applied to experimental exposure durations of less than 30 minutes or greater than 8 hours in order to determine the 4-hour LC_{50} for C&L purposes.

Nowadays a modification of Haber's Law is used ($C^n.t = k$) as for many substances it has been shown that n is not equal to 1 (Haber's Law). In case extrapolation of exposure duration is required, the *n* value should be considered. If this *n* value is not available from literature, a default value may be used. It is recommended to set n = 3 for extrapolation to shorter duration than the duration for which the LC_{50} or EC_{50} was observed and to set n = 1 for extrapolation to longer duration (ACUTEX TGD, 2006), also taking the range of approximately 30 minutes to 8 hours into account.

Experimentally, when concentration-response data are needed for specific purposes, OECD TG 403 (EU B.2) or the CxT approach could be taken into consideration. The OECD TG 403/(EU B.2 will

result in a concentration-response curve at a single exposure duration, the CxT approach will result in a concentration-time-response curve, taking different exposure durations into account. The CxT approach (under consideration for the revision of OECD TG 403) uses two animals per CxT combination and exposure durations may vary from about 15 minutes up to approximately 6 hours. This approach may provide detailed information on the concentration-time-response relationship in particular useful for risk assessment and determination of NOAEL/LOAEL.

R.7.4.4.2 Human data on acute toxicity

When available, epidemiological studies, case reports, information from medical surveillance or volunteer studies may be crucial for acute toxicity and can provide evidence of effects that are undetectable in animal studies (e.g. symptoms like nausea or headache). Nevertheless, the conduct of human studies is not recommended.

Such data could also be useful to identify particular sensitive sub-populations like new born, children, patients with diseases (in particular with chronic respiratory, e. g. asthma, BPOC).

Additional guidance should be provided on the reliability and the relevance of human studies because there are no standardised guidelines for such studies (except for odour threshold determination) and these are not usually conducted according to GLP. Such guidance is provided in Section R.4.3.3.

R.7.4.4.3 Exposure considerations on acute toxicity

Particular attention should be addressed to the potential routes of exposure in humans to select the appropriate testing strategy.

Generic aspects of data waivers based on exposure considerations are presented in Section R.5.1. Information on the role of exposure information in the testing strategies for acute toxicity is presented in Section R.7.4.6.

R.7.4.4.4 Remaining uncertainty on acute toxicity

In most cases, remaining uncertainties will exist due to the absence of valid human acute toxicity data, and so appropriate assessment factors should be applied. Toxicokinetic data could help in deriving chemical-specific interspecies assessment factors. As acute toxicity testing does not usually include clinical chemistry, haematology and detailed histopathology and functional observations, an additional assessment factor may need to be applied when a NOAEL or LOAEL from these studies is used to derive DNELs (for more guidance on the setting of DNELs for acute toxicity, see Chapter R.8, Appendix R.8-8).

R.7.4.5 Conclusions on acute toxicity

R.7.4.5.1 Concluding on suitability for Classification and Labelling

In order to achieve classification and labelling, Annex VI of the Dangerous Substances Directive 67/548/EEC[24] must be applied. The criteria for classification are based on specific 'cut offs' based on the LD_{50} or LC_{50}, although determination of a precise LD_{50} or LC_{50} value is not essential for classification purposes. This is because the LD_{50}/LC_{50} is not an absolute value (Schütz, 1969) since many factors influence its reproducibility (Zbinden and Flury-Roversi, 1981).

Ideally, classification and labelling should be achieved using data generated from studies conducted in accordance with officially adopted OECD test guidelines, or test methods incorporated for the time being into Annex V of Directive 67/548/EEC[25]. Such studies will permit identification of the LD_{50}, LC_{50}, the discriminating dose (fixed dose procedures), or a range of exposure where lethality and/or severe toxicity is expected (acute toxic class methods). For materials of low toxicity (no mortalities expected at the upper dose limit) testing is restricted to this dose level (the limit test) and if absence of mortalities is confirmed, classification of the substance with respect to acute toxicity is not required.

In the Up-and-Down Procedure (OECD TG 425), where individual animals are dosed sequentially, estimation of the LD_{50} with a confidence interval is possible and this can be used for classification purposes. Data generated in the fixed dose/concentration procedures (OECD TG 420, draft 433 and 434 and EU B.1 bis) and the acute toxic class methods (OECD TG 423, draft TG 436 and EU B.1 tris) are equally sufficient for classification purposes. In the fixed dose/concentration procedures, the discriminating dose is identified as the dose causing evident toxicity but not mortality, and must be one of the four dose levels specified in the test method. Evident toxicity is a general term describing clear signs of toxicity such that at the next highest dose level, either severe pain and enduring signs of severe distress, moribund status or probable mortality can be expected in most animals. In the acute toxic class methods, the range of exposure where death is expected is determined by testing at one or more of the four fixed doses. The OECD and EU guidelines for fixed dose procedure and acute toxic class methods include flow charts that allow conclusions to be drawn with respect to GHS classification. In addition the flow charts in the acute toxic class methods allow identification of LD_{50} or LC_{50} cut offs. In the absence of GLP compliant data generated in accordance with OECD or EU methods, all other available information should be considered. Each individual set of data (e.g. a non-GLP study) must be assessed for reliability and relevance as stated in Section R.7.4.4 and any unsuitable data (i.e. that considered unreliable or not relevant) should be disregarded. When experimental data for acute toxicity are available in several animal species, scientific judgement should be used in selecting the most relevant data from among the valid, well-performed tests. When equally reliable data from several species are available, priority should be given to the data relating to the most sensitive species, unless there are reasons to believe that this species is not an appropriate model for humans. If definitive classification and labelling cannot be achieved from any individual source, but multiple sets of data all lead to the same conclusion, then, the WoE approach might be sufficient to classify and a robust proposal detailing this should be put forward.

---

[24] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

[25] The new Test Methods Regulation is currently (February 2008) under adoption and contains all the test methods previously included in Annex V to Directive 67/548/EEC

Where evidence is available from both humans and animals and there is a conflict between the findings, the quality and reliability of the evidence from both sources shall be evaluated in order to resolve the question of classification. Generally, data of good quality and reliability in humans shall have precedence over other data. However, well designed and conducted epidemiological studies may lack the sufficient number of subjects to detect relatively rare, but nevertheless important, effects. Also, the interpretation of many studies is hampered by difficulties in identifying and taking account of confounding factors. Positive results from well-conducted animal studies are not necessarily negated by the lack of positive human experience but require an assessment of the robustness and quality of both the human and animal data.

If the existing data are contradictory, not concordant or insufficient to reliably determine the appropriate classification and labelling of the substance, additional *in vitro* studies, QSARs, read-across should be considered before conducting any OECD or EU compliant *in vivo* study. In that way *in vitro* data could have a supporting role in a read-across or chemical grouping approach. Study data, which permit an assessment of dose response relationship, should be considered for risk assessment and classification and labelling.

Of particular importance in classifying for inhalation toxicity is the use of well-articulated values in the high toxicity categories for dusts and mists. Inhaled particles between 1 and 4 microns mean mass aerodynamic diameter (MMAD) will deposit in all regions of the rat respiratory tract. This particle size range corresponds to a maximum dose of about 2 mg/L (draft OECD GD 39). In order to achieve applicability of animal experiments to human exposure, dusts and mists would ideally be tested in this range in rats. The cut off values in the table for dusts and mists allow clear distinctions to be made for materials with a wide range of toxicities measured under varying test conditions.

Currently, non-animal test data (e.g. *in vitro*, QSARs and read-across data) cannot be used as stand-alone for classification and labelling purposes, but can be used for classification to support a read-across argument. In future they might be used in different purposes when such methods have been formally validated and incorporated into official test guidelines, and when classification systems have been adapted to take account of such data.

R.7.4.5.2 Concluding on suitability for Chemical Safety Assessment

For chemical safety assessment, both standard OECD/EU test guideline data and all applicable data considered both reliable and relevant should be used. A quantitative rather than qualitative assessment is preferred to conclude on the risk posed by a substance with regards to acute toxicity dependent on the data available and the potential exposure to the substance during the use pattern/lifecycle of the substance. If quantitative data are not available, the nature and the severity of the specific acute toxic effects can be used to make specific recommendations with respect to handling and use of the substance.

Information on acute toxicity is not normally limited to availability of a LD_{50} or LC_{50} value. Additional information which is important for the chemical safety assessment will be both qualitative and quantitative and will include parameters such as the nature and severity of the clinical signs of toxicity, local irritant effects, the time of onset and reversibility of the toxic effects, the occurrence of delayed signs of toxicity, body weight effects dose response relationships (the slope of the dose response curve), sex-related effects, specific organs and tissues affected, the highest non-toxic and lowest lethal dose (adapted from ECETOC Monograph No. 6, 1985).

If a NOAEL can be identified this can be used in determination of a DNEL. However, depending upon the nature of the acute toxicity information available, this may not always be possible. For

instance, data from an OECD/EU test method may permit calculation of an LD_{50}/LC_{50} value, or identification of the range of exposure where lethality is expected, or the dose at which evident toxicity is observed, but may not provide information on the dose level at which no adverse effects on health are observed. If the data permits construction of a dose-response curve, then derivation of the NOAEL may be possible. When a limit test has been conducted, and no adverse effects on health have been observed, then the limit dose can be regarded as the NOAEL. If adverse effects on health are seen at the limit dose then it is unlikely that lower dose levels will have been investigated and in this case identification of the NOAEL will not be possible. If data is available for several species, then the most sensitive species should be chosen for the purposes of the Chemical Safety Assessment, provided it is the most relevant to humans.

If human data on acute toxicity is available, it is unlikely that this will be derived from carefully controlled studies or from a significant number of individuals. In this situation, it may not be appropriate to determine a DNEL from this data alone, but the information should certainly be considered in the WoE and may be used to confirm the validity of animal data. In addition, human data should be used in the risk assessment process to be able to determine DNEL for particular sensitive sub-populations like new-born, children or those in poor health (patients).

More extensive guidance on the setting of DNELs for acute toxicity, see Chapter R.8, Appendix R.8-8.

The anticipated effects from physico-chemical properties and bioavailability data on the acute toxicity profile of the substance must also be considered in the Chemical Safety Assessment.

R.7.4.5.3 Information not adequate

A WoE approach, comparing available adequate information with the tonnage-triggered information requirements by REACH, may result in the conclusion that the requirements are not fulfilled.

In absence of data from test guidelines or equivalent methods, data from other endpoints could be helpful for the determination of acute toxicity potential. For example, data could be provided by subchronic toxicity or neurotoxicity studies, as in general the design of these studies includes a pilot study to determine dose of departure for the main test. In order to proceed with further information gathering the following testing strategy can be adopted.

R.7.4.6 Integrated Testing Strategy (ITS) for acute toxicity

R.7.4.6.1 Objective / General principles

The main objective of this Integrated Testing Strategy (ITS) is to provide advice on how the REACH Annex VII and VIII information requirements for acute toxicity can be met using the most humane methods. If the ITS is followed, the information generated will be sufficient to make a classification decision with respect to acute toxicity hazard and may provide data for the risk assessment and DNEL derivation. In addition, assessment of acute toxicity may provide information that is valuable for the conduct of repeated dose toxicity studies, such as identification of target organ toxicity and dose selection.

By adhering to the criteria outlined in the previous chapters, informed decisions may be made on whether sufficient data already exist to cover the objectives, or whether further testing is required.

If further testing is deemed necessary, the use of the most appropriate study in accordance with the REACH proposal is considered rather than a *one study fits all* approach. An overarching principle is that all data requirements are met in the most efficient and humane manner so that animal usage and costs are minimized.

R.7.4.6.2 Preliminary considerations

The standard information requirements for acute toxicity under the REACH regulations are given in Section R.7.4.2.

According to REACH, acute toxicity studies should not be conducted if a substance is known to be corrosive. However, if there are health concerns regarding exposure to non-corrosive concentrations, then acute toxicity assessment may be considered appropriate. In such cases, a specific protocol should be developed as standard LC_{50} or any other *in vivo* acute toxicity testing cannot be performed. For example, *in vitro* data on basal cytotoxicity could be used to establish the most appropriate range of concentrations to be tested.

Regardless of tonnage level, before any testing is triggered, careful consideration of existing toxicological data, exposure characteristics and current risk management procedures is recommended to ascertain whether the fundamental objectives of the ITS have already been met. This consideration should take account of discussions that have taken place under other regulatory schemes, such as ESR, DPD, BPD and the EU hazard classification scheme. If it is concluded that further testing is required, then a series of decision points are defined to help shape the scope of an appropriate testing program.

The following four-stage process has been developed for clear decision-making:

Stage 1. gather existing information according to Annex VI

Stage 2. consider information needs according to the relevant Annex VII to X

Stage 3. identify data gaps (and adequacy of all available data for classification and labelling and/or risk assessment, or to fulfil the criteria for waiving)

Stage 4. generate new data / propose testing strategy

R.7.4.6.3 Testing strategy for acute toxicity (see Figure R.7.4-1)

Stage 1. Gathering of existing information

The starting point of the ITS is the review of existing data (e.g. human or animal data, physico-chemical properties, (Q)SARs, *in vitro* test data). For non-corrosive substances, the results of skin and eye irritation and skin sensitisation studies (Annex VII) may provide useful information on the potential for systemic toxicity.

In the ITS, all existing human and test data (e.g. from clinical reports, poisoning cases, animal studies, corrosivity, physico-chemical properties) should be considered. Some information from the existing data e.g. *in vitro* studies (*de novo in vitro* basal cytotoxicity and dermal penetration studies), systemic effects observed in other studies, route of human exposure, physico-chemical properties, dermal or respiratory toxicity of structurally-related substances, might primarily be used for the selection of either an acute *in vivo* inhalation test or an acute *in vivo* dermal test. No specific

reference is made to valid (Q)SAR models/approaches or to valid *in vitro* methods, but such data should be assessed when available or generated.

Section R.7.4.3 presents a detailed discussion of the sources that may provide relevant information for the assessment of acute toxicity.

Stage 2. Considerations on information needs

A detailed evaluation of the existing information collated in Stage 1 is conducted to allow an informed decision on the testing needs to fulfil the REACH requirements. It is important to ensure that the available data are relevant and reliable to fulfil these requirements.

It should be noted that if a substance is predicted to be corrosive then further consideration should be given as to whether or not an acute oral test can be justified (in particular in relation with animal welfare considerations). Justifications for conducting a study must be provided in order to minimise the animal use. If the substance is considered likely to be corrosive, no acute toxicity testing should normally be conducted (see above). Where information on corrosivity is not available then *in vitro* corrosivity tests should be conducted.

The standard information requirements for acute toxicity under the REACH regulations are given in Section R.7.4.2.

When acute toxicity via a second route is required, the choice of the second route (dermal or inhalation) depends on the nature of the substance and the likely route of human exposure. However, information on only one route of exposure may be sufficient and justified (based on physico-chemical, toxicokinetic or human data and review of all possible exposure scenarios; for example with gases only inhalation route could be evaluated as no relevant human exposure may occur by oral or dermal route; for liquid with high viscosity, no testing by inhalation route should be conducted).

If human exposure is possible via inhalation, or if physico-chemical properties indicate that such exposure may occur, then testing via this route for acute toxicity should be conducted. Data from skin/eye irritation, skin sensitisation and acute oral toxicity should be used as indicators to help testing via inhalation (for example, substance with only potential local toxicity; choice of exposure concentrations). If no systemic effects are shown during acute oral testing, then the requirement to conduct inhalation testing should be considered on a case-by-case basis.

Consideration of the need for assessment of acute dermal toxicity should be given if the inhalation route is not considered appropriate. In some cases, it may be possible to draw conclusions about the potential for acute dermal toxicity without further testing, on the basis of the data available from acute oral toxicity and/or dermal absorption studies. Evidence for the potential of high dermal absorption should be considered on a case-by-case basis taking into account physico-chemical properties e.g. Log Kow, water solubility, molecular weight and melting point of the substance. Testing for acute dermal toxicity is indicated if:

- Systemic toxicity is observed in skin/eye irritation and/or skin sensitisation studies;
- Death is observed in an acute oral toxicity test and there is potential for dermal absorption;
- Systemic toxicity is observed in an acute oral toxicity test and there is potential for high dermal absorption (determined following e.g. OECD TG 428, EU B.45)
- There is the potential for high dermal exposure (case-by-case basis)

Stage 3. Identification of data gaps / adequacy of data

The purpose of this step is to identify what additional information is required in order to classify the substance and to perform a risk assessment.

The available information may include data generated using study protocols that differ from the standard regulatory tests. The evaluation should include whether the available information meets or exceeds the data requirements from standard regulatory study protocols. Therefore it may be possible that the tonnage-driven minimum needs can be met through combined data obtained from several sources.

At this stage, it is also necessary to verify if the available information is adequate for hazard characterisation. For this process, all relevant information should be taken account of in a weight of evidence assessment. Quantitative data on the dose response relationship for the critical toxicological effects and/or estimations of the either the LC_{50}/LD_{50} values or the Discriminating Dose will be important for assessing the hazard classification and can be used in the risk assessment. Information from testing for other toxicological endpoints (e.g. repeated dose toxicity) may also be useful for the risk assessment (see also Chapter R.8, Appendix R.8-8). Mathematical modelling should be considered for estimating a threshold exposure level (e.g. benchmark dose), as an alternative to generating additional *in vivo* data.

For the inhalation route, standard protocols involve a 4-hour exposure. If data for other time periods are available (e.g. for 0.5 to 8 hours), extrapolation to a 4-hour exposure period can be achieved using a modification of Haber's Law ($C^n.t = k$). If this «n» value is not available from the literature, a default value may be used; it is recommended to set n = 3 for extrapolation to shorter duration than the duration for which the LC_{50} or EC_{50} was observed and to set n = 1 for extrapolation to longer duration (ACUTEX TGD, 2006). Experimentally, the value of n can be determined using the CxT approach (draft revision OECD TG 403).

If the data and subsequent decisions are deemed consistent with an adequate hazard characterisation and are sufficient to classify the substance or to conduct a risk assessment, then no further testing for acute toxicity is recommended.

In some cases, the substance may be excluded from acute toxicity testing if it does not appear as scientifically necessary (Annex XI). This might be the case for example if

- A WoE analysis demonstrates that the available information is sufficient for an adequate hazard characterisation and the exposure to the substance is adequately controlled;
- The substance is not bio-available via a specific route and possible local effects are adequately characterised (example, no dermal absorption for dermal route)
- For inhalation route, no testing is required if it is not technically possible to generate a testing atmosphere, the vapour pressure is very low (<0.1 Pa at 20°C) or the particle size is > 100 μm

Finally, the conclusion that no further testing is required may be reached when the data meet the requirements for classification for toxic effects or if the substance has already been classified for acute toxic effects.

Where evidence is available from both humans and animals and there is a conflict between the findings, the evidence should be evaluated towards understanding the toxicological basis for these divergent findings. Issues relating to the quality and reliability of the data should also be taken into

account. Generally, data of good quality and reliability in humans shall take precedence over other data. However, well-designed and conducted epidemiological studies may lack a sufficient number of subjects to detect relatively rare but still significant effects, to assess potentially confounding factors. Positive results from well-conducted animal studies are not necessarily negated by the lack of positive human experience but require an assessment of the robustness and quality of both the human and animal data.

If the remaining data are contradictory, not concordant or insufficient to determine reliably the appropriate classification and labelling of the substance, additional *in vitro* studies, QSARs, read-across should be considered before conducting any OECD compliant *in vivo* study. Study data, which permit an assessment of dose response relationship, should be considered particularly valuable for risk assessment purposes.

Stage 4. Generation of new data / proposal for testing strategy

If sufficient data for risk assessment and classification purposes are already available, no further testing will be required. If data gaps need to be filled, new data shall be generated (Annexes VII & VIII). Due to animal welfare considerations, new tests on animals should only be performed as a last resort when all other sources of information have been exhausted.

The standard OECD guidelines should normally be used as these provide the necessary information on acute toxicity hazard in a way that balances the need to protect human health with animal welfare concerns (see Section R.7.4.3 and the above guidance for Stage 3).

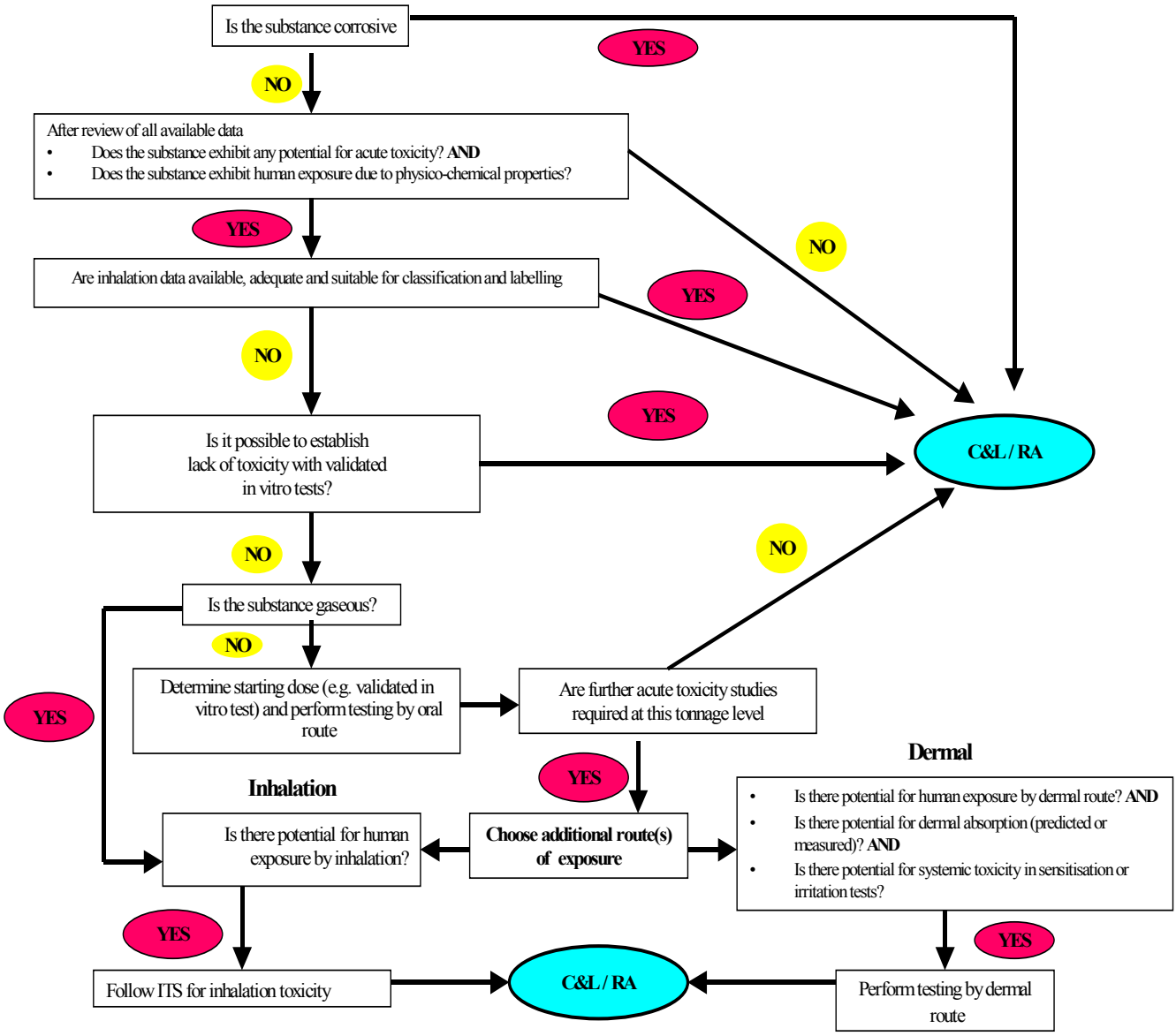
The route of exposure to be used for acute toxicity evaluation depends on the nature of the substance (e.g. gas or not, molecular weight, log K_{ow}) and should reflect the most likely route of human exposure. If any specific human exposure may be identified, further testing for risk assessment should be considered as proposed in Annex VIII. If any human exposure by inhalation is identified, then the testing strategy by inhalation should be proposed (Figure R.7.4-2).

First considerations should be based on defining the potential of the substance for acute toxicity. For such a question, information may be provided by existing data from SARs, QSARs, chemical categories approaches and available *in vitro* and *in vivo* data. If no potential for toxicity is shown, then no further testing is required and a decision on classification can be taken. Such information may also provide relevant information in risk assessment considerations.

Following the general testing strategy, dose selection appears to be an important aspect in order to select the most appropriate starting point. When validated *in vitro* tests are available, as shown by the joint ECVAM-ICCVAM study, these may provide relevant results, and help the dose selection for oral route testing (see Section R.7.4.4.1).

For substances in the ≥10 t/y tonnage band, testing by the dermal route should be considered if a human exposure is identified, or if results from physico-chemical properties and in particular skin irritation/sensitisation tests show any dermal absorption or any systemic toxicity. Depending on such information, dermal testing should be conducted or not following standard protocols (see Section R.7.4.3).

Figure R.7.4-1 ITS for acute toxicity endpoint



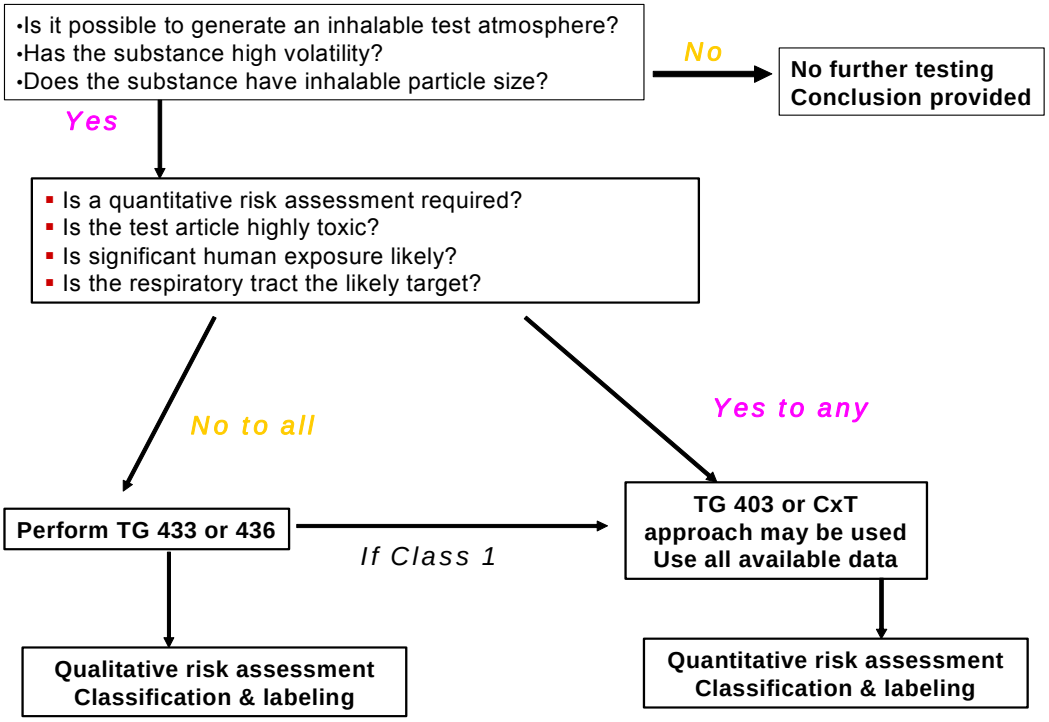
(*) if the substance is corrosive but there are health concerns regarding exposure to non-corrosive concentrations, then acute toxicity assessment may be considered appropriate
(**) Testing by inhalation may be required if the substance is a gas, a liquid or a solid with a high vapour pressure, or a solid with inhalable particle size (particular substances in powder form nanoparticles, fibres…)

A specific testing strategy (Figure R.7.4-2) is proposed for the inhalation route. Primary considerations should be based on the in(ability) to generate a suitable atmosphere depending on the physico-chemical properties (for example, low volatility, solid, particle size >100 µm (see also Section R.7.4.4.1). In this situation, no human exposure may be identified and no further testing is required.

Wherever possible, assessment of acute inhalation toxicity should be conducted in accordance with OECD TG's 433 and 436 (official adoption in process) since they have been designed to use less animals than OECD TG 403 and EU B.2. In addition, OECD TG 433 does not require mortality as

endpoint. However, in some circumstances, i.e. if a dose response curve is needed for risk assessment purposes, testing according to OECD TG 403, EU B.2 or the CxT approach may be considered appropriate (see also draft OECD Guidance Document 39).

Figure R.7.4-2 ITS for acute inhalation toxicity endpoint (see also draft OECD GD 39)



R.7.4.7 References on acute toxicity

ACUTEX project. 2006. Technical guidance document (TGD) methodology to develop AETLs. (www.acutex.info)

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

Cope AC, Hancock EM (1939). Substituted Vinyl Barbituric Acids III. Derivatives Containing a Danish EPA (2001). Report on the Advisory list for self-classification of dangerous substances, Environmental Project no. 636, 2001. http://www.mst.dk/udgiv/publications/2001/87-7944-694-9/html/default_eng.htm

Dialkylvinyl Group having Five or More Carbon Atoms. J Am Chem Soc 61, 776-779

Cronin MTD, Dearden JC (1995). QSAR in toxicology. 2. Prediction of acute mammalian toxicity and interspecies correlations. Quant Struct-Act Relat 14, 117-120.

Cronin MTD (1996). Quantitative Structure-Activity Relationship (QSAR) Analysis of the Acute Sublethal Neurotoxicity of Solvents. Toxicology in vitro 10, 103-110

Cronin MTD, Dearden JC, Walker JD, Worth, AP (2003). Quantitative structure-activity relationships for human health effects: commonalities with other endpoints. Environ Toxicol Chem 22, 1829-1843

Duff T, Carter S, Feldman G, McEwan G, Pfaller W, Rhodes P, Ryan M, Hawksworth G. Transepithelial resistance and inulin permeability as endpoints in vitro nephrotoxicity testing. Altern Lab Anim. 2002 Dec;30 Suppl 2:53-9

ECETOC. 1985. Acute toxicity Tests, LD_{50} (LC_{50}) Determinations and Alternatives, Monograph No.6

ECETOC. 2004. Workshop on the use of human data in risk assessment; workshop report n°3; 23-24 February 2004, Cardiff.

ECVAM workshop report 16, 1996

European Commission. 2005. EU Technical Guidance Document in Support of Commission Directive 93/67/EEC on Risk Assessment for new notified substances and commission regulation (EC) No 1488/94 on Risk Assessment for existing substances. http://ecb.jrc.it

European Commission. 2005. European Technical Guidance Document on Risk Assessment, Human Risk Characterisation (revised chapter, final draft November 2005).

Frantik, E, Hornychova M, Horvath M (1994). Relative acute neurotoxicity of solvents: isoeffective air concentrations of 48 compounds evaluated in rats and mice. Environ Res 66, 173-185.8.

Hansch C, Kurup (2003). A. QSAR of chemical polarisability and nerve toxicity.hit3 2. J Chem Inf Comp Sci, 43, 1647-1651

Hostýnek,JJ (1998). Structure-Activity Relationships in Percutaneous Absorption. In: Dermatotoxicology Methods The Laboratory Worker's *Vade Mecum.* Taylor & Francis.

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), National Toxicology Program (NTP), Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Public Health Service, Department of Health and Human Services (2006). Peer Review Panel Report: The Use of In Vitro Basal Cytotoxicity Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing, http://iccvam.niehs.nih.gov/methods/acutetox/inv_nru_brd.htm

King LA (1985). Fergusons principle and the prediction of fatal drug levels in blood. Human Toxicol 4, 273-2

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

Lessigiarska I, Worth AP, Netzeva TI. (2005). Comparative review of QSARs for acute toxicity. EUR 21559 EN

Lessigiarska I, Worth AP, Netzeva TI, Dearden JC, Cronin MTD (2006). Quantitative structure-activity-activity and quantitative structure-activity investigations of human and rodent toxicity. Chemosphere, in press

Moridani MY, Siraki A, O'Brien PJ (2003). Quantitative structure toxicity relationships for phenols in isolated rat hepatocytes. Chem Biol Interact 145, 213-223

OECD. Organisation for Economic Co-operation and Development. Guidelines for the Testing of Chemicals 402, 403, 420,423,425, 433,434, 436.

OECD. Organization for Economic Co-operation and Development. Guidance Document n°19 on the recognition, assessment, and use of clinical signs as humane health endpoints for experimental animals used in safety evaluation. November 2000.

OECD. Organization for Economic Co-operation and Development. Draft Guidance Document n°39 on acute inhalation toxicity testing.

Schütz, E. (1969). On acute oral toxicity tests. Amer. Perf. Cosmet., 84,41

Selassie CD, Shusterman AJ, Kapur S, Verma RP, Zhang L, Hansch C (1999). On the Toxicity of Phenols to Fast Growing Cells. A QSAR Model for a Radical-Based Toxicity. Perkin Transactions 2, 2729-2733

Tsakovska I, Lessigiarska I, Netzeva T, Worth A. (2006). Review of QSARs for mammalian toxicity, 22486 EN

UNECE. 2005. Globally Harmonised System of Classification and Labelling of Chemicals (GHS). http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html

Zbinden, G. and Flury-Roversi, M. (1981). Significance of the LD50 test for the toxicological evaluation of chemical substances. Arch. Toxicol., 47, 77.

R.7.5 Repeated dose toxicity

R.7.5.1 Introduction

Repeated dose toxicity studies provide information on possible adverse general toxicological effects likely to arise from repeated exposure to a substance. Furthermore, these studies may provide information on e.g. reproductive toxicity and carcinogenicity, even though they are not specifically designed to investigate these endpoints.

Organs and tissues investigated in repeated dose toxicity studies include vital organs such as heart, brain, liver, kidneys, pancreas, spleen, immune system, lungs etc. Effects examined may include changes in morphology, physiology, growth or life span, behaviour which result in impairment of functional capacity or impairment of capacity to compensate for additional stress or increase in the susceptibility to the harmful effects of other environmental influences. Therefore, it is important that the possible adverse general toxicological effects are assessed for chemical substances that may be present in the environment.

R.7.5.1.1 Definition of repeated dose toxicity

The term *repeated dose toxicity* comprises the general toxicological effects occurring as a result of repeated daily dosing with, or exposure to, a substance for a part of the expected lifespan (sub-acute or sub-chronic exposure) or for the major part of the lifespan, in case of chronic exposure.

The term *general toxicological effects* (in this report often referred to as *general toxicity*) includes effects on, e.g. body weight and/or body weight gain, absolute and/or relative organ and tissue weights, alterations in clinical chemistry, urinalysis and/or haematological parameters, functional disturbances in the nervous system as well as in organs and tissues in general, and pathological alterations in organs and tissues as examined macroscopically and microscopically. Repeated dose toxicity studies may also examine parameters, which have the potential to identify specific manifestations of toxicity such as e.g., neurotoxicity, immunotoxicity, endocrine-mediated effects, reproductive toxicity and carcinogenicity.

An *adverse effect* is a change in the morphology, physiology, growth, development, reproduction or life span of an organism, system, or (sub) population that results in an impairment of functional capacity, or an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences (OECD, 2003).

A chemical substance may induce systemic and/or local effects.

- A *local effect* is an effect that is observed at the site of first contact, caused irrespective of whether a substance is systemically available.
- A *systemic effect* is defined as an effect that is normally observed distant from the site of first contact, i.e., after having passed through a physiological barrier (mucous membrane of the gastro-intestinal tract or of the respiratory tract, or the skin) and becomes systemically available.
- It should be noted, however, that toxic effects on surface epithelia may reflect indirect effects as a consequence of systemic toxicity or secondary to systemic distribution of the substance or its active metabolite(s).

R.7.5.1.2 Objective of the guidance on repeated dose toxicity

The objectives of assessing repeated dose toxicity are to evaluate:

- whether exposure of humans to a substance has been associated with adverse toxicological effects occurring as a result of repeated daily exposure for a part of the expected lifetime or for the major part of the lifetime; these human studies potentially may also identify populations that have higher susceptibility;
- whether administration of a substance to experimental animals causes adverse toxicological effects as a result of repeated daily exposure for a part of the expected lifespan or for the major part of the lifespan; effects that are predictive of possible adverse human health effects;
- the target organs, potential cumulative effects and the reversibility of the adverse toxicological effects;
- the dose-response relationship and threshold for any of the adverse toxicological effects observed in the repeated dose toxicity studies;
- the basis for risk characterisation and classification and labelling of substances for repeated dose toxicity.

R.7.5.2 Information requirements for repeated dose toxicity

Section R.2.1 provides general guidance on the information requirements of REACH. For repeated dose toxicity, all available information relevant for the endpoint needs to be evaluated and classification considered at each tonnage level. The following standard information requirements on repeated dose toxicity are specified in REACH Annexes VII-X:

In **Annex VII** (≥ 1 t/y), no test requirements on repeated dose toxicity are specified additional to the available information relevant for repeated dose toxicity.

In **Annex VIII** (≥ 10 t/y), a short-term repeated dose toxicity study (28 days) is usually required, in one species, male and female, using the most appropriate route of administration, having regard to the likely route of human exposure.

In **Annex IX** (≥ 100 t/y), a sub-chronic repeated dose toxicity study (90-days) is usually required, in one species (90-day study: rodent), male and female, and a short-term repeated dose toxicity study (28 days) is the minimum requirement, using the most appropriate route of administration, having regard to the likely route of human exposure. It should be noted that the 28-day test is not required at this tonnage level if already provided as part of Annex VIII requirements or if the 90-day study is proposed at this tonnage level.

In **Annex X** (≥ 1000 t/y), no specific test requirements additional to those required in Annexes VIII-IX for repeated dose toxicity is required at this tonnage level.

Column 1 of the REACH Annexes VII to X establishes the standard information required for all chemical substances and Column 2 lists specific rules according to which the required standard information requirements for individual endpoints may be modified (adapted) by waiving requirement for certain information, or in certain cases, defining the need for additional or different information. (see Section R.2.1 for further details).

In addition to the specific rules for adaptation listed in column 2 of the Annexes VII to X, the required standard information may also be adapted according to Annex XI, which specifies general

rules for adaptation of the standard testing requirements set out in Annexes VII-X in cases where 1) testing does not appear scientifically necessary, 2) testing is technically not possible, and 3) testing may be omitted based on the exposure scenarios developed in the CSA (substance-tailored exposure-driven testing) (see Section R.5.1 (Exposure based waiving)).

It should also be noted that the introductory sections to Annexes VII-X point at a specific adaptation to the standard information requirements as *in vivo* testing shall be avoided with corrosive substances at concentration/dose levels causing corrosivity.

Factors that can influence the standard information requirements include the results of other toxicity studies, immediate disintegration of the substance, accumulation of the substance or its metabolites in certain tissues and organs, failure to identify a NOAEL in the required test at a given tonnage level, toxicity of particular concern, exposure route, structural relationships with a known toxic substance, physico-chemical properties of the substance, and use and human exposure patterns. These adaptations are detailed in the stepwise ITS presented in Section R.7.5.6.

R.7.5.3 Information and its sources on repeated dose toxicity

Toxicological information, including repeated dose toxicity, can be obtained from unpublished studies, data bases and publications such as books, scientific journals, criteria documents, monographs and other publications (see Chapter R.3 for further general guidance). Information relevant for repeated dose toxicity can also be obtained from data on other endpoints, structural analogues and physico-chemical properties.

Before new tests are carried out to determine the hazardous properties of a chemical substance, all available information, shall be assessed, according to REACH Annex VI, step 1. (See Chapter R.4 for general guidance on evaluation of information).

R.7.5.3.1 Non-human data on repeated dose toxicity

a) Non-testing data on repeated dose toxicity

Physico-chemical data

The physico-chemical properties of a chemical substance are essential elements in deciding on the appropriate administration route to be applied in experimental *in vivo* repeated dose toxicity studies as well as to decide on exemption from testing in cases where testing is technically not possible.

(Q)SAR models

The OECD has recently prepared a report on the use of (Q)SAR in the various member countries (OECD, 2006), which provides clear insight in how these tools are being used in the various OECD member countries. A review conducted by ECETOC on the use of (Q)SARs within current regulatory decision-making frameworks in EU, North America, and Japan, and within industry concluded that applicability of currently available (Q)SARs for chronic mammalian toxicity, certainly as a stand-alone approach, was very limited at that time (ECETOC 2003).

The ECB has started building a freely accessible inventory of evaluated (Q)SAR models which help to identify valid (Q)SARs for regulatory purposes (see also cross cutting guidance on (Q)SARs). If there are any models relevant for the underlying endpoint these will be included in the ECB inventory.

More extensive guidance on the availability and application of (Q)SARs is available in Section R.6.1.

Structurally or mechanistically related substance(s) (read-across/chemical category)

The concept of grouping, including both read-across and the related chemical category concept has been developed under the OECD HPV program (OECD 2007a. This is an approach which might be used to fill data gaps without the need for conducting tests when specific conditions, as specified in REACH Annex XI Section 1.3, are met.

Extensive guidance on the application of chemical categories/read across is available in Section R.6.2.

Testing data on repeated dose toxicity

In vitro data

Currently, no available alternatives to animal testing are accepted for regulatory purposes for detecting toxicity after repeated exposure. Numerous *in vitro* systems have been developed over the last decades and have been discussed and summarized in recent ECVAM reports on repeated dose toxicity testing (Worth & Balls 2002, Prieto et al., 2005, and Prieto et al., 2006). At present, the *in vitro* models listed in these reports are at research and development level and cannot be used for repeated dose toxicity predictive purposes, although they are very useful to study individual types of organ toxicity or in assessing mechanistic aspects of target organ toxicity, on the tissue, cellular and molecular level. Some of the drawbacks are for instance the limited possibilities of current cell culture systems to account for kinetics and biotransformation, and the difficulty to derive from *in vitro* systems values such as NOAELs. Further development and optimisation of current *in vitro* systems as well as the selection of endpoints relevant to general as well as cell-type-specific mechanisms of toxicity or expression of toxic effects *in vivo* is ongoing. New technologies such as genomics, transcriptomics, proteomics and metabolomics could help in the identification of specific markers of toxicity that occur early in the process of long-term toxic responses and that are mechanistically linked to the underlying pathology. A recent ECVAM workshop report (Prieto et al., 2006) includes a proposed approach to assess repeated dose toxicity *in vitro* by integrating physiologically-based kinetic (PBK) modelling, the use of biomarkers, and omics technologies. However, this integrated approach is still under development and evaluation and is not ready for regulatory purposes.

The latest information on the status of alternative methods that are under development can be obtained from the ECVAM website (current address: http://ecvam.jrc.it) and other international centres for validation of alternative methods.

Human *in vitro* data, particularly on kinetics and metabolism, may assist in study interpretation thereby avoiding the need for unnecessary animal experimentation.

At present, available *in vitro* test data from well-characterised target organ and target system models on, e.g. mode of action(s) / mechanism(s) of toxicity may be useful in the interpretation of observed repeated dose toxicity.

Animal data

The most appropriate data on repeated dose toxicity for use in hazard characterisation and risk assessment are primarily obtained from studies in experimental animals conforming to internationally agreed test guidelines. In some circumstances repeated dose toxicity studies not conforming to conventional test guidelines may also provide relevant information for this endpoint.

The information that can be obtained from the available EU/OECD test guideline studies for repeated dose toxicity is briefly summarised below. Table R.7.5-2 summarises the parameters examined in these OECD test guideline studies in more detail to facilitate overview of the similarities and differences between the various studies. It should be noted that the test guidelines given in Annex V to Directive 67/548/EEC[26] (http://ecb.jrc.it/testing-methods/) are generally comparable to the OECD test guidelines (http://www.oecd.org/env/testguidelines). Further details of the study protocols are described in the respective test guidelines.

Repeated dose 28-day toxicity studies:

Separate guidelines are available for studies using oral administration (EU B.7 / OECD TG 407), dermal application (EU B.9 / OECD TG 410), or inhalation (EU B.8 / OECD TG 412). The principle of these study protocols is identical although the OECD TG 407 protocol includes additional parameters compared to those for dermal and inhalation administration, enabling the identification of a neurotoxic potential, immunological effects or reproductive organ toxicity.

The 28-day studies provide information on the toxicological effects arising from exposure to the substance during a relatively limited period of the animal's life span.

Repeated dose 90-day toxicity studies:

Separate guidelines are available for studies using oral administration (OECD TG 408/409 / EU B.26/B.27 in rodent/non-rodent species, respectively), dermal application (OECD TG 411/EU B.28), or inhalation (OECD TG 413/EU B.29). The principle of these study protocols is identical although the revised OECD TG 408 protocol includes additional parameters compared to those for dermal and inhalation administration, enabling the identification of a neurotoxic potential, immunological effects or reproductive organ toxicity.

The 90-day studies provide information on the general toxicological effects arising from sub-chronic exposure (a prolonged period of the animal's life span) covering post-weaning maturation and growth well into adulthood, on target organs and on potential accumulation of the substance.

Chronic toxicity studies:

The chronic toxicity studies (OECD TG 452/EU B.30) provide information on the toxicological effects arising from repeated exposure over a prolonged period of time covering the major part of the animal's life span. The duration of the chronic toxicity studies should be at least 12 months.

The combined chronic toxicity / carcinogenicity studies (OECD TG 453/EU B.33) include an additional high-dose satellite group for evaluation of pathology other than neoplasia. The satellite group should be exposed for at least 12 months and the animals in the carcinogenicity part of the study should be retained in the study for the majority of the normal life span of the animals.

[26] All the test methods previously included in Annex V to Directive 67/548/EEC will be incorporated in a new Test Methods (TM) Regulation that is currently (February 2008) under adoption. The TM Regulation will be adapted to technical progress whenever a new test method has been developed, scientifically validated and accepted for regulatory use by the National Coordinators of the Member states

Ideally, the chronic studies should allow for the detection of general toxicity effects (physiological, biochemical and haematological effects etc.) but could also inform on neurotoxic, immunotoxic, reproductive and carcinogenic effects of the substance. However, in 12-month studies, non-specific life shortening effects, which require a long latent period or are cumulative, may possibly not be detected in this study type. In addition, the combined study will allow for detection of neoplastic effects and a determination of a carcinogenic potential and the life-shortening effects.

The combined repeated dose toxicity study with the reproduction/ developmental toxicity screening test:

The combined repeated dose toxicity / reproductive screening study (OECD TG 422[27]) provides information on the toxicological effects arising from repeated exposure (generally oral exposure) over a period of about 6 weeks for males and approximately 54 days for females (a relatively limited period of the animal's life span) as well as on reproductive toxicity. For the repeated dose toxicity part, the OECD TG 422 is in concordance with the OECD TG 407/EU B.7 except for use of pregnant females and longer exposure duration in the OECD TG 422 compared to the OECD TG 407/EU B.7.

Neurotoxicity studies:

The neurotoxicity study in rodents (OECD TG 424/EU B.43) has been designed to further characterise potential neurotoxicity observed in repeated dose systemic toxicity studies. The neurotoxicity study in rodents will provide detailed information on major neuro-behavioural and neuro-pathological effects in adult rodents.

Delayed neurotoxicity studies of organophosphorus substances:

The delayed neurotoxicity study (OECD TG 419/ EU Annex B.38) is specifically designed to be used in the assessment and evaluation of the neurotoxic effects of organophosphorus substances. This study provides information on the delayed neurotoxicity arising from repeated exposure over a relatively limited period of the animal's life span.

Other studies providing information on repeated dose toxicity:

Although not aiming at investigating repeated dose toxicity per se, other available OECD/EU test guideline studies involving repeated exposure of experimental animals may provide useful information on repeated dose toxicity. These studies are summarised in Table R.7.5-1.

It should be noted that the repeated dose toxicity studies, if carefully evaluated, may provide information on potential reproductive toxicity and on carcinogenicity (e.g., pre-neoplastic lesions).

The one- and two-generation studies (OECD TG 415/416/EU B.34/B.35) may provide information on the general toxicological effects arising from repeated exposure over a prolonged period of time (about 90 days for parental animals) as clinical signs of toxicity, body weight, selected organ weights, and gross and microscopic changes of selected organs are recorded.

The prenatal developmental toxicity study (OECD TG 414/EU B.31), the reproduction/developmental toxicity screening study (OECD TG 421[28]) and the developmental neurotoxicity study (draft OECD TG 426[28]) may give some indications of general toxicological

[27] To date there is no corresponding EU testing method available.

[28] To date there is no corresponding EU testing method available.

effects arising from repeated exposure over a relatively limited period of the animals life span as clinical signs of toxicity and body weight are recorded.

The carcinogenicity study (OECD TG 451/EU B.32) will, in addition to information on neoplastic lesions, also provide information on the general toxicological effects arising from repeated exposure over a major portion of the animal's life span as clinical signs of toxicity, body weight, and gross and microscopic changes of organs and tissues are recorded.

Table R.7.5-1 Overview of other *in vivo* test guideline studies giving information on repeated dose toxicity

Test	Design	Endpoints (general toxicity)
OECD TG 416 (EU B.35) Two-generation reproduction toxicity study	Exposure before mating for at least one spermatogenic cycle until weaning of 2nd generation At least 3 dose levels plus control At least 20 parental males and females per group	Clinical observations Body weight and food/water consumption Gross necropsy (all parental animals) Organ weights (reproductive organs, brain, liver, kidneys, spleen, pituitary, thyroid, adrenal glands, and known target organs) Histopathology (reproductive organs, previously identified target organ(s) - at least control and high-dose groups
OECD TG 415 (EU B.34) One-generation reproduction toxicity Study	Exposure before mating for at least one spermatogenic cycle until weaning of 1st generation At least 3 dose levels plus control At least 20 parental males and females per group	As in TG 416
OECD TG 414 (EU B.31) Prenatal developmental toxicity study	Exposure at least from implantation to one or two days before expected birth At least 3 dose levels plus control At least 20 pregnant females per group	Clinical observations Body weight and food/water consumption Macroscopical examination all dams for any structural abnormalities or pathological changes, which may have influenced the pregnancy
OECD TG 421[29] Reproduction/ developmental toxicity screening test	Exposure from 2 weeks prior to mating until at least post-natal day 4 At least 3 dose levels plus control At least 8-10 parental males and females per group	Clinical observations Body weight and food/water consumption Gross necropsy (adult animals, special attention to reproductive organs) Organ weights (all adult males: testes, epididymides) Histopathology (reproductive organs in at least control and high-dose groups)
OECD TG 426[29] Developmental neurotoxicity study (draft)	Exposure at least from implantation throughout lactation (PND 20) At least 3 dose levels plus control At least 20 pregnant females per group	Clinical observations Body weight and food/water consumption
OECD TG 451 (EU B.32) Carcinogenicity studies	Exposure for majority of normal life span At least 3 dose levels plus control At least 50 males and females per group	Clinical observations (special attention to tumour development) Body weight and food consumption Gross necropsy

[29] To date there is no corresponding EU testing method available.

Test	Design	Endpoints (general toxicity)
		Histopathology (all groups - all grossly visible tumours or lesions suspected of being tumours; at least control and high-dose groups - brain, pituitary, thyroid, parathyroid, thymus, lungs, heart, salivary glands, liver, spleen, kidneys, adrenals, oesophagus, stomach, duodenum, jejunum, ileum, cecum, colon, rectum, uterus, urinary bladder, lymph nodes, pancreas, gonads, accessory sex organs, female mammary gland, skin, musculature, peripheral nerve, spinal cord, sternum with bone marrow and femur, eyes)

R.7.5.3.2 Human data on repeated dose toxicity

Human data adequate to serve as the sole basis for the hazard and dose-response assessment are rare. When available, reliable and relevant human data are preferable over animal data and can contribute to the overall *Weight of Evidence*. However, human volunteer studies are not recommended due to practical and ethical considerations involved in deliberate exposure of individuals to chemicals.

The following types of human data may already be available, however:

- Analytical epidemiology studies on exposed populations. These data may be useful for identifying a relationship between human exposure and effects such as biological effect markers, early signs of chronic effects, disease occurrence, or long-term specific mortality risks. Study designs include case control studies, cohort studies and cross-sectional studies.
- Descriptive or correlation epidemiology studies. They examine differences in disease rates among human populations in relation to age, gender, race, and differences in temporal or environmental conditions. These studies may be useful for identifying priority areas for further research but not for dose-response information.
- Case reports describe a particular effect in an individual or a group of individuals exposed to a substance. Generally case reports are of limited value for hazard identification, especially if the exposure represents single exposures, abuse or misuse of certain substances.
- Controlled studies in human volunteers. These studies, including low exposure toxicokinetic studies, might also be of use in risk assessment.

Meta-analysis. In this type of study data from multiple studies are combined and analysed in one overall assessment of the relative risk or dose-response curve.

R.7.5.3.3 Exposure considerations on repeated dose toxicity

Information on exposure, use and risk management measures should be collected in accordance with Article 10 and Annex VI (Section 3) of REACH.

Such information may lead to adaptation of the extent and nature of information needed on repeated dose toxicity under REACH; three types of *adaptations* are possible due to exposure considerations: exposure-based waiving of a study, exposure-based triggering of further studies, or definition of appropriate exposure route.

More detailed guidance of exposure-based adaptations of the repeat dose toxicity information requirements is given in Sections R.7.5.4 (evaluation of available information) and R.7.5.6 (Integrated testing strategy).

R.7.5.4 Evaluation of available information on repeated dose toxicity

General guidance on how to evaluate the available information is given in Chapter R.4.

R.7.5.4.1 Non-human data on repeated dose toxicity

Non-testing data on repeated dose toxicity

Physico-chemical properties

The physico-chemical properties of a chemical substance under registration should always be considered before any new experimental *in vivo* repeated dose toxicity studies are undertaken.

The physico-chemical properties of a substance can indicate whether it is likely that the substance can be absorbed following exposure to a particular route and whether it (or an active metabolite) is likely to reach the target organ(s) and tissue(s). The physico-chemical properties are thus essential elements in deciding on the appropriate administration route to be applied in experimental *in vivo* repeated dose toxicity studies (see Section R.7.5.4.3).

The physico-chemical properties are also important in order to judge whether testing is technically possible. Testing for repeated dose toxicity may, as specified in Annex XI Section 2 of REACH, be omitted if it is technically not possible to conduct the study as a consequence of the properties of the substance, e.g. very volatile, highly reactive or unstable substances cannot be used, or mixing of the substance with water may cause danger of fire or explosion. The Annex further emphasises that the guidance given in the test methods referred to in REACH Article 13 (3), more specifically on the technical limitations of a specific method, shall always be respected.

Additional generic guidance on the use of physico-chemical properties is provided e.g. in Section R.7.12 on toxicokinetics.

Read-across to structurally or mechanistically similar substances (SAR)

The potential toxicity of a substance, for which no data are available on a specific endpoint can, in some cases, be evaluated by read-across from structurally or mechanistically related substances for which experimental data exists. The read-across approach is based on the principle that structurally and/or mechanistically related substances may have similar toxicological properties. Note that there are no formal criteria to identify structural alerts for repeated dose toxicity or for read-across to closely related substances.

Based on structural similarities between different substances, the repeated dose toxicity potential of one substance or a group of substances can be extended (read-across) to a substance, for which there are no or limited data on this endpoint.

A mechanism of toxicity or mode of action identified for a substance and/or group of substances and causally related to adverse effects in a target organ can be extended (read-across) to a substance for which a similar mechanism or mode of action has been identified, but where no or limited data on repeated dose toxicity are available. In such cases, the substance under evaluation may reasonably be expected to exhibit the same pattern of toxicity in the target organ(s) and tissue(s).

The chemical category concept has been developed under the OECD HPV programme (OECD 2004) as an approach to fill data gaps without the need for conduction of tests. A chemical category is a group of chemicals whose physico-chemical and toxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity. In the category approach, not every substance needs to be tested for every endpoint. However, the information finally compiled for the category must prove adequate to support a hazard assessment, a risk assessment and a classification for the category and its members. That is, the final data set must allow one to assess the untested endpoints, ideally by interpolation between and among the category members.

When analogue data are used to fill the data gaps for repeated dose toxicity, the data for the analogues must be compared and discussed in relation to the substance under evaluation in order to shed light on the similarities and differences in the toxicological profile of the substance under evaluation and its analogue(s).

Specific guidance regarding use of analogues is available in Section R.6.2 in order to decide on when further *in vivo* repeated dose toxicity studies shall be proposed (Annex VIII) or may be proposed (Annex X) as well as to decide on when analogue data can replace *in vivo* testing (Annex XI Section 1.3).

(Q)SAR

A (Q)SAR analysis for a substance may give indications for a specific mechanism to occur and identify possible organ or systemic toxicity upon repeated exposure. The reliability, applicability and overall scope of (Q)SAR science to identify chemical hazard and assist in risk assessment have been evaluated by various groups and organizations. Guidance on this issue is presented in Section R.6.1 of this document and in OECD Monograph No. 69. (OECD 2007b).

Overall, (Q)SAR approaches are currently not well validated for repeated dose toxicity and consequently no firm recommendations can be made concerning their routine use in a testing strategy in this area. There are a large number of potential targets/mechanisms associated with repeated dose toxicity that today cannot be adequately covered by a battery of (Q)SAR models. Therefore, a negative result from current (Q)SAR models without other supporting evidence cannot be interpreted as demonstrating a lack of a toxicological hazard or a need for hazard classification. Another limitation of QSAR modelling is that dose-response information, including the N(L)OAEL, is not provided. Similarly, a validated QSAR model might identify a potential toxicological hazard, but because of limited confidence in this approach, such a result would not be adequate to support hazard classification.

In some cases, QSAR models could be used as part of a *Weight of Evidence* approach, when considered alongside other data, provided the applicability domain is appropriate. Also, QSAR's can be used as supporting evidence when assessing the toxicological properties by read-across within a substance grouping approach, providing the applicability domain is appropriate. Positive and negative QSAR modelling results can be of value in a read-across assessment and for classification purposes.

Testing data on repeated dose toxicity

In vitro data

As mentioned earlier in Section R.7.5.3.1 available *in vitro* data, at present, is not useful on its own for regulatory decisions such as risk assessment and C&L. However, such data may be helpful in the assessment of repeated dose toxicity, for instance to detect local target organ effects and/or to clarify the mechanisms of action. Since, at present, there are not validated and regulatory accepted

in vitro methods, the quality of each of these studies and the adequacy of the data provided should be carefully evaluated.

Generic guidance is given in Chapters R.4 and R.5 for judging the applicability and validity of the outcome of various study methods, assessing the quality of the conduct of a study, reproducibility of data and aspects such as vehicle, number of replicates, exposure/incubation time, GLP-compliance or comparable quality description.

Animal data

The basic concept of repeated dose toxicity studies to generate data on target organ toxicity following sub-acute to chronic exposure is to treat experimental animals for 4 weeks, 13 weeks or longer. These studies are mentioned in Section R.7.5.3.1 and summarised in Table R.7.5-2. In addition, other studies performed in experimental animals may provide useful information on repeated dose toxicity. While at this time most alternative methods remain in the research and development stage and are not ready as surrogates for sub-chronic/chronic animal studies there are opportunities to improve data collection for risk assessment providing greater efficiency and use of fewer animals and better use of resources. Although not required by REACH, other opportunities include early development of kinetic data, in conjunction with early repeat dose toxicity testing thus ensuring that the maximum amount of information is drawn from the animal studies and for use in the risk assessment process.

The number of repeated dose toxicity studies available for a substance under registration is likely to be variable, ranging from none, a dose-range finding study, a 28-day repeated dose toxicity guideline study, to a series of guideline studies for some substances, including sub-chronic and/or chronic studies. There may also be studies employing different species and routes of exposure. In addition, special toxicity studies investigating further the nature, mechanism and/or dose-relationship of a critical effect in a target organ or tissue may also have been performed for some substances.

The following general guidance is provided for the evaluation of repeated dose toxicity data and the development of the *Weight of Evidence*:

- Studies on the most sensitive animal species should be selected as the significant ones, unless toxicokinetic and toxicodynamic data show that this species is less relevant for human risk assessment.
- Studies using an appropriate route, duration and frequency of exposure in relation to the expected route(s), frequency and duration of human exposure have greater weight.
- Studies enabling the identification of a NOAEL, and a robust hazard identification have a greater weight.
- Studies of a longer duration should be given greater weight than a repeated dose toxicity study of a shorter duration in the determination of the most relevant NOAEL.
- If sufficient evidence is available to identify the critical effect(s) (with regard to the dose-response relationship(s) and to the relevance for humans), and the target organ(s) and/or tissue(s), greater weight should be given to specific studies investigating this effect in the identification of the NOAEL. The critical effect can be a local as well as a systemic effect.

While data available from repeated dose toxicity studies not performed according to conventional guidelines and/or GLP may still provide information of relevance for risk assessment and classification and labelling such data require extra careful evaluation. REACH Annex XI

specifically identifies circumstances where use of existing studies not carried out according to GLP or test methods referred to in Article 13(3) (guideline studies) can replace *in vivo* testing performed in accordance with Article 13(3). Data from non-guideline studies shall be considered to be equivalent to data generated by corresponding test methods referred to in Article 13(3) if the following conditions are met:

- adequate for the purpose of classification and labelling and/or risk assessment,
- adequate and reliable coverage of the key parameters foreseen to be investigated in the corresponding test methods referred to in Article 13(3),
- exposure duration comparable to or longer than the corresponding test methods referred to in REACH Article 13(3) if exposure duration is a relevant parameter, and
- adequate and reliable documentation of the study is provided.

In all other situations, non-guideline studies may contribute to the overall weight of the evidence but cannot stand alone for a hazard and risk assessment of a substance and thus, cannot serve as the sole basis for an assessment of repeated dose toxicity as well as for exempting from the standard information requirements for repeated dose toxicity at a given tonnage level, i.e. cannot be used to identify a substance as being adequately controlled in relation to repeated dose toxicity.

If sufficient information from existing studies is available on the repeated dose toxicity potential of a substance in order to perform a risk assessment as well as to conclude on classification and labelling for repeated dose toxicity (R48), no further *in vivo* testing is needed. The existing information is considered sufficient when, based on a *Weight of Evidence* analysis, the critical effect(s) and target organ(s) and tissue(s) can be identified, the dose-response relationship(s) and NOAEL(s) and/or LOAEL(s) for the critical effect(s) can be established, and the relevance for human beings can be assessed.

It should be noted that potential effects in certain target organs (e.g., the thyroid) following repeated exposure may not be observed within the span of the 28-day study. Attention is also drawn to the fact that the protocols for the oral 28-day and 90-day studies include additional parameters compared to those for the 28-day and 90-day dermal and inhalation protocols.

Where it is considered that the existing data as a whole is inadequate to provide a clear assessment of this endpoint, the need for further testing should be considered in view of all available relevant information on the substance, including use pattern, the potential for human exposure, physico-chemical properties, and structural alerts. The testing strategy is presented in Section R.7.5.6.3.

Specific investigations such as studies for neurotoxicity or immunotoxicity are also elements in the testing strategy presented in REACH.

Regarding neurotoxicity and immunotoxicity, standard oral 28-day and 90-day toxicity studies include endpoints capable of detecting such effects. Indicators of neurotoxicity include clinical observations, a functional observational battery, motor activity assessment and histopathological examination of spinal cord and sciatic nerve. Indicators of immunotoxicity include changes in haematological parameters, serum globulin levels, alterations in immune system organ weights such as spleen and thymus, and histopathological changes in immune organs such as spleen, thymus, lymph nodes and bone marrow. Where data from standard oral 28-day and 90-day studies identify evidence of neurotoxicity or immunotoxicity other studies may be necessary to further investigate the effects. It should be noted that endpoints capable of detecting neurotoxicity and immunotoxicity are not examined in the standard 28-day and 90-day dermal or inhalation repeated dose toxicity studies.

More focus has also been put on endocrine disrupters during the latest decade. In relation to hazard and risk assessment, there are currently no test strategies or methods available, which specifically detect all effects, which have been linked to the endocrine disruption mechanism. It should be noted that work is on-going with the purpose of updating the present oral 28-days study (OECD TG 407/EU B.7) with more emphasis to be placed on detection of endocrine effects.

If data are not available from an oral standard 28-day repeated dose toxicity guideline study (OECD TG 407/EU B.7), the minimum repeated dose toxicity data requirement (28-day study) at tonnage levels from 10 t/y may in certain circumstances be met by results obtained from *the combined repeated dose toxicity study with the reproduction / developmental toxicity screening test* (OECD TG 422[30]). An advantage of this approach is obtaining information on repeated dose toxicity and reproductive toxicity in a single study providing an overall saving in the number of animals used for testing. In addition, the number of animals is higher (10 per sex compared to 5 per sex in the standard oral 28-day study) and the dosing period is longer in the combined study than in the standard oral 28-day study. Therefore, more information on repeated dose toxicity could be expected from the combined study. Potential complications in using the combined study include selecting adequate dose levels to examine adequately both repeated dose toxicity and reproductive toxicity. In addition, interpretation of the results may be complicated due to differences in sensitivity between pregnant and non-pregnant animals, and an assessment of the general toxicity may be more difficult especially when serum and histopathological parameters are not evaluated at the same time in the study. Consequently, where the combined study is used for the assessment of repeated dose toxicity, the use of data obtained from such a study should be clearly indicated. Despite such complications, the use of the combined study is recommended for the initial hazard assessment of the repeated dose toxicity potential of a substance when this study is relevant also for reproductive toxicity assessment.

In general, results from toxicological studies requiring repeated administration of a test substance (see also Section R.7.5.3.1) such as *reproduction and developmental toxicity studies* as well as *carcinogenicity studies* can contribute to the assessment of repeated dose toxicity. However, such toxicological studies rarely provide the information obtained from a standard repeated dose toxicity study and therefore, cannot stand alone as the sole basis for the assessment of repeated dose toxicity or for exempting from the standard information requirements for repeated dose toxicity at a given tonnage level.

Studies such as *acute toxicity and irritation studies* as well as *in vivo genotoxicity studies* contribute limited information to the overall assessment of the repeated dose toxicity. However, such studies may be useful in deciding on the dose levels for use in repeated dose toxicity.

Guidance on the dose selection for repeated dose toxicity testing (see also Table R.7.5-2) is provided in detail in the EU and OECD test guidelines. Unless limited by the physical-chemical nature or biological effects of the test substance, the highest dose level should be chosen with the aim to induce toxicity but not death or severe suffering.

Although not required by REACH, toxicokinetic studies may be helpful in the evaluation and interpretation of repeated dose toxicity data, for example in relation to accumulation of a substance or its metabolites in certain tissues or organs as well as in relation to mechanistic aspects of repeated dose toxicity and species differences. Toxicokinetic information can also assist in the selection of the dose levels. When conducting repeated dose toxicity studies it is necessary to ensure that the observed treatment-related toxicity is not associated with the administration of excessive

[30] To date there is no corresponding EU testing method available.

high doses causing saturation of absorption and detoxification mechanisms. The results obtained from studies using excessive doses causing saturation of metabolism are often of limited value in defining the risk posed at more relevant and realistic exposures where a substance can be readily metabolised and cleared from the body. It is suggested that a key resource in designing better repeated dose toxicity studies is to select appropriate dose levels based on results from useful metabolic and toxicokinetic investigations. Further details on the application of toxicokinetic information in the design and evaluation of repeated dose toxicity studies is available in Section R.7.12 on toxicokinetics.

Table R.7.5-2 Overview of *in vivo* repeated dose toxicity test guideline studies

Test	Design	Endpoints
OECD TG 407 (EU B.7) Repeated dose 28-day oral toxicity study in rodents	Exposure for 28 days At least 3 dose levels plus control At least 5 males and females per group Preferred rodent species: rat	Clinical observations Functional observations (4th exposure week – sensory reactivity to stimuli of different types, grip strength, motor activity) Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total and differential leucocyte count, platelet count, blood clotting time/potential) Clinical biochemistry Urinalysis (optional) Gross necropsy (full, detailed, all animals) Organ weights (all animals - liver, kidneys, adrenals, testes, epididymides, thymus, spleen, brain, heart) Histopathology (full, at least control and high-dose groups - all gross lesions, brain, spinal cord, stomach, small and large intestines, liver, kidneys, adrenals, spleen, heart, thymus, thyroid, trachea and lungs, gonads, accessory sex organs, urinary bladder, lymph nodes, peripheral nerve, a section of bone marrow)
OECD TG 410 (EU B.9) Repeated dose dermal toxicity: 21/28-day study	Exposure for 21/28 days At least 3 dose levels plus control At least 5 males and females per group Rat, rabbit or guinea pig	Clinical observations Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total and differential leucocyte count, clotting potential) Clinical biochemistry Urinalysis (optional) Gross necropsy (full, detailed, all animals) Organ weights (all animals - liver, kidneys, adrenals, testes) Histopathology (full, at least control and high-dose groups - all gross lesions, normal and treated skin, liver, kidney)
OECD TG 412 (EU B.8) Repeated dose inhalation toxicity: 28-day or 14-day study	Exposure for 28 or 14 days At least 3 concentrations plus control At least 5 males and females per group Rodents: preferred species - rat	Clinical observations Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total and differential leucocyte count, clotting potential) Clinical biochemistry Urinalysis (optional) Gross necropsy (full, detailed, all animals) Organ weights (all animals - liver, kidneys, adrenals, testes) Histopathology (full, at least control and high-dose groups - all gross lesions, lungs, liver, kidney,

Test	Design	Endpoints
		spleen, adrenals, heart)
OECD TG 408 (EU B.26) Repeated dose 90-day oral toxicity study in rodents	Exposure for 90 days At least 3 dose levels plus control At least 10 males and females per group Preferred rodent species: rat	Clinical observations Ophthalmological examination Functional observations (towards end of exposure period – sensory reactivity to stimuli of different types, grip strength, motor activity) Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total and differential leucocyte count, platelet count, blood clotting time/potential) Clinical biochemistry Urinalysis Gross necropsy (full, detailed, all animals) Organ weights (all animals - liver, kidneys, adrenals, testes, epididymides, uterus, ovaries, thymus, spleen, brain, heart) Histopathology (full, at least control and high-dose groups - all gross lesions, brain, spinal cord, pituitary, thyroid, parathyroid, thymus, oesophagus, salivary glands, stomach, small and large intestines, liver, pancreas, kidneys, adrenals, spleen, heart, trachea and lungs, aorta, gonads, uterus, accessory sex organs, female mammary gland, prostate, urinary bladder, gall bladder (mouse), lymph nodes, peripheral nerve, a section of bone marrow, and skin/eyes on indication)
OECD TG 409 (EU B.27) Repeated dose 90-day oral toxicity study in non-rodents	Exposure for 90 days At least 3 dose levels plus control At least 4 males and females per group Preferred species: dog	Clinical observations Ophthalmological examination Body weight and food/water consumption Haematology (as in TG 408) Clinical biochemistry Urinalysis Gross necropsy (full, detailed, all animals) Organ weights (as in TG 408 - additional: gall bladder, thyroid, parathyroid) Histopathology (as in TG 408 – additional: gall bladder, eyes)
OECD TG 411 (EU B.28) Subchronic dermal toxicity: 90-day study	Exposure for 90 days At least 3 dose levels plus control At least 10 males and females per group Rat, rabbit or guinea pig	Clinical observations Ophthalmological examination Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total and differential leucocyte count, clotting potential) Clinical biochemistry Urinalysis Gross necropsy (full, detailed, all animals)

Test	Design	Endpoints
		Organ weights (all animals - liver, kidneys, adrenals, testes) Histopathology (full, at least control and high-dose groups - all gross lesions, normal and treated skin, and essentially the same organs and tissues as in TG 408)
OECD TG 413 (EU B.29) Subchronic inhalation toxicity: 90-day study	Exposure for 90 days At least 3 concentrations plus control At least 10 males and females per group Rodents: preferred species - rat	Clinical observations Ophthalmological examination Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total and differential leucocyte count, clotting potential) Clinical biochemistry Urinalysis Gross necropsy (full, detailed, all animals) Organ weights (all animals - liver, kidneys, adrenals, testes) Histopathology (full, at least control and high-dose groups - all gross lesions, respiratory tract, and essentially the same organs and tissues as in TG 408)
OECD TG 452 (EU B.30) Chronic toxicity studies	Exposure for at least 12 months At least 3 dose levels plus control Rodents : At least 20 males and females per group Non-rodents: At least 4 males and females per group Preferred rodent species: rat Preferred non-rodent species: dog	Clinical observations, including neurological changes Ophthalmological examination Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total leucocyte count, platelet count, clotting potential) Clinical biochemistry Urinalysis Gross necropsy (full, detailed, all animals) Organ weights (all animals - brain, liver, kidneys, adrenals, gonads, thyroid/parathyroid (non-rodents only)) Histopathology (full, at least control and high-dose groups - all grossly visible tumours and other lesions, as well as essentially the same organs and tissues as in the 90-day studies (TG 408/409))
OECD TG 453 (EU B.33) Combined chronic toxicity / carcinogenicity studies	Exposure for at least 12 months (satellite groups) or majority of normal life span (carcinogenicity part) At least 3 dose levels plus control At least 50 males and females per group Satellite group: At least 20 males and females per group	Essentially as in TG 452

Test	Design	Endpoints
	Preferred species: rat	
OECD TG 422[31] Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test	Exposure for a minimum of 4 weeks (males) or from 2 weeks prior to mating until at least post-natal day 4 (females – at least 6 weeks of exposure) At least 3 dose levels plus control At least 10 males and females per group	Clinical observations as in TG 407 Functional observations as in TG 407 Body weight and food/water consumption Haematology as in TG 407 Clinical biochemistry Urinalysis (optional) Gross necropsy (full, detailed, all adult animals) Organ weights (testes and epididymides - all males; liver, kidneys, adrenals, thymus, spleen, brain, heart - in 5 animals of each sex per group, i.e. as in TG 407) Histopathology (ovaries, testes, epididymides, accessory sex organs, all gross lesions - all animals in at least control and high-dose groups; brain, spinal cord, stomach, small and large intestines, liver, kidneys, adrenals, spleen, heart, thymus, thyroid, trachea and lungs, urinary bladder, lymph nodes, peripheral nerve, a section of bone marrow - in 5 animals of each sex in at least control and high-dose groups, i.e. as in TG 407)
OECD TG 424 (EU B.43) Neurotoxicity study in rodents	Exposure for at least 28 days Dose levels: not specified At least 10 males and females per group Preferred rodent species: rat Generally oral route of administration	Detailed clinical observations Functional observations (sensory reactivity to stimuli of different types, grip strength, motor activity, more specialized tests on indication) Ophthalmological examination Body weight and food/water consumption Haematology (haematocrit, haemoglobin, erythrocyte count, total and differential leucocyte count, platelet count, blood clotting time/potential) Clinical biochemistry Histopathology: at least 5 animals/sex/ group) for neuropathological examinations (brain, spinal cord, and peripheral nerves); remaining animals to be used either for specific neurobehavioural, neuropathological, neurochemical or electrophysiological procedures that may supplement the histopathology or alternatively, for routine pathological evaluations according to the guidelines for standard repeated dose toxicity studies
OECD TG 419 (EU B.38) Delayed neurotoxicity of organophosphorus	Exposure for 28 days At least 3 dose levels plus control At least 12 birds per group Species: domestic laying hen	Detailed clinical observations Body weight and food/water consumption Clinical biochemistry (NTE activity, acetylcholinesterase activity

[31] To date there is no corresponding EU testing method available

Test	Design	Endpoints
substances: 28-day repeated dose study		Gross necropsy (all animals) Histopathology (neural tissue)

R.7.5.4.2 Human data on repeated dose toxicity

Human data in the form of epidemiological studies or case reports can contribute to the hazard identification process as well as to the risk assessment process itself. Criteria for assessing the adequacy of epidemiology studies include an adequate research design, the proper selection and characterisation of the exposed and control groups, adequate characterisation of exposure, sufficient length of follow-up for the disease as an effect of the exposure to develop, valid ascertainment of effect, proper consideration of bias and confounding factors, proper statistical analysis and a reasonable statistical power to detect an effect. These types of criteria have been described in more detail (Swaen, 2006 and can be derived from Epidemiology Textbooks (Checkoway *et al*, 1989; Hernberg, 1991; Rothman, 1998).

The results from human experimental studies are often limited by a number of factors, such as a relatively small number of subjects, short duration of exposure, and low dose levels resulting in poor sensitivity in detecting effects.

In relation to hazard identification, the relative lack of sensitivity of human data may cause particular difficulty. Therefore, negative human data cannot be used to override the positive findings in animals, unless it has been demonstrated that the mode of action of a certain toxic response observed in animals is not relevant for humans. In such a case a full justification is required. It is emphasised that testing with human volunteers is strongly discouraged, but when there are good quality data already available they can be used in the overall *Weight of Evidence*.

R.7.5.4.3 Exposure considerations for repeated dose toxicity

Three types of *adaptations* from testing are possible due to exposure considerations: exposure-based waiving of a study, exposure-based triggering of further studies, or selection of appropriate exposure route. More information on exposure-based waiving is available in Section R.5.1. More detailed guidance of exposure-based adaptations of the testing for repeated dose toxicity is given below and in Section R.7.5.6 (Integrated Testing Strategy).

Comparison of exposure and effect data should consider the existing (or most likely expected) *exposure patterns* for humans (e.g. daily exposure during life-time or repeated short or medium periods of exposures) and the most adequate DNEL (Derived No Effect Level) that reflects the specific exposure route and time pattern for each human population group at exposure. For instance, short-term exposure estimates should be compared to a descriptor of short-term toxicity whereas repeated daily exposure estimates should be compared to a corresponding descriptor of chronic toxicity. In all cases actually experienced daily human exposures are to be used in this comparison instead of daily exposures obtained by averaging over exposed and non-exposed days.

Concerning repeated dose toxicity testing the oral route is the preferred one. However, dependent on the physico-chemical properties of a substance as well as on the most relevant route of human exposure, the dermal or the inhalation route could also be appropriate as specified in REACH Annex VIII and IX.

The dermal route is appropriate if the physico-chemical properties suggest potential for a significant rate of absorption through the skin. The inhalation route is appropriate if exposure of humans via inhalation is the most relevant route of human exposure taking into account the vapour pressure of the substance and/or the possibility of exposure to aerosols, particles or droplets of an inhalable size.

According to Annex VIII-X further studies shall be proposed by the registrant or may be required by the Agency for example if there is particular concern regarding exposure, e.g. use in consumer products leading to exposure levels which are:

- close to the dose levels at which toxicity to humans may be expected (Annex VIII) i.e. a dose lower than, but in the vicinity of, the dose levels at which toxicity to humans may be expected
- high relative to the dose levels at which toxicity to humans may be expected (Annex IX), i.e. exposure levels higher than the dose levels at which toxicity to humans may be expected
- close to the dose levels at which toxicity is observed (Annex X); i.e. a dose lower than, but in the vicinity of, the dose levels at which toxicity is observed from animal studies.

Any of the exposure-triggered studies proposed by the registrant or required by the Agency should be considered on a case-by-case basis.

Various types of exposure considerations are possible for *waiving* of repeated dose toxicity studies. For instance, it is stated in REACH Article 13 and Annex XI:3 that testing in accordance with Annex VIII, Sections 8.6 and 8.7 (i.e. repeated dose toxicity and reproductive toxicity), Annex IX and X may be omitted based on the exposure scenario(s) developed in the Chemical Safety Report. Adequate justification and documentation shall in all cases be provided (see Section R.5.1.).

Further, the sub-chronic toxicity study (90-days study) does not need to be conducted according to Annex IX of REACH if: "the substance is unreactive, insoluble and not inhalable and there is no evidence of absorption and no evidence of toxicity in a 28-days *limit test*, particularly if such a pattern is *coupled with limited human exposure*. In order to omit the study the prerequisites interpreted above have to be considered jointly since the word "and" is used in between them. In addition, limited human exposure would strengthen the possibility for waiving.

The interpretation of *un-reactive* can be that it relates to the inherent chemical reactivity and as such, is an indicator of lack of local effects and mutagenicity, *insoluble and not inhalable* can be interpreted as indicators of low exposure potential and should be further defined, and *no evidence of absorption* that there has to be evidence for lack of absorption in order to omit the study. Further *no evidence of toxicity in a 28-days limit test* can be interpreted as it has to be at least a 28-days limit test available in order to waive the 90-days study, and this 28-days study should not show any sign of toxicity at 1000 mg/kg.

Limited exposure should consider the level of exposure, the frequency and/or the duration of exposure. Therefore, limited exposure must be considered on a case-by-case basis.

Finally, according to REACH Annex VIII testing of repeated dose toxicity (28-days study) does not need to be conducted if: *relevant human exposure can be excluded*.

Relevant human exposure depends on the inherent properties of the substance, if the population comes into contact with the substance or not, and how the substance is used. Thus, waiving might be considered on a case-by-case basis.

The concept of the Threshold of Toxicological Concern (TTC) might be applied to reduce the use of animals and other evaluation resources (Kroes et al., 2004); Use of the TTC concept may also be seen as a driving force for deriving exposure information of adequate quality. However, there are a number of limitations or drawbacks that should be taken into consideration in deciding if the concept is to be applied for industrial chemicals and further discussions on the cut-off values are needed before integration into into the guidance (see Appendix R.7.1-1; TemaNord, 2005*)*.

R.7.5.4.4 Remaining uncertainty on repeated dose toxicity

The key requirement for a CSA is the DNELs per exposure scenario (box 5 of Figure R.7.5-1). The DNEL for repeated dose toxicity is the threshold of the critical effect derived in a *Weight of Evidence* assessment of the available repeated dose toxicity data and an overall assessment factor (AF) that takes into account any uncertainty. The following elements contribute to the uncertainty in the determination of a threshold for the critical effects and the selection of the AF (further guidance on deriving a DNEL and application of AFs is provided in Chapter R.8).

Threshold of the critical effect

In the determination of the overall threshold for repeated dose toxicity all relevant information is evaluated to determine the lowest dose that induces an adverse effect (i.e. LOAEL or LOAEC) and the highest level with no biologically or statically significant adverse effects (i.e. NOAEL or NOAEC). In this assessment all toxicological responses are taken into account and the critical effect is identified. The uncertainty in the threshold depends on the strength of the data and is largely determined by the design of the underlying experimental data. Parameters such as group size, study type/duration or the methodology need to be taken into account in the assessment of the uncertainty in the threshold of the critical effect(s).

The NOAEL is typically used as the starting point for the derivation of the DNEL. In case a NOAEL has not been achieved, a LOAEL may be used, provided the available information is sufficient for a robust hazard assessment and for Classification and Labelling. The Bench Mark Dose (BMD) may also be used as the starting point for the derivation of the DNEL (Chapter R.8).

The selection of NOAEL or LOAEL is usually based on the dose levels used in the most relevant toxicity study, without considering the shape of the dose response curve. Therefore, the NOAEL/LOAEL may not reflect the true threshold for the adverse effect. On the other hand, the BMD is a statistical approach for the determination of the threshold and relies on the dose response curve. Alternatively, mathematical curve fitting techniques or statistical approaches exist to determine the threshold for an adverse effect. The use of such approaches (e.g. Benchmark Dose) to estimate the threshold should be considered on a case-by-case basis. For further guidance see Chapter R.8

Overall AF

Variability in sensitivity across and within species is another source of uncertainty for repeated dose toxicity. These inter- and intraspecies differences, respectively, are linked with variations in the toxicokinetics and dynamics of a substance. Information derived from non-testing, *in vitro* or *in vivo* methods may lead to an improvement of the understanding of the relevance of animal data for human risk assessment and may lead to a replacement of adopted standard default AF for these differences.

The quality of the whole database should be assessed for reliability and consistency across different studies and endpoints and taking into account the quality of the testing method, size and power of

the study design, biological plausibility, dose-response relationships and statistical association. Missing test data might be substituted by non-testing data obtained from physico-chemical properties, read-across to structurally or mechanistically related substances (SAR/chemical category) or by quantitative structure-activity relationships (QSARs). Also *in vitro* data might be used to fill in data gaps as well as *in vivo* non-standard animal experimental tests. Such data in combination with toxicity tests according to standard OECD/EU guidelines may in some cases lead to an improved understanding to the toxicological effect resulting in a reduction in the overall uncertainty. On the other hand information solely based on *in-vitro* and non-testing data are at present insufficient to act as a surrogate for repeated dose toxicity data and the uncertainty is sufficiently large that such information is unsuitable for use in a CSA and for classification and labelling. In the case of chemical categories information from non-testing methods or *in vitro* data may used to fulfil the data requirements on repeated dose toxicity and lead to improvement in the overall reliability and consistency for the read-across within a category of substances.

Since the adequacy and/or completeness of different data may vary, lack of quality and completeness of the overall database should be compensated for with an assessment factor for remaining uncertainty.

Besides AF addressing these differences (inter- and intraspecies, quality of the whole database), other uncertainties relating to differences between human and animal exposure conditions (e.g. route, and duration), and dose response characteristics are taken into account in the more extensive guidance on deriving a DNEL (see Section R.8.4.3).

Other considerations

Another situation may arise when testing is not technically possible, a waiving option indicated in Annex XI(2) (see also Chapter R.5). In such cases approaches such as QSAR, category formation and read-across may be helpful in the hazard characterisation; they should also be considered for information that might be suitable as a surrogate for a dose descriptor. Alternatively, generic threshold approaches, e.g. the Threshold of Toxicological Concern, TTC might be considered for the starting point of a risk characterisation (see Appendix R.7.1-1).

R.7.5.5 Conclusions on repeated dose toxicity

The evaluation of all available toxicological information for repeated dose toxicity (step 3 in Figure R.7.5-1 should include an assessment whether the available information as a whole (i.e. testing and non-testing, and relevant information from studies addressing other endpoints) meets the tonnage driven data requirements necessary to fulfil the REACH requirements. A *Weight of Evidence* approach should be used in assessing the database for a substance. This approach requires a critical evaluation of the entire body of available data for consistency and biological plausibility. Potentially relevant studies should be judged for quality and studies of high quality given more weight than those of lower quality. When both epidemiological and experimental data are available, similarity of effects between humans and animals is given more weight. If the mechanism or mode of action is well characterised, this information is used in the interpretation of observed effects in either human or animal studies. *Weight of Evidence* is not to be interpreted as simply tallying the number of positive and negative studies, nor does it imply an averaging of the doses or exposures identified in individual studies that may be suitable as starting points for risk assessment. The study or studies used for the starting point are identified by an informed and expert evaluation of all the available evidence.

The available repeated dose toxicity data should be evaluated in detail for a characterisation of the health hazards upon repeated exposure. In this process an assessment of all toxicological effect(s), their dose-response relationships and possible thresholds are taken into account. The evaluation should include an assessment of the severity of the effect, whether the observed effect(s) are adverse or adaptive, if the effect is irreversible or not or if it is a precursor to a more significant effect or secondary to general toxicity. Correlations between changes in several parameters, e.g. between clinical or biochemical measurements, organ weights and (histo)pathological effects, will be helpful in the evaluation of the nature of effects. Further guidance to this issue can be found in publications of the International Programme on Chemical Safety (IPCS 1994, 1999) and ECETOC (2002).

The effects data are also analysed for indications of potential serious toxicity of target organs or specific organ systems (e.g. neurotoxicity or immunotoxicity), delayed effects or cumulative toxicity. Furthermore, the evaluation should take into account the study details and determine if the exposure conditions and duration and the parameters studied are appropriate for an adequate characterisation of the toxicological effect(s).

If an evaluation allows the conclusion that the information of the repeated dose toxicity is adequate for a robust characterisation of the toxicological hazards, including an estimate of a dose descriptor (NOAEL/LOAEL/BMD), and the data are adequate for risk assessment and classification and labelling, no further testing will be necessary unless there are indications for further risk, according to column 2 of Annexes VIII-X of REACH.

Another consideration to be taken into account is whether the study duration has been appropriate for an adequate expression of the toxicological effects. If the critical effect involves serious specific system or target organ toxicity (e.g. haemolytic anaemia, neurotoxicity or immunotoxicity), delayed effects or cumulative toxicity and a threshold has **not** been established dose extrapolation may not be appropriate and further studies are required. In this case a specialised study is likely to be more appropriate for an improved hazard characterisation and should be considered instead of a standard short-term rodent or sub-chronic toxicity test at this stage.

In the identification of the NOAEL, other factors need to be considered such as the severity of the effect, the presence or absence of a dose- and time-effect relationship and/or a dose- and time-response relationship, the biological relevance, the reversibility, and the normal biological variation of an effect that may be shown by representative historical control values (IPCS, 1990).

R.7.5.5.1 Concluding on suitability for Classification and Labelling

In order to conclude on the suitability for classification and labelling (C&L), the data requirements in Annex VI of the dangerous substances Directive 67/548/EEC[32] have to be considered (box 4 in Figure R.7.5-1).

A decision on classification and labelling will affect downstream events/Directives under REACH. Therefore, it is important that the data are adequate for checking against the classification criteria in order to ensure safe use under REACH.

Basically the following conclusions can be obtained from the assessment of adequacy for C&L for repeated dose toxicity:

[32] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

1. Data are considered adequate for the purpose of C&L and can be checked against the criteria (boxes 6 and 11 in Figure R.7.5-1)[33].
2. Data are considered as inadequate for the purpose of C&L and cannot be checked against the criteria (inconclusive or lacking data). In this case testing should be considered in relation to the risk management of the substance.

R.7.5.5.2 Concluding on suitability for Chemical Safety Assessment

In order to be suitable for CSA (box 5 of Figure R.7.5-1) appropriate DNELs have to be established for each exposure scenario. Typically, the derivation of the DNEL takes into account a dose descriptor, modification of the starting point and application of assessment factors (see Chapter R.8).

Identification of the so-called dose descriptor: i.e. an appropriate threshold dose for the critical effect as the starting point for DNEL derivation, i.e. a NOAEL or BMD. If a NOAEL can not be identified, the LOAEL may be used instead provided the data are adequate for a robust hazard assessment.
It is to be noted that the dose descriptor should be route-specific. Thus, in case only animal data with oral exposure are available and humans are exposed mainly via skin and/or inhalation, a DNEL for dermal route and/or DNEL for inhalation route are needed: i.e. route-to-route extrapolation is needed, if allowed. Guidance for this route-to-route extrapolation is provided in Section R.8.4.2.

If this route-to-route extrapolation is not allowed, route-specific information is needed, possibly including testing, as a last resort (see Section R.7.5.6.3).

Derivation of a DNEL from this dose descriptor by applying AFs (to address uncertainty in the available data) is described elsewhere (see Section R.8.4.3; see also Section R.7.5.4.4).

R.7.5.5.3 Information not adequate

A *Weight of Evidence* approach comparing available adequate information with the tonnage-triggered information requirements by REACH may result in the conclusion that the requirements are not fulfilled. In order to proceed in further information gathering the testing strategy described in Section R.7.5.6.3 can be adopted.

R.7.5.6 Integrated Testing Strategy (ITS) for repeated dose toxicity

R.7.5.6.1 Objective / General principles

The objective in this testing strategy is to give guidance on a stepwise approach to hazard identification with regard repeated dose toxicity. A principle of the strategy is that the results of one

[33] It should be noted that although the exposure assessment and risk characterisation need not to be performed, when a substance is not classified (see Part A, section A.1.2), for potency-based endpoints like repeated dose toxicity, there could still potentially be a risk. Therefore one might consider performing an exposure assessment and risk characterisation on voluntary basis, to ensure safe handling and use.

study are evaluated before another study is initiated. The strategy seeks to ensure that the data requirements are met in the most efficient and humane manner so that animal usage and costs are minimised.

The core objectives of the Integrated Testing Strategy (ITS) for repeated dose toxicity are to generate sufficient information to allow:

- Characterisation of the hazard profile and the dose-response of a substance upon repeated exposure
- Performance of a chemical safety assessment for repeated dose toxicity

Information generated in this strategy should be suitable for Classification and Labelling according to the criteria given in Annex VI to Directive 67/548/EEC[34].

In addition, information from repeated dose toxicity studies can give valuable information to other endpoints based on repeated exposure (e.g. reproductive and developmental toxicity), and are valuable for other *in vivo* studies.

R.7.5.6.2 Preliminary considerations

On the basis of the objectives outlined above, a framework has been developed so that informed decisions can be made on the need for further testing. If generation of further data is deemed necessary, the information needs should be met efficiently in terms of resources and animal use. This means the use of the most appropriate study type in accordance with the tonnage-driven requirements stipulated by the REACH information requirements and taking into account modifications due to considerations of exposure, grouping and category formation. The data requirements may be increased or decreased taking into account exposure considerations or the level of concern noted during any of the stages in the testing strategy.

Testing for repeated dose toxicity is not required for chemicals produced at tonnage levels less than 10 tonne per annum (t/y). At higher production volumes, standard data requirements are, in general, increased with each tonnage band (see Section R.7.5.2); maintaining flexibility to adopt the most appropriate testing regime for any single chemical is a key component of the ITS. However, regardless of whether testing for repeated dose toxicity is required or not at a specific tonnage level, all existing test data, and all other available and relevant information on the substance should be collected.

R.7.5.6.3 Testing strategy for repeated dose toxicity

In order to proceed in further information gathering the following testing strategy is out-lined (step 4 in Figure R.7.5-1).

Before testing is initiated the available information should be scrutinised for evidence that may indicate severe effects, serious specific system or target organ toxicity (e.g. neurotoxicity or immunotoxicity), delayed effects or cumulative toxicity (boxes 8, 9 and 12 in Figure R.7.5-1). These indications may provide a trigger for specialised study protocols instead of the standard

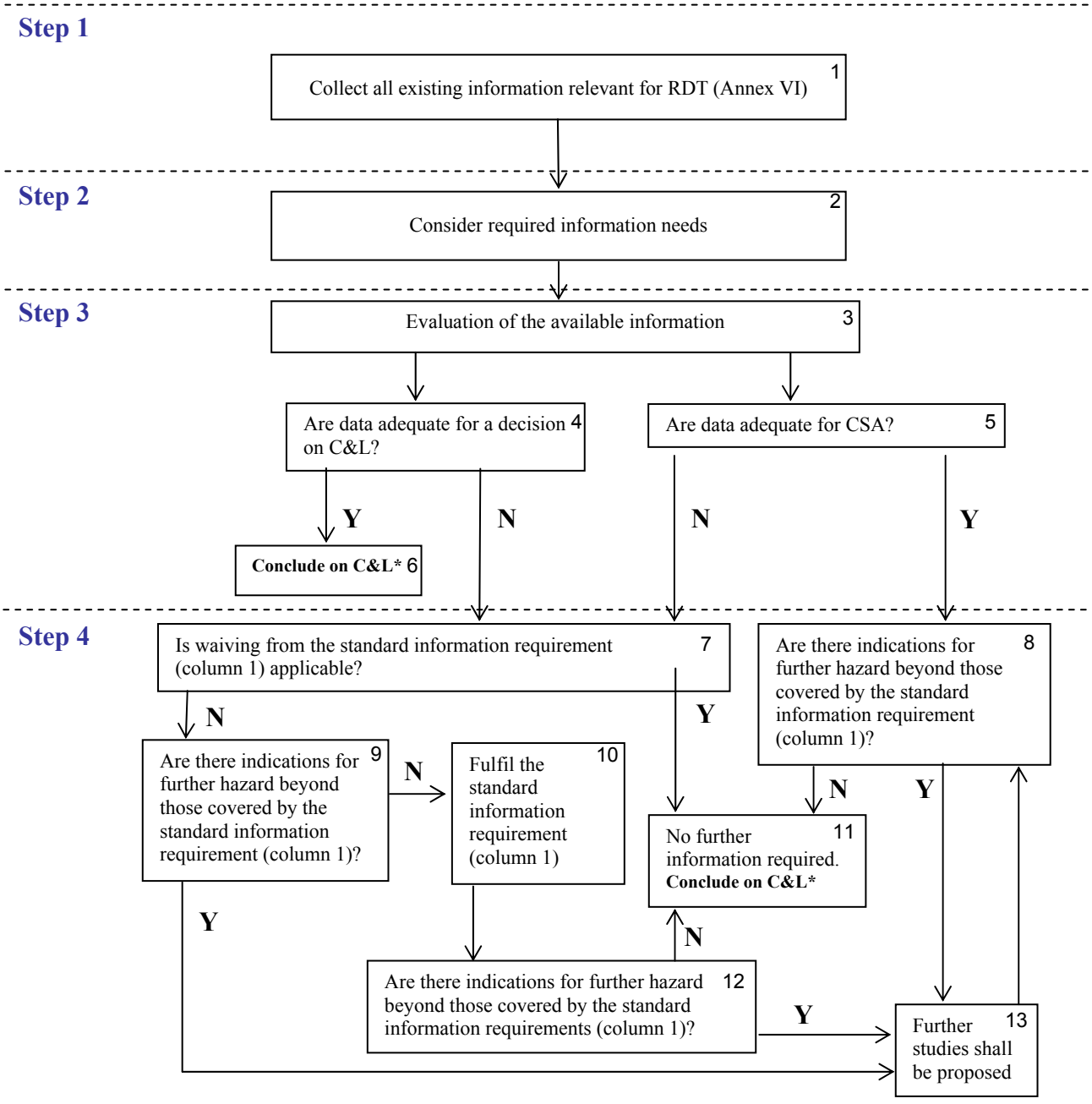
[34] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

protocols for the short-term and/or (sub)chronic toxicity (box 13 in Figure R.7.5-1). These specific protocols should be designed on a case-by-case basis, such that they enable an adequate characterisation of these hazards, including the dose-response, threshold for the toxic effect and an understanding of the nature of the toxic effects. An example of such an approach is given in Appendix R.7.5-1 Testing strategy for specific system/organ toxicity.

Annexes VII-X of the REACH regulation provide the standard information requirements in Column 1 (box 10 of Figure R.7.5-1) and specify triggering and waiving possibilities for the specific endpoints in Column 2. Different descriptors used for repeated dose toxicity in these annexes varying from *limited* (Annex IX) to *no relevant exposure* (Annex VIII). In addition, Annex XI of the REACH regulation contains basic approaches, or rules for adaptation of the standard testing regime, set out in Annexes VII-IX (see Chapter R.5; for waiving see box 7 in Figure R.7.5-1).

Exposure considerations at this stage may trigger a need for additional data if the applications include wide dispersive uses to a large population (e.g. consumer products) and if a particular concern exists for a low margin of exposure (box 13 in Figure R.7.5-1). The data to be generated at this stage should aim to improve the risk quotient and could therefore be a trigger for an improved exposure characterisation or an improved hazard characterisation. In the latter case the required information might include a special study leading to an improved characterisation of the critical toxic endpoint thereby decreasing the uncertainty in the NOAEL for repeated dose toxicity. An example of such a testing approach applied to neurotoxicity is given in Appendix R.7.5-1.

Figure R.7.5-1 Integrated Testing Strategy for repeated dose toxicity



Utilisation of the different tests at each of the different tonnage levels is summarised below:

10 t/y or more (Annex VIII)

At this tonnage level a short-term (28-day) toxicity test (OECD TG 407/EU B.7) is usually required. The use of a combined repeated dose toxicity study with the reproduction/developmental toxicity screening test (OECD TG 422[35]) is recommended if an initial assessment of repeated dose toxicity and reproductive toxicity is required. The route of exposure in these tests is oral unless the predominant route of human exposure or the physico-chemical properties indicate that the dermal or

[35] To date there is no corresponding EU testing method available.

inhalational route may be a more appropriate route of exposure to assess the repeated dose toxicity test (requiring OECD TG 410 or 412/EU B.9 or B.8).

If the results of a short-term rodent toxicity study (OECD TGs 407; 410, 412, 422) are adequate for a dose response characterisation and C&L and risk assessment, and if there are no indications for further risks, no further testing is required (see Section R.7.5.5.2 for a detailed discussion of the criteria for a robust hazard characterisation).

At this tonnage level the short-term toxicity study (28 days) does not need to be conducted if:

- a reliable sub-chronic (90 days) or chronic toxicity study is available, provided that an appropriate species, dosage, and route of administration were used; or
- where a substance undergoes immediate disintegration and there are sufficient data on the cleavage products; or
- relevant human exposure can be excluded in accordance with Annex XI Section 3.

It should be noted that any of the rules for adaptation according to Annex XI also apply (see Chapter R.5). For further details see this section under Annex XI (below).

According to REACH (Annex IX, 8.6.2), the sub-chronic toxicity study (90 days) shall be proposed by the registrant if:

- the frequency and duration of human exposure indicates that a longer term study is appropriate;

and one of the following conditions is met:

- other available data indicate that the substance may have a dangerous property that cannot be detected in a short-term toxicity study; or
- appropriately designed toxicokinetic studies reveal accumulation of the substance or its metabolites in certain tissues or organs which would possibly remain undetected in a short-term toxicity study but which are liable to result in adverse effects after prolonged exposure.

REACH also specifies that further studies shall be proposed by the registrant or may be required by the Agency in accordance with Article 40 or 41 in case of:

- failure to identify a NOAEL in the 28 or the 90 days study, unless the reason for the failure to identify a NOAEL is absence of adverse toxic effects; or
- toxicity of particular concern (e.g., serious/severe effects); or
- indications of an effect for which the available evidence is inadequate for toxicological and/or risk characterisation. In such cases it may also be more appropriate to perform specific toxicological studies that are designed to investigate these effects (e.g., immunotoxicity, neurotoxicity); or
- the route of exposure used in the initial repeated dose study was inappropriate in relation to the expected route of human exposure and route-to-route extrapolation cannot be made; or
- particular concern regarding exposure (e.g. use in consumer products leading to exposure levels which are close to the dose levels at which toxicity to humans may be expected); or

- effects shown in substances with a clear relationship in molecular structure with the substance being studied, were not detected in the 28 or the 90 days study.

It should be pointed out that a failure to identify a NOAEL does not lead to a data gap in every case and should not trigger additional studies by default. If the data are sufficient for a robust hazard assessment and for Classification and Labelling, the LOAEL may be used as the starting point for the CSA (see also Sections R.7.5.4.4 and R.7.5.5 and Chapter R.8).

A specialised study is likely to be more appropriate for an improved hazard characterisation and should be considered instead of a standard short-term rodent or sub-chronic toxicity test at this stage.

100 t/y or more (Annex IX)

At this tonnage level, the following information is required (REACH Annex IX, Sections 8.6.1 and 8.6.2):

- a short-term study (28 day) in a single rodent species is the minimum requirement. The default route of exposure in these tests is oral (OECD TG 407/EU B.7; TG 422[36]) unless the predominant route of human exposure or the physico-chemical properties indicates that the dermal or inhalational route (OECD TG 410, 412/EU B.9, B.8) is a more appropriate route of exposure in the repeated dose toxicity tests.

- a sub-chronic toxicity study (90-day) in a single rodent species is usually required. The default route of exposure in these tests is oral (OECD TG 408/EU B.26) unless the predominant route of human exposure or the physico-chemical properties indicates that the dermal or inhalational route (OECD TG 411, 413/EU B.28, B.29) is a more appropriate route of exposure in the repeated dose toxicity tests.

According to REACH, at this tonnage level the sub-chronic toxicity study (90 days) does not need to be conducted if:

- a reliable short-term toxicity study (28 days) is available showing severe toxicity effects according to the criteria for classifying the substance as R48, for which the observed NOAEL-28 days, with the application of an appropriate assessment factor, allows the extrapolation towards the NOAEL-90 days for the same route of exposure; or

- a reliable chronic toxicity study is available, provided that an appropriate species and route of administration were used; or

- a substance undergoes immediate disintegration and there are sufficient data on the cleavage products (both for systemic effects and effects at the site of uptake); or

- the substance is unreactive, insoluble and not inhalable and there is no evidence of absorption and no evidence of toxicity in a 28-day *limit test*, particularly if such a pattern is coupled with limited human exposure;

It should be noted that any of the rules for adaptation according to Annex XI also apply. For further details see the section on Annex XI below.

In case human exposure is limited or different in frequency and duration from that used in the test protocol for repeated dose toxicity, the sub-chronic toxicity study may not be necessary if the data

[36] To date there is no corresponding EU testing method available.

for the short-term toxicity study are adequate for a robust hazard characterisation, a risk assessment and classification and labelling. This adaptation requires full justification by the registrant.

In case the weight of the evidence indicates that the available information is adequate to characterise the short-term toxicity and sufficiently robust for proper dose-selection of the 90-day study, a dedicated 28-day study is not necessary at this stage.

No further testing is required if the available data, which may include a sub-chronic rodent toxicity study (OECD TG 408, 411, 413/EU B.26, B.28, B.29) are adequate for a dose response characterisation and C&L and risk assessment.

In case data are inadequate for hazard characterisation and risk assessment further studies shall be proposed by the registrant or may be required by the Agency in accordance with REACH Articles 40 or 41: According to REACH Annex IX Section 6.6.2 such a situation may arise if there is:

- failure to identify a NOAEL in the 90 days study unless the reason for the failure to identify a NOAEL is absence of adverse toxic effects; or
- toxicity of particular concern (e.g. serious/severe effects); or
- indications of an effect for which the available evidence is inadequate for toxicological and/or risk characterisation; In such cases it may also be more appropriate to perform specific toxicological studies that are designed to investigate these effects (e.g. immunotoxicity, neurotoxicity); or
- particular concern regarding exposure (e.g. use in consumer products leading to exposure levels which are high relative to the dose levels at which toxicity to humans occurs)

A specialised study is likely to be more appropriate for an improved hazard characterisation and should be considered instead of a standard short-term rodent or sub-chronic toxicity test. An example of such an approach given in Appendix R.7.5-1.

It should be pointed out that a failure to identify a NOAEL does not lead to a data gap in every case and should not be a default trigger for additional studies. If the data are sufficient for a robust hazard assessment or for Classification and Labelling, the LOAEL may be used as the starting point for the CSA (see also Sections R.7.5.4.4 and R.7.5.5 and Chapter R.8).

1000 t/y or more (Annex X)

There is no default testing requirement for repeated dose toxicity at this tonnage level beyond those recommended for the level 100 t/y or more (see above). However, in accordance with REACH Articles 40 and 41, if the frequency and duration of human exposure indicates that a long-term study is appropriate and one of the following conditions is met a long-term repeated toxicity test (≥12 months) may be proposed:

- serious or severe toxicity effects of particular concern were observed in the 28-days or 90-days study for which available evidence is inadequate for toxicological evaluation or risk characterisation; or
- effects shown in substances with clear relationship in molecular structure with the substance being studied were not detected in the 28-days or 90-days study; or
- the substance may have a dangerous property that cannot be detected in a 90-days study.

In addition, further studies shall be proposed by the registrant or may be required by the Agency in accordance with REACH Articles 40 or 41, in case of:

- toxicity of particular concern (e.g. serious/severe effects); or
- indications of an effect for which the available evidence is inadequate for toxicological evaluation and/or risk characterisation; In such cases it may also be more appropriate to perform specific toxicological studies that are designed to investigate these effects (e.g. immunotoxicity, neurotoxicity); or
- particular concern regarding exposure (e.g. use in consumer products leading to exposure levels which are close to the dose levels at which toxicity is observed).

In some cases a specialised study might the most appropriate study in case an improved hazard characterisation is necessary and should be considered instead of a standard sub-chronic or chronic toxicity test. An example of such an approach given in Appendix R.7.5-1.

No further testing is required if the results of a sub-chronic rodent toxicity study (OECD TG 408, 410, 411, 412, 413 or EU B.26, B.9, B.28, B.8, B.29) are adequate for a robust hazard characterisation and suitable for risk assessment and classification and labelling (see step 3 Identify data gaps for a detailed discussion of the criteria for a robust hazard characterisation).

Also, the testing requirements can be adapted if any of the rules according to REACH Annex XI apply: For further details see this Section under *REACH Annex XI* (below).

As there is no standard test requirement at this tonnage level, column 2 also had no waiving options.

REACH Annex XI adaptations of the standard testing regime for repeated dose toxicity

General guidance on the application of the Annex XI adaptations to information requirements is given in Chapter R. 5. For repeated dose toxicity the following additional guidance applies.

Testing does not appear scientifically necessary

Some substances may be excluded from testing for repeated dose toxicity if it does not appear scientifically necessary (Annex XI Section 1). This might be the case for example if:

- a *Weight of Evidence* analysis demonstrates that the available information is sufficient for an adequate hazard characterisation, and a CSA where the exposure to the substance is adequately controlled;
- a substance is not bio-available via a specific route and possible local effects have been adequately characterised;
- the vapour pressure is sufficiently low that inhalational exposures are unlikely to be of significance, or if human exposure is limited to dusts or aerosols unlikely to be inhalable
- for substances belonging to a group or a category of substances that have a common functionality and/or breakdown products or sufficient information for a qualitative and quantitative understanding of the toxicological properties, testing of all individual category members may not be necessary (Annex XI Section 1.5). The criteria for application of read-across for a category of substances and detailed guidance can be found in Sections R.4.3.2 and R.6.2.

Testing is technically not possible

There may also be cases where it is technically not possible to conduct a repeated dose toxicity test (Annex XI Section 2). This might be the case if

- The substance ignites in air at ambient conditions.
- The substance undergoes immediate disintegration. In such a case the information requirements for the cleavage products should be assessed following an approach similar to that outlined in this document.
- The substance is corrosive in the dose range of interest for the study. Also, for reasons of animal welfare such studies should be avoided.

Substance-tailored exposure-driven testing

Exposure considerations may also lead to adaptation of the testing requirements (Annex XI Section 3). This might be the case if:

Testing requirements may be adapted based on a substance-specific exposure-assessment according to Annex XI Section 3. In this case testing for short-term repeated dose toxicity (Annex VIII, 8.6.1) may be waived at the 10-100 tonnage level if relevant human exposure can be excluded (see Section R.7.5.4.3).

Human exposure is limited at the tonnage level of 100 t/y or more (Annexes IX and X). The need for a sub-chronic study should be considered if the substance is only handled in industrial or commercial installations using closed systems and/or handled only as preparations at low concentrations.

Appendix 1 to Section R.7.5

Appendix R.7.5-1 Testing strategy for specific system/organ toxicity.

Content of Appendix 7.5-1

1. General aspects
2. Structure-activity considerations
3. Assessment of available information or results from initial testing
4. Recommendations from the WHO/FAO Joint Meeting of Experts on Pesticide Residues (JMPR)
5. Further neurotoxicity testing

Mechanisms of respiratory irritation

1. General aspects

For some specific system/organ effects the testing methods of EU Annex V or the OECD may not provide for adequate characterisation of the toxicity. There may be indications of such effects in the standard studies for systemic toxicity, or from SAR. For adequate characterisation of the toxicity and, hence, the risk to human health, it may be necessary to conduct studies using other published test methods, *in-house* methods or specially designed tests. Some references are given in Error! Reference source not found.. Before initiating a study to investigate specific organ/system toxicity, it is important that the study design is presented to the Agency, in order that the need for (and scope/size of) studies using live animals should be particularly carefully considered.

Specific investigation of organ/systemic toxicity is to some extent undertaken as part of the repeated dose toxicity tests conducted according to test guidelines of the OECD and Annex V to Directive 67/548/EEC[37]. Specific investigation (or further investigation) of any organ/system toxicity (e.g. immune, endocrine or nervous system) may sometimes be necessary and should be addressed on a case-by-case basis. As an example of a testing strategy the approach for neurotoxicity is given below.

Definition of neurotoxicity

Neurotoxicity is the induction by a chemical of adverse effects in the central or peripheral nervous system, or in sense organs. It is useful for the purpose of hazard and risk assessment to differentiate sense organ-specific effects from other effects which lie within the nervous system. A substance is considered *neurotoxic* if it induces a reproducible lesion in the nervous system or a reproducible pattern of neural dysfunction.

[37] All the test methods previously included in Annex V to Directive 67/548/EEC will be incorporated in a new Test Methods (TM) Regulation that is currently (February 2008) under adoption. The TM Regulation will be adapted to technical progress whenever a new test method has been developed, scientifically validated and accepted for regulatory use by the National Coordinators of the Member states

The starting point for the testing strategy are the REACH requirements specified in Annex VIII, IX and X and detailed in Section R.7.5.6.3 Depending on the tonnage level, these requirements may trigger a 28-day and/or a 90-day test (e.g. OECD TG 407, 408/EU B.7, B.26). These protocols include a number of nervous system endpoints (e.g. clinical observations of motor and autonomous nervous system activity, histopathology of nerve tissue), which should be regarded as the starting point for evaluation of a substance potential to cause neurotoxicity. It should be recognised that the standard 28-/90-day tests only measure some aspects of nervous system structure and function e.g. Functional Observational Battery, while other aspects, e.g. learning and memory and sensory function is not or only superficially tested. SAR considerations may prompt the introduction of additional parameters to be tested in standard toxicity tests or the immediate request of studies such as delayed neurotoxicity (OECD TG 418 or 419/EU B.37 or B.38,; see below).

If there are no indications of neurotoxicity from available information i.e. adequately performed repeated dose toxicity tests, other testing systems (e.g. *in vitro*), non-testing systems ((Q)SAR and read-across) or human data, it will not be necessary to conduct any special tests for neurotoxicity.

The approach presented below is a hierarchical, step-wise strategy to investigate the potential neurotoxicity of a substance. It should be pointed out that the requirements outlined in steps 1 and 2 are met by the tonnage-based information requirements in Annex VIII, IX and X of REACH.

2. Structure-activity considerations

Structural alerts are only used as a positive indication of neurotoxic potential. Substance classes with an alert for neurotoxicity may include organic solvents (for chronic toxic encephalopathy); organophosphorus compounds (for delayed neurotoxicity), and carbamates (for cholinergic effects). Several estimation techniques are available, one of which is the rule-based DEREK (Deductive Estimation of Risk from Existing Knowledge) system. The rulebase comprises the following hazards and structural alerts: Organophosphate (for direct and indirect anticholinesterase activity); N-methyl or N,N-dimethyl carbamate (for direct anticholinesterase activity); gamma-diketones (for neurotoxicity).

3. Assessment of available information or results from initial testing

Signs of neurotoxicity in standard acute or repeated dose toxicity tests may be secondary to other systemic toxicity or to discomfort from physical effects such as a distended or blocked gastrointestinal tract. Nervous system effects seen at dose levels near or above those causing lethality should not be considered, in isolation, to be evidence of neurotoxicity. In acute toxicity studies where high doses are administered, clinical signs are often observed which are suggestive of effects on the nervous system (e.g. observations of lethargy, postural or behavioural changes), and a distinction should be made between specific and non-specific signs of neurotoxicity.

Neurotoxicity may be indicated by the following signs: morphological (structural) changes in the central or peripheral nervous system or in special sense organs; neurophysiological changes (e.g. electroencephalographic changes); behavioural (functional) changes; neurochemical changes (e.g. neurotransmitter levels).

A *Weight of Evidence* approach should be taken into account for the assessment of the neurotoxicity and the type, severity, number and reversibility of the effect should be considered. A consistent pattern of neurotoxic findings rather than a single or a few unrelated effects should be taken as persuasive evidence of neurotoxicity.

It is important to ascertain whether the nervous system is the primary target organ. The reversibility of neurotoxic effects should also be considered. The potential for such effects to occur in exposed humans (i.e. the exposure pattern and estimated level of exposure are *acute*) should be considered in the risk characterisation. Reversible effects may be of high concern depending on the severity and nature of effect. In this context it should be kept in mind that effects observed in experimental animals that appear harmless might be of high concern in humans depending on the setting in which they occur (e.g. sleepiness in itself may not be harmful, but in relation to operation of machinery it is an effect of high concern). Furthermore the possibility that a permanent lesion has occurred cannot be excluded, even if the overt effect is transient. The nervous system possesses reserve capacity, which may compensate for the damage, but the resulting reduction in the reserve capacity should be regarded as an adverse effect. Irreversible neurotoxic effects are of high concern and usually involve structural changes, though, at least in humans, lasting functional effects (e.g. depression, involuntary motor tremor) are suspected to occur as a result of neurotoxicant exposure, apparently without morphological abnormalities.

For the evaluation of organophosphate pesticides, the WHO/FAO Joint Meeting of Experts on Pesticide Residues (JMPR) has published recommendations on "Interpretation of Cholinesterase Inhibition" (FAO, 1998; 1999). The applicability of these recommendations, outlined below, could also be extended to other substances that inhibit cholinesterase. It should be pointed out that for substances that may have a structural alert for cholinesterase inhibition, the measurement of acetylcholinesterase activity as recommended by JMPR can be included in the list of parameters for the standard 28- or 90 day testing protocols required by REACH, irrespective of the route of exposure.

4. Recommendations from the WHO/FAO Joint Meeting of Experts on Pesticide Residues (JMPR)

The inhibition of brain acetylcholinesterase activity and clinical signs are considered to be the primary endpoints of concern in toxicological studies on compounds that inhibit acetylcholinesterases. Inhibition of erythrocyte acetylcholinesterase is also considered to be an adverse effect, insofar as it is used as a surrogate for brain and peripheral nerve acetylcholinesterase inhibition, when data on the brain enzyme are not available. The use of erythrocyte acetylcholinesterase inhibition as a surrogate for peripheral effects is justified for acute exposures resulting in greater acetylcholinesterase inhibition in erythrocytes than in the brain. However, reliance on inhibition of erythrocytic enzyme in studies of repeated doses might result in an overestimate of inhibition on peripheral tissues, because of the lower rate of resynthesis of the enzyme in erythrocytes than in the nervous system. Plasma acetylcholinesterase inhibition is considered not relevant. Regarding brain and erythrocyte acetylcholinesterase inhibition, the experts defined that statistically significant inhibition by 20% or more represents a clear toxicological effect and any decision to dismiss such findings should be justified. JMPR also agreed on the convention that statistically significant inhibition of less than 20% or statistically insignificant inhibition above 20% indicate that a more detailed analysis of the data should be undertaken. The toxicological significance of these findings should be determined on a case-by-case basis. One of the aspects to consider is the dose-response characteristic.

5. Further neurotoxicity testing

If the data acquired from the standard systemic toxicity tests required by REACH provide indications of neurotoxicity which are not adequate for a hazard assessment, risk characterisation or

classification and labelling, the nature of further investigation will need to be considered. If a 90-day study is triggered to meet the requirements of Annex IX following a standard 28-day study, a number of endpoints assessing the nervous system endpoints should be included,. irrespective of the administration route. In some cases, it may be necessary to conduct a specific study such as a neurotoxicity test using the OECD TG 424 with possible inclusion of a satellite group for assessment of reversibility of effects. The OECD TG 424 is intended for confirmation or further characterisation of potential neurotoxicity identified in previous studies. The OECD guideline allows for a flexible approach, in which the number of simple endpoints which duplicate those already examined during standard testing may be minimised, and where more effort is put into in-depth investigation of more specific endpoints by inclusion of more specialised tests. Adjustment of dose levels to avoid confounding by general toxicity should be considered.

If data from standard toxicity studies are clearly indicative of specific neurotoxicity, e.g. neurotoxicity occurring at lower dose levels than systemic toxicity, further specific neurotoxicity testing is required to confirm and extend the findings from the general toxicity studies and to establish an NOAEL for neurotoxicity. Again, the neurotoxicity test according to OECD TG 424 is considered appropriate for this situation.

Certain substances and/or certain effects are best investigated in particular species. Pyridine derivatives are neurotoxic to humans and primates but not to rats. Among other neurotoxic compounds, organophosphorus compounds are a group with known delayed neurotoxic properties, which need to be assessed in a specified test for delayed neurotoxicity, to be performed preferentially in the adult laying hen according to EU B.37 or OECD TG 418 (Delayed neurotoxicity of organophosphorus substances following acute exposure) and B.38 or OECD TG 419 (Delayed neurotoxicity of organophosphorus substances: 28-day repeated dose study). Such studies are specifically required for biocidal substances of similar or related structures to those capable of inducing delayed neurotoxicity. If anticholineesterase activity is detected, a test for response to reactivating agent may be required.

Standard exposure conditions may not always be adequate for neurotoxicity studies. The duration of exposure needed to induce specific neurotoxic effects in an animal experiment will depend on the underlying mechanism of action. Short-term peak exposures can be important for certain types of substance/effect. When the test compound is administered as a bolus via the intravenous, subcutaneous or oral route it is essential to determine the time-effect course, and to perform measurements of neurotoxicity parameters preferentially at the time of peak effect.

For example, the neurotoxicity associated with short-term exposure to some volatile organic solvents has largely been identified following human exposure - particularly occupational exposure. Acute inhalation studies, using protocols designed to detect the expected effects, are ideal for such substances/effects. For some neurotoxic substances a long exposure period is necessary to elicit neurotoxicity.

The most appropriate methods for further investigation of neurotoxicity should be determined on a case-by-case basis, guided by the effects seen in the standard systemic toxicity tests and/or from SAR-based predictions. Extensive coverage of methods which may be used is given in OECD (2004a), IPCS (1986) and ECETOC (1992), and some are summarised in the Table R.7.5-3.

Table R.7.5-3 Methods for investigation of neurotoxicity

Effect	Methods available	References*
Morphological changes	Neuropathology. Gross anatomical techniques. Immunocytochemistry. Special Stains	Krinke, 1989; Odonoghue, 1989; Mattson et al., 1990
Physiological changes	Electrophysiology (e.g. nerve conduction velocity (NCV), Electroencephalogram (EEG), evoked potentials	Fox et al., 1982; Rebert, 1983; Mattson and Albee, 1988
Behavioural changes	Functional observations. Sensory function tests. Motor function tests (e.g. locomotor activity). Cognitive function tests	Robbins, 1997; Tilson et al., 1980; Cabe and Eckerman, 1982; Pryor et al., 1983 Moser and McPhail, 1990; Moser 1995
Biochemical changes	Neurotoransmitter analysis. Enzyme/protein activity. Measures of cell integrity.	Dewar and Moffet, 1977; Damstra and Bondy, 1982; Cooper et al., 1986; Costa, 1998.

*Given in full in ECETOC (1982), IPCS (1986) or Mitchell (1982)

R.7.5.7 References on repeated dose toxicity

Checkoway H, Pearce NE, Crawford-Brown DJ. Research methods in Occupational Epidemiology Oxford University Press 1989.

ECETOC (1992). Evaluation of the neurotoxic potential of chemicals. European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), Monograph No. 18, Brussels.

ECETOC (2002). Recognition of, and Differentiation between, Adverse and Non-adverse Effects in Toxicology Studies. Technical Report No. 85, The European Centre for Ecotoxicology and Toxicology of Chemicals, December 2002.

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, The European Centre for Ecotoxicology and Toxicology of Chemicals, September 2003.

FAO (1998). Pesticide Residues in Food. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide. Residues in Food and the Environment and the WHO Core Assessment Group.

FAO (1999). Plant Production and Protection Paper, No. 148, 17-19.

Hernberg S. Introduction to Occupational Epidemiology. Lewis Publishers 1991.

IPCS (1986). Principles and Test Methods for the Assessment of Neurotoxicity Associated with Exposure to Chemicals. World Health Organisation (WHO), International Programme on Chemical Safety (IPCS), Environmental Health Criteria 60, Geneva.

Kroes, R., Renwick, AG, Cheeseman, M. et al., 2004. Structure-based thresholds of toxicological concerns (TTC): Guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42, 65-83.

Mitchell CL (ed) (1982). Nervous System Toxicology. Raven Press, New York, NY.

OECD (2003) Descriptions of Selected Key Generic Terms used in Chemical Hazard/Risk Assessment; OECD Series on Testing and Assessment: No 44. (ENV/JM/MONO(2003)15). 30 Oct 2003.

OECD (2004a). Guidance Document for Neurotoxicity Testing. Organization for Economic Cooperation and Development (OECD), Environment Directorate, OECD Environmental Health and Safety Publications, Series on Testing and Assessment No 20. Paris.

OECD (2006). 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. (ENV/JM/MOM(2006)25). 11 August 2006

OECD (2007a) Manual for Investigation of HPV Chemicals.
http://www.oecd.org/document/7/0,2340,en_2649_34379_1947463_1_1_1_1,00.html

OECD (2007b) Guidance Document on the Validation of (Quantitative)Structure-Activity Relationships [(Q)Sar] Models, Monograph No. 69, (ENV/JM/MONO(2007)2). 30 Mar 2007

Prieto P, Clemedson C, Meneguz A, Pfaller W, Sauer UG and Westmoreland C. (2005). 3.6 Subacute and subchronic toxicity. ATLA **33**, Suppl. 1, 109-116.

Prieto P, Baird AW, Blaauboer BJ, Castell Ripoll JV, Corvi R, Dekant W, Dietl P., Gennari A, Gribaldo A, Griffin JL, Hartung T, Heindel JJ, Hoet P, Jennings P, Marocchio L, Noraberg J, Pazos P, Westmoreland C, Wolf A, Wright J, Pfaller W. (2006). The Assessment of Repeated Dose

Toxicity In Vitro: A Proposed Approach. The Report and Recommendations of ECVAM Workshop 56. ATLA. ATLA 34, 315-341.

Rennen MAJ, De Heer C and Houben GF (1999). Prediction of Local Effects upon Dermal and Respiratory Repeated Exposure to Chemicals. Organisation for Applied Scientific Research (TNO), TNO Nutrition and Food. Research Institute, TNO Report V98.1267, Zeist, The Netherlands.

Rothman KJ, Greenland S. Modern Epidemiology Lippincott-Raven 1998.

Swaen GMH. A framework for using epidemiological data for risk assessment. Human and Experimental Toxicology 2006; 25:147-155,

TemaNord 2005: 559Threshold of Toxicological Concern (TTC), Literature review and applicability ISBN 92-893-1196-7

WHO, 1990. IPCS Environmental Health Criteria 104; Pesticide Residues in food, principles for the toxicological assessment,World Health Organization, Geneva 1990

WHO, 1994. IPCS Environmental Health Criteria 170; Guidance values for human exposure limits, World Health Organization, Geneva 1994

WHO, 1999. IPCS Environmental Health Criteria 210, Principles for the assessment of Risks to human health from exposure to chemicals. World Health Organization, Geneva 1999

WHO 2004. IPCS Risk assessment terminology. World Health Organization, Geneva, 2004

Worth, A.P. & Balls, M., eds (2002). Alternative (non-animal) methods for chemicals testing: current status and future prospects. A report prepared by ECVAM and the ECVAM working group on chemicals. ALTA 30, Suppl. 1, 71.

R.7.6 Reproductive and developmental toxicity

R.7.6.1 Introduction

At the population level the property of reproductive toxicity is of obvious high concern because the continuance of the human species is dependent on the integrity of the reproductive cycle. Similarly, to the individual an impairment of the ability to reproduce and the occurrence of developmental disorders are self-evidently serious health conditions. Therefore it is important that the potential hazardous properties with respect to reproduction are established for chemicals with relevant human exposure that may be present in the environment, at the workplace and in consumer products.

R.7.6.1.1 Definition of reproductive toxicity

The term *reproductive toxicity* is used to describe the adverse effects induced (by a substance) on sexual function and fertility in adult males and females, developmental toxicity in the offspring and effects on or mediated via lactation, as defined in Part 3 of the Globally Harmonised System of Classification and Labelling of Chemicals System (GHS) (United Nations 2005). In practical terms, reproductive toxicity is characterised by multiple diverse endpoints, which relate to impairment of male and female reproductive functions or capacity (*fertility*) and the induction of non-heritable harmful effects on the progeny (*developmental toxicity*). Effects on male or female fertility include adverse effects on libido, sexual behaviour, any aspect of spermatogenesis or hormonal or physiological response, which would interfere with the capacity to fertilise, fertilisation itself or the development of the fertilised ovum up to and including implantation. Developmental toxicity includes any effect interfering with normal development, both before and after birth. It includes effects induced or manifested either pre- or postnatally. This includes embryotoxic/foetotoxic effects such as reduced body weight, growth and developmental retardation, organ toxicity, death, abortion, structural defects (teratogenic effects), functional effects, peri- and postnatal defects, and impaired postnatal mental or physical development up to and including normal pubertal development.

R.7.6.1.2 Objective of the guidance on reproductive toxicity

To provide guidance to all stakeholders, in order to establish:

- whether exposure of humans to the substance of interest has been associated with reproductive toxicity and/or
- whether, on the basis of information other than human data, it can be predicted that the substance will cause reproductive toxicity in humans.
- whether the pregnant female is potentially more susceptible to general toxicity;
- the dose-response relationship for any adverse effects on reproduction.

Substance-related adverse effects on reproduction are always of potential concern, but it is important, where possible, to distinguish between a specific effect on reproduction as a consequence of an intrinsic property of the substance and an adverse reproductive effect which is a non-specific consequence to general toxicity (e.g. marked changes in bodyweight, marked reductions in food or water intake, maternal stress, see Section R.7.6.4.1 for further discussion).

R.7.6.2 Information requirements for reproductive toxicity

The standard data requirements for reproductive toxicity under the REACH Regulations are as follows:

- A reproduction/developmental toxicity screening test (OECD TGs 421 or 422)[38], usually required at the Annex VIII tonnage.
- A prenatal developmental toxicity study (EU B.31, OECD TG 414) in one species, usually required at the REACH Annex IX level. A study in a second species should be considered at either Annex IX or at Annex X level.
- A two-generation reproduction toxicity study[39] (OECD TG 416, EU B.35) in one species, usually required at the Annex X level.

However, according to column 2 specific rules (see Annexes VIII-X of the REACH legislation) and to Annex XI these tonnage-related standard data requirements can be adapted, either as reduced (a data waiver) or deferred testing or as the need for extended testing, as detailed in the stepwise Integrated Testing Strategy presented in Section R.7.6.6. Factors that can influence the testing requirements include structural relationships with other chemicals, the results of other toxicity studies, presence of mutagenic and carcinogenic properties, available data from humans exposed to the substance, concerns for endocrine disruption and the use and human exposure patterns.

This guidance provides advice on how the registrant can meet the information requirements of REACH, thereby providing data on the hazardous properties that can be used for classification (include a PBT assessment) and in the risk assessment.

R.7.6.3 Information on reproductive toxicity and its sources

Relevant information on reproductive toxicity can be obtained from various of sources, which are indicated below.

R.7.6.3.1 Non-human data on reproductive toxicity

Non-testing data on reproductive toxicity

Information of relevance to reproductive toxicity can be inferred from the physico-chemical characteristics of a substance.

Information on SARs (chemical grouping or read-across) and (Q)SAR models may be available.

Testing data on reproductive toxicity

In vitro data

Currently there is no officially adopted EU or OECD test guideline for *in vitro* tests of relevance to reproductive toxicity. Three tests have recently been subjected to an extensive

[38] To date there are no corresponding EU testing methods available.

[39] A proposed *F1-extended one-generation study* may replace OECD TG 416 as a definitive study for reproductive toxicity in the near future, subject to gaining regulatory acceptance in the EU

multicentre validation study in the EU (Genschow *et al.* 2002) and have been declared to be scientifically validated tests for use in assessing embryotoxic potential according to the European Centre for the Validation of Alternative Methods (ECVAM) procedures:

- embryonic stem cell test (EST, Genschow *et al*. 2004)
- limb bud micromass culture (Spielmann *et al*. 2004)
- whole embryo culture (WEC, Piersma *et al*. 2004)

Recently, *in vitro* tests for detecting a potential to affect endocrine activity have become available (Nordic Chemicals Group, 2005). Most of the assays that are relevant to reproductive toxicity are designed to assess the ability of a chemical to bind and activate or block the androgen receptor (AR) or the oestrogen receptor (ER). These include cell-free or whole cell binding assays, cell proliferation assays and transcription assays. Also, tests for detecting the ability to interfere with steroidogenesis are currently being developed.

The latest information on the status of alternative methods that are under development can be obtained from the ECVAM website (current address: ecvam.jrc.cec.eu.int) and other international centres for validation of alternative methods.

Animal data

Data may be available from a wide variety of animal studies, which give different amounts of direct or indirect information on the potential reproductive toxicity of a substance; e.g.:

- screening studies (such as OECD TGs 421 or 422)[40]
- other short-term *in vivo* screening tests (e.g. Chernoff/Kavlock tests see Hardin *et al.* 1987, uterotrophic and Hershberger assays)
- one- or two- (or multi-) generation studies (such as B.35, OECD TGs 415 or 416,or EU B.34 or a 'F1-extended one-generation study, as proposed by the ILSI Agricultural Chemical Safety Assessment Project)
- prenatal developmental toxicity tests (such as EU B.31, OECD TG 414)
- developmental neurotoxicity studies (such as draft OECD TG 426)[41]
- peri-postnatal studies
- male or female fertility studies of non-standard design
- repeated-dose toxicity studies, if relevant parameters are included, for example semen analysis, oestrous cyclicity and/or reproductive organ histopathology
- dominant lethal assay (EU B.22, OECD TG 478)
- mechanistic and toxicokinetic studies
- studies in non-mammalian species

[40] To date there are no corresponding EU testing methods available.

[41] To date there is no corresponding EU testing method available.

R.7.6.3.2 Human data on reproductive toxicity

Epidemiological studies, conducted in the general population or in occupational cohorts, may provide information on possible associations between exposure to a chemical and adverse effects on reproduction. Clinical data and case reports (e.g. biomonitoring after accidental substance release) may also be available.

R.7.6.4 Evaluation of available information for reproductive toxicity

The generic guidance on the process of judging and ranking the available data for its adequacy (reliability and relevance) completeness and remaining uncertainty is provided in Chapter R.4. This generic guidance is relevant to reproductive toxicity.

R.7.6.4.1 Non-human data on reproductive toxicity

Non-testing data on reproductive toxicity

Physico-chemical properties

It may be possible to infer from the physico-chemical characteristics of a substance whether it is likely to be absorbed following exposure by a particular route and, furthermore, whether it (or an active metabolite) is likely to cross the placental, blood-brain or blood-testes barriers, or be secreted in milk. Information on the physico-chemical properties may contribute to a *Weight of Evidence* assessment.

Additional generic guidance on this topic is provided in Section R.7.6.6 (see also Section R.4.4).

Read-across to structurally or mechanistically similar substances (SAR)

The concept of structure-activity relationships (SAR) offers approaches for estimating the reproductive toxicity potential of a substance. By grouping substances with similar structures there is an opportunity for the toxicity potential of well-investigated substances to be extended to substances for which there are no or incomplete data. This is particularly the case where the toxicity profile (or lack thereof) can be associated with structural characteristics and reproductive toxicity potential may be extrapolated or interpolated across a homologous series or category. Such an approach has been endorsed under the chemical category concept, which has been developed under the OECD HPVC program (OECD 2004) and further elaborated for the context of REACH as an approach to fill data gaps with a reduced requirement for testing.

Another consideration relates to a substance for which a mechanism of toxicity has been identified that is causally related to reproductive toxicity. In such cases, substances with a similar mechanism identified in other screening tests (e.g. repeated-dose toxicity tests or screens for endocrine activity) may reasonably be expected to exhibit the same pattern of reproductive toxicity. Further testing may be required, on a case-by-case basis, to support a read-across proposal.

Additional generic guidance on this topic, including reporting formats, is provided in Section R.6.2.6.

(Q)SAR

There are a large number of potential targets/mechanisms associated with reproductive toxicity that, on the basis of current knowledge, cannot be adequately covered by a battery of QSAR models. Unlike some toxicological endpoints for which specific structural alerts have been identified (e.g. mutagenicity, sensitisation), there are currently no formal criteria to identify structural alerts for reproductive toxicity.

QSAR approaches are currently not well validated for reproductive toxicity and consequently no firm recommendations can be made concerning their routine use in a testing strategy in this area. Therefore, a negative result from current QSAR models cannot be interpreted as demonstrating the absence of a reproductive hazard unless there is other supporting evidence. Another limitation of QSAR modelling is that dose-response information, for example the N(L)OAEL, required for risk assessment is not provided.

However, a positive result in a validated QSAR model could provide a trigger (alert) for further testing but because of limited confidence in this approach such a result would not normally be adequate as a primary support for a hazard classification decision.

Additionally, QSAR models could be used as part of a *Weight of Evidence* approach, when considered alongside other data, provided the applicability domain is appropriate. Also, QSARs can be used as supporting evidence when assessing the toxicological properties by read-across within a substance grouping approach, providing the applicability domain is appropriate. Positive and negative QSAR modelling results can be of value in a read-across assessment.

Additional generic guidance on QSARs is provided in Section R.6.1.

Testing data on reproductive toxicity

In vitro data

In vitro testing is a rapidly developing field, with significant recent improvements particularly in developmental toxicity and the detection of a potential to affect endocrine activity, which holds much promise for the future. The design of alternatives to *in vivo* testing for reproductive toxicity is especially challenging in view of the complexity of the reproductive process and large number of potential targets/mechanisms associated with this broad area of toxicity.

At the present time *in vitro* approaches have many limitations, for example the lack of capacity for biotransformation of the test substance (Coecke et al 2006). Consequently, no firm recommendations can be made for the exclusive use of *in vitro* methods in a testing strategy for reproductive toxicity. The combination of assays in a tiered and/or battery approach may improve predictivity, but the *in vivo* situation remains more than the sum of the areas modelled by a series of *in vitro* assays (see Piersma 2006 for review). Therefore, a negative result for a substance with no supporting information cannot be interpreted with confidence as demonstrating the absence of a reproductive hazard. Another limitation of in vitro tests is that a N(L)OAEL and other dose-response information required for a risk assessment is not provided.

However, a positive result in a validated *in vitro* test could provide a justification for further testing, dependent on the effective concentration and taking account of what is known about the toxicokinetic profile of the substance. However, because of limited confidence in this

approach at this time, such a result in isolation would not be adequate to support hazard classification.

Additionally, validated and non-validated *in vitro* tests, provided the applicability domain is appropriate, could be used with other data in a *Weight of Evidence* assessment approach to gathering the information required to support a classification decision and risk assessment. *In vitro* techniques can be used in mechanistic investigations, which can also provide support for regulatory decisions. Also, *in vitro* tests can be used as supporting evidence when assessing the toxicological properties by read-across within a substance grouping approach, providing the applicability domain is appropriate. Positive and negative *in vitro* test results can be of value in a read-across assessment. Generic guidance is given in Chapters R.4 and R.5 for judging the applicability and validity of the outcome of various study methods.

Notably, the recent validation study of the three most promising tests for detection of developmental effects, the embryonic stem cell test, the limb bud micromass culture and the whole embryo culture, showed that these had high predictivity for the limited number of strongly embryotoxic chemicals included in the study (Genschow *et al***.** 2002, Piersma 2006, Spielmann *et al*. 2006). However, a number of weaknesses in the design of both the validation study and of the *in vitro* tests have been identified, such as the limited number and range of substances tested and absence of a biotransformation system, which have lead to the conclusion that the tests currently have limited value in a regulatory context. Nevertheless, as discussed above, the results of these tests can have a role, when considered alongside other data, in a *Weight of Evidence* assessment and in support of read-across approaches, and can serve as a trigger for further testing. The results of other *in vitro* tests for developmental toxicity should be assessed with reference to the generic guidance given in Section R.4.3.1.1.

The currently available *in vitro* testing approaches, focusing on the AR and ER binding and transcription have the following limitations. Endocrine disruption may occur via mechanisms other than through the AR or ER such as alterations in hormone synthesis or transport, actions on other receptors and altered metabolism, endpoints for which *in vitro* tests are not currently available. Furthermore, many *in vitro* test systems lack metabolic capability or the range of chemicals that can be tested is restricted due to of problems with solubility in the testing medium. Nevertheless, for certain classes of chemicals that do not require metabolic activation or deactivation, or the metabolites are known and tested, *in vitro* testing may offer practical advantages in terms of speed and cost over *in vivo* screening. Overall, positive *in vitro* test results may indicate a potential to affect endocrine activity *in vivo* by a mechanism relevant for humans, particularly if the *in vitro* activity is high, and may therefore provide a justification for *in vivo* testing. However, negative *in vitro* test results do not provide a reliable indication of a lack of potential to cause reproductive toxicity because of these limitations.

Animal data

Repeated-dose toxicity studies

Although not aimed directly at investigating reproductive toxicity, repeated-dose toxicity studies (e.g. EU B.7, OECD TG 407) may reveal clear effects on reproductive organs in adult animals. However, if these findings occur in the presence of marked systemic toxicity (up to the highest dose level tested in a repeated-dose study) may lower concerns for effects

on fertility and can contribute to decisions on further testing requirements. However, this does not rule out the possibility that the substance may have the capacity to affect fertility.

The observation of effects on reproductive organs in repeated-dose toxicity studies may also be sufficient for identifying a N(L)OAEL for use in the risk assessment. It should, however, be noted that the sensitivity of repeated-dose toxicity studies for detecting effects on reproductive organs may be less than reproductive toxicity studies because of the lower number of animals per group. In addition, a number of cases have demonstrated that effects on the reproductive system may occur at lower doses during the development of foetuses and young animals than in adults. Consequently, in cases where there are substantiated indications for adverse effects on the reproductive organs of adult animals the use of an increased assessment factor in the risk assessment process may be considered. Alternatively, further studies, for example a screening test (OECD TG 421)[42] or a two-generation study (EU B.35, OECD TG 416) may be triggered based on a *Weight of Evidence* assessment. Some effects seen in repeated-dose toxicity studies may be difficult to interpret, for example changes in sex hormone level, and should be investigated further as part of studies that may be required to meet standard REACH information requirements (for example EU B.26, OECD TG 408 or other repeated-dose toxicity studies), rather than serve as a trigger for the immediate conduct of a two-generation study.

Repeated-dose toxicity studies may also provide indications to evaluate the need to investigate developmental neurotoxicity endpoints.

***In vivo* assays for endocrine disruption**

The endocrine system has a critical role in the control of all aspects of the reproductive cycle and therefore endocrine disruption is a potential mechanism for reproductive toxicity.

A number of new *in vivo* assays are under development and may be available for a chemical (see Hass *et al*, 2004 for a detailed discussion). However, none of these assays are standard REACH information requirements and they do not have a role in the ITS (see Section R.7.6.6). The performance of these *single endpoint* assays is not favoured, unless there is strong scientific justification, as they provide only limited information in relation to the numbers of animals used.

The uterotrophic (OECD 2003a) and Hershberger (OECD 2003b) assays, presently being internationally evaluated under the OECD Test Guideline Program, appear reliable in identifying substances with oestrogenic or (anti)androgen modes of action. These studies involve dosing of immature or ovarectomised/castrated animals, and the weighing of oestrogen/ androgen dependent tissues (e.g. uterus or prostate).

A negative result in the uterotrophic assay, in a thorough dose-response study, indicates that the test substance is not an ER-ligand *in vivo*. Equally, a negative result in the Hershberger assay indicates that the test substance is neither an AR-ligand nor a 5-alpha reductase inhibitor *in vivo*. A test compound found negative in these assays may, however, still have endocrine disrupting properties as well as a potential for reproductive toxicity mediated through other mechanisms. Nevertheless, the uterotrophic and Hershberger assays provide *in vivo* NOEL/LOELs for the endpoints examined.

[42] To date there is no corresponding EU testing method available.

A number of assays in experimental animals may provide information on the ability of a substance to act on the production of steroids, and the pubertal assays and the intact male assay provide information about the potency of the compound *in vivo* (US-EPA 2002). Effects on the various endpoints included in these assays can be considered adverse and/or as representing an effect on a mechanism relevant for humans.

In summary, while these *in vivo* assays are considered predictive for hazard identification and risk assessment, and give indications of effects that may be seen in a more comprehensive study, they are not definitive studies. Positive and negative results in the uterotrophic or Hershberger assays, as well as pubertal assays, may be used in combination with other evidence to satisfy the data needs for the classification and risk assessment for effects on reproduction. Positive effects may also provide justification for the conduct of further higher tier testing, such as the two-generation study (EU B.35, OECD TG 416).

As part of the OECD test guideline development program, work is being conducted with the aim of updating the repeated-dose 28 day oral toxicity study (EU B.7, OECD TG 407, reviewed by Gelbke *et al* (2006) to ensure that chemicals acting through (anti)estrogenic, (anti)androgenic and (anti)thyroid mechanisms can be identified. The enhancements include additional parameters based on the respective target organs of the male and female reproductive tracts and the thyroid. Initial validation studies indicate that an enhanced design can reliably identify substances with a strong potential to act through endocrine modes of action on the gonads and thyroid. A negative result with respect to endocrine activity in such a study up to the highest dose tested provides some evidence of the absence of potent effects. However, effects of lower potency cannot be ruled out and therefore a negative result does not provide reassurance of the absence of the capability to cause reproductive toxicity via the mechanism of endocrine disruption. Notably in this context, prolongation of exposure from 28 days up to 90 days is unlikely to improve the detectability of endocrine effects (Gelbke *et al*, 2006). Evidence of endocrine disruption seen in a repeated-dose toxicity study provides a trigger for the conduct of a more comprehensive study, for example a two-generation study (EU B.35, OECD TG 416).

***In vivo* reproductive toxicity tests**

The available OECD test guidelines (or drafts) specifically designed to investigate reproductive toxicity are shown in Table R.7.6-1.

The purpose of Reproduction/Developmental Toxicity Screening Test (OECD TGs 421 and 422) is to provide information of the effects on male and female reproductive performance such as gonadal function, mating behaviour, conception, development of conceptus and parturition. The observation of clear evidence of adverse effects on reproduction or on reproductive organs in these tests may be sufficient to meet the information needs for a classification and risk assessment (using an appropriate assessment factor), and providing a N(L)OAEL from which a DNEL can be identified. If so, there may be no requirement for the conduct of a two-generation study at higher tonnage levels (see the Testing Strategy in Section R.7.6.6 for more information). However, the results should be interpreted with caution because OECD TGs 421/422 are screening assays that were not designed as an alternative or a replacement of the definitive reproductive toxicity studies (OECD TGs 414 and 416, EU B.31 and B.35). These screening tests are not meant to provide complete information on all aspects of reproduction and development. In particular, the post-natal effects associated with prenatal exposure (such as undetected malformations affecting viability or functional effects) or effects resulting from post-natal or lactational exposure are not covered in these studies. Furthermore, the exposure duration in these studies may not be

sufficient to detect all effects on the spermatogenic cycle, although it is likely that in practice the 2-week exposure period will be sufficient to detect the majority of testicular toxicants (Ulbrich and Palmer, 1995). However, the number of animals per dose group is limited which may affect the statistical power of the study to detect an effect. These screening tests may in some cases give indications for reproductive effects (e.g. fertility and post natal effects) that cannot be investigated in a prenatal developmental toxicity study (OECD TG 414, EU B.31). A negative result in a screening study may lower concerns for reproductive toxicity, but this will not provide reassurance of the absence of this hazardous property. However, a negative result can provide the basis for a DNEL in relation to reproductive toxicity derived from the highest dose level used in the study and using an assessment factor that takes account of the limitations of this study; but note that such a DNEL will be relevant only at the Annex VIII level. An evaluation of the OECD TG 421 or TG 422 has confirmed that these tests are useful for initial hazard assessment and can contribute to decisions on further test requirements (Reuter et al 2003, Gelbke et al 2004).

The two-generation study (OECD TG 416, EU B.35) is a general test which allows evaluation of the effects of the test substance on the complete reproductive cycle including libido, fertility, development of the conceptus, parturition, post-natal effects in both dams (lactation) and offspring and the reproductive capacity of the offspring. The two-generation study has conventionally been preferred to the one-generation study (OECD TG 415, EU B.34) in the testing of chemicals because the latter does not test for potential effects on all phases of the reproductive cycle. Post weaning development, maturation and the reproductive capacity of the offspring are not assessed. Consequently some adverse effects, for example oestrogenic- or antiandrogenic-mediated alterations in testicular development, may not be detected. The ILSI Agricultural Chemical Safety Assessment Project has proposed a *F1-extended one-generation study* (as described by Cooper et al 2006). If properly validated and accepted in the EU this could be used in place of the two-generation study as the preferred definitive study to test for reproductive toxicity. This flexible study addresses the main limitation of OECD TG 415 (EU B.34) by incorporating additional post-natal evaluations, which include clinical pathology, a functional observation battery, immunotoxicity endpoints, oestrous cyclicity and semen analysis, and using an extended F1 generation dosing period (to PND day 70) endpoints addressing developmental neurotoxicity. The study has a shortened F0 male premating dosing period, justified by the observation of no differences in the detection rates for adverse effects on fertility between 4- and 9-week pre-mating dosing periods in a number of studies (reviewed by Ulbrich and Palmer 1995).

The prenatal developmental toxicity study (OECD TG 414, EU B.31) provides a focussed evaluation of potential effects on prenatal development, although only effects that are manifested before birth can be detected.
Positive results in these studies will be relevant to hazard classification and the human health risk assessment, unless there is information to show that effects seen in these studies could not occur in humans. N(L)OAELs can be identified from OECD TGs 414 (EU B.31), 415 (EU B.34), 416 (EU B.35), draft 426 and the F1-extended one-generation study.

Developmental neurotoxicity studies (e.g. draft OECD TG 426) are designed to provide information on the potential functional and morphological hazards to the nervous system arising in the offspring from exposure of the mother during pregnancy and lactation. These studies investigate changes in behaviour due to effects on the central nervous system (CNS) and the peripheral nervous system. As behaviour also may be affected by the function of other organs such as liver, kidneys and the endocrine system, toxic effects on these organs in

offspring may also be reflected in general changes in behaviour. No single test is able to reflect the entire complex and intricate function of behaviour. For testing behaviour, therefore, a range of parameters, a *test battery*, is used to identify changes in individual functions.

In exceptional cases when relevant triggers are met testing for developmental neurotoxicity effects should be considered. Relevant triggers could be if the substance has been shown to (1) cause structural abnormalities of the central nervous system, (2) cause clear signs of behavioural or functional adverse effects of nervous system involvement in adult studies e.g. repeated-dose toxicity studies or (3) have a mode of action that has been closely linked to neurotoxic or developmental neurotoxicity effects e.g. cholinesterase inhibition or thyroid effects. However, in the case of (3) targeted testing on the specific mode of action in developing animals may provide sufficient information for regulatory purposes.

The DNT test protocol (draft OECD TG 426, developmental neurotoxicity, not a REACH standard information requirement) is designed to be performed as an independent study. However, observations and measurements described in the protocol can also be added on to a two-generation reproduction study (EU B.35, OECD TG 416). An advantage of this approach is that fewer animals are needed compared to running both studies separately. However, when the developmental neurotoxicity study is incorporated within or attached to another study, it is imperative to preserve the integrity of both study types.

Positive results in a developmental neurotoxicity study will be relevant to hazard classification and the human health risk assessment, providing a N(L)OAEL, unless there is information to show that effects seen in these studies could not occur in humans.

See Nordic Chemicals Group (2005), ECETOC (2002) and WHO (2001) for more detailed reviews of how to interpret the test guidelines mentioned in this report, including a discussion of their strengths and limitations.

Table R.7.6-1 Overview of *in vivo* OECD test guidelines for reproductive toxicity

Test	Design	Endpoints
OECD TG 416 Two-Generation study	Exposure before mating for at least one spermatogenic cycle until weaning of 2nd generation 3 dose levels plus control N = 20 parental males and females	Fertility Oestrus cyclicity and sperm quality Pregnancy outcome, e.g. dystocia Growth, development and viability Anogenital distance if triggered Sexual maturation Histopathology and weight of reproductive organs, brain and target organs Recommended: motor activity, sensory function, reflex ontology in F_1 generation
OECD TG 415 One-Generation Study (not a standard REACH information requirement)	Exposure before mating for at least one spermatogenic cycle until weaning of 1st generation 3 dose levels plus control N = 20 parental males and females	Fertility Growth, development and viability Histopathology and weight of reproductive organs, brain and target organs
OECD TG 414 Prenatal Developmental Toxicity Study (Teratology study)	At least from implantation to one or two days before expected birth 3 dose levels plus control N = 20 pregnant females	Implantation, resorptions Foetal growth Morphological variations and malformations
OECD TG 426 Developmental Neurotoxicity Study (draft, not a standard REACH information requirement)	At least from implantation throughout lactation (PND 20) 3 dose levels plus control N = 20 pregnant females	Birth and pregnancy length Growth, development and viability Physical and functional maturation Behavioural changes due to CNS and PNS effects Brain weights and neuropathology
OECD TG 421 and 422 Reproduction/ Developmental toxicity screening test	From 2 weeks prior to mating until at least day 4 postnatally 3 dose levels plus control N = 8-10 parental males and females	Fertility Pregnancy length and birth Foetal and pup growth and survival until day 4 OECD TG 422 combines reproduction/developmental screen with repeated-dose toxicity investigations that are in concordance with the requirements of OECD TG 407

Developmental effects should be considered in relation to adverse effects occurring in the parents. Since adverse effects in pregnancy or postnatally may result as a secondary consequence of maternal toxicity, reduced food or water intake, maternal stress, lack of maternal care, specific dietary deficiencies, poor animal husbandry, intercurrent infections etc., it is important that the effects observed should be interpreted in conjunction with possible concomitant maternal toxicity (ECB 2004, Fleeman *et al*. 2005, Cappon *et al*. 2005). The nature, severity and dose-response of all effects observed in progeny and parental animals should be considered and compared together to achieve a balanced integrated assessment of available data on all endpoints relevant for reproductive toxicity.

R.7.6.4.2 Human data on reproductive toxicity

Epidemiological data require a detailed critical appraisal that includes an assessment of the adequacy of controls, the quality of the health effects and exposure assessments, and of the influence of bias and confounding factors. Epidemiological studies, case reports and clinical data may provide sufficient hazard and dose-response evidence for classification of chemicals as reproductive toxicants in Category 1 and for risk assessment, including the identification of a N(L)OAEL. In such cases, there will normally not be a need to test the chemical. However, convincing human evidence of reproductive toxicity for a specific chemical is rarely available because it is often impossible to identify a population suitable for study that is exposed only to the chemical of interest. Human data may provide limited evidence of reproductive toxicity that indicates a need for further studies of the chemical; the test method selected should be based on the potential effect suspected.

When evidence of a reproductive hazard has been derived from animal studies it is unlikely that the absence of evidence of this hazard in an exposed human population will negate the concerns raised by the animal model. This is because there will usually be methodological and statistical limitations to the human data. For example, statistical power calculations indicate that a prospective study with well-defined exposure during the first trimester with 300 pregnancies could identify only those developmental toxins that caused at least a 10-fold increase in the overall frequency of malformations; a study with around 1000 pregnancies would have power to identify only those developmental toxins that caused at least a 2-fold increase (EMEA/CHMP Guideline, 2006). Extensive, high quality and preferable prospective, data are necessary to support a conclusion that there is *no risk* from exposure to the chemical.

R.7.6.4.3 Exposure considerations for reproductive toxicity

General information on the pattern and extent of human exposure to the substance must be considered, as this may influence the data requirements with respect to reproductive toxicity. Generic aspects of data waivers based on exposure considerations are presented in Section R.5.1.3. There are rules for waiving certain reproductive information requirements that include criteria relating to human exposure levels in REACH Annexes IX and X. Furthermore, all the reproductive toxicity tests (and also most other *in vivo* toxicity) may be omitted at any of the tonnage levels based on exposure scenarios developed in the Chemical Safety Report according to REACH Annex XI Section 3. The influence of human exposure on the reproductive toxicity ITS is discussed in more detail in Section R.7.6.6.

R.7.6.4.4 Remaining uncertainty on reproductive toxicity

The adequacy and reliability of the various types of data that may be available, or could be generated using the Integrated Testing Strategy (see Section R.7.6.6), as a basis for a decision on classification and for a risk assessment are described in Sections R.7.6.4.1 and R.7.6.4.2.

R.7.6.5 Conclusions on reproductive toxicity

Reproductive toxicity endpoints should be considered collectively, using a *Weight of Evidence* approach to establish the most relevant endpoint and its NOAEL or Critical Effect Dose to be used in risk assessment.

A *Weight of Evidence* assessment involves the consideration of all data that is available and may be relevant to reproductive toxicity, as listed in Section R.7.6.3. There can be no firm rules to the conduct of a *Weight of Evidence* assessment as this process involves expert judgment and because the mix and reliability of information available for a particular substance will probably be unique. Also, the *Weight of Evidence* assessment should consider all toxicity endpoints together, and not look at reproductive toxicity in isolation.

One example of a *Weight of Evidence* assessment is the pooling of information from several *in vivo* reproductive toxicity studies. Individually, these studies may have deficiencies, such as brief reporting, small group size, limited range of endpoints evaluated, the dose levels or the dosing schedule was not appropriate for a comprehensive evaluation of potential effects on the reproductive cycle, the study was not in compliance with GLP. However, taking account of their reliability and relevance and consistency of findings, collectively these studies could provide a level of information similar to that of the EU or OECD test guideline studies, and therefore meet the tonnage-related information requirements needed for the classification decision and risk assessment.

R.7.6.5.1 Concluding on Classification and Labelling

In order to conclude on a proper C&L, all the available information needs to be taken into account, and considerations should be given to both Annex VI of the Directive 67/548/EEC[43] and the various remarks (as they relate to classification and labelling) made throughout this guidance document.

R.7.6.5.2 Concluding on suitability for Chemical Safety Assessment

In order to be suitable for CSA appropriate DNELs have to be established for each exposure scenario. Typically, the derivation of the DNEL takes into account a dose descriptor, modification of the starting point and application of assessment factors - see Chapter R.8 and Appendix R.8-12.

R.7.6.6 Integrated Testing Strategy (ITS) for reproductive toxicity

R.7.6.6.1 Objective / General principles

Fundamentally based on a *Weight of Evidence* approach, the Integrated Testing Strategy (ITS) has been developed around two core objectives:

- to have sufficient information to support risk assessment.
- to have adequate information to consider whether classification as a reproductive toxicant is warranted.

With these objectives underpinning each stage of the process, the ITS was designed to permit informed decisions on reproductive toxicity potential in a step-by-step tiered manner, within the production tonnage related data requirements framework of REACH Annexes VII to X and

[43] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

influenced by toxicological factors (termed *alerts*, see Section R.7.6.6.4) or exposure considerations that may increase or decrease concerns for reproductive toxicity.

By adhering to the criteria outlined above, the ITS will enable decisions to be made at the relevant tonnage level on the need for further testing or whether sufficient information already exist to meet the agreed objectives. Furthermore, if further testing is deemed necessary, the use of the most appropriate study in accordance with the REACH proposal is considered rather than a *one study fits all* approach. An overarching principle is that all data requirements are met in the most efficient and humane manner so that animal usage and costs are minimised.

R.7.6.6.2 Preliminary considerations

Consistent with the parameters defined within the REACH programme (Annex VII-XI), testing for reproductive toxicity is not required for chemicals produced at tonnage levels <10 tonnes per annum (t/y), although all available information relevant to reproductive toxicity must be evaluated, and classification for this area of toxicity should be considered. At higher production volumes, standard data requirements are, in general, proportional to the tonnage level (≥10 t/y, ≥100 t/y or ≥1000 t/y) although maintaining flexibility to adopt the most appropriate testing regime for any single chemical is a key component of the ITS.

However, regardless of tonnage level, before any testing is triggered, careful consideration of all the available toxicological data, exposure characteristics and current risk management procedures is necessary to ascertain whether the fundamental objectives of the ITS (see above) have already been met. This consideration should take account of discussions that have taken place under other regulatory schemes, such as the EU Existing Substances Regulation (ESR), pesticides and the EU hazard classification scheme. If it is concluded that further testing is required, then a series of decision points are defined to help shape the scope of an appropriate testing programme. To satisfy these multiple objectives, the ITS provides a three-stage process for clear decision-making, relevant for all tonnage levels ≥10 t/y.

Stage 1. a series of preliminary questions to consider before deciding on the scope of further reproductive toxicity testing that may be required. Therefore, dependent on the outcome of this analysis, it is possible that some chemicals may not progress beyond Stage 1.

Stage 2. evaluation of the available toxicology database and consideration of reproductive toxicity alerts. This evaluation should consider data for substances with a similar structure or causing toxicity via a similar mode of action. The aim of this stage is to determine the scope of reproductive and/or developmental toxicity testing necessary to satisfy the REACH information requirements. It is possible that, following this review coupled to a *Weight of Evidence* analysis in Stage 1 or if sufficient data for risk assessment/risk management and classification purposes already are available, no further testing may be necessary.

Stage 3. describes the relevant reproductive and developmental toxicity tests upon which classification, labelling and risk assessment decisions will be based for chemicals progressing beyond Stages 1 and 2.

R.7.6.6.3 Testing strategy for reproductive toxicity

Stage 1. Questions to consider before deciding whether any testing for reproductive toxicity potential is required (relevant for all tonnage levels ≥10 t/y)

Stage 1.1. Has the substance already been classified for effects on fertility as Reproductive Toxicity Category 1 or Cat 2: R60 and development as Reproductive Toxicity Category 1 or Category 2: R61?
If the answer is no, proceed to Stage 1.2. If the answer is yes, and the available data are adequate to support a robust risk assessment, then no further testing for reproductive toxicity will be necessary. If the available data are not adequate to support a robust risk assessment then proceed to Stage 2.

Stage 1.2. Is the substance classified as a genotoxic carcinogen (Carcinogen Category 1 and Mutagen Category 3 or Carcinogen Category 2 and Mutagen Category 3) or a germ cell mutagen (Mut. Cat. 1 or Cat. 2)?
If the answer is no, proceed to Stage 1.3. If the answer is yes, it is important to establish that appropriate risk management measures addressing potential carcinogenicity, genotoxicity and reproductive toxicity have been implemented and therefore further specific testing for reproductive and/or developmental toxicity will not be necessary. Exceptionally, appropriate risk management measures may not be in place and a Stage 2 review of the available data should be considered.

Stage 1.3. Does the substance exhibit (a) low toxicological activity and (b) negligible systemic absorption and (c) no or no significant human exposure?
At the ≥100 and ≥1000 t/y levels, no further testing for reproductive toxicity will be required if all three criteria (a, b and c, above) are met; otherwise proceed to the stage 2 analysis. In addition, testing will not be required if the application of a parallel exposure-based information waiving provision in Annex XI Section 3 of REACH (*Substance-tailored exposure-driven testing*) is justified.
However, these three criteria do not apply at the >10 t/y level. At this level, no further testing for reproductive toxicity will be necessary only if the application of the exposure-based information waiving provision in Annex XI Section 3 of REACH is justified; otherwise proceed to the stage 2 analysis.

For further discussion see Section R.7.6.6.5.

Stage 2. Conduct a detailed review of all existing toxicological data to identify any specific alerts and testing requirements for reproductive and/or developmental toxicity

Substances may be excluded from further testing at Stage 3 (for more details on criteria for decision making, see Section R.7.6.6.4), this can only be achieved if sufficient data exist to conclude[44] that the substance does not present a reproductive toxicity hazard or that further data are unlikely to change a classification in Reproductive Toxicity Category 3. In the latter case, a thorough scientific justification is needed.

≥10 t/y

[44] Data adequate for Classification and Labelling and risk assessment

Before any testing is conducted, all substances at this tonnage level will be subject to a thorough data review. If sufficient data are available to permit a conclusion on reproductive and developmental toxicity potential, then no further testing is required. If there is insufficient data or alerts exist, then a testing strategy for reproductive and/or developmental toxicity in Stage 3 will be recommended. It should be pointed out that the observation of no adverse effects on the reproductive organs in a repeated-dose toxicity, such as a 28- or 90-day toxicity study, may justify a lower priority for further testing for effects on fertility. However, this would not provide sufficient data to justify a lower priority for testing for effects on development.

≥100 and ≥1000 t/y

For substances at these tonnage levels progressing beyond Stage 1, the standard data requirements include the definitive OECD tests for reproductive toxicity (for details see Stage 3 below). However, before any specific reproductive toxicity testing is undertaken, all substances at these tonnage levels will be subject to a thorough Stage 2 data review. If sufficient data exist to permit a robust conclusion on reproductive toxicity potential, then no further testing is required. If there is insufficient data or alerts exist, then a reproductive toxicity testing strategy for Stage 3 will be recommended.

Stage 3 Reproductive toxicity tests triggered by tonnage level or alerts identified in Stage 1 and 2

Four internationally harmonised guideline studies are listed in the REACH Annexes that can be used at Stage 3 to provide the necessary information to support a robust classification and risk assessment and to identify N(L)OAELs. However, it will not usually be necessary to assess all chemicals reaching Stage 3 in all four tests. Instead, individual chemical testing requirements will be customised based on the nature of alerts identified in Stages 1 and 2 and the tonnage level of that substance.
The tests listed in the REACH annexes are:

- Reproduction/developmental toxicity screening test (OECD TG 421) OR the combined repeat dose toxicity study with the reproduction/ developmental toxicity screening test (OECD TG 422)
- Prenatal developmental toxicity study (OECD TG 414, EU B.31) in a first species and possibly second species
- Two-generation reproduction study (OECD TG 416, EU B.35)[45]

A brief description of the study protocols considered in the ITS is presented in Table R.7.6-1 in Section R.7.6.4.1. Utilisation of these tests at each of the three tonnage levels is summarised below.

≥ 10 t/y

At this tonnage level, progression beyond Stages 1 and 2 will trigger a reproduction/ developmental toxicity screening test (OECD TG 421/422) as the standard information requirement, if there is no evidence from available information on structurally related substances, QSAR estimates or from *in vitro* methods that the substance may be a

[45] As discussed earlier (Section R.7.6.4.1), a proposed *F1-extended one-generation study* may replace OECD TG 416 as a definitive study for reproductive toxicity in the near future, subject to gaining regulatory acceptance in the EU.

developmental toxicant. If this test provides no alerts for reproductive and developmental toxicity, then dependent on the *Weight of Evidence* from Stages 1 and 2, further testing for reproductive toxicity will not be required at this tonnage level. Similarly, if a clear and unequivocal reproductive and/or developmental toxicity effect is observed in these tests which is deemed sufficient to enable a scientifically robust decision on classification and risk assessment, then no further testing beyond the OECD TG 421 or 422 is recommended at this tonnage level. If a 28-day study (EU B.7, OECD TG 407) is not already available, the conduct of the OECD TG 422 is preferred to TG 421 for animal welfare reasons, as the former also includes an investigation of repeated-dose toxicity equivalent to that of the 28-day toxicity study, thus eliminating the need to conduct the 28-day study (see Section R.7.5). If an alert for reproductive and/or developmental toxicity is generated from an OECD TG 421 or OECD TG 422 study but is deemed insufficient for a classification assessment then the regulatory actions should take account of the additional uncertainty. For example, the DNEL identification may require the application of a larger Assessment Factor; exceptionally, further testing may be required on a case-by-case basis. The specific testing requirement will be dependent on the nature of the alerts.

However, dependent on the nature of the alert(s) observed in Stages 1 and 2, it may be more appropriate to conduct a two-generation reproduction study[5] (EU B.35, OECD TG 416) or a prenatal developmental toxicity study (EU B.31, OECD TG 414) instead of the screening study. In general, it should be noted that the OECD TG 414 (EU B.31) study does not incorporate post-natal parameters and therefore it is advisable not to bypass the screening study when data of a prenatal developmental toxicity study is either available or a respective study is triggered. This is because the screening study will provide information on the viability and postnatal development of the offspring which can be important to the developmental toxicity assessment, as well as information on many other aspects of reproduction that would not otherwise be available. However, if the prenatal developmental toxicity test was positive, there would be less need for the screening test. If an OECD TG 414 (EU B.31) study has been performed, it will be important to establish whether these data are sufficient to enable a clear regulatory decision and to assess whether the results of further testing for developmental toxicity in a second species are likely to influence regulatory decisions. Testing in a second species will not normally be required at this tonnage level if the study is negative. Additional guidance on the acquisition of information on potential developmental toxicity from two animal species is provided in Section R.7.6.6.4 It should be noted that although the OECD TG 414 study does not incorporate post-natal parameters, some findings might raise concerns for post-natal effects such as pup survival and in such cases follow-up testing in a two-generation reproductive toxicity study (EU B.35, OECD TG 416) in the most relevant species (usually the rat) may be appropriate. Alternatively, such effects could initially be investigated in an OECD TG 421/422 test that has been modified to include an extended postnatal observation period.

≥ 100 t/y

At this tonnage level, progression beyond Stage 1 and 2 will trigger a prenatal developmental toxicity study (OECD TG 414), conducted in the most relevant species (see Section R.7.6.6), as a standard data requirement and, in case of an alert for this test, a two-generation reproduction study (EU B.35, OECD TG 416). Additionally, the ≥10 t/y standard data requirement for a reproduction/developmental toxicity screening test (OECD TG 421/422) will need to be met. However, this screening test will not be

necessary if a two-generation study is proposed at the >100 t/y level; also this test will not be required if adverse effects on reproductive organs have been observed in existing repeated-dose studies and these findings are sufficient to support classification for effects on fertility and the risk assessment.

As for ≥10 t/y substances, following completion of the OECD TG 414 (EU B.31) study it will be important to establish whether these data are sufficient to enable a clear regulatory decision and to assess whether the results of further testing for developmental toxicity are likely to influence regulatory decisions. Guidance on the investigation of developmental toxicity in a second species is presented in Section R.7.6.6. As outlined above in Stage 2 for this tonnage level, a detailed review of the available data will be conducted to identify any reproductive toxicity alerts. This review coupled to the data emerging from the OECD TG 414 (EU B.31) study will form the basis of a Weight of Evidence assessment of the requirement for a two-generation reproductive toxicity study (EU B.35, OECD TG 416). If specific triggers are present as discussed in Section R.7.6.4.1 the need for inclusion of the optional developmental neurotoxicity endpoints should be evaluated. The conduct of OECD TG 416 (EU B.35) should also be considered if it is anticipated that the ≥1000 t/y supply tonnage threshold will be reached in the near future.

REACH Annex IX specific rules for adaptation states that the need to perform a OECD TG 416 (EU B.35) study in a second species, either at this tonnage level or the next, should be considered, based on the outcome of the first test and any other data. However, the two-generation study is very rarely conducted in a species other than the rat, and it is envisaged that a second species study could not be justified.

≥ 1000 t/y

At this tonnage level, progression beyond Stage 1 and 2 will trigger a prenatal developmental toxicity study (EU B.31, OECD TG 414) and a two-generation reproductive toxicity study (EU B.35, OECD TG 416) in the most relevant species as a standard data requirement. The need for a developmental toxicity study (EU B.31, OECD TG 414) in a second species should be evaluated, following the guidance presented above in Section R.7.6.6.3.
If specific triggers are present as discussed in Section R.7.6.4.1, inclusion of optional developmental neurotoxicity endpoints should be considered. The reproduction/ developmental toxicity screening test (OECD TG 421/422) , a standard data requirement at the ≥10 t/y level, will not be needed if a two-generation is conducted because this study provides a superior level of information.

R.7.6.6.4 Elements of the ITS

Alerts from existing toxicological database and their implications for further testing, classification/labelling and risk assessment

Challenging the existing toxicity database from a reproductive toxicity perspective.

An *alert* is any factor, with the exclusion of convincing evidence derived from the definitive reproductive toxicity studies (i.e. OECD TG 414 and 416), that is present in the existing toxicological database, whether based on theoretical considerations or from experimental or observational data, that raises concerns that a substance may be reproductive toxicant.

As part of the Stage 2 data review the following questions should be asked:

- are there alerts for reproductive toxicity?
- are the data sufficient/adequate for assessing the classification and labelling and risk assessment without further testing, irrespective of the presence or absence of alerts?
- if the data are insufficient, what study (or studies) is most appropriate? This decision must take account of both the standard tonnage related information requirements of REACH, the nature of the alert(s) and *Weight of Evidence* as well as human exposure considerations.
- is there any knowledge of the chemical, chemical groups or categories, that would indicate special features to be included in the study design? If so, what?

From a scientific perspective, it is not possible to generate an exhaustive and rigid list of alerts that would automatically trigger a particular study or have clearly defined implications for classification and risk assessment. Instead, alerts mentioned in this report should be viewed as a helpful guide of indicators that would provide input to the regulatory decision-making process – in other words, contribute to a *Weight of Evidence* analysis requiring expert judgement, that leads to the most appropriate testing and regulatory outcome.

Section R.7.6.4, which discusses the information that may be available for a substance, provides many examples of alerts and their implications for testing, classification and risk assessment.

Consideration of existing reproductive studies not required under REACH

Although the REACH standard information requirements refer to a specific series of reproductive studies, it is recognised that there may be other studies already performed that could address some of the endpoints covered by these standard protocols, reducing the need for new animal testing. These could include one-generation studies (for example EU B.34, OECD 415 or the previously discussed *F_1-extended one-generation study*), non-GLP studies, or non-guideline investigations such as the NTP continuous breeding study (Chapin and Sloane, 1997). The available data should be evaluated to assess their suitability for use, taking account of the robustness of design, and quality. As an example, a one-generation study (EU B.34, OECD TG 415) and repeated-dose toxicity study that includes oestrous cycle monitoring and semen analysis may already have been performed. In this case, the level of information available, though not equivalent to that provided by a two-generation study, could be sufficient using a *Weight of Evidence* analysis for classification and risk assessment.

In summary, the information requirements set out in the REACH annexes should be treated as endpoints to be evaluated rather than studies to be conducted. Thus, relevant existing studies that do not conform to the OECD test guidelines referred to the REACH Annexes but nevertheless provide an equivalent level of information can be used to meet the REACH information requirements.

Selection of Species for Assessment of Prenatal Developmental Toxicity

The purpose of the prenatal developmental toxicity study (EU B.31, OECD TG 414) is to identify effects upon organogenesis and foetal growth prior to parturition. For a comprehensive assessment of developmental toxicity according to Annexes IX and X information from two species, one rodent (usually the rat) and one non-rodent (usually the

rabbit) should be considered. When considering the use of two species, care should be given to deciding the order in which these studies are performed. Since most acute, repeated-dose, and toxicokinetic studies are conventionally conducted in the rat, it is advisable that the first developmental toxicity study should also be conducted in this species. Findings from previous studies may be useful in dose selection, or the identification of additional endpoints for evaluation. In addition, the outcome of the prenatal developmental toxicity study may be helpful in the interpretation of other reproductive toxicity studies (e.g. OECD TG 421/422), for which the rat is generally the favoured species.

Although the OECD TG 414 (EU B.31) is designed specifically to identify developmental toxicity, information on this endpoint can also be obtained from observations of the offspring in a one- or two-generation study, which will almost always have been conducted in the rat. So, if a generation study is available, a prenatal developmental toxicity study (EU B.31, OECD TG 414) in the rat may not provide any additional information that would have an influence on the classification decision or risk assessment, and therefore the conduct of this study in the rat may not always be necessary.

If the outcome of this first developmental toxicity study is positive, this may be enough for classification and risk assessment; if this is so, a study in a second species will not be required. Further investigations may be warranted, on a case-by-case basis, if the outcome of the first study is equivocal or if the relevance of the findings to humans is unclear. At ≥1000 t/y, a study in a second species will normally be required when the first study is negative, unless *Weight of Evidence* assessment or specific data e.g. toxicokinetic data provide scientific justification not to conduct the study in a second species. This could be the case if available data demonstrate that for example the rat is the most relevant species for extrapolating to humans or if the rabbit is not a suitable model for testing for developmental toxicity.

R.7.6.6.5 Exposure considerations (and substances of low toxicological activity and with negligible systemic absorption) for reproductive toxicity

Exposure considerations may be used to justify the waiver of certain data requirements or, exceptionally, the conduct of reproductive toxicity testing that is additional to the REACH Annex VIII, IX and X information requirements.

Upgraded testing requirements

The use pattern or the exposures to a substance may indicate a need for additional information requirements, on a case-by-case basis. For example, there may be serious concerns that human exposures, particularly to consumers, are close to the levels at which toxicity might be expected. Such concerns for human health may be satisfactorily addressed by improved risk management measures and therefore additional information on hazard would be of limited value. Thus, proposals to refine a risk assessment with the use of information obtained from new *in vivo* testing that is in excess of the REACH tonnage-related information requirements can be justified only in exceptional circumstances.

Reduced testing requirements: ≥ 10 t/y

As stated in REACH Annex VIII specific rules for adaptation the OECD TG 421/422 study listed as a standard information requirement does not need to be conduced *if relevant human exposure can be excluded* in accordance with Annex XI Section 3. This clause states that tests may be omitted based on exposure scenarios developed in the Chemical Safety Report.

The criteria defining what constitutes adequate justification for omitting these tests under Annex XI Section 3 are not currently available, but will be adopted by the Commission within 18 months of REACH coming into force.

Reduced testing requirements: ≥ 100 t/y and ≥ 1000 t/y

According the REACH Annex IX and X specific rules for adaptation (mainly column 2), the reproductive toxicity studies listed as standard information requirements do not need to be conducted if the three following criteria are met:

1. The substance is of low toxicological activity (no evidence of toxicity seen in any of the tests available) and
2. It can be proven from toxicokinetic data that no systemic absorption occurs via relevant routes of exposure (e.g. plasma/blood concentrations below detection limit using a sensitive method and absence of the substance and of metabolites of the substance in urine, bile or exhaled air) and
3. There is no or no significant human exposure.

At least two cases pertain, the first being no human exposure (e.g., substances only produced and used in closed systems) and the second being no significant human exposure. Whether a human exposure is significant depends on the reproductive toxicity potency of the substance relative to exposure (consequence of a risk) and might be decided on the basis of other information indicating e.g. the probability of a risk. E.g.: At least substances used in closed systems fall under this criterion, but other possibilities may be identified as well e.g. industrial and commercial uses for substances exclusively used in preparations in very low concentrations or substances, uses of substances in consumer products which are completely chemically reacted during manufacturing, integrated in a matrix and characterised by very low migration.

In addition to the REACH Annex IX and X specific rules for adaptation, there is the parallel exposure-based provision in Annex XI Section 3 of the REACH Regulation (*Substance-tailored exposure-driven testing); all the reproductive toxicity tests (and also most other in vivo toxicity*) may be omitted at any of the tonnage levels based on exposure scenarios developed in the Chemical Safety Report. As stated above, the criteria defining what constitutes adequate justification for omitting these tests under Section 3 are not currently available.

R.7.6.7 References on reproductive toxicity

Cappon GD, Fleeman TL, Chapin RE, Hurtt ME (2005) Effects of feed restriction during organogenesis on embryo-foetal development in the rabbit. Birth Defects Res B Dev Reprod Toxicol74, 424-430

Chapin RE, Sloane RA. (1997) Reproductive assessment by continuous breeding: evolving study design and summaries of ninety studies. Environ Health Perspect 105 Suppl 1, 199-205

Coecke S et al (2006) Metabolism: a bottleneck in vitro toxicological test development. The report and recommendations of ECVAM workshop 54. ATLA 34, 49-84

Cooper RL, Lamb IV JC, Barlow SM, Bentley K, Brady AM, Doerrer NG, Eisenbrandt DL, Fenner-Crisp PA, Hines RN, Irvine LFH, Kimmel CA, Koeter H, Li AA, Makris SL (2006) A tiered approach to life stages testing for agricultural chemical safety assessment. Crit Rev Toxicol 36, 69-98

ECETOC: Guidance on Evaluation of reproductive toxicity data, Monograph No. 31, 2002.

EMEA (European Agency for the Evaluation of Medicinal Products) (2006) Guideline on risk assessment of medicinal products on human reproduction and lactation: from data to labelling. EMEA/CHMP/203927. Available at http.//www.emea.eu.int

ECB (European Chemical Bureaux) (2004) Document ECBI/30/04: Expert discussion on the classification of substances toxic to reproduction, Ispra, 4-5 March 2004

Genschow E, Spielmann H, Scholz G, Seiler A, Brown N, Piersma A, Brady M, Clemann N, Huuskonen H, Paillard F, Bremer S, Becker K (2002) The ECVAM international validation study on *in vitro* embryotoxicity tests: results of the definitive phase and evaluation of prediction models. ATLA 30, 151-176.

Fleeman TL, Cappon GD, Chapin RE, Hurtt ME (2005) The effects of feed restriction during organogenesis on embryo-foetal development in the rat. Birth Defects Res B Dev Reprod Toxicol 74, 442-449.

Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N, Bremer S, Becker K (2004) Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. ATLA 32, 209-244

Gelbke H-P, Fleig H, Meder M on behalf of German Chemical Industry Association (2004) SIDS Reprotoxicity screening test update: testing strategy and use. Reg Toxicol Pharm 39, 81-86.

Gelbke HP, Hofmann A, Owens JW, Freyberger A (2006), The enhancement of the subacute repeat dose toxicity test OECD TG 407 for the detection of endocrine active chemicals: comparison with toxicity tests of longer duration. Arch Toxicol, published online 20.9.2006

Hardin BD, Becker RA, Kavlock RJ, Seidenberg JM, Chernoff N (1987) Workshop on the Chernoff/Kavlock preliminary developmental toxicity test. Teratog Carcinog Mutagen 7, 119-27

Nordic Chemicals Group (2005). Information Strategy for Reproductive Toxicity. Nordic Project on Integrated Information Strategies. Step 2 project report (draft 5. April 2005).

OECD (2003a), OECD Draft report of the validation of the rat Uterotrophic bioassay: Phase 2. testing of potent and weak oestrogen agonist by multiple laboratories. ENV/JM/TG/EDTA (2003)1.

OECD (2003b). OECD Draft report of the OECD validation of the rodent Hershberger bioassay: Phase 2. testing of androgen agonist, androgen antagonist and 5 alpha-reductase inhibitor in dose response studies by multiple laboratories. ENV/JM/TG/EDTA (2003)5.

OECD (2004) Manual for Investigation of HPV Chemicals. Available at http://www.oecd.org/document/7/0,2340,en_2649_34379_1947463_1_1_1_1,00.html

Piersma AH, Genschow E, Verhoef A, Spanjersberg MQ, Brown NA, Brady M, Burns A, Clemann N, Seiler A, Spielmann H (2004) Validation of the postimplantation rat whole-embryo culture test in the

International ECVAM validation study on three in vitro embryotoxicity tests. ATLA 32, 275-307

Piersma AH (2006) Alternatives to animal testing in developmental toxicity testing. Basic Clin Toxicol Pharmacol 98, 427-431

Reuter U, Heinrich-Hirsch B, Hellwig J, Holzum B, Welsch F (2003) Evaluation of OECD screening tests 421 (reproduction/developmental toxicity screening test) and 422 (combined repeated-dose toxicity study with the reproduction/developmental toxicity screening test). Reg Toxicol Pharm 38, 17-26.

Spielmann H, Genschow E, Brown NA, Piersma AH, Verhoef A, Spanjersberg MQ, Huuskonen H, Paillard F, Seiler A (2004) Validation of the rat limb bud micromass test in the international ECVAM validation study on three in vitro embryotoxicity tests. ATLA 32, 245-274

Spielmann H, Seiler A, Bremer S, Hareng L, Hartung T, Ahr H, Faustman E, Haas U, Moffat GJ, Nau H, Vanparys P, Piersma A, Riego Sintes J, Stuart J (2006) The practical application of three validated *in vitro* embryotoxicity tests. The report and recommendations of an ECVAM/ZEBET workshop (ECVAM workshop 57). ATLA 34, 527-538

Ulbrich B, Palmer AK (1995) Detection of effects on male reproduction- a literature survey. J Amer Coll Toxicol 14, 292-327

World Health Organization (WHO) (2001) Principles for evaluating health risks to reproduction associated with exposure to chemicals. Environmental Health Criteria 225. World Health Organization, Geneva

Worth A, Balls M (2002) The Principles of Validation and the ECVAM Validation Process. ATLA 30 Suppl 2, 15-21

United Nations (2005) Globally Harmonised System of Classification and Labelling of Chemicals (GHS). Part 3 Health Hazards, page 175. Available at http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html

US-EPA (2002). Draft detailed review paper on steroidogenesis screening assays and endocrine disruptors. May 2002.

R.7.7 Mutagenicity and carcinogenicity

R.7.7.1 MUTAGENICITY

R.7.7.1.1 Definition of mutagenicity

In the risk assessment of substances it is necessary to address the potential effect of *mutagenicity*. It can be expected that some of the available data will have been derived from tests conducted to investigate potentially harmful effects on genetic material (*genotoxicity*). Hence, both the terms *mutagenicity* and *genotoxicity* are used in this document.

The chemical and structural complexity of the chromosomal DNA and associated proteins of mammalian cells, and the multiplicity of ways in which changes to the genetic material can be effected make it difficult to give precise, discrete definitions.

Mutagenicity refers to the induction of permanent transmissible changes in the amount or structure of the genetic material of cells or organisms. These changes may involve a single gene or gene segment, a block of genes or chromosomes. The term clastogenicity is used for agents giving rise to structural chromosome aberrations. A clastogen can cause breaks in chromosomes that result in the loss or rearrangements of chromosome segments. Aneugenicity (aneuploidy induction) refers to the effects of agents that give rise to a change (gain or loss) in chromosome number in cells. An aneugen can cause loss or gain of chromosomes resulting in cells that have not an exact multiple of the haploid number. For example, three number 21 chromosomes or trisomy 21 (characteristic of Down syndrome) is a form of aneuploidy.

Genotoxicity is a broader term and refers to processes which alter the structure, information content or segregation of DNA and are not necessarily associated with mutagenicity. Thus, tests for genotoxicity include tests which provide an indication of induced damage to DNA (but not direct evidence of mutation) via effects such as unscheduled DNA synthesis (UDS), sister chromatid exchange (SCE), DNA strandbreaks, DNA adduct formation or mitotic recombination, as well as tests for mutagenicity.

R.7.7.1.2 Objective of the guidance on mutagenicity

The aims of testing for genotoxicity are to assess the potential of substances to induce genotoxic effects which may lead to cancer or cause heritable damage in humans. Genotoxicity data are used in risk characterisation and classification of substances.

Alterations to the genetic material of cells may occur spontaneously or be induced as a result of exposure to ionising or ultraviolet radiation, or genotoxic substances. In principle, human exposure to substances that are mutagens may result in increased frequencies of mutations above background.

Mutations in somatic cells may be lethal or may be transferred to daughter cells with deleterious consequences for the affected organism (*e.g.* cancer may result when they occur in proto-oncogenes, tumour suppressor genes and/or DNA repair genes) ranging from trivial to detrimental or lethal.

There is considerable evidence of a positive correlation between the mutagenicity of substances *in vivo* and their carcinogenicity in long-term studies with animals. Genotoxic carcinogens are chemicals for which the most plausible mechanism of carcinogenic action involves genotoxicity.

Heritable damage to the offspring, and possibly to subsequent generations, of parents exposed to substances that are mutagens may follow if mutations are induced in parental germ cells. To date, all known germ cell mutagens are also mutagenic in somatic cells *in vivo*. Substances that are mutagenic in somatic cells may produce heritable effects if they, or their active metabolites, reach the genetic material of germ cells. Conversely, substances that do not induce mutations in somatic cells *in vivo* would not be expected to be germ cell mutagens.

R.7.7.2 Information requirements on mutagenicity

The information requirements on mutagenicity are described by REACH Annexes VI to XI, that specify the information that shall be submitted for registration and evaluation purposes. The information is thus required for substances produced or imported in quantities of >1 t/y (tons per annum). When a higher tonnage level is reached, the requirements of the corresponding Annex have to be considered. However, factors including not only production volume but also pre-existing toxicity data, information about the identified use of the substance and exposure of humans to the substance will influence the precise information requirements. The REACH Annexes shall thus be considered as a whole, and in conjunction with the overall requirements of registration, evaluation and the duty of care.

Column 1 of the Annexes VII to X of REACH inform on the standard information requirements for substances produced or imported in quantities of > 1 t/y, >10 t/y, >100 t/y, and >1000 t/y, respectively.

Column 2 of the Annexes VII-X list specific rules according to which the required standard information may be omitted, replaced by other information, provided at a different stage or adapted in another way. If the conditions are met under which column 2 of this Annex allows adaptations, the fact and the reasons for each adaptation should be clearly indicated in the registration.

The standard information requirements for mutagenicity and the specific rules for adaptation of these requirements are presented in Table R.7.7-1.

Table R.7.7-1 REACH information requirements for mutagenicity

COLUMN 1 **STANDARD INFORMATION REQUIRED**	**COLUMN 2** **SPECIFIC RULES FOR ADAPTATION FROM COLUMN 1**
Annex VII: 1. *In vitro* gene mutation study in bacteria.	Further mutagenicity studies shall be considered in case of a positive result.
Annex VIII: 1. *In vitro* cytogenicity study in mammalian cells or *in vitro* micronucleus study. 2. *In vitro* gene mutation study in mammalian cells, if a negative result in Annex VII, 1 and Annex VIII, 1.	1. The study does not usually need to be conducted - if adequate data from an *in vivo* cytogenicity test are available or - the substance is known to be carcinogenic category 1 or 2 or mutagenic category 1, 2 or 3. 2. The study does not usually need to be conducted if adequate data from a reliable *in vivo* mammalian gene mutation test are available. Appropriate *in vivo* mutagenicity studies shall be considered in case of a positive result in any of the genotoxicity studies in Annex VII or VIII.
Annex IX:	If there is a positive result in any of the *in vitro* genotoxicity studies in Annex VII or VIII and there are no results available from an *in vivo* study already, an appropriate *in vivo* somatic cell genotoxicity study shall be proposed by the registrant. If there is a positive result from an *in vivo* somatic cell study available, the potential for germ cell mutagenicity should be considered on the basis of all available data, including toxicokinetic evidence. If no clear conclusions about germ cell mutagenicity can be made, additional investigations shall be considered.
Annex X:	If there is a positive result in any of the *in vitro* genotoxicity studies in Annex VII or VIII, a second *in vivo* somatic cell test may be necessary, depending on the quality and relevance of all the available data. If there is a positive result from an *in vivo* somatic cell study available, the potential for germ cell mutagenicity should be considered on the basis of all available data, including toxicokinetic evidence. If no clear conclusions about germ cell mutagenicity can be made, additional investigations shall be considered.

In addition to these specific rules, the required standard information set may be adapted according to the general rules contained in Annex XI. In this case as well, the fact and the reasons for each adaptation should be clearly indicated in the registration.

In some cases, the rules set out in Annex VII to XI may require certain tests to be undertaken earlier than or in addition to the tonnage-triggered requirements. See for further guidance on testing requirements Section R.7.7.6.

R.7.7.3 Information and its sources on mutagenicity

To be able to evaluate the mutagenic potential of a substance in a comprehensive way, information is required on its capability to induce gene mutations, structural chromosome aberrations (clastogenicity) and numerical chromosome aberrations (aneugenicity). Many test methods are available by which such information can be obtained. Non-testing methods, such as SAR, QSAR and read-across approaches, may also provide information on the mutagenic potential of a substance.

Typically, *in vitro* tests are performed with cultured bacterial cells, human or other mammalian cells. The sensitivity and specificity of tests will vary with different classes of substances and, if adequate data are available for the class of substance to be tested, can guide the selection of the most appropriate test systems to be used. In order to detect mutagenic effects also of substances that need to be metabolically activated to become mutagenic, an exogenous metabolic activation system is usually added in *in vitro* tests. For this purpose the post-mitochondrial 9000 x g supernatant (S-9 fraction) of whole liver tissue homogenate containing a high concentration of metabolising enzymes is most commonly employed. In the case when information is required on the mutagenic potential of a substance *in vivo,* several test methods are available. In *in vivo* tests whole animals are used, in which metabolism and toxicokinetic mechanisms in general exist as natural components of the test animal. It should be noted that species-specific differences in metabolism are known. Therefore, different genotoxic responses may be obtained.

Some test methods have an officially adopted EU/OECD guideline for the testing procedure, although for many test methods this is not the case. Furthermore, modifications to OECD protocols have been developed for various classes of substances and may serve to enhance the accuracy of test results. Use of such modified protocols is a matter of expert judgement and will vary as a function of the chemical and physical properties of the substance to be evaluated. Commonly used non-guideline *in vivo* tests employ methods by which any tissue of an animal can be examined for effects on the genetic material, giving the possibility to examine site-of-contact tissues (*i.e.,* skin, epithelium of the respiratory or gastro-intestinal tract) in genotoxicity testing. In addition, test methods developed over the past decades in *Drosophila* and in various species of plants and fungi are available.

R.7.7.3.1 Non-human data on mutagenicity

Non-testing data on mutagenicity

Non-test information about the mutagenicity of a substance can be derived in a variety of ways, ranging from simple inspection of the chemical structure through various read-across techniques, the use of expert systems, metabolic simulators, to *global* or *local* (Q)SARs. The usefulness of such techniques varies with the amount and nature of information available, as well as with the specific regulatory questions under consideration.

Regarding substances for which testing data exist, non-test information can be used in the total *Weight of Evidence* approach, to help confirm results obtained in specific tests, or to help develop a better understanding of mutagenicity mechanisms. The information may be useful in deciding if, or what, additional testing is required. At the other extreme, where no testing data are available, similar alternative sources of information may assist in setting test priorities. In cases where no testing is likely to be done (low exposure, <1 t/y) they may be the only options available to establish a hazard profile.

Weight of Evidence approaches that use expert judgement to include test results for close chemical analogues are ways of strengthening regulatory positions on the mutagenicity of a substance. Methods that identify general *structural alerts* for genotoxicity such as the Ashby-Tennant super-mutagen molecule (Ashby and Tennant, 1988) may also be useful.

There are hundreds of (Q)SAR models available in the literature for predicting test results for genotoxic endpoints for closely related structures, *i.e.* the one presented by Chung *et al*. (1997). These are known as *local* (Q)SARs. When essential features of the information domain are clearly represented, these models may constitute the best predictive tools for estimating a number of mutagenic/genotoxic endpoints. However, quality of reporting varies from model to model and predictivity must be assessed case-by-case on the basis of clear documentation.

Generally, (Q)SAR models that contain putative mechanistic descriptors are preferred; however many models use purely structural descriptors. While such models may be highly predictive, they rely on statistical methods and the toxicological significance of the descriptors may be obscure.

Another type of (Q)SAR model for mutagenicity attempts to predict (within their domain) diverse (non-congeneric) groups of substances. These are termed *global* (Q)SARs and are far more ambitious than the more simple local models. Global (Q)SARs are all computer programs which in essence first divide chemicals into local (Q)SARs and then make a conventional prediction.

Most global models for mutagenicity are commercial and some of the suppliers of these global models consider the data in their modelling sets to be proprietary. Proprietary means that the training set data used to develop the (Q)SAR model is hidden from the user. In other cases it means that it may not be distributed beyond use by regulatory authorities.

The most common genotoxicity endpoint for global models has been to predict results of the Ames test. Examples of widely used models for this endpoint include an expert system called DEREK, artificial intelligence modules from MULTICASE, the topological system, TOPKAT and the OASIS system which includes a metabolic simulator.

There are models for many other mutagenicity endpoints. For example, the Danish EPA has developed a (Q)SAR database that contains predictions from, in addition to assorted Ames models, models on the following *in vitro* endpoints: chromosomal aberrations (CHO and CHL cells), mouse lymphoma/*tk*, CHO/hprt gene- mutation assays and UDS (rat hepatocytes) information. *In vivo* models include *Drosophila* SLRL, mouse micronucleus, rodent dominant lethal, mouse SCE in bone marrow and mouse Comet assay data. All models were derived using MULTICASE software. This information can be accessed for free over the internet from the ECB (http://ecbqsar.jrc.it/). The Danish database contains predictions for over 166,000 chemicals and includes a flexible system for chemical structure and parameter searching.

Another free source of information on mutagenicity is the Enhanced NCI Database Browser (http://cactus.nci.nih.gov) sponsored by the U.S. National Cancer Institute. It contains predictions for over 250,000 chemicals for mutagenicity as well as other non-mutagenic endpoints, some of which may provide valuable mechanistic information (for example alkylating ability or microtubule formation inhibition). It is also searchable by a wide range of parameters and structure combinations. Modelling was done using PASS (Prediction of Activity Spectra for Substances).

Neither of these two examples is perfect, in part due to commercial considerations, but they illustrate a trend towards predictions of multiple endpoints and may assist those making *Weight of Evidence* decisions regarding the mutagenic potential of untested substances.

Further information on mutagenicity models (and other endpoints) can be found in the OECD Database on Chemical Risk Assessment Models, where they have been assembled as part of an effort to identify tools for use in research and development of chemical substances (www.oecd.fr).

The guidance on (Q)SARs (Section R.6.1) explains basic concepts of (Q)SARs and gives generic guidance on validation, adequacy and documentation for regulatory purposes. The guidance in Sections R.6.2 and R.6.1 describe a stepwise approach for the use of read-across/grouping and (Q)SARs.

Testing data on mutagenicity

Test methods preferred for use are listed below. Some of these have officially adopted EU/OECD guidelines, the others are regarded as scientifically acceptable for genotoxicity testing.

In vitro data

Table R.7.7-2 *In vitro* test methods

Test method	**GENOTOXIC ENDPOINTS measured/ PRINCIPLE OF THE TEST METHOD**	**EU/OECD guideline**
Bacterial reverse mutation test	Gene mutations/The test uses amino-acid requiring strains of bacteria to detect (reverse) gene mutations (point mutations and frameshifts).	EU: B.12/13 OECD: 471
In vitro mammalian cell gene mutation test – *hprt* test	Gene mutations/The test identifies chemicals that induce gene mutations in the *hprt* gene of established cell lines.	EU: B.17 OECD: 476
In vitro mammalian cell gene mutation test – Mouse lymphoma assay	Gene mutations and structural chromosome aberrations/The test identifies chemicals that induce gene mutations in the *tk* gene of the L5178Y mouse lymphoma cell line. If colonies in a *tk* mutation test are scored using the criteria of normal growth (large) and slow growth (small) colonies, gross structural chromosome aberrations may be measured, since mutant cells that have suffered the most extensive genetic damage have prolonged doubling times and are more likely to form small colonies.	EU: B.17 OECD: 476
In vitro mammalian chromosome aberration test	Structural and numerical chromosome aberrations/The test identifies chemicals that induce chromosome aberrations in cultured mammalian established cell lines, cell strains or primary cell cultures. An increase in polyploidy may indicate that a chemical has the potential to induce numerical chromosome aberrations	EU: B.10 OECD: 473
In vitro micronucleus test	Structural and numerical chromosome aberrations/The test identifies chemicals that induce micronuclei in the cytoplasm of interphase cells. These micronuclei may originate from acentric fragments or whole chromosomes, and the test thus has the potential to detect both clastogenic and aneugenic chemicals.	EU: none OECD: 487 (draft)

The ECVAM Scientific Advisory Committee (ESAC) has endorsed a statement of validity on the *in vitro* micronucleus test in November 2006, and the official adoption is expected in 2007.

As noted earlier, accepted modifications to the standard test protocols have been developed to enhance test sensitivity to specific classes of substances. Expert judgement should be applied to judge whether any of these are appropriate for a given substance being registered. For example,

protocol modifications for the Ames assay might be appropriate for substances such as gases, volatile liquids and petroleum oil derived products.

Animal data

Somatic cells

Table R.7.7-3 I*n vivo* test methods, somatic cells

Test method	GENOTOXIC ENDPOINTS measured/ PRINCIPLE OF THE TEST METHOD	EU/OECD guideline
In vivo mammalian bone marrow chromosome aberration test	Structural and numerical chromosome aberrations/The test identifies chemicals that induce structural chromosome aberrations in the bone-marrow cells of animals, usually rodents. An increase in polyploidy may indicate that a chemical has the potential to induce numerical chromosome aberrations.	EU: B.11 OECD: 475
In vivo mammalian erythrocyte micronucleus test	Structural and numerical chromosome aberrations/The test identifies chemicals that cause micronuclei in erythroblasts sampled from bone marrow and/or peripheral blood cells of animals, usually rodents. These micronuclei may originate from acentric fragments or whole chromosomes, and the test thus has the potential to detect both clastogenic and aneugenic chemicals.	EU: B.12 OECD: 474
Unscheduled DNA synthesis (UDS) test with mammalian liver cells *in vivo*	DNA repair/The test identifies chemicals that induce DNA repair (measured as unscheduled ”DNA” synthesis) in liver cells of animals, commonly rats. The test is usually based on the incorporation of tritium labelled thymidine into the DNA by repair synthesis after excision and removal of a stretch of DNA containing a region of damage.	EU: B.39 OECD: 486
Transgenic animal models	Gene mutations/The tests can measure gene mutations in any tissue of an animal and may, therefore, also be used in specific site of contact tissues.	EU: none OECD: none
In vivo alkaline single-cell gel electrophoresis assay for DNA strand breaks (Comet assay)	DNA strand breaks/The test can measure DNA strand breaks in any tissue of an animal and may, therefore, also be used in specific site of contact tissues.	EU: none OECD: none

A detailed review of transgenic animal model assays including recommendations on the conduct of such assays in somatic cells has been produced for the OECD (Lambert *et al.,* 2005).

Protocols for conducting the *in vivo* alkaline single-cell gel electrophoresis assay for DNA strand breaks (Comet assay) developed by an expert panel that met at the 2nd International Workshop on Genotoxicity Testing (IWGT, under the umbrella of the International Association of Environmental Mutagen Societies) are available (Tice *et al.,* 2000), as are recommendations for conducting this test developed by an expert panel who met in conjunction with the 4th International Comet Assay

Workshop (Hartmann *et al.,* 2003). An international validation study on the *in vivo* alkaline single-cell gel electrophoresis assay is foreseen to start end of 2006 and will be coordinated by the Japanese Centre for the Validation of Alternative Methods (JaCVAM).

Germ cells

Testing in germ cells will be conducted only on very rare occasions (see Section R.7.7.6).

Table R.7.7-4 *In vivo* test methods, germ cells

Test method	GENOTOXIC ENDPOINTS measured/ PRINCIPLE OF THE TEST METHOD	EU/OECD guideline
Mammalian spermatogonial chromosome aberration test	Structural and numerical chromosome aberrations/The test measures structural chromosome aberrations in mammalian, usually rodent, spermatogonial cells and is, therefore, expected to be predictive of induction of heritable mutations in germ cells. An increase in polyploidy may indicate that a chemical has the potential to induce numerical chromosome aberrations.	EU: B.23 OECD: 483
Rodent dominant lethal test	Structural and numerical chromosome aberrations/The test measures dominant lethal effects causing embryonic or foetal death resulting from inherited dominant lethal mutations induced in germ cells of an exposed parent, usually the male. It is generally accepted that dominant lethals are due to structural and numerical chromosome aberrations. Rats or mice are recommended as the test species.	EU: B.22 OECD: 478
Transgenic animal models	Gene mutations/ The tests measure gene mutations in spermatocytes of an animal and may, therefore, be used to obtain information about the mutagenic activity of a chemical in germ cells.	EU: none OECD: none
In vivo alkaline single-cell gel electrophoresis assay for DNA strand breaks (Comet assay)	DNA strand breaks/ The test measures DNA strand breaks in spermatocytes of an animal and may, therefore, be used to obtain information about the DNA-damaging activity of a chemical in germ cells.	EU: none OECD: none

A detailed review of transgenic animal model assays including recommendations on the conduct of such assays in germ cells has been produced for the OECD (Lambert *et al.,* 2005).

R.7.7.3.2 Human data on mutagenicity

Occasionally, studies of genotoxic effects in humans exposed by, for example, accident, occupation or participation in clinical studies (e.g. from case reports or epidemiological studies) may be

available. Generally, cells circulating in blood are investigated for the occurrence of various types of genetic alterations.

R.7.7.4 Evaluation of available information on mutagenicity

Genotoxicity is a complex endpoint and requires evaluation by expert judgement. For both steps of the effects assessment, *i.e.* hazard identification and dose (concentration)-response (effect) assessment, it is very important to evaluate the data with regard to their adequacy and completeness. The evaluation of adequacy shall address the reliability and relevance of the data in a way as outlined in the introductory chapter. The completeness of the data refers to the conclusion on the comparison between the available adequate information and the information that is required under the REACH proposal for the applicable tonnage level of the substance. Such a conclusion relies on *Weight of Evidence* approaches, mentioned in Annex XI Section 1.2 of REACH, which categorise available information based on the methods used: *guideline tests*, *non-guideline tests*, and other types of information which may justify adaptation of the standard testing regime. Such a *Weight of Evidence* approach also includes an evaluation of the available data as a whole, *i.e.* both *over and across* toxicological endpoints.

This approach provides a basis to decide whether further information is needed on endpoints for which specific data appear inadequate or not available, or whether the requirements are fulfilled.

R.7.7.4.1 Non-human data on mutagenicity

Non-testing data for mutagenicity

In a more formal approach, documentation can include reference to a related chemical or group of chemicals that leads to the conclusion of concern or lack of concern. This can either be presented according to scientific logic (read-across) or sometimes as a mathematical relationship of chemical similarity.

If well-documented and applicable (Q)SAR data are available, they should be used to help reach the decision points described in the section below. In many cases the accuracy of such methods will be sufficient to help, or allow either a testing or a specific regulatory decision to be made. In other cases the uncertainty may be unacceptable due to the severe consequences of a possible error. This may be driven by many factors including high exposure potential or toxicological concerns.

Chemicals for which no test-data exist present a special case in which reliance on non-testing data may be absolute. Many factors will dictate the acceptability of non-testing methods in reaching a conclusion based on no tests at all. It is yet to be established whether weight-of evidence decisions based on multiple genotoxicity and carcinogenicity estimates can equal or exceed those obtained by one or two *in vitro* tests. This must be considered on a case-by-case basis.

Testing data on mutagenicity

Evaluation of genotoxicity test data should be made with care. Regarding *positive* findings, responses generated only at highly toxic/cytotoxic concentrations should be interpreted with caution, and the presence or absence of a dose-response relationship should be considered.

Particular points to take into account when evaluating *negative* test results include:

- the doses or concentrations of test substance used (were they high enough?)

- was the test system used sensitive to the nature of the genotoxic changes that might have been expected? For example, some *in vitro* test systems will be sensitive to point mutations and small deletions but not to mutagenic events that create large deletions.
- the volatility of the test substance (were concentrations maintained in tests conducted *in vitro*?)
- for studies *in vitro*, the possibility of metabolism not being active in the system including those in extra-hepatic organs
- was the test substance taken up by the test system used for *in vitro* studies?
- for studies *in vivo*, is the substance reaching the target organ? (taking also toxicokinetic data into consideration, e.g. rate of hydrolysis and electrophilicity may be factors that need to be considered)

Contradictory results between different test systems should be evaluated with respect to their individual significance. Examples of points to be considered are as follows:

- conflicting results obtained in non-mammalian systems and in mammalian cell tests may be addressed by considering possible differences in substance uptake, metabolism or in the organisation of genetic material. Although the results of mammalian tests may be considered of higher significance, additional data may be needed to resolve contradictions
- if the results of indicator tests (*e.g*. DNA binding; SCE) are not supported by results obtained in tests for mutagenicity, the results of mutagenicity tests are generally of higher significance
- if contradictory findings are obtained *in vitro* and *in vivo*, in general, the results of *in vivo* tests indicate a higher degree of reliability. However, for evaluation of *negative* results *in vivo*, it should be considered whether there is adequate evidence of target tissue exposure
- the sensitivity and specificity of different test systems varies for different classes of substances. If available testing data for other related substances permits assessment of the performance of difference assays for the class of substance under evaluation, the result from the test system known to produce more accurate responses would be given higher priority

Conflicting results may be also available from the same test, performed by different laboratories or on different occasions. In this case, expert judgement should be used to reach an overall evaluation of the data. In particular, the quality of each of the studies and of the data provided should be evaluated, with special consideration of the study design, reproducibility of data, dose-effect relationships, and biological relevance of the findings. The purity of the test substance may also be a factor to take into account. In the case where an EU/OECD guideline is available for a test method, the quality of a study using the method is regarded as being higher if it was conducted in compliance with the requirements stated in the guideline. Furthermore, studies compliant with GLP may be regarded as being of a higher quality.

When making an assessment of the potential mutagenicity of a substance, or considering the need for further testing, data from various tests and genotoxic endpoints may be found. Both the strength and the weight of the evidence should be taken into account. The strongest evidence will be provided by modern, well-conducted studies with internationally established test protocols. For each test type and each genotoxic endpoint, there should be a separate *Weight of Evidence* analysis. It is not unusual for positive evidence of mutagenicity to be found in just one test type or for only one endpoint. In such cases the positive and negative results for different endpoints are not conflicting, but illustrate the advantage of using test methods for a variety of genetic alterations to increase the probability of identifying substances with mutagenic potential. Hence, results from

methods testing different genotoxic endpoints should not be combined in an overall *Weight of Evidence* analysis, but should be subjected to such analysis separately.

R.7.7.4.2 Human data on mutagenicity

Human data have to be assessed carefully on a case-by-case basis. The interpretation of such data requires considerable expertise. Attention should be paid especially to the adequacy of the exposure information, confounding factors, co-exposures and to sources of bias in the study design or incident. The statistical power of the test may also be considered.

R.7.7.4.3 Remaining uncertainty on mutagenicity

Reliable data can be generated from well-designed and conducted studies *in vitro* and *in vivo*. However, due to the lack of human data available, a certain level of uncertainty remains when extrapolating these testing data to the effect in humans.

R.7.7.5 Conclusions on mutagenicity

R.7.7.5.1 Concluding on suitability for Classification and Labelling

In order to conclude on an appropriate classification and labelling position with regard to mutagenicity, the available data should be considered using the criteria according to Annex VI to Directive 67/548/EEC[46].

R.7.7.5.2 Concluding on suitability for Chemical Safety Assessment

Considerations on dose response shapes and mode of action of mutagenic substances in test systems

Considerations of the dose-response relationship and of possible mechanisms of action are important components of a risk assessment. The default assumption for genotoxic chemicals, in the absence of mechanistic evidence to the contrary, is that they have a linear dose-response relationship. However, both direct and indirect mechanisms of genotoxicity can be non-linear or thresholded and, consequently, sometimes this default assumption may be inappropriate.

Examples of mechanisms of genotoxicity that may be demonstrated to lead to non-linear or thresholded dose-response relationships include extremes of pH, ionic strength and osmolarity, inhibition of DNA synthesis, alterations in DNA repair, overloading of defence mechanisms (anti-oxidants or metal homeostatic controls), interaction with microtubule assembly leading to aneuploidy, topoisomerase inhibition, high cytotoxicity, metabolic overload and physiological perturbations (e.g. induction of erythropoeisis). Assessment of the significance to be assigned to genotoxic responses mediated by such mechanisms would include an assessment of whether the underlying mechanism can be induced at substance concentrations that can be expected to occur under relevant *in vivo* conditions.

[46] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

In general, several doses are tested in genotoxicity assays. Determination of experimental dose-effect relationships may be used to assess the genotoxic potential of a substance, as indicated below. It should be recognised that not all of these considerations may be applicable to *in vivo* data.

- a dose-related increase in genotoxicity is one of the relevant criteria for identification of positive findings. In practice, this will be most helpful for *in vitro* tests, but care is needed to check for cytotoxicity or cell cycle delay which may cause deviations from a dose-response related effect in some experimental systems

- genotoxicity tests are not designed in order to derive no effect levels. However, the magnitude of the lowest dose with an observed effect (i.e. the Lowest Observed Effect Dose or LOED) may, on certain occasions, be a helpful tool in risk assessment. This is true specifically for genotoxic effects caused by thresholded mechanisms, like, e.g. aneugenicity. Further, it can give an indication of the mutagenic potency of the substance in the test at issue. Modified studies, with additional dose points and improved statistical power may be useful in this regard

- unusual shapes of dose-response curves may contribute to the identification of specific mechanisms of genotoxicity. For example, extremely steep increases suggest an indirect mode of action or metabolic switching which could be confirmed by further investigation

Considerations on genetic risks associated with human exposure to mutagenic substances

There are no officially adopted methods for estimating health risks associated with (low) exposures of humans to mutagens. In fact, most – if not all tests used today – are developed and applied to identify mutagenic properties of the substance, i.e. identification of the mutagenic hazard *per se*. In today's regulatory practice, the assessment of human health risks from exposure to mutagenic substances is considered to be covered by assessing and regulating the carcinogenic risks of these agents. The reason for this is that mutagenic events underlie these carcinogenic effects. Therefore, mutagenicity data is not used for deriving dose descriptors for risk assessment purposes and the reader is referred to this aspect in Section R.7.7.8 (Carcinogenicity) for guidance on how to assess the chemical safety for mutagenic substances.

R.7.7.5.3 Information not adequate

A *Weight of Evidence* approach, comparing available adequate information with the tonnage-triggered information requirements by REACH, may result in the conclusion that the requirements are not fulfilled. In order to proceed in gathering further information, the following testing strategy can be adopted:

R.7.7.6 Integrated Testing Strategy (ITS) for mutagenicity

R.7.7.6.1 Objective / General principles

This testing strategy describes a flexible, stepwise approach for hazard identification with regard to the mutagenic potential of substances, so that sufficient data may be obtained for adequate risk characterisation including classification and labelling. It serves to help minimise the use of animals and costs as far as is consistent with scientific rigour. (A flow chart and a summary of the testing strategy are presented in Figure R.7.7-1 and Figure R.7.7-2 respectively). As noted later in this section, deviations from this strategy may be considered if existing data for related substances

indicate that alternate testing strategies yield results with greater sensitivity and specificity for mutagenicity *in vivo*.

The strategy defines a level of information that is considered sufficient to provide adequate reassurance about the potential mutagenicity of most substances. As described below, this level of information will be required for most substances at the Annex VIII tonnage level specified in REACH, although circumstances are described when the data may be required for substances at Annex VII.

For some substances, relevant data from other sources/tests may also be available (e.g. physico-chemical, toxicokinetic, and toxicodynamic parameters and other toxicity data; data on well-investigated, structurally similar, chemicals). These should be reviewed because, sometimes, they may indicate that either more or less genotoxicity studies are needed on the substance than defined by standard information requirements; i.e., they may allow tailored testing/selection of test systems. For example, bacterial mutagenesis assays of inorganic metal compounds are frequently negative due to limited capacity for uptake of metal ions. The high prevalence of false negatives for metal compounds might suggest that mutagenesis assays with mammalian cells, as opposed to bacterial cells, would be the preferred starting point for testing for this class of Annex VII substances.

In summary, a key concept of the strategy is that initial genotoxicity tests and testing protocols should be selected with due consideration to pre-existing data that has established the most accurate testing strategy for the class of compound under evaluation. Even then, initial testing may not always give adequate information and further testing may sometimes be considered necessary in the light of all available relevant information on the substance, including its use pattern. Further testing will normally be required for substances which give rise to positive results in any of the *in vitro* tests.

If negative results are available from an adequate evaluation of genotoxicity from existing data in appropriate test systems, there may be no requirement to conduct additional genotoxicity tests.

Substances for which there is a formal agreement to classify them in category 1, 2 or 3 for mutagenicity and/or category 1 or 2 for carcinogenicity will usually not require additional testing in order to meet the requirements of Annexes VII-X. In cases where a registrant is unsure of the formal position on the classification of a substance, or wishes to make a classification proposal themselves, advice should be sought from an appropriate regulatory body before proceeding with any further testing.

R.7.7.6.2 Preliminary considerations

For a comprehensive coverage of the potential mutagenicity of a substance, information on gene mutations (base substitutions and deletions/additions), structural chromosome aberrations (breaks and rearrangements) and numerical chromosome aberrations (loss or gain of chromosomes, defined as aneuploidy) is required. This may be obtained from available data or tests on the substance itself or, sometimes, by prediction using appropriate *in silico* techniques (e.g. chemical grouping, read-across or (Q)SAR approaches).

It is important that whatever is known of the physico-chemical properties of the test substance is taken into account before devising an appropriate testing strategy. Such information may impact upon both the selection of test systems to be employed and/or modifications to the test protocols used. The chemical structure of a substance can provide information for an initial assessment of mutagenic potential. The need for special testing requirements in relation to photomutagenicity may be indicated by the structure of a molecule, its light absorbing potential or its potential to be

photoactivated. By using expert judgement, it may be possible to identify whether a substance, or a potential metabolite of a substance, shares structural characteristics with known mutagens or non-mutagens. This can be used to justify a higher or lower level of priority for the characterisation of the mutagenic potential of a substance. Where the level of evidence for mutagenicity is particularly strong, it may be possible to make a conclusive hazard assessment in accordance with Annex I of REACH without additional testing on the basis of structure-activity relationships alone.

In vitro tests are particularly useful for gaining an understanding of the potential mutagenicity of a substance and they have a critical role in this testing strategy. They are not, however, without their limitations. Animal tests will, in general, be needed for the clarification of the relevance of positive findings and in case of specific metabolic pathways that cannot be simulated adequately *in vitro*.

The toxicokinetic and toxicodynamic properties of the test substance should be considered before undertaking, or appraising, animal tests. Understanding these properties will enable appropriate protocols for the standard tests to be developed, especially with respect to tissue(s) to be investigated, the route of substance administration and the highest dose tested. If little is understood about the systemic availability of a test substance at this stage, toxicokinetic investigations or modelling may be necessary.

Certain substances in addition to those already noted may need special consideration, such as highly electrophilic substances that give positive results *in vitro*, particularly in the absence of metabolic activation. Although these substances may react with proteins and water *in vivo* and thus be rendered inactive towards many tissues, they may be able to express their mutagenic potential at the initial site of contact with the body. Consequently, the use of test methods that can be applied to the respiratory tract, upper gastrointestinal tract and skin may be appropriate. It is possible that specialised test methods will need to be applied in these circumstances, and that these may not have recognised, internationally valid, test guidelines. The validity and utility of such tests and the selection of protocols should be assessed by appropriate experts or authorities on a case-by-case basis.

A substance giving an equivocal test result should be reinvestigated immediately, normally using the same test method, but varying the conditions to obtain conclusive results. Wherever possible, clear results should be obtained for one step in the strategic procedure before going on to the next. In cases where this does not prove to be possible, a further test should be conducted in accord with the strategy.

Tests need not be performed if it is not technically possible to do so, or if they are not considered necessary in the light of current scientific knowledge. Scientific justifications for not performing tests required by the strategy should always be documented. It is preferred that tests as described in OECD Guidelines or EU Directive 67/548/EEC are used where possible. Alternatively, for other tests, up-to-date protocols defined by internationally recognised groups of experts, e.g. International Workshop on Genotoxicity Testing (IWGT, under the umbrella of the International Association of Environmental Mutagen Societies), may be used provided that the tests are scientifically justified. It is essential that all tests be conducted according to rigorous protocols in order to maximise the potential for detecting a mutagenic response, to ensure that negative results can be accepted with confidence and that results are comparable when tests are conducted in different laboratories. At the time of writing this guidance, regulatory guidelines are still to be established for some of the *in vitro* and *in vivo* tests included in the testing strategy described below. If one of these tests is to be conducted, consultation on the protocol with an appropriate expert or authority is advisable.

If a registrant wishes to undertake any tests for substances at the Annex IX or X tonnage levels that require the use of vertebrate animals, then there is a need to make a proposal to the European

Chemicals Agency first. Testing may only be undertaken when an agreement has been reached with the Agency.

R.7.7.6.3 Testing strategy for mutagenicity

Standard information requirement at Annex VII

A preliminary assessment of mutagenicity is required for substances at the REACH Annex VII tonnage level. All available information should be included but, as a minimum, there should normally be data from a gene mutation test in bacteria unless existing data for analogous substances indicates this would be inappropriate. For substances with significant toxicity to bacteria, not taken up by bacteria, or for which the gene mutation test in bacteria cannot be performed adequately, an *in vitro* mammalian cell gene mutation test may be used as an alternative test.

When the result of the bacterial test is positive, it is important to consider the possibility of the substance being genotoxic in mammalian cells. The need for further test data to clarify this possibility at the Annex VII tonnage level will depend on an evaluation of all the available information relating to the genotoxicity of the substance.

Standard information requirement at Annex VIII

For a comprehensive coverage of the potential mutagenicity of a substance, information on gene mutations, and structural and numerical chromosome aberrations is required for substances at the Annex VIII tonnage level of REACH.

In order to ensure the necessary minimum level of information is provided, at least one further test is required in addition to the gene mutation test in bacteria. This should be an *in vitro* mammalian cell test capable of detecting both structural and numerical chromosome aberrations. In REACH Annex VIII this is referred to as an *in vitro* cytogenicity study or an *in vitro* micronucleus test in mammalian cells.

There are essentially two different methods that can be viewed as options for this first mammalian cell test.

- An *in vitro* chromosome aberration test (OECD 473), i.e. a cytogenetic assay for structural chromosome aberrations using metaphase analysis. It may be possible to present some preliminary information from this test on potential aneugenicity by recording the incidence of hyperdiploidy and polyploidy. If this preliminary information presents a possible concern about aneugenicity, this indicates the need for specific investigations to assess potential aneugenicity of the chemical. An alternative option would be to conduct an *in vitro* micronucleus test.

- An *in vitro* micronucleus test (OECD TG 487, in preparation). This is a cytogenetic assay that has the advantage of detecting not only structural chromosomal aberrations but also aneuploidy. In advance of an OECD guideline being finalised, the draft guideline may be used as the standard reference for this test. Use of fluorescence *in situ* hybridisation with probes for centromeric DNA or stains for kinetochore proteins can enable aneugens to be distinguished from clastogens. This may sometimes be useful for risk characterisation.

Other *in vitro* tests may be acceptable as the first mammalian cell test, but care should be taken to evaluate their suitability for the substance being registered and their reliability as a screen for chemicals that cause structural and/or numerical chromosome aberrations. A supporting rationale should be presented for a registration with any of these other tests.

It is possible to present existing data from an *in vivo* cytogenetics test (i.e. a study or studies conducted previously) as an alternative to the first *in vitro* mammalian cell test. For instance, if an adequately performed *in vivo* micronucleus test is available already it may be presented as an alternative.

An *in vitro* gene mutation study in mammalian cells (OECD TG 476) is the second part of the standard information set required for registration at the Annex VIII tonnage level. For substances that have been tested already, this information should always be presented as part of the overall *Weight of Evidence* for mutagenicity. For other substances, this second *in vitro* mammalian cell test will normally only be required when the results of the bacterial gene mutation test and the first study in mammalian cells (i.e. an *in vitro* cytogenicity study or an *in vitro* micronucleus test, as specified in Annex VIII of REACH) are negative. This is to detect *in vitro* mutagens that give negative results in the other two tests.

Under specific circumstances it may be possible to omit the second *in vitro* study in mammalian cells, *i.e.* if it can be demonstrated that this mammalian cell test will not provide any further useful information about the potential *in vivo* mutagenicity of a substance, then it does not need to be conducted. This should be evaluated on a case-by-case basis as there may be classes of compound for which conclusive data can be provided to show that the sensitivity of the first two *in vitro* tests cannot be improved by the conduct of the third test.

The *in vitro* mammalian cell gene mutation test will not usually be required if adequate information is available from a reliable *in vivo* study capable of detecting gene mutations. Such studies may include a transgenic rodent assay, a comet assay, or a liver UDS test.

Provided the *in vitro* tests have given negative results, normally, no *in vivo* tests will be required to fulfil the standard information requirements at Annex VIII. However, there may be rare occasions when it is appropriate to conduct testing *in vivo*, for example when it is not possible technically to perform satisfactory tests *in vitro*. Substances which, by virtue of, for example, their physico-chemical characteristics, chemical reactivity or toxicity cannot be tested in one or more of the *in vitro* tests should be considered on a case-by-case basis.

Requirement for testing beyond the standard levels specified for Annexes VII and VIII

Introductory comments

Concerns raised by positive results from *in vitro* tests can justify further testing. The chemistry of the substance, data on analogous substances, toxicokinetic and toxicodynamic data, and other toxicity data will also influence the timing and pattern of further testing.

Testing beyond the standard set of *in vitro* tests is first directed towards investigating the potential for mutagenicity in somatic cells *in vivo*. Positive results in somatic cells *in vivo* constitute the trigger for consideration of investigation of potential expression of genotoxicity in germ cells.

Substances that are negative in the standard set of in vitro tests

In general, substances that are negative in the full set of *in vitro* tests specified in REACH Annexes VII and VIII are considered to be non-genotoxic. There are only a very limited number of chemicals that have been found to be genotoxic *in vivo*, but not in the standard *in vitro* tests. Most of these are pharmaceuticals designed to affect pathways of cellular regulation, including cell cycle regulation, and this evidence is judged insufficient to justify routine *in vivo* testing of industrial chemicals. However, occasionally, knowledge about the metabolic profile of a substance may indicate that the standard *in vitro* tests are not sufficiently reassuring and a further *in vitro* test, or an *in vivo* test, may be needed in order to ensure mutagenicity potential is adequately explored (e.g. use of an

alternative to rat liver S-9 mix, a reducing system, a metabolically active cell line like HepG2 cells or genetically engineered cell lines might be judged appropriate).

Substances for which an in vitro test is positive

REACH Annex VII substances for which only a bacterial gene mutation test has been conducted and for which the result is positive should be studied further, according to the requirements of Annex VIII.

Regarding Annex VIII, when both the mammalian cell tests are negative but there was a positive result in the bacterial test, it will be necessary to decide whether any further testing is needed on a case-by-case basis. For example, suspicion that a unique positive response observed in the bacterial test was due to a specific bacterial metabolism of the test substance could be explored further by investigation *in vitro.* Alternatively, an *in vivo* test may be required (see below).

In REACH Annex VIII, following a positive result in an *in vitro* mammalian cell mutagenicity test, adequately conducted somatic cell *in vivo* testing is required to ascertain if this potential can be expressed *in vivo*. It is recommended that the first test *in vivo* should be initiated as soon as possible. In cases where it can be sufficiently deduced that a positive *in vitro* finding is not relevant for *in vivo* situations (e.g. due to the effect of the test substances on pH or cell viability: see also Section R.7.7.4.1), *in vivo* testing will not be necessary.

At Annexes IX and X, if there is a positive result in any of the *in vitro* studies from Annex VII or VIII and there are no results available from an *in vivo* study already, an appropriate *in vivo* somatic cell genotoxicity study should be proposed.

Before any decisions are made about the need for *in vivo* testing, a review of the *in vitro* test results and all available information on the toxicokinetic and toxicodynamic profile of the test substance is needed. A particular *in vivo* test should be conducted only when it can be reasonably expected from all the properties of the test substance and the proposed test protocol that the specific target tissue will be adequately exposed to the test substance and/or its metabolites. If necessary, a targeted investigation of toxicokinetics should be conducted before progressing to *in vivo* testing (e.g. a preliminary toxicity test to confirm that absorption occurs and that an appropriate dose route is used).

In the interest of ensuring that the number of animals used in genotoxicity tests is kept to a minimum, both males and females should not automatically be used. In accord with standard guidelines, testing in one sex only is possible when the substance has been investigated for general toxicity and no sex-specific differences in toxicity have been observed. If the test is performed in a laboratory with substantial experience and historical data, it should be considered whether a concurrent positive control and a concurrent negative control for all time points (*e.g.* for both the 24h and 48 h time point in the micronucleus assay) will really be necessary (Hayashi *et al*. 2000).

For test substances with adequate systemic availability (*i.e*. evidence for adequate availability to the target cells) there are several options for the *in vivo* testing.

- A rodent bone marrow or mouse peripheral blood micronucleus test (OECD TG 474) or a rodent bone marrow clastogenicity study (OECD TG 475). Potential species-specific effects may influence the choice of species and test method used.

- A Comet (single cell gel electrophoresis) assay, which detects DNA strand breaks. This assay has the advantage of not being restricted to bone marrow cells. In principle every organ can be sampled. Although there is not yet an OECD guideline for this test, published guidance documents with respect to the design and performance of the test are available. Other DNA

strand breakage assays may be presented as alternatives to the comet assay. All DNA strand break assays should be considered as surrogate tests, they do not necessarily detect permanent changes to DNA.

- A test for gene mutations in a transgenic rodent model, *e.g.* using *lacI*, *lacZ* or *cII* as reporter gene present in every tissue. Although there is not yet an OECD guideline for this test, published guidance documents with respect to the design and performance of the test are available.

- A rat liver Unscheduled DNA synthesis (UDS) test. The UDS test is an indicator test measuring DNA repair. The UDS test should be considered as a surrogate test for an *in vivo* gene mutation test.

The *in vivo* genotoxicity test may be incorporated, if appropriate scientifically, into a short-term repeated dose toxicity test (28 days), for example, if this is to be performed to meet the requirements of the REACH Annex VIII tonnage level.

Any one of these tests may be conducted, but this has to be decided using expert judgement on a case-by-case basis. The nature of the original *in vitro* response(s) (i.e. gene mutation, structural or numerical chromosome aberration) should be considered when selecting the *in vivo* study. For example, if the test substance showed evidence of *in vitro* clastogenicity, then it would be most appropriate to follow this up with either a micronucleus test or chromosomal aberration test or a Comet assay. However, if a positive result were obtained in the *in vitro* micronucleus test, the rodent micronucleus test would be appropriate to best address clastogenic and aneugenic potential.

The rat liver UDS test may be appropriate for substances that appear preferentially to induce gene mutations, although the Comet and transgenic tests are also suitable. These latter test systems offer greater flexibility, most notably the possibility of selecting a range of tissues for study on the basis of what is known of the toxicokinetics and toxicodynamics of the substance. It should be realised that the UDS and Comet tests are indicator assays detecting putative DNA lesions. In contrast, the transgenic test measures permanent mutations.

Additionally, evidence for *in vivo* DNA adduct formation in somatic cells together with *in vitro* test data may sometimes be sufficient to conclude that a substance is an *in vivo* somatic cell mutagen. In such cases, positive *in vitro* test results may not trigger further *in vivo* somatic tissue testing, and the substance would be classified at least as a category 3 mutagen. The possibility for effects in germ cells would need further investigation (see Section R.7.7.6.3).

Non-standard studies supported by published literature may sometimes be more appropriate and informative than established assays. Guidance from an appropriate expert or authority should be sought before undertaking novel studies. Furthermore, additional data that support or clarify the mechanism of action may justify a decision not to test further.

For substances that are short-lived, reactive, *in vitro* mutagens, or for which no indications of systemic availability have been presented, an alternative strategy involving studies to focus on tissues at initial sites of contact with the body should be considered. Expert judgement should be used on a case-by-case basis to decide which tests are the most appropriate. The main options are the *in vivo* Comet assay, gene mutation tests with transgenic rodents, and DNA adduct studies. For any given substance, expert judgement, based on all the available toxicological information, will indicate which of these tests are the most appropriate. The route of exposure should be selected that best allows assessment of the hazard posed to humans. For insoluble substances, the possibility of release of active molecules in the gastrointestinal tract may indicate that a test involving the oral route of administration is particularly appropriate.

At the time this guidance was drafted (2007), test guidelines were still being prepared for some of the *in vivo* tests mentioned above. In the absence of such a guideline, expert advice or appropriate alternative guidance, e.g., from the International Workshops on Genotoxicity Testing (IWGT, under the umbrella of the International Association of Environmental Mutagen Societies) should be sought about the conduct of these tests.

If the first *in vivo* test is negative, the need for a further *in vivo* somatic cell test should be considered. The second *in vivo* test should only then be proposed if it is required to make a conclusion on the genotoxic potential of the substance under investigation; i.e. if the *in vitro* data show the substance to have potential to induce both gene and chromosome mutations and the first *in vivo* test has not addressed this comprehensively. In this regard, on a case-by-case basis, attention should be paid to the quality and relevance of all the available toxicological data, including the adequacy of target tissue exposure.

For a substance giving negative results in adequately conducted, appropriate *in vivo* test(s), as defined by this strategy, it will normally be possible to conclude that the substance is not an *in vivo* mutagen.

Substances that give positive results in an in vivo test for genotoxic effects in somatic cells

Substances that have given positive results in cytogenetic tests *in vitro* and in such tests in somatic cells *in vivo* can be studied further to establish whether they specifically act as aneugens and thresholds for their genotoxic activity can be identified, if this has not been established adequately already. This should be done using *in vitro* methods and will be helpful in risk evaluation.

The potential for substances that give positive results in *in vivo* tests for genotoxic effects in somatic cells to affect germ cells should always be considered. The same is true for substances otherwise classified as category 3 mutagens. The first step is to make an appraisal of all the available toxicokinetic and toxicodynamic properties of the test substance. Expert judgement is needed at this stage to consider whether there is sufficient information to conclude that the substance poses a mutagenic hazard to germ cells. If this is the case, it can be concluded that the substance may cause heritable genetic damage and no further testing is justified. Consequently, the substance is classified as a category 2 mutagen. If the appraisal of mutagenic potential in germ cells is inconclusive, additional investigation will be necessary. In the event that additional information about the toxicokinetics of the substance would resolve the problem, toxicokinetic investigation (i.e. not a full toxicokinetic study) tailored to address this is required.

If germ cell testing is to be undertaken, and this should be in exceptional circumstances, expert judgement should be used to select the most appropriate test strategy. Internationally recognised guidelines are available for investigating clastogenicity in rodent spermatogonial cells and for the dominant lethal test. Dominant lethal mutations are believed to be primarily due to structural or numerical chromosome aberrations.

Alternatively, other methods can be used if deemed appropriate by expert judgement. These may include the Comet assay, gene mutation tests with transgenic animals, or DNA adduct analysis. At the time of writing this guidance, a test determining changes in hypervariable tandem repetitive regions (minisatellites) that are scattered throughout the chromosomes is being developed as a germ cell mutagenicity test (Somers *et al.,* 2002; Yauk, 2004). Mutations in the offspring of treated animals are detected as fragment length polymorphisms after restriction enzyme analysis. In principle, it is the potential for effects that can be transmitted to the progeny that should be investigated, but tests used historically to investigate transmitted effects (the heritable translocation

test and the specific locus test) use very large numbers of animals. They are rarely used and should not normally be conducted for industrial substances.

In order to minimise animal use, the possibility to combine germ cell genotoxicity tests and reproductive toxicity tests may be considered.

Figure R.7.7-1 Flow chart of the mutagenicity testing strategy

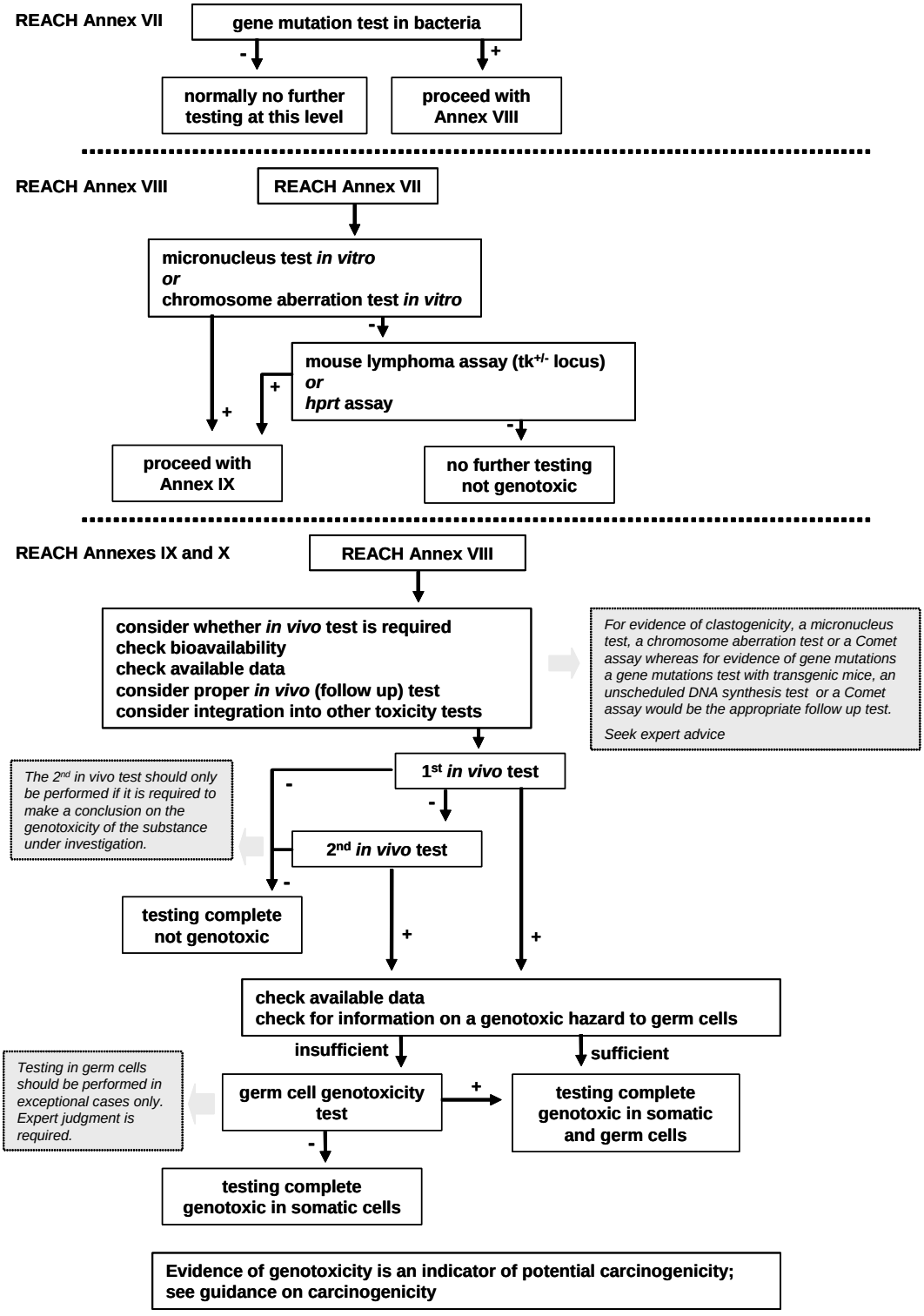


Table R.7.7-5

	GM bact	Cytvitro	GMvitro	Cytvivo	GMvivo	Standard information required General follow up procedure	Conclusion	Specific rules for adaptation [for detailed guidance, incl. timing of the tests, see text]	Comments
1	neg					Annex VII: no further tests are required. Annexes VIII, IX & X: conduct a CAbvitro/MNTvitro, if this is negative, a GMvitro.	Annex VII: not genotoxic		Annexes VIII, IX & X: Select further tests in such a way that all the tests, together with other available information, enable thorough assessment for gene mutations and effects on chromosome structure and number.
2	neg	neg				Annex VII: no further tests are required. Annexes VIII, IX & X: conduct a GMvitro.	Annex VII: not genotoxic		Annexes VIII, IX & X: Select tests in such a way that all the tests, together with other available information, enable a thorough assessment for gene mutations and effects on chromosome structure and number.
3	neg		neg			Annex VII: no further tests are required. Annexes VIII, IX & X: conduct a CAbvitro/MNTvitro	Annex VII: not genotoxic		Annexes VIII, IX & X: Select tests in such a way that all the tests, together with other available information, enable a thorough assessment for gene mutations and effects on chromosome structure and number.

	GM bact	Cytvitro	GMvitro	Cytvivo	GMvivo	Standard information required General follow up procedure	Conclusion	Specific rules for adaptation [for detailed guidance, incl. timing of the tests, see text]	Comments
4	neg	neg	neg			Annexes VII, VIII, IX & X: no further tests are required.	not genotoxic		The available metabolic evidence may, on rare occasions, indicate that in vitro testing is inadequate; in vivo testing is needed. Seek expert advice. Annexes VIII, IX & X: Select tests in such a way that all the tests, together with other available information, enable a thorough assessment for gene mutations and effects on chromosome structure and number.
5	pos					Annexes VII, VIII, IX & X: Complete in vitro testing with a CAbvitro/MNTvitro .			Consider need for further tests to understand the in vivo mutagenicity hazard, to make a risk assessment, and to determine whether C&L is justified.

	GM bact	Cytvitro	GMvitro	Cytvivo	GMvivo	Standard information required General follow up procedure	Conclusion	Specific rules for adaptation [for detailed guidance, incl. timing of the tests, see text]	Comments
6	pos	neg				Annexes VII & VIII: Complete in vitro testing by conducting a GMvitro Annexes IX & X: If systemic availability cannot be ascertained reliably, it should be investigated before progressing to in vivo tests. Select adequate somatic cell in vivo test to investigate gene mutations in vivo (UDSvivo, Tg or Comet) If necessary seek expert advice.		Suspicion that a positive response observed in the GMbact was due to a specific bacterial metabolism of the test substance could be explored further by investigation in vitro.	Ensure that all tests together with other available information enable thorough assessment for gene mutations and effects on chromosome structure and number. Consider on a case-by-case basis need for further tests to understand the in vivo mutagenicity hazard, to make a risk assessment, and to determine whether C&L is justified.
7	neg	pos				Annexes VII, VIII, IX & X: If systemic availability cannot be ascertained reliably, it should be investigated before progressing to in vivo tests. Select adequate somatic cell in vivo test to investigate structural or numerical chromosome aberrations (MNTvivo or Comet for in vitro clastogens and/or aneugens or CAbvivo for in vitro-clastogens) If necessary seek expert advice.			Ensure that all tests together with other available information enable thorough assessment for gene mutations and effects on chromosome structure and number. Consider need for further tests to understand the in vivo mutagenicity hazard, to make a risk assessment and to determine whether C&L is justified.

	GM bact	Cytvitro	GMvitro	Cytvivo	GMvivo	Standard information required General follow up procedure	Conclusion	Specific rules for adaptation [for detailed guidance, incl. timing of the tests, see text]	Comments
8	pos	pos				Annexes VII, VIII, IX & X: If systemic availability cannot be ascertained with acceptable reliability, it should be investigated before progressing to in vivo tests. Select adequate somatic cell in vivo tests to investigate both structural or numerical chromosome aberrations and gene mutations. If necessary seek expert advice.		Generally, both genotoxic endpoints should be investigated. If the first in vivo test is positive, a second in vivo test to confirm the other genotoxic endpoint need not be conducted. If the first in vivo test is negative, a second in vivo test is required if the first test did not address the endpoints comprehensively.	Ensure that all tests together with other available information enable thorough assessment for gene mutations and effects on chromosome structure and number. Consider need for further tests to understand the in vivo mutagenicity hazard, to make a risk assessment, and to determine whether C&L is justified.
9	pos	neg			Neg	Annexes VII, VIII, IX & X: no further tests are required.	not genotoxic		Further in vivo test may be necessary pending on the quality and relevance of available data.
	neg	pos		neg					
10	pos	neg			Pos	Annexes VII, VIII, IX & X: No further testing in somatic cells is needed. Germ cell mutagenicity tests shall be considered. If necessary seek expert advice on implications of all available data on toxicokinetics and toxicodynamics and on the choice of the proper germ cell mutagenicity test.	genotoxic	Expert judgement is needed at this stage to consider whether there is sufficient information to conclude that the substance poses a mutagenic hazard to germ cells. If this is the case, it can be concluded that the substance may cause heritable genetic damage and no further testing is justified.	If the appraisal of mutagenic potential in germ cells is inconclusive, additional investigation may be necessary. Risk assessment and C&L can be completed.
	neg	pos		pos					
	neg	neg	pos		Pos				

11	pos	pos	(pos)	pos		Annexes VII, VIII, IX & X: No further testing in somatic cells is needed. Germ cell mutagenicity tests shall be considered. If necessary seek expert advice on implications of all available data on toxicokinetics and toxicodynamics and on the choice of the proper germ cell mutagenicity test.	genotoxic	Expert judgement is needed at this stage to consider whether there is sufficient information to conclude that the substance poses a mutagenic hazard to germ cells. If this is the case, it can be concluded that the substance may cause heritable genetic damage and no further testing is justified.	If the appraisal of mutagenic potential in germ cells is inconclusive, additional investigation may be necessary. Risk assessment and C&L can be completed.
	pos	pos	(pos)		Pos				
12	pos	pos	(pos)	neg		Annexes VII, VIII, IX & X: Select adequate somatic cell in vivo tests to investigate both structural or numerical chromosome aberrations and gene mutations. If necessary seek expert advice.			
	pos	pos	(pos)		Neg				
13	pos	pos	(pos)	neg	Neg	Annexes VII, VIII, IX & X: no further tests are required.	not genotoxic	Further in vivo test may be necessary pending on the quality and relevance of available data.	Risk assessment and C&L can be completed.
14	pos	pos	(pos)	neg	Pos	Annexes VII, VIII, IX & X: No further testing in somatic cells is needed. Germ cell mutagenicity tests shall be considered. If necessary seek expert advice on implications of all available data on toxicokinetics and toxicodynamics and on the choice of the proper germ cell mutagenicity test.	genotoxic	Expert judgement is needed at this stage to consider whether there is sufficient information to conclude that the substance poses a mutagenic hazard to germ cells. If this is the case, it can be concluded that the substance may cause heritable genetic damage and no further testing is justified.	If the appraisal of mutagenic potential in germ cells is inconclusive, additional investigation will be necessary. Risk assessment and C&L can be completed.
	pos	pos	(pos)	pos	Neg				

Abbreviations: pos: positive; neg: negative; pos/neg: the follow up is independent from the result of this test; GM_{bact}: gene mutation test in bacteria (Ames test); Cyt_{vitro}:cytogenetic assay in mammalian cells; CAb_{vitro}, *in vitro* chromosome aberration test; MNT_{vitro}, *in vitro* micronucleus test; GM_{vitro}:gene mutation assay in mammalian cells; Cyt_{vivo}:cytogenetic assay in experimental animals; GM_{vivo}:gene mutation assay in experimental animals;

CAb_{vivo}, *in vivo* chromosome aberration test (bone marrow); MNT_{vivo}, *in vivo* micronucleus test (erythrocytes); UDS_{vivo}, *in vivo* unscheduled DNA synthesis test: Tg, *in vivo* gene mutation test with transgenic mice: Comet, Comet assay.

R.7.7.7 References on mutagenicity

Ashby J and Tennant RW (1988) Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. Mutat Res 204: 17-115.

Chung KT, Kirkovsky L, Kirkovsky A and Purcell WP (1997) Review of mutagenicity of monocyclic aromatic amines: quantitative structure-activity relationships. Mutat Res. 387: 1-16.

Hartmann A, Agurell E, Beevers C, Brendler-Schwaab S, Burlinson B, Clay P, Collins A, Smith A, Speit G, Thybaud V and Tice RR (2003) Recommendations for conducting the *in vivo* alkaline Comet assay. Mutagenesis 18: 45-51.

Kazius J, McGuire R and Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem. 48: 312-20.

Lambert IB, Singer TM, Boucher SE and Douglas GR (2005) Detailed review of transgenic rodent mutation assays. Mutat Res 590: 1-280.

Somers CM, Yauk C, White PA, Parfett CLJ, Quinn JS (2002) Air pollution induces heritable DNA mutations. Proc. Natl. Acad. Sci. USA 99: 15904-15907.

Tice RR, Agurell E, Anderson D, Burlinson B, Hartmann A, Kobayashi H, Miyamae Y, Rojas E, Ryu J-C and Sasaki F (2000) Single cell gel/Comet assay: guidelines for in vitro and in vivo genetic toxicology testing. Environ Mol Mutagen 35: 206-221.

Yauk C (2004) Advances in the application of germline tandem repeat instability for in situ monitoring. Mutat Res, 566, 169-182.

R.7.7.8 CARCINOGENICITY

R.7.7.8.1 Definition of carcinogenicity

Chemicals are defined as carcinogenic if they induce tumours, increase tumour incidence and/or malignancy or shorten the time to tumour occurrence. Benign tumours that are considered to have the potential to progress to malignant tumours are generally considered along with malignant tumours. Chemicals can induce cancer by any route of exposure (e.g., when inhaled, ingested, applied to the skin or injected), but carcinogenic potential and potency may depend on the conditions of exposure (e.g., route, level, pattern and duration of exposure). Carcinogens may be identified from epidemiological studies, from animal experiments and/or other appropriate means that may include (Quantitative) Structure-Activity Relationships ((Q)SAR) analyses and/or extrapolation from structurally similar substances (read-across). Each strategy for the identification of potential carcinogens is discussed in detail later in this report. The determination of the carcinogenic potential of a chemical is based on a *Weight of Evidence* approach. Classification criteria are given in the (EU Directive 67/548/EEC).[47]

[47] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

The process of carcinogenesis involves the transition of normal cells into cancer cells via a sequence of stages that entail both genetic alterations (i.e. mutations[48]) and non-genetic events. Non-genetic events are defined as those alterations/processes that are mediated by mechanisms that do not affect the primary sequence of DNA and yet increase the incidence of tumours or decrease the latency time for the appearance of tumours. For example; altered growth and death rates, (de)differentiation of the altered or target cells and modulation of the expression of specific genes associated with the expression of neoplastic potential (e.g. tumour suppressor genes or angiogenesis factors) are recognised to play an important role in the process of carcinogenesis and can be modulated by a chemical agent in the absence of genetic change to increase the incidence of cancer.

Carcinogenic chemicals have conventionally been divided into two categories according to the presumed mode of action: genotoxic or non-genotoxic[48]. Genotoxic modes of action involve genetic alterations caused by the chemical interacting directly with DNA to result in a change in the primary sequence of DNA. A chemical can also cause genetic alterations indirectly following interaction with other cellular processes (e.g., secondary to the induction of oxidative stress). Non-genotoxic modes of action include epigenetic changes, i.e., effects that do not involve alterations in DNA but that may influence gene expression, altered cell-cell communication, or other factors involved in the carcinogenic process. For example, chronic cytotoxicity with subsequent regenerative cell proliferation is considered a mode of action by which tumour development can be enhanced: the induction of urinary bladder tumours in rats may, in certain cases, be due to persistent irritation/inflammation, tissue erosion and regenerative hyperplasia of the urothelium following the formation of bladder stones. Other modes of non-gentoxic action can involve specific receptors (e.g., PPARα, which is associated with liver tumours in rodents; or tumours induced by various hormonal mechanisms). As with other nongenotoxic modes of action, these can all be presumed to have a threshold.

R.7.7.8.2 Objective of the guidance on carcinogenicity

The objective of investigating the carcinogenicity of chemicals is to identify potential human carcinogens, their mode(s) of action, and their potency.

With respect to carcinogenic potential and potency the most appropriate source of information is directly from human epidemiology studies (e.g. cohort, case control studies). In the absence of human data, animal carcinogenicity tests may be used to differentiate carcinogens from non-carcinogens. However, the results of these studies subsequently have to be extrapolated to humans, both in qualitative as well as quantitative terms. This introduces uncertainty, both with regard to potency for as well as relevance to humans, due to species specific factors such as differences in chemical metabolism and toxicokinetics and difficulties inherent in extrapolating from the high doses used in animal bioassays to those normally experienced by humans.

Once a chemical has been identified as a carcinogen, there is a need to elucidate the underlying mode of action, i.e. whether the chemical is directly genotoxic or not. In risk assessment a distinction is made between different types of carcinogens (see above).

For genotoxic carcinogens exhibiting direct interaction with DNA it is not generally possible to infer the position of the threshold from the *no-observed-effect level* on a dose-response curve, even though a biological threshold below which cancer is not induced may exist.

[48] For a definition and for background information on the terms mutagnicity and genotoxicity see Section R.7.7.1.1.

For non-genotoxic carcinogens, *no-effect-thresholds* are assumed to exist and to be discernable (e.g. if appropriately designed studies of the dose response for critical non-genotoxic effects are conducted). No effect thresholds may also be present for certain carcinogens that cause genetic alterations via indirect effects on DNA following interaction with other cellular processes (e.g. carcinogenic risk would manifest only after chemically induced alterations of cellular processes had exceeded the compensatory capacity of physiological or homeostatic controls). However, in the latter situation the scientific evidence needed to convincingly underpin this indirect mode of genotoxic action may be more difficult to achieve. Human studies are generally not available for making a distinction between the above mentioned modes of action; and a conclusion on this, in fact, depends on the outcome of mutagenicity/genotoxicity testing and other mechanistic studies. In addition to this, animal studies (e.g. the carcinogenicity study, repeated dose studies, and experimental studies with initiation-promotion protocols) may also inform on the underlying mode of carcinogenic action.

The cancer hazard and mode of action may also be highly dependent on exposure conditions such as the route of exposure. A pulmonary carcinogen, for example, can cause lung tumours in rats following chronic inhalation exposure, but there may be no cancer hazard associated with dermal exposure. Therefore, all relevant effect data and information on human exposure conditions are evaluated in a *Weight of Evidence* approach to provide the basis for regulatory decisions.

R.7.7.9 Information requirements on carcinogenicity

For the endpoint of carcinogenicity, standard information requirements are specifically described for substances produced or imported in quantities of ≥1000 t/y (Annex X). The precise information requirements will differ from substance to substance, according to the toxicity information already available and details of use and human exposure for the substance in question. The REACH Annexes VI to XI should be considered as a whole and in conjunction with the overall requirements of registration and evaluation.

Column 2 of Annex X lists specific rules according to which the required standard information may be omitted, replaced by other information, provided at a different stage or adapted in another way. If the conditions are met for adaptations under column 2 of this Annex, the fact and the reasons for each adaptation should be clearly indicated in the registration.

The standard information requirements for carcinogenicity and the specific rules for adaptation of these requirements are presented in Table R.7.7-6.

Table R.7.7-6 Standard information requirements for carcinogenicity and the specific rules for adaptation of these requirements

COLUMN 1 STANDARD INFORMATION REQUIRED	COLUMN 2 SPECIFIC RULES FOR ADAPTATION FROM COLUMN 1
Annexes VII-IX	
Annex X: 1. Carcinogenicity study.	1. A carcinogenicity study may be proposed by the registrant or may be required by the Agency in accordance with Articles 40 or 41 if: - the substance has a widespread dispersive use or there is evidence of frequent or long-term human exposure; and - the substance is classified as mutagen category 3 or there is evidence from the repeated dose study(ies) that the substance is able to induce hyperplasia and/or pre-neoplastic lesions. If the substance is classified as mutagen category 1 or 2, the default presumption would be that a genotoxic mechanism for carcinogenicity is likely. In these cases, a carcinogenicity test will normally not be required.

R.7.7.10 Information and its sources on carcinogenicity

There are many different sources of information that may permit inferences to be drawn regarding the potential of chemicals to be carcinogenic to humans. Clearly, these sources not only allow the identification of potential carcinogenic activity, but in case a substance is identified as a likely carcinogen they should also be informative with respect to the underlying mode of action as well as probable carcinogenic potency. The requirements of REACH call for proper classification and labelling, as well as for a quantitative assessment of risk that permits conclusions to be drawn regarding conditions under which safe use of the chemical may occur: i.e. the data should allow concluding on threshold or non-threshold mode of action, and on some dose descriptor (characterising the dose-response), preferably in quantitative terms.

It is noted (and indicated below), that the various sources inform differently on the aspects of hazard identification, mode of action, or carcinogenic potency.

R.7.7.10.1 Non-human data on carcinogenicity

Non-testing data on carcinogenicity

The capacity for performing the standard rodent cancer bioassay is limited by economic, technical and animal welfare considerations, such that an increased emphasis is being placed on the development of alternative, non-animal testing methods. However, carcinogenicity predictions through use of non-testing data currently represent an extreme challenge due to the multitude of possible mechanisms. Prediction of carcinogenicity in humans is especially problematic.

Although significant challenges remain, a broad spectrum of non-testing techniques exist for elucidating mechanistic, toxicokinetic or toxicodynamic factors important in understanding the carcinogenic process. These range from expert judgement, to the evaluation of structural similarities

and analogues (i.e. read-across and grouping), to the use of (Q)SAR models for carcinogenicity. Such information may assist with priority setting, hazard identification, elucidation of the mode of action, potency estimation and/or with making decisions about testing strategies based on a *Weight of Evidence* evaluation.

Genotoxicity remains an important mechanism for chemical carcinogenesis and its definitive demonstration for a chemical is often decisive for the choice of risk assessment methodology. A commentary about non-testing options for genotoxicity is provided in Section R.7.7.1. It has long been known that certain chemical structures or fragments can be associated with carcinogenicity, often through DNA-reactive mechanisms. Useful guidance for structures and fragments that are associated with carcinogenicity via DNA reactive mechanisms has been provided by the US Food and Drug Administration's "Guideline for Threshold Assessment, Appendix I, Carcinogen Structure Guide" (US FDA, 1986); the Ashby-Tennant "super-mutagen model" (*e.g.*, Ashby and Tennant, 1988); and subsequent builds on this model (*e.g.*, Ashby and Paton, 1993; Munro *et al*., 1996). Additional information on structural categories can be found in the "IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Man" (IARC, 2006).

Models predicting test results for genotoxic endpoints for closely related structures are known as *local* or congeneric (Q)SARs. These congeneric models are less common for carcinogenicity than for mutagenicity. Franke *et al.* (2001) provide an example of such a model for a set of genotoxic carcinogens.

The situation is far more complex for non-genotoxic carcinogenicity due to the large number of different mechanisms that may be involved. However, progress is being made in predicting activity for classes of compounds that exert effect via binding to oestrogen receptors, induction of peroxisomal proliferation, and binding to tubulin proteins. Although many potentially useful models exist, their applicability will be highly dependant on the proposed mechanism and chemical class.

Several *global* (non-congeneric) models exist which attempt to predict (within their domain) the carcinogenic hazard of diverse (non-congeneric) groups of substances (*e.g.* Matthews and Contrera, 1998). These models may also assist in screening, priority-setting, deciding on testing strategies and/or the assessment of hazard or risk based on *Weight of Evidence*. Most are commercial and include expert systems such as Onco-Logic® (currently made available by US-EPA) and DEREK, artificial intelligence systems from MULTICASE, and the TOPKAT program. Historically, the performance of such models has been mixed and is highly dependent on the precise definition of carcinogenicity among those substances used to develop and test the model. These have been reviewed by ECETOC (2003) and Cronin *et al*. (2003).

Free sources of carcinogenicity predictions include the Danish EPA (Q)SAR database (accessible through the European Commission's Chemicals Bureau: ECB http://ecbqsar.jrc.it/). Predictions in this database for 166,000 compounds include eight MULTICASE FDA cancer models, a number of genotoxicity predictions, rodent carcinogenic potency, hepatospecificity, oestrogenicity and aryl hydrocarbon (AH) receptor binding. Another source of carcinogenicity predictions is the Enhanced NCI database "*Browser*", which is sponsored by the US National Cancer Institute. This has 250,000 chemical predictions within it (http://cactus.nci.nih.gov), including general carcinogenicity, mutagenicity and additional endpoints, which may be of potential mechanistic interest in specific cases.

Further information on carcinogenicity models is available in the OECD Database on Chemical Risk Assessment Models where they are listed in an effort to identify tools for research and development of chemical substances. (http://www.olis.oecd.org/comnet/env/models.nsf/-MainMenu?OpenForm).

The guidance on the Grouping of Chemicals and on (Q)SARs (see Sections R.6.2 and R.6.1, respectively) explains basic concepts of grouping and (Q)SARs and gives generic guidance on validation, adequacy and documentation for regulatory purposes. The guidance also describes a stepwise approach for the use of read-across/grouping and (Q)SARs.

It is noted that all the above mentioned sources may potentially inform on possible carcinogenic hazard and on the underlying mode of action, as well as on carcinogenic potency.

Testing data on carcinogenicity

In vitro data

The following *in vitro* data, which provide direct or indirect information useful in assessing the carcinogenic potential of a substance and (potentially) on the underlying mode(s) of action, may be available. No single endpoint or effect in and of itself possesses unusual significance for assessing carcinogenic potential but must be evaluated within the context of the overall toxicological effects of a substance under evaluation as described in Section R.7.7.11.1. Except as noted, standardised protocols do not exist for most of the *in vitro* endpoints noted. Rather, studies are conducted in accordance with expert judgement using protocols tailored to the specific substance, target tissue and cell type or animal species under evaluation.

genotoxicity studies: the ability of substances to induce mutations or genotoxicity (as defined in Section R.7.7.1) can be indicative of carcinogenic potential. However, correlations between mutagenicity/genotoxicity and carcinogenesis are stronger when effects are observed in appropriately designed *in vivo* as opposed to *in vitro* studies.

***in vitro* cell transformation assay results:** such assays assess the ability of chemicals to induce changes in the morphological and growth properties of cultured mammalian cells that are presumed to be similar to phenotypic changes that accompany the development of neoplastic or pre-neoplastic lesions *in vivo* (OECD, 2006). The altered cells detected by such assays may other targeted mechanisms of action

possess, or can subsequently acquire, the ability to grow as tumours when injected into appropriate host animals. As in vitro assays, cell transformation assays are restricted to the detection of effects of chemicals at the cellular level and will not be sensitive to carcinogenic activity mediated by effects exerted at the level of intact tissues or organisms.

mechanistic studies, e.g. on:

possess, or can subsequently acquire, the ability to grow as tumours when injected into appropriate host animals. As in vitro assays, cell transformation assays are restricted to the detection of effects of chemicals at the cellular level and will not be sensitive to carcinogenic activity mediated by effects exerted at the level of intact tissues or organisms.

mechanistic studies, e.g. on:

- cell proliferation: sustained cell proliferation can facilitate the growth of neoplastic/pre-neoplastic cells and/or create conditions conducive to spontaneous changes that promote neoplastic development.
- altered intercellular gap junction communication: exchange of growth suppressive or other small regulatory molecules between normal and neoplastic/pre-neoplastic cells through gap junctions is suspected to suppress phenotypic expression of neoplastic potential. Disruption of gap junction function, as assessed by a diverse array of assays for fluorescent dye transfer or the

exchange of small molecules between cells, may attenuate the suppression of neoplastic potential by normal cells.

- hormone- or other receptor binding; a number of agents may act through binding to hormone receptors or sites for regulatory substances that modulate the growth of cells and/or control the expression of genes that facilitate the growth of neoplastic cells. Interactions of this nature are diverse and generally very compound specific.

other targeted mechanisms of action

- immunosuppressive activity: neoplastic cells frequently have antigenic properties that permit their detection and elimination by normal immune system function. Suppression of normal immune function can reduce the effectiveness of this *immune surveillance* function and permit the growth of neoplastic cells induced by exogenous factors or spontaneous changes.
- ability to inhibit or induce apoptosis: apoptosis, or programmed cell death, constitutes a sequence of molecular events that results in the death of cells, most often by the release of specific enzymes that result in the degradation of DNA in the cell nucleus. Apoptosis is integral to the control of cell growth and differentiation in many tissues. Induction of apoptosis can eliminate cells that might otherwise suppress the growth of neoplastic cells; inhibition of apoptosis can permit pre-neoplastic/neoplastic cells to escape regulatory controls that might otherwise result in their elimination.
- ability to stimulate angiogenesis or the secretion of angiogenesis factors: the growth of pre-neoplastic/neoplastic cells in solid tumours will be constrained in the absence of vascularisation to support the nutritional requirements of tumour growth. Secretion of angiogenesis factors stimulates the vascularisation of solid tumour tissue and enables continued tumour growth.

Animal data

A wide variety of study categories may be available, which may provide direct or indirect information useful in assessing the carcinogenic potential of a substance to humans. They include:

carcinogenicity studies (conventional long-term or life-time studies in experimental animals): Carcinogenicity testing is typically conducted using rats and mice, and less commonly in animals such as the Guinea pig, Syrian hamster and occasionally mini-pigs, dogs and primates. The standard rodent carcinogenicity bioassay would be conducted using rats or mice randomly assigned to treatment groups. Exposures to test substances may be via oral, inhalation or dermal exposure routes. The selection of exposure route is often dictated by *a priori* assumptions regarding the routes of exposure relevant to humans and/or other data sources (e.g. epidemiology studies or repeated dose toxicity studies in animals) that may indicate relevance of a given exposure route. Standardised protocols for such studies have been developed and are well validated (*e.g.* OECD TGs 451, 453 or US-EPA 870.4200).

short and medium term bioassay data (e.g., mouse skin tumour, rat liver foci model, neonatal mouse model): multiple assays have been developed that permit the detection and quantitation of putative pre-neoplastic changes in specific tissues. The induction of such *pre-neoplastic foci* may be indicative of carcinogenic potential. Such studies are generally regarded as adjuncts to conventional cancer bioassays, and while less validated and standardised, are applicable on a case-by-case basis for obtaining supplemental mechanistic and dose response information that may be useful for risk assessment (Enzmann et al, 1998).

genetically engineered (transgenic) rodent models (*e.g.*, *Xpa*$^{-/-}$, *p53*$^{+/-}$, *ras*H2 or Tg.AC): animals can be genetically engineered such that one or more of the molecular changes required for the

multi-step process of carcinogenesis has been accomplished (Tennant et al., 1999). This can increase the sensitivity of the animals to carcinogens and/or decrease the latency with which spontaneous or induced tumours are observed. The genetic changes in a given strain of engineered animals can increase sensitivity to carcinogenesis in a broad range of tissues or can be specific to the changes requisite for neoplastic development in one or only a limited number of tissues (Jacobson-Kram, 2004; Pritchard et al., 2003; ILSI/HESI 2001). Data from these models may be used in a *Weight of Evidence* analysis of a chemical's carcinogenicity.

genotoxicity studies *in vivo*: the ability of substances to induce mutations or genotoxicity (as defined in Section R.7.7.1.1) can be indicative of carcinogenic potential. There is, in general, a good correlation between positive genotoxicity findings *in vivo* and animal carcinogenicity bioassay results

repeated dose toxicity tests: can identify tissues that may be specific targets for toxicity and subsequent carcinogenic effects. Particular significance can be attached to the observation of pre-neoplastic changes (e.g. hyperplasia or metaplasia) suspected to be conducive to tumour development and may assist in the development of dose-effect relationships (Elcombe *et al*, 2002).

studies on the induction of sustained cell proliferation: substances can induce sustained cell proliferation via compensatory processes that continuously regenerate tissues damaged by toxicity. Some substances can also be tissue-specific mitogens, stimulating cell proliferation in the absence of overt toxic effects. Mitogenic effects are often associated with the action of tumour promoters. Both regenerative cell proliferation and mitogenic effects can be necessary, but not sufficient, for tumour development but have sufficiently different mechanistic basis that care should be exercised in assessing which is occurring (Cohen and Ellwein, 1991; Cohen et al., 1991).

studies on immunosuppressive activity: as noted earlier, suppression of normal immune surveillance functions can interfere with normal immune system functions that serve to identify and eliminate neoplastic cells.

studies on toxicokinetics: can identify tissues or treatment routes that might be the targets for toxicity and can deliver data on exposure and metabolism in specific organs. Linkages to subsequent carcinogenic impacts may or may not exist, but such data can serve to focus carcinogenesis studies upon specific tissue types or animal species.

other studies on mechanisms/modes of action, e.g. OMICs studies (toxicogenomics, proteomics, metabonomics and metabolomics): carcinogenesis is associated with multiple changes in gene expression, transcriptional regulation, protein synthesis and other metabolic changes. Specific changes diagnostic of carcinogenic potential have yet to be validated, but these rapidly advancing fields of study may one day permit assessment of a broad array of molecular changes that might be useful in the identification of potential carcinogens.

It is noted that the above tests differently inform on hazard identification, mode of action or carcinogenic potency. For example, conventional bioassays are used for hazard identification and potency estimation (i.e. derivation of a dose descriptor), whereas studies using genetically engineered animals are informative on potential hazard and possibly mode of action, but less on carcinogenic potency as they are considered to be highly sensitive to tumour induction.

R.7.7.10.2 Human data on carcinogenicity

Human data may provide direct information on the potential carcinogenicity of the substance. Relevant human data of sufficient quality, if available, are preferable to animal data as no

extrapolations between species, or from high to low dose are necessary. Epidemiological data will not normally be available for new substances but may well be available for substances that have been in use for many decades. For substances in common use prior to the implementation of modern occupational hygiene measures, the intensity of human exposures to some carcinogens was sufficient to produce highly significant, dose-dependent increases in cancer incidence.

A number of basic epidemiological study designs exist and include cohort, case-control and registry based correlational (e.g. ecological) studies. The most definitive epidemiological studies on chemical carcinogenesis are generally cohort studies of occupationally exposed populations, and less frequently the general population. Cohort studies evaluate groups of initially healthy individuals with known exposure to a given substance and follow the development of cancer incidence or mortality over time. With adequate information regarding the intensity of exposure experienced by individuals, dose dependent relationships with cancer incidence or mortality in the overall cohort can be established. Case-control studies retrospectively investigate individuals who develop a certain type of cancer and compare their chemical exposure to that of individuals who did not develop disease. Case control studies are frequently nested within the conduct of cohort studies and can help increase the precision with which excess cancer can be associated with a given substance. Correlational or ecological studies evaluate cancer incidence/mortality in groups of individuals presumed to have exposure to a given substance but are generally less precise since measures of the exposure experienced by individuals are not available. Observations of cancer clusters and case reports of rare tumours may also provide useful supporting information in some instances but are more often the impetus for the conduct of more formal and rigorous cohort studies.

Besides the identification of carcinogens, epidemiological studies may also provide information on actual exposures in representative (or historical) workplaces and/or the environment and the associated dose-response for cancer induction. Such information can be of much value for risk characterisation.

Although instrumental in the identification of known human carcinogens, epidemiology studies are often limited in their sensitivity by a number of technical factors. The extent and/or quality of information that is available regarding exposure history (e.g. measurements of individual exposure) or other determinants of health status within a cohort is often limited. Given the long latency between exposure to a carcinogen and the onset of clinical disease, robust estimates of carcinogenic potency can be difficult to generate. Similarly, occupational and environmentally exposed cohorts often have co-exposures to carcinogenic substances that have not been documented (or are incompletely documented). This can be particularly problematic in the study of long established industry sectors (e.g. base metal production) now known to entail co-exposures to known carcinogens (e.g. arsenic) present as trace contaminants in the raw materials being processed.. Retrospective hygiene and exposure analyses for such sectors are often capable of estimating exposure to the principle materials being produced, but data documenting critical co-exposures to trace contaminants may not be available. Increased cancer risk may be observed in such settings, but the source of the increased risk can be difficult to determine. Finally, a variety of lifestyle confounders (smoking and drinking habits, dietary patterns and ethnicity) influence the incidence of cancer but are often inadequately documented for purposes of adequate confounder control. Thus, modest increases in cancer at tissue sites known to be impacted by confounders (e.g. lung and stomach) can be difficult to interpret.

Techniques for biomonitoring and molecular epidemiology are developing rapidly. These newly developed tools promise to provide information on biomarkers of individual susceptibility, critical target organ exposures and whether effects occur at low exposure levels. Such ancillary information may begin to assist in the interpretation of epidemiology study outcomes and the definition of dose response relationships. For example, monitoring the formation of chemical adducts in haemoglobin

molecules (Birner et al., 1990; Albertini et al., 2006), the urinary excretion of damaged DNA bases (Chen, H.J. and Chiu, W.L. (2005), and the induction of genotoxicity biomarkers (micronuclei or chromosome aberrations; Boffetta et al., 2007) are presently being evaluated and/or validated for use in conjunction with classical epidemiological study designs. Such data are usually restricted in their application to specific chemical substances but such techniques may ultimately become more widely used, particularly when combined with animal data that defines potential mechanisms of action and associated biomarkers that may be indicative of carcinogenic risk. Monitoring of the molecular events that underlie the carcinogenic process may also facilitate the refinement of dose response relationships and may ultimately serve as early indicators of potential cancer risk. However, as a generalisation, such biomonitoring tools have yet to demonstrate the sensitivity requisite for routine use.

R.7.7.10.3 Exposure considerations for carcinogenicity

Information on exposure, use and risk management measures should be collected in accordance with Article 9 and Annex VI of REACH.

It is indicated in REACH Annex X a carcinogenicity study may be required by the European Chemicals Agency (or proposed by the registrant) when the substance has a widespread dispersive use or there is evidence of frequent or long-term human exposure. Preliminary toxicokinetic studies may be required first to address specific questions regarding potential target tissues and relevant exposure routes relevant for the chemical of concern.

On the other hand, investigations on the carcinogenic properties of a chemical can be deferred, if it can be demonstrated to the satisfaction of the Agency that the chemical is used only in a closed system and that human exposures are negligible, (i.e. risk reduction measures on the substance are already equivalent to those applied to high potency carcinogenic substances of category 1 and 2.Reasons for this could include the presence of other substances for which strict exposure regimes are implemented or enforced). The rationale for exemption from testing, of course, needs to be clearly documented upon registration.

Also, considerations on exposure may influence the search for information, e.g. applicable to the actual route of exposure. For example, if from exposure scenarios it is clear that only a single specific route is involved, toxicity data for this route is of higher relevance in data gathering and evaluation than for the other routes. Also, the involvement of inhalation exposure to particles will prioritise toxicity information needs in order to allow a proper hazard evaluation and risk assessment.

R.7.7.11 Evaluation of available information on carcinogenicity

This particular endpoint is complex and requires evaluation by expert judgement.

Note that the objective of this evaluation is to acquire information on the carcinogenic potential of the substance: i.e. is the substance carcinogenic or not, and, if so, what is the underlying mode of action (thresholded or not), and what is its carcinogenic potency (i.e. there is a need to define a dose descriptor).

An evaluation on the above mentioned properties requires a combining of various types of information, as indicated indicated in Section R.7.7.10 (and below). Such an evaluation needs a *Weight of Evidence* approach for arriving at conclusions, i.e. a careful gathering, sorting and

weighing of the various pieces of information available. This exercise is particularly complex and, therefore, requires expert judgement input.

R.7.7.11.1 Non-human data on carcinogenicity

Non-testing data for carcinogenicity

To date little experience is available for the evaluation of substances on non-testing data, since the use of non-testing data for regulatory decisions is rather new. Therefore, at every stage in the assessment for potential chemical toxicity, specialist judgement is essential. It is recognised though, that non-testing data may potentially inform on all carcinogenic properties, i.e. including mode of action and potency.

Documentation should include reference to a related chemical or groups of chemicals that give rise to concern or lack of concern. This can either be presented according to scientific logic (read-across) or as a mathematical relationship of chemical similarity.

In some cases, the carcinogenic potential posed by a substance can be assessed based upon analysis of the relative concentrations of constituents believed to present a risk in a complex mixture. For example, the classification of certain complex coal- and oil-derived substances as carcinogens can varies as a function of the content of marker carcinogens (benzene, 1,3-butadiene and benzene), whereas for others it depends on the level of polycyclic aromatic hydrocarbons measured following DMSO solvent extraction. (see Annex I of EU Directive 67/548/EEC). When properly validated, such chemical extraction and analysis techniques are highly predictive of the outcomes that would be obtained in animal carcinogenicity studies.

If well documented and applicable, (Q)SARs can be used to help reach the decision points described in the section below. The accuracy of such methods may be sufficient to help or allow either a testing or a specific regulatory decision to be made. Expert judgment is needed to make this determination.

Chemicals for which no test-data exist present a special case in which reliance on non-testing methods may be absolute. Many factors will dictate the acceptability of non-testing methods in reaching a conclusion based on no tests at all. A *Weight of Evidence* evaluation of carcinogenicity based on multiple genotoxicity and carcinogenicity estimates (*e.g.* from (Q)SAR models) may in some cases equal or exceed the decision basis which could be obtained by experimentally testing a chemical in one or two *in vitro* tests. This must be considered on a case-by-case basis by the registrant.

Further guidance on the use of Grouping of Chemicals and on (Q)SARs both for a qualitative (i.e. classification and labelling) as well as a quantitative assessment (i.e. identifying some dose descriptor value) is provided in Sections R.4.3.2 and R.6.2, respectively, and also includes basic concepts used, validation status, adequacy and documentation needs for regulatory purposes.

Testing data on carcinogenicity

In vitro data

In vitro data can only give preliminary information about the carcinogenic potential of a substance and possible underlying mode(s) of action. For example, *in vitro* genotoxicity studies may provide information about whether or not the substance is likely to be genotoxic *in vivo*, and thus a potential genotoxic carcinogen (see Section R.7.7.1), and herewith on the potential mode of action underlying carcinogenicity: with or without a threshold.

Besides genotoxicity data other *in vitro* data (described in Section R.7.7.10.1) such as *in vitro* cell transformation can help to decide, in a *Weight of Evidence* evaluation, whether a chemical possesses a carcinogenic potential. Cell transformation results in and of themselves do not inform as to the actual underlying mode(s) of action, since they are restricted to the detection of effects exerted at the level of the single cell and may be produced by mechanistically distinct processes.

Studies can also be conducted to evaluate the ability of substances to influence processes thought to facilitate carcinogenesis. Many of these endpoints are assessed by experimental systems that have yet to be formally validated and/or are the products of continually evolving basic research. Formalised and validated protocols are thus lacking for the conduct of these tests and their interpretation. Although it is difficult to give general guidance on each test due to the variety and evolving nature of tests available, it is important to consider them on a case-by-case basis and to carefully consider the context on how the test was conducted.

A number of the test endpoints evaluate mechanisms that may contribute to neoplastic development, but the relative importance of each endpoint will vary as a function of the overall toxicological profile of the substance being evaluated. It should further be noted that there are significant uncertainties associated with extrapolating *in vitro* data to an *in vivo* situation. Such *in vitro* data will, in many instances, provide insights into the nature of the *in vivo* studies that might be conducted to define carcinogenic potential and/or mechanisms.

Animal data

In vivo data can give direct information about the carcinogenic potential of a substance, possible underlying mode(s) of action, and its potency.

Testing for carcinogenicity is conventionally carried out in groups of rats or mice according to standard test protocols or guidelines (e.g. OECD TGs 451, 453 or US-EPA 870.4200) and a conclusion is based on a comparison of the incidence, nature and time of occurrence of neoplasms in treated animals and controls.

Knowledge of the historic tumour incidence for the strain of animal used is important (laboratory specific data are preferable). Also attention to the study design used is essential because of the requirement for statistical analyses. The quality, integrity and thoroughness of the reported data from carcinogenicity studies are essential to the subsequent analysis and evaluation of studies. A qualitative assessment of the acceptability of study reports is therefore an important part of the process of independent evaluation. Sources of guidance in this respect can be found in IEH (2002), CCCF (2004) and OECD (2002). If the available study report does not include all the information required by the standard test guideline, judgement is required to decide if the experimental procedure is or is not acceptable and if essential information is lacking.

The final design of a carcinogenicity bioassay may deviate from OECD guidelines if expert judgement and experience in the testing of analogous substances supports the modification of protocols. Such modifications to standard protocols can be considered as a function of the specific properties of the material under evaluation.

Carcinogenicity data may sometimes be available in species other than those specified in standard test guidelines (e.g., Guinea pig, Syrian hamster and occasionally mini-pigs, dogs and primates). Such studies may be in addition to, or instead of, studies in rats and mice and they should be considered in any evaluation.

Data from non-conventional carcinogenicity studies, such as short- and medium-term carcinogenicity assays with neonatal or genetically engineered (transgenic) animals, may also be

available (CCCF, 2004; OECD, 2002). Genetically engineered animals possess mutations in genes that are believed to be altered in the multi-step process of carcinogenesis, thereby enhancing animal sensitivity to chemically induced tumours. A variety of transgenic animal models exist and new models are continually being development. The genetic alteration(s) in a specific animal model can be those suspected to facilitate neoplastic development in a wide range of tissue types or the alterations can be in genes suspected to be involved in tissue specific aspects of carcinogenesis. The latter must be applied with recognition of both their experimental nature and the specific mechanistic pathways they are designed to evaluate. For example, a transgenic animal model sensitive to mesothelioma induction would be of limited value in the study of a suspected liver carcinogen. While such animal model systems hold promise for the detection of carcinogens in a shorter period of time and using fewer animals, their sensitivity and specificity remains to be determined. Due to a relative lack of validation, such assays have not yet been accepted as alternatives to the conventional lifetime carcinogenicity studies, but may be useful for screening purposes or to determine the need for a rodent 2-year bioassay. Several evaluations of these types of study have been published (e.g., Jacobson-Kram, 2004; Pritchard et al., 2003; ILSI/HESI (2001).

When data are available from more than one study of acceptable quality, consistency of the findings should be established. When consistent, it is usually straightforward to arrive at a conclusion, particularly if the studies were in more than one species or if there is a clear treatment-related incidence of malignant tumours in a single study. If a single study only is available and the test substance is not carcinogenic, scientific judgement is needed to decide on whether (a) this study is relevant or (b) additional information is required to provide confidence that it should not be considered to be carcinogenic.

Study findings also may not clearly demonstrate a carcinogenic potential, even when approved study guidelines have been followed. For example, there may only be an increase in the incidence of benign tumours or of tumours that have a high background incidence in control animals. Although less convincing than an increase in malignant and rare tumours, and recognising the potential over-sensitivity of this model (Haseman, 1983; Ames and Gold, 1990), a detailed and substantiated rationale should be given before such positive findings can be dismissed as not relevant.

Repeated dose toxicity studies may provide helpful additional information to the *Weight of Evidence* gathered to determine whether a substance has the potential to induce cancer, and for potential underlying modes of action (Elcombe *et al*., 2002). For example, the induction of hyperplasia (either through cytotoxicity and regenerative cell proliferation, mitogenicity or interference with cellular control mechanisms) and/or the induction of pre-neoplastic lesions may contribute to the *Weight of Evidence* for carcinogenic potential. Toxicity studies may also provide evidence for immunosuppressive activity, a condition favouring tumour development under conditions of chronic exposure.

Finally, toxicokinetic data may reveal the generation of metabolites with relevant structural alerts. It may also give important information as to the potency and relevance of carcinogenicity and related data collected in one species and its extrapolation to another, based upon differences in absorption, distribution, metabolism and or excretion of the substance. Species specific differences mediated by such factors may be demonstrated through experimental studies or by the application of toxicokinetic modelling.

Positive carcinogenic findings in animals require careful evaluation and this should be done with reference to other toxicological data (e.g. *in vitro* and/or *in vivo* genotoxicity studies, toxicokinetic data, mechanistic studies, (Q)SAR evaluations) and the exposure conditions (e.g., route). Such comparisons may provide evidence for (a) specific mechanism(s) of action, a significant factor to

take into account whenever possible, that may then be evaluated with respect to relevance for humans.

A conceptual framework that provides a structured and transparent approach to the *Weight of Evidence* assessment of the mode of action of carcinogens has been developed (see Sonich-Mullin *et al.*, 2001; Boobis et al., 2006). This framework should be followed when the mechanism of action is key to the risk assessment being developed for a carcinogenic substance and can be particularly critical in a determination of whether a substance induces cancer via genotoxic or nongenotoxic mechanisms.

For example, a substance may exhibit limited genotoxicity *in vivo* but the relevance of this property to carcinogenicity is uncertain if genotoxicity is not observed in tissues that are the targets of carcinogenesis, or if genotoxicity is observed via routes not relevant to exposure conditions (e.g. intravenous injection) but not when the substance is administered via routes of administration known to induce cancer. In such instances, the apparent genotoxic properties of the substance may not be related to the mechanism(s) believed to underlie tumour induction. For example, oral administration of some inorganic metal compounds will induce renal tumours via a mechanism believed to involve organ specific toxicity and forced cell proliferation. Although genotoxic responses can be induced in non-target tissues for carcinogenesis via intravenous injection, there is only limited evidence to suggest that this renal carcinogenesis entails a genotoxic mechanism (IARC, 2006). The *burden of proof* in drawing such mechanistic inferences can be high but can have a significant impact upon underlying assumptions made in risk assessment.

In general, tumours induced by a genotoxic mechanism (known or presumed) are, in the absence of further information, considered to be of relevance to humans even when observed in tissues with no direct human equivalent. Tumours shown to be induced by a non-genotoxic mechanism are, in principle, also considered relevant to humans but there is a recognition that some non-genotoxic modes of action do not occur in humans (see OECD 2002). This includes, for example, some specific types of rodent kidney, thyroid, urinary bladder, forestomach and glandular stomach tumours induced by rodent-specific modes of action, i.e., by mechanisms/modes of action not operating in humans or operative in humans under extreme and unrealistic conditions. Reviews are available for some of these tumour types providing a detailed characterisation that includes the key biochemical and histopathological events that are needed to establish these rodent-specific mechanisms that are not relevant for human health (see Technical Publication Series by IARC). Recently, the IPCS has developed a framework and provided some examples on how to evaluate the relevance to humans of a postulated mode of action in animals (ILSI RSI, 2003; Boobis *et al.*, 2006).

The information available for substances identified as carcinogenic based on testing and/or non-testing data should be further evaluated in an effort to identify underlying mode(s) of action and potency in order to subsequently allow a proper quantitative assessment of risk (see Section R.7.7.12.2). As already pointed out, the use of non-standard animal models (e.g. transgenic or neonatal animals) needs careful evaluation by expert judgement as to how to apply the results obtained for hazard and risk assessment purposes; it is not possible to provide guidance for such evaluations.

R.7.7.11.2 Human data on carcinogenicity

Epidemiological data may potentially be used for hazard identification, exposure estimation, dose response analysis, and risk assessment. The degree of reliability for each study on the carcinogenic potential of a substance should be evaluated using accepted causality criteria, such as that of Hill

(1965). Particular attention should be given to exposure data in a study and to the choice of the control population. Often a significant level of uncertainty exists around identifying a substance unequivocally as being carcinogenic because of inadequate reporting of exposure data. Chance, bias and confounding factors can frequently not be ruled out. A clear identification of the substance, the presence or absence of concurrent exposures to other substances and the methods used for assessing the relevant dose levels should be explicitly documented. A series of studies revealing similar excesses of the same tumour type, even if not statistically significant, may suggest a positive association, and an appropriate joint evaluation (meta-analysis) may be used in order to increase the sensitivity, provided the studies are sufficiently similar for such an evaluation. When the results of different studies are inconsistent, possible explanations should be sought and the various studies judged on the basis of the methods employed.

Interpretation of epidemiology studies must be undertaken with care and include an assessment of the adequacy of exposure classification, the size of the study cohort relative to the expected frequency of tumours at tissue sites of special concern and whether basic elements of study design are appropriate (e.g. a mortality study will have limited sensitivity if the cancer induced has a high rate of successful treatment). A number of such factors can limit the sensitivity of a given study – unequivocal demonstration that a substance is not a human carcinogen is difficult and requires detailed and exact measurements of exposure, appropriate cohort size, adequate intensity and duration of exposure, sufficient follow-up time and sound procedures for detection and diagnosis of cancers of potential concern. Conversely, excess cancer risk in a given study can also be difficult to interpret if relevant co-exposures and confounders have not been adequately documented. Efforts are ongoing to improve the sensitivity and specificity of traditional epidemiological methods by combining cancer endpoints with data on established pre-neoplastic lesions or molecular indicators (biomarkers) of cancer risk.

Once identified as a carcinogenic substance on the basis of human data, well-performed epidemiology studies may be valuable for providing information on the relative sensitivity of humans as compared to animals, and/or may be useful in demonstrating an upper bound on the human cancer risk. Identification of the underlying mode(s) of action – needed for the subsequent risk assessment (see Section R.7.7.12.2) – quite often depends critically on available testing and/or non-testing information.

R.7.7.11.3 Exposure considerations for carcinogenicity

Exposure considerations may lead to adaptation of the evaluation of available information, and / or of the testing strategy.

As indicated before, waiving of carcinogenicity studies may apply, e.g. when it can be demonstrated that the substance is only produced and used in closed systems, which among other reasons may be due to the presence of other substances for which strict exposure regimes are implemented or enforced. On the other hand, a carcinogenicity study may be required (by the Agency or proposed by the registrant) when the substance has a widespread dispersive use or there is evidence of frequent or long-term human exposure, and information on its carcinogenic properties cannot be obtained by others means (from available effect information). Preliminary toxicokinetic studies may be required first to identify the potential target tissues and exposure routes that would guide the design of appropriate studies for the chemical of concern.

In the former case, i.e. when the substance is produced and used in closed systems only, conclusions on safe use and handling can be verified by use of read-across to risk assessments of structurally related carcinogens or to the so-called Threshold of Toxicological Concern (TTC) concept (see

Appendix R.7-1): this concept identifies a *de minimis* exposure value for all chemicals, including genotoxic carcinogens, below which there is no appreciable risk to human health for any chemical. If it can be demonstrated that exposures are below these values, there is good reason for not performing the required tests. Clearly, good quality exposure information is essential in all these cases.

R.7.7.11.4 Remaining uncertainty on carcinogenicity

As indicated in the previous sections, adequate human data for evaluating the carcinogenic properties of a chemical are most often not available, and alternative approaches have to be used.

As also indicated in the previous sections and the Section R.7.7.1, test systems for identifying genotoxic carcinogens are reasonably well developed and adequately cover this property. There is also agreement that animal carcinogens which act by a genotoxic mode of action may reasonably be regarded as human carcinogens unless there is convincing evidence that the mechanisms by which mutagenicity and carcinogenicity are induced in animals are not relevant to humans. Unclear, however, and herewith introducing some uncertainty, is the relationship between carcinogenic potency in animals and in humans.

There is, on the other hand, a shortage of sensitive and selective test systems to identify non-genotoxic carcinogens, apart from the carcinogenicity bioassay. In the absence of non-testing information on the carcinogenicity of structurally related chemicals, indications for possible carcinogenic properties may come from existing repeated dose toxicity data, or from *in vitro* cell transformation assays. However, whereas the former source of data will have a low sensitivity (*e.g.* in case of a 28-day study), there is a possibility that the latter may lead to an over-prediction of carcinogenic potential. Insufficient data are available to provide further general guidance in this regard.

Non-genotoxic carcinogens may be difficult to identify in the absence of animal carcinogenicity test data. However, it could be argued that current conservative (cautious) risk assessment methodology should cover the risk for carcinogenic effects via this mode of action as well: i.e. current risk assessments for many non-genotoxic carcinogens are based on NOAELs for precursor effects or target organ toxicity with the application of conservative assessment factors to address uncertainty. For example, see the risk assessment for coumarin (EFSA, 2004; Felter *et al*., 2006). Such a risk assessment is not performed, though, in case this substance is not classified as dangerous for any other properties.

Once identified as a non-genotoxic carcinogen (from testing or non-testing data) there may be uncertainty as to the human relevance of this observation, i.e. to the human relevance of the underlying mode of action. In the absence of specific data on this, observations in the animal are taken as relevant to humans. However, additional uncertainty will exist for the relationship between carcinogenic potency in animals and in humans; this uncertainty, though, will be addressed in the procedure for deriving human standards (ILSI RSI 2003).

Finally, conventional assays of carcinogenicity in animals have been found to be insensitive for some well-established human carcinogenic substances (e.g. asbestos and arsenic compounds). These substances can be shown to be carcinogenic when the test conditions are modified, thus illustrating that there will always be a possibility that a chemical could pose a carcinogenic hazard in humans but be missed in conventional animal studies. This is also true for other toxicological endpoints and should be taken into account by risk managers, especially when making decisions

about the acceptability of scenarios showing particularly high exposures to workers and/or consumers.

R.7.7.12 Conclusions on carcinogenicity

R.7.7.12.1 Concluding on suitability for Classification and Labelling

In order to conclude on an appropriate classification and labelling position with regard to carcinogenicity, the available data should be considered using the criteria and guidance associated with the (EU Directive 67/548/EEC)[49].

R.7.7.12.2 Concluding on suitability for Chemical Safety Assessment

Besides the identification of a chemical as a carcinogenic agent from either animal data, epidemiological data or both, dose response assessment is an essential further step in order to characterise carcinogenic risks for certain exposure conditions or scenarios. A critical element in this assessment is the identification of the mode of action underlying the observed tumour-formation, as already explained in Section R.7.7.11.1: i.e. whether this induction of tumours is thought to be via a genotoxic mechanism or not.

In regulatory work, it is generally assumed that in the absence of data to the contrary an effect-threshold cannot be identified for genotoxic carcinogens exhibiting direct interaction with DNA, i.e., it is not possible to define a *no-effect level* for carcinogenicity induced by such agents. However, in certain cases even for these compounds a threshold for carcinogenicity may be identified in the low-dose region: e.g. it has in certain cases been clearly demonstrated that an increase in tumours did not occur at exposures below those associated with local chronic cytotoxicity and regenerative hyperplasia. It is also recognised that for certain genotoxic carcinogens causing genetic alterations, a practical threshold may exist for the underlying genotoxic effect. For example, this has been shown to be the case for aneugens (agents that induce aneuploidy – the gain or loss of entire chromosomes to result in changes in chromosome number), or for chemicals that cause indirect effects on DNA that are secondary to another effect (e.g., through oxidative stress that overwhelms natural antioxidant defence mechanisms).

Non-genotoxic carcinogens exert their effects through mechanisms that do not involve direct DNA-reactivity. It is generally assumed that these modes of actions are associated with threshold doses, and it may be possible to define no-effect levels for the underlying toxic effects of concern. There are many different modes of action thought to be involved in non-genotoxic carcinogenicity. Some appear to involve direct interaction with specific receptors (e.g. oestrogen receptors), whereas appear to be non-receptor mediated. Chronic cytotoxicity with subsequent regenerative cell proliferation is considered a mode of action by which tumour development can be induced: the induction of urinary bladder tumours in rats, for example, may, in certain cases, be due to persistent irritation/inflammation/erosion and regenerative hyperplasia of the urothelium following the formation of bladder stones which eventually results in tumour formation. Specific cellular effects,

[49] Directive 67/548/EEC will be repealed and replaced with the EU Regulation on classification, labelling and packaging of substances and mixtures, implementing the Globally Harmonized System (GHS).

such as inhibition of intercellular communication, have also been proposed to facilitate the clonal growth of neoplastic/pre-neoplastic cells.

The identification of the mode of action of a carcinogen is based on a combination of results in genotoxicity tests (both *in vitro* and *in vivo*) and observations in animal experiments, e.g. site and type of tumour and parallel observations from pathological and microscopic analysis. Epidemiological data seldom contribute to this.

Once the mode of action of tumour-formation is identified as having a threshold or not, a dose descriptor has to be derived for the purpose of allowing to conclude on chemical safety assessment. For threshold mechanisms the No-Observed-Adverse-Effect-Level (NOAEL) or Lowest-Observed-Adverse-Effect-Level (LOAEL) (see general introduction for definition and derivation of these descriptors) for tumour-formation or for the underlying (toxic) effect should be established to allow the derivation of a so-called Derived-No-Effect-Level (DNEL) (Chapter R.8), that subsequently is used in the safety assessment to establish safe exposure levels.

If the mode of action of tumour formation is identified as non-thresholded, dose descriptors such as T25, BMD10 or BMDL10 (general introduction for definition and derivation these descriptors) are to be established, that allow the derivation of a so-called Derived-Minimal-Effect-Level (DMEL; for guidance see Section R.8.5), that subsequently is used in the safety assessment to establish exposure levels of minimal concern.

Though mainly derived from animal data, epidemiological data may also occasionally provide dose descriptors that allow derivation of a DNEL or DMEL, e.g. Relative Risk (RR) or Odds Ratio (OR).

Substance-specific data for carcinogenicity normally will be absent, especially for the lower tonnage level substances. As indicated in Section R.7.7.11.1, non-testing data (read-across, grouping and/or (Q)SAR) may occasionally be considered sufficient to conclude on this endpoint, i.e. for classification, but also for establishing the underlying mode of action and for estimating the carcinogenic potency. This may introduce some additional uncertainty, especially with respect to the dose descriptor value, the addressing of which requires expert judgement; it is noted that experience to date on this is extremely limited. Guidance on read-across and/or grouping, and the use of (Q)SAR is provided in Sections R.6.2 and R.6.1.respectively.

R.7.7.12.3 Information not adequate

A *Weight of Evidence* approach comparing available adequate information with the tonnage-tiered information requirements by REACH may result in the conclusion that the information/data requirements are not fulfilled. In order to proceed in further information gathering, the following testing strategy can be adopted.

R.7.7.13 Integrated Testing Strategy (ITS) for carcinogenicity

R.7.7.13.1 Objective / General principles

The objective of this strategy is to describe where required how carcinogenicity should be assessed for all substances subject to registration under REACH: i.e. to identify substances with carcinogenic properties, their associated underlying mode of action, and their potency. Guidance is provided especially for those substances lacking pre-existing epidemiological or toxicological data on carcinogenicity.

The strategy provides the rationale for deciding whether or not a standard animal carcinogenicity study or any other further testing is required. It is recognised that standard carcinogenicity tests take considerable time to conduct and report, are expensive, and involve the use of a large number of animals. Consequently, it is preferable that decisions about the potential carcinogenicity of substances under REACH be taken as frequently as possible without the conduct of such tests.

The strategy recognises that the available information will differ from substance to substance. This may include various different types of toxicity information for the substance in question and/or for its analogues/structurally related chemicals. Details about the use and human exposure potential of the substance will also be available. All this will have an impact on the need for further data acquisition. Proposals for conducting a carcinogenicity test should be made with regard to the potential risk to human health and with consideration of the actual or intended production and/or use pattern.

REACH only specifies a carcinogenicity test for substances at the Annex X tonnage level (≥1000 t/y; see Section R.7.7.9). However, REACH also requires that carcinogenic substances at all tonnage levels be identified as substances of high concern, taking into account information from all available relevant sources (see Section R.7.7.10).

At the tonnage levels below 1000 t/y, the main concern is for those chemicals that are genotoxic. Chemicals may cause cancer secondary to other forms of toxicity, but protection of human health against the underlying toxicity (e.g., as identified from a repeat-dose toxicity study) will also protect against cancer that is secondary to that toxicity. It is noted, though, that some of these non-genotoxic carcinogens, when not classified for any other property and not identified as such in (limited) repeated dose toxicity studies will go unidentified; this also regards the risks associated with human exposures.

Finally, the strategy recognises that the carcinogenic process is a complex multi-step process. Chemically-induced cancer may be induced by any number of different pathways or modes of action and this allows for a variety of different approaches to carcinogenicity assessment. Substances that have the potential to act as genotoxic carcinogens can be identified by *in vitro* and *in vivo* mutagenicity tests, as described in Section R.7.7.1. Carcinogens that act by non-genotoxic modes of action are more difficult to identify because comparable, well-validated, short-term tests for the potentially numerous modes of actions involved are generally not available, and those tests that are available are not required as part of the standard information requirements of REACH.

A flow chart of the strategy is presented in Figure R.7.7-2.

R.7.7.13.2 Preliminary considerations

As a starting point, there will be the information collected with respect to mutagenicity. If they are available, test and non-test data from a literature search and, if possible, from members of an applicable chemical category or (Q)SAR analysis should be taken into account.

For substances for which there is no concern for mutagenic activity, and no other toxicological indicators of concern for carcinogenicity (i.e. for the substance itself or for structurally-related substances), there is no need for further consideration of its carcinogenic potential. This applies equally to those substances at the Annex X tonnage level as to those at lower tonnage levels.

If, however, for non-genotoxic substances toxicological indicators of concern are available (e.g. hyperplastic or pre-neoplastic lesions in repeated dose toxicity studies of the substance itself and/or of closely related substances), they should be investigated further on a case-by-case basis. Any

decision on further testing is dependent upon the type and strength of the indications for carcinogenicity, the potential mechanism of action and their relevance to humans, and the type and level of human exposure (see Section R.7.7.10.2).

If no conclusion can be drawn regarding the potential genotoxicity of the substance then, in general, it will be determined on a case-by-case basis when and how the carcinogenic potential should be explored further. Again, this will then depend on the type and strength of the indications for carcinogenicity, the potential mechanism(s) of action, and the type and level of human exposure.

At least for substances at the higher tonnage levels, subchronic and/or chronic studies may provide additional important information on possible carcinogenic effects. There may, for example, be indications of peroxisomal proliferation or of hyperplastic or pre-neoplastic responses, including dose-response characteristics. These should be investigated further on the already indicated case-by-case basis, depending on the type and strength of the indications for carcinogenicity, the potential mechanism of action and relevance to humans, and the type and level of human exposure.

It may be appropriate on occasions to propose other tests to be undertaken, e.g. to test a read-across option with available non-testing data. These could include short-term tests, such as those for *in vitro* cell transformation or cell proliferation, or medium-term tests, like genetically engineered (transgenic) or neonatal models. It may well be that data generated in this way supports this read-across to available non-testing data, and herewith provides sufficient confidence in a read-across derived estimate of the carcinogenic potency for the substance and also for the magnitude of the risks associated with experienced exposure levels. The data generated may also weaken or even disprove the basis for read-across. It is noted that experience to date on this is very limited (as indicated in Section R.7.7.11.1). Guidance on read-across and/or grouping is provided in Section R.6.2.

As validated testing procedures are not yet available and published in the OECD test guideline programme, it is essential that appropriate expert advice is sought regarding the application and suitability of any of these other tests.

Substances for which concern for carcinogenicity is solely based on positive genotoxicity data will, in a first step, be evaluated according to the approach outlined for identification of the genotoxicity hazard (see Section R.7.7.5).

Formally, for a substance classified as a category 1 or 2 mutagen, a carcinogenicity study will not normally be required (see Section R.7.7.9); *i.e.* it will be regarded as a genotoxic carcinogen. In order to allow an assessment of the magnitude of potential cancer risks associated with the prevailing human exposures, it may well be that available non-testing data (read-across, grouping, (Q)SAR) provide a sufficiently helpful estimate of the carcinogenic potency of the substance (i.e. by read-across) from which risks can be assessed. Guidance on read-across and/or grouping, and the use of (Q)SAR is provided in Sections R.6.2 and R.6.1, respectively.

In case such an approach is not possible, an estimate of acceptable exposure conditions may alternatively be obtained by use of the available data from animal toxicity studies: i.e. by identifying the minimal toxic dose in sub-chronic studies (if available, as some surrogate value for the dose descriptor) and by applying a large assessment factor; see for further guidance Gold *et al*. (2003). It is stressed that expert judgement is definitively needed here.

On very rare occasions, a case may be made to perform a carcinogenicity study in animals for substances that have been classified for mutagenicity in categories 1 or 2. Such a case would have to explain why the study was critically important; e.g. in the context of the clarification of carcinogenic risk associated with human exposures.

For substances classified as category 3 mutagens, and for which there is no carcinogenicity study, there should first be an evaluation of whether classification in category 2 for mutagenicity is possible. If such a classification is made, then the approach described above can be followed with regards to carcinogenicity. Occasionally, it may be established that classification as a category 2 mutagen is not appropriate. In such instances, it should not be assumed automatically that the substance has carcinogenic potential. However, unless there is clear evidence to indicate the contrary, it is expected that these substances will be regarded as genotoxic carcinogens.

As the previous paragraph implies, mutagenic potential *in vivo* is not always a reliable indicator of carcinogenic potential. If repeated dose toxicity studies indicate that pre-neoplastic changes (e.g. hyperplasia, precancerous lesions) occur, then the probability that carcinogenic activity will be expressed is increased. Non-testing data such as read-across and (Q)SAR may also contribute to this evaluation.

For substances at the REACH Annex X tonnage level, the need for or waiving of a standard animal test should be clearly explained, taking into account all the available toxicological and hygiene information on the substance and/or other relevant substances. For example, if it can be demonstrated that the substance is used only in a closed system and that human exposures are negligible, it is possible to propose no further testing for carcinogenicity.

It is recommended that when a carcinogenicity bioassay is required, study design and test protocol are well considered prior to delivering the test-proposal (e.g. OECD TG 453). Particular consideration, based on all the available data, should be given to the selection of the species and strain to be used in the carcinogenicity test, the route of exposure and dose level selection. It is also recommended that when a carcinogenicity test is to be conducted, an investigation of chronic toxicity should, whenever possible, form part of the study protocol. Finally, the limited value of a mouse assay as second species should be considered in this (Doe *et al.*, 2006).

The approaches outlined below may be used in the assessment of the potential carcinogenic risk of a substance to humans, and to help decide whether or not a carcinogenicity test will be required and, if so, when.

R.7.7.13.3 Testing strategy for carcinogenicity

As for other endpoints, the following three steps apply for the assessment of carcinogenicity (i.e. the hazard, underlying mode of action, and potency) for substances at each of the tonnage levels specified in Annexes VII to X of REACH.

(I) Gather and assess all available test and non-test data from read-across/proper chemical category and suitable predictive models. Examine the *Weight of Evidence* that relates to carcinogenicity.

(II) Consider whether the standard information requirements are met.

(III) Ensure that the information requirements of Annexes VII and VIII are met; make proposals to conform with Annexes IX and X.

Further details about the procedures to follow at each of the different tonnage levels are described below.

Substances at Annexes VII, VIII and IX

A definitive assessment of carcinogenicity is usually not possible from the data available at the Annex VII, VIII and IX tonnage levels. However, for all substances, any relevant test data that are already available, together with information from predictive techniques such as read-across or chemical grouping, should be used to form a judgement about this important hazard endpoint.

The minimum information to be provided at the Annex VII, VIII and IX tonnage levels in relation to this endpoint is equivalent to that required for the mutagenicity endpoint (see Section R.7.7.2): positive results from *in vitro* mutagenicity studies provide an alert for possible carcinogenicity, and need confirmation via further testing *in vitro* and/or *in vivo* mutagenicity testing. As such, this will not lead to classification of a substance as a carcinogen, but this evidence should be taken into account in risk assessment: substances shown to be *in vivo* mutagens should be assumed to be potentially carcinogenic.

Furthermore, the results of repeated dose toxicity studies and /or reproductive/ developmental toxicity tests may be informative about a possible carcinogenic potential: hyperplasia or other pre-neoplastic effects may be observed in these studies. These observations may also be informative on potential mode(s) of action underlying the carcinogenic effect.

Although the criteria for carcinogenicity classification may not be met in the absence of substance-specific carcinogenicity data, the evidence from the available information alerting to possible carcinogenicity should be taken into account in the risk assessment for this endpoint: ways that allow an assessment of the magnitude of potential cancer risks associated with human exposures without performing the assay are indicated in indicated in Section R.7.7.13.2. (see Section for derivation of DMEL and DNEL values Chapter R.8).

It is important to note that at the tonnage levels below 1000 t/y, the main concern is for those chemicals that are genotoxic. The repeated dose toxicity studies mentioned above may indicate cancers which are secondary to other forms of toxicity. For those the protection of human health against the underlying toxicity will also protect against cancer that is secondary to the toxicity. It is noted, though, that some of these non-genotoxic carcinogens, when not classified for any other property and not identified as such in (limited) repeated dose toxicity studies will go unidentified; this also regards the risks associated with human exposures.

Substances at Annex X

All substances at this tonnage should be evaluated for carcinogenicity.

All relevant data from all toxicity studies should be assessed to see whether a sufficiently reliable assessment about the carcinogenicity of the substance is possible, including alternative means, if needed: i.e. predictive techniques such as chemical grouping and read-across, and the use of (Q)SARs. On some occasions, it may be proposed to supplement these predictive approaches with *in vitro* or alternative shorter-term *in vivo* investigations in order to circumvent the need for a carcinogenicity study. This should usually be in the context of adding to the *Weight of Evidence* that a substance may be carcinogenic.

Formally, if the substance is classified as a category 1 or 2 mutagen (GHS category 1), a carcinogenicity study will not normally be required. For a substance classified as a category 3 mutagen (GHS category 2) it should first be established whether a case should be made for a higher level of classification.

For risk assessment, all the substances are then regarded as genotoxic carcinogens unless there is scientific evidence to the contrary. Ways that allow an assessment of the magnitude of potential

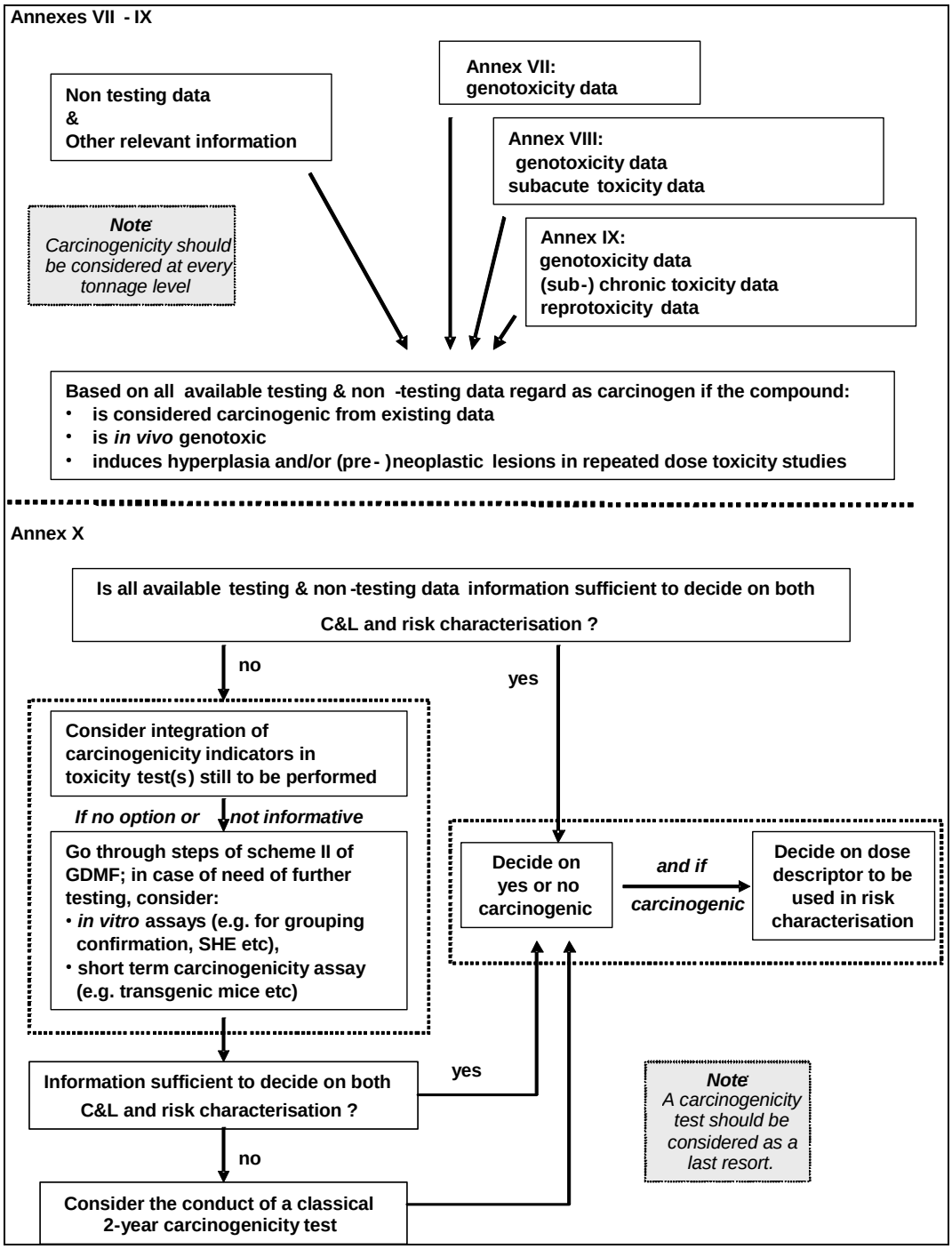
cancer risks associated with human exposures without performing the assay are indicated in Section R.7.7.13.2. (see Chapter R.8 for derivation of DMEL and DNEL values).

A carcinogenicity study may, on occasion, be justified. If there are clear suspicions that the substance may be carcinogenic, and available information (from both testing and non-testing data) are not conclusive in this, both in terms of hazard and potency, then the need for a carcinogenicity study should be explored. In particular, such a study may be required for substances with a widespread, dispersive use or for substances producing frequent or long-term human exposures. However, it should be considered only as a last resort.

It is noted, though, that some of non-genotoxic carcinogens, i.e. when not classified for any other property and not identified as potential carcinogens in (limited) repeated dose toxicity studies will go unidentified; this also regards the risks associated with human exposures.

If, in any case there is a need for further testing, the registrant must prepare and submit a well-considered test proposal (see Section R.7.7.6.2), and a time schedule for fulfilling the information requirements.

Figure R.7.7-2 Integrated Testing Strategy for carcinogenicity



R.7.7.14 References on carcinogenicity

Albertini, R., Bird, M., Doerrer, N., Needham, L., Robison, S., Sheldon, L. and Zenick, H. (2006). The use of biomonitoring data in exposure and human health risk assessments. Environ. Health Perspect. 114: 1755 – 1762.

Ames, BN, and Gold, LS (1990) Chemical carcinogenesis: too many rodent carcinogens. Proc.Natl.Acad.Sci.USA 87, 7772-7776.

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

Ashby J and Tennant RW (1988) Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. Mutat Res 204: 17-115.

Boffetta. P., van der Hel, O., Norppa, H., Fabianova, E., Fucic, A., Gundy, S., Lazutka, J., Cebulska-Wasilewska, A., Puskailerova, D., Znaor, A., Kelecsenyi, Z., Kurtinaitis, J., Forni, A., Vermeulen, R, and bonassie, S. (2007). Chromosomal aberrations and cancer risk: Results of a cohort study from Central Europe. Am. J. Epidemiol. 165: 36 – 43.

Boobis, A.R., Cophen, S.M., Dellarco, V., McGregor, D. Meek, M.E., Vickers, C., Willcocks, D. and Farland, W. (2006). IPCS framework for analysing the relevance of a cancer mode of action for humans, Crit. Rev. Toxicol. 36: 781 – 792.

Chen, H.J. and Chiu, W.L. (2005). Association between cigarette smoking and urinary excretion of 1,N2-ethenoguanine measured by isotope dilution liquid chromatography-electrospray ionization/tandem mass spectrometry. Chem. Res. Toxicol. 18: 1593 – 1599.

Cohen, S.M. and Ellwein, L.B. (1991). Genetic errors, cell proliferation and carcinogenesis. Cancer Res. 51: 6493 – 6505.

Cohen, S. M., Purtilo, D.T. and Ellwein, L.B. (1991). Pivotal role of increased cell proliferation in human carcinogenesis. Mod. Pathol. 4:371 – 375.

Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment (2004) Guidance on a Strategy for the Risk Assessment of Chemical Carcinogens.

Cronin M.T.D., Jaworska J.S., Walker J.D., Comber M.H.I, Watts C.D. & Worth A.P. (2003) Use of quantitative structure-activity relationships in international decision-making frameworks to predict health effects of chemical substances. Environmental Health Perspectives 22, 1391-1401.

Doe, J.E., Lewis, R.W. and Botham (2006). Comments on a scientific and animal welfare assessment of the OECD health effects test guidelines for the safety testing of chemicals under the European Union REACH system. Altern. Lab. Anim. 34: 111 – 114.

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

EFSA (European Food Safety Authority) (2004) Opinion of the Scientific Panel on Food Additives, Flavouring, Processing Aids and Materials in Contact with Food (AFC) on a Request from the Commission Related to Coumarin. Question Number EFSA-Q-2003-118. Adopted on 6 October 2004. The EFSA Journal 104, 1–36. Available from: <http://www.efsa.eu.int/science/afc/afc_opinions/726_en. html>.

Enzmann, H., Bomhard, E., Iatropoulos, M., Ahr, H.J. Schleuter, G. and Williams, G.M (1998). Short- and intermediate-term carinogencitiсy testing – a review. Food. Chem.. Toxicol. 36:979 – 995.

Elcombe CR, Odum J, Foster JR, Stone S, Hasmall S, Soames AR, Kimber I, and Ashby J (2002) Prediction of rodent non-genotoxic carcinogenesis: evaluation of biochemical and tissue changes in rodents following exposure to nine non-genotoxic NTP carcinogens. Env. Health Perspect.,110, 363-370.

Felter, SP, Vassallo, JD, Carlton, BD, and Daston, GP. (2006) A safety assessment of coumarin taking into account species-specificity of toxicokinetics. Fd. Chem. Toxicol., 44: 462–475.

Franke R, Gruska A, Giuliani A and Benigni R (2001) Prediction of rodent carcinogenicity of aromatic amines: a quantitative structure-activity relationships model. Carcinogenesis 22: 1561-1571.

Gold LS, Gaylor DW, and Slone TH (2003) Comparison of cancer risk estimates based on a variety of risk assessment methodologies. Regulatory Toxicology and Pharmacology 37, 45-53.

Haseman JK (1983) A re-examination of false-positive rates for carcinogenesis studies. Fundam. Appl.Toxicology 3, 334-339.

Hill AB (1965) The environment and disease: Association or causation? Proc R Soc Med 58: 295-300.

IARC (International Agency for Research on Cancer) (2006). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Preamble. Lyon. Available online: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf

IARC (International Agency for Research on Cancer) (1999). Species Differences in Thyroid, Kidney and Urinary Bladder Carcinogenesis. Capen CC, Dybing E, Rice JM & Wilbourn JD (Eds). IARC Scientific Publications 147, Lyon.

IARC (International Agency for Research on Cancer) (2006). Monograph 87: Inorganic and Organic Lead Compounds. IARC, Lyon.

ILSI/HESI (2001) Alternatives to carcinogenicity testing project. Toxicol. Pathol. 29 (supplement), 1- 351.

ILSI RSI (2003) A framework for human relevance analysis of information on carcinogenic modes of action. Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, Longfellow DG, Pastoor T, Seed J, Patton DE. Crit Rev in Toxicology, 33(6), 591-653.

Institute for Environment and Health (2002) Assessment of Chemical Carcinogens: Background to general principles of a weight of evidence approach, ISBN 1899110 37 2.

Jacobson-Kram, D, Sistare, FD, Jacobs, AC. (2004) Use of transgenic mice in carcinogenicity hazard assessment. Toxicologic Pathology, 32(Suppl. 1), 49-52.

Matthews, E.J., Contrera, J.F. (1998). A New Highly Specific Method for Predicting the Carcinogenic Potential of Pharmaceuticals in Rodents Using Enhanced MCASE QSAR-ES Software. Regulatory Toxicology and Pharmacology 28: 242-264.

Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., (1996). Thresholds of toxicological concern based on structure–activity relationships. Drug Metabolism Reviews 28, 209–217.

OECD (2002) Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies, OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 35 and Series on Pesticides No.14.

OECD (2006) draft OECD review paper on cell transformation assays for detection of cehmical carcinogens. OECD Environment, Health and Safety Publ.Serioes oj Testing and Asessment No.31, Paris (France).

Pritchard, JB, French, JE, Davis, BJ, Haseman, JK. 2003. The role of transgenic mouse models in carcinogen identification. Environmental Health Perspectives, 111(4), 444-454.

Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P, Grant D, Hartley M, Knaap A, Kroese D, Mangelsdorf I, Meek E, Rice JM & Younes M (2001) IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Reg Toxicol Pharmacol 34: 146–152.

Tennant, R.W., Stasiewicz, S., Mennear, J., French, J.E. and Spalding, J.W. (1999). Genetically altered mouse models for identifying carcinogens. In: McGregor, D.B., Rice, J.M., Venitt, S. (eds.). The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. Lyon, France: International Agency for Research on Cancer.

US FDA (1986) General principles for evaluating the safety of compounds used in foodproducing animals. Appendix 1. Carcinogen structure guide. Washington: U.S.Food and Drug Administration.