Guidance on information requirements and chemical safety assessment

Chapter R.7b: Endpoint specific guidance

May 2008
(version 1.1)

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/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).

DOCUMENT HISTORY

Version	Comment	Date
Version 1	First edition	May 2008
Version 1.1	Re-introduction of lost pieces of Appendix 7.8-5 "Assessment of available information on endocrine and other related effects"	August 2008

Convention for citing the REACH regulation

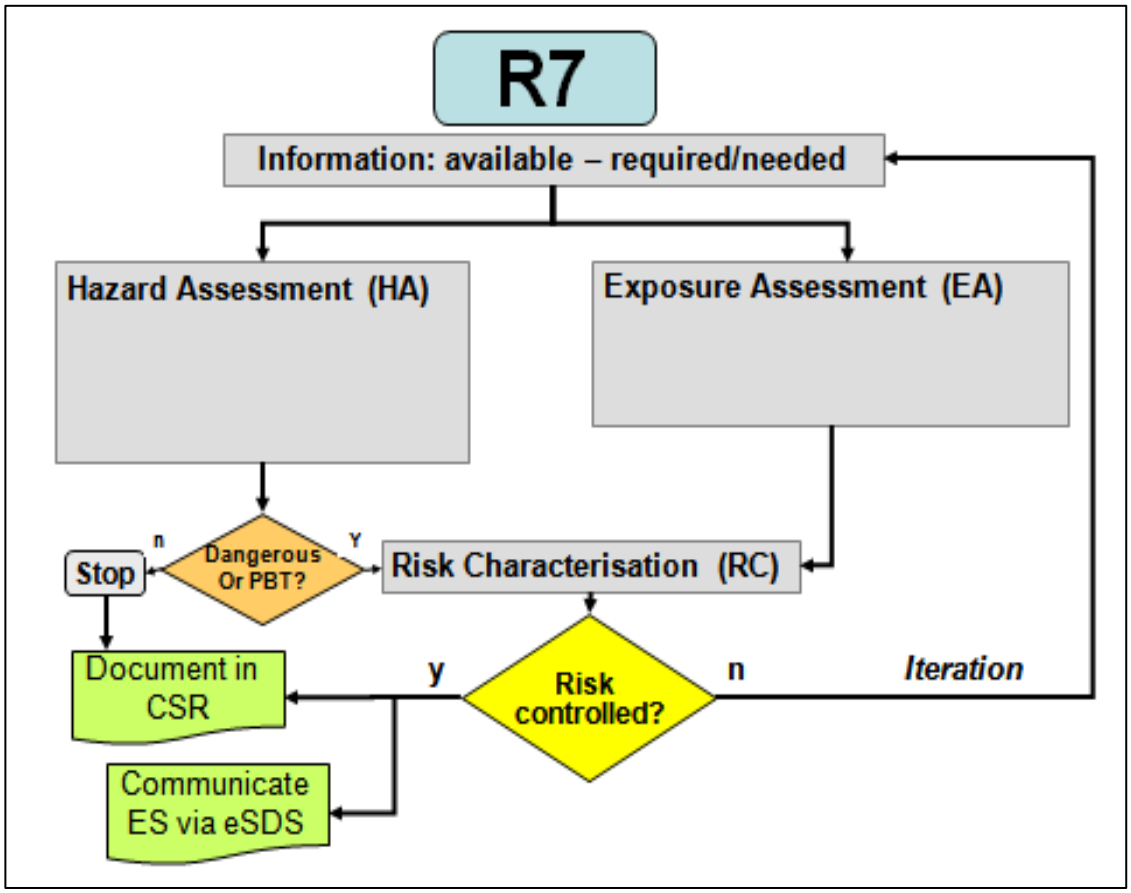
Where the REACH regulation is cited literally, this is indicated by text in italics between quotes.

Table of Terms and Abbreviations

See Chapter R.20

Pathfinder

The figure below indicates the location of part R.7b within the Guidance Document



CONTENTS

FIGURES

TABLES

APPENDICES

R.7.8 Aquatic toxicity; long-term toxicity to sediment organisms

R.7.8.1 Introduction to Aquatic pelagic toxicity

Information on aquatic toxicity is used to assess hazard and risk to freshwater and marine organisms living in the water column. In addition, the data obtained from testing on freshwater species may also serve as basis for assessment of effects in marine environment as well as for extrapolation of the measured effects to other compartments within the aquatic ecosystem (e.g. sediment) and soil.

Related endpoints are (i) mammalian long-term/reproductive toxicity, where information on endocrine activity obtained in toxicological studies may also be relevant for fish and (ii) degradation, where information on possible (fast) primary degradation would lead to inclusion of metabolites in hazard assessment of the parent compound.

R.7.8.1.1 Definition of aquatic pelagic toxicity

Aquatic toxicity refers to intrinsic property of a substance to be detrimental to an organism in short-term and/or long-term exposure to that substance.

In general, it is assumed that the aquatic toxicity is mainly related to the waterborne exposure of a substance and expressed as external concentration of that substance in test water. There may be cases where food uptake is the predominant route of exposure (i.e. for lipophilic substances). These effects are measured by employment of dietary studies.

Some attempts have been made to relate toxic effects to internal concentration of substances in the exposed organisms, e.g. by using body burden approach. This approach has to be further developed and verified/validated before its application for regulatory purposes (for details see Section R.7.8.10).

Acute toxicity related to waterborne exposure is generally expressed in terms of a concentration which is lethal to 50% of the test organisms (lethal concentration, LC_{50}), causes a measurable adverse effect to 50% of the test organisms (e.g. immobilization of daphnids), or leads to a 50% reduction in test (treated) organism responses from control (untreated) organism responses (e.g. growth rate in algae) following an exposure in the range of hours to days, expressed as effective concentration, EC_{50}.

Chronic toxicity related to waterborne exposure refers to the potential or actual properties of a substance to cause adverse effects to aquatic organisms during exposures which are determined in relation to the life-cycle of the organism. Such chronic effects usually include a range of sublethal endpoints and are generally expressed in terms of NOEC (No Observed Effect Concentration), LOEC (Lowest Observed Effect Concentration), ECx or MATC (Maximal Acceptable Toxicant Concentration). Further guidance on these terms is given in Chapter R.10.

Observable endpoints in chronic studies typically include survival, growth and/or reproduction. Chronic toxicity exposure durations can vary widely depending on test endpoint measured and test species used.

Although data from standard toxicity tests (internationally harmonised test guidelines) are preferred, adverse effects in the water environment may also be predicted from other information sources.

R.7.8.1.2 Objective of the guidance on aquatic pelagic toxicity

The main objective is to provide guidance to registrants on aquatic pelagic toxicity testing and to develop an Integrated Testing Strategy (ITS) for aquatic toxicity aiming at gathering data and information on substances to enable the environmental hazard assessment, i.e. for use in classification and labelling and derivation of the PNECwater (Predicted No Effect Concentration for water) and for determination of the toxicity (T) criterion in the PBT assessment.. The PNECwater is compared with the Predicted Environmental Concentration in water (PECwater) to decide whether there is a risk or not to pelagic organisms from the exposure to the substance.

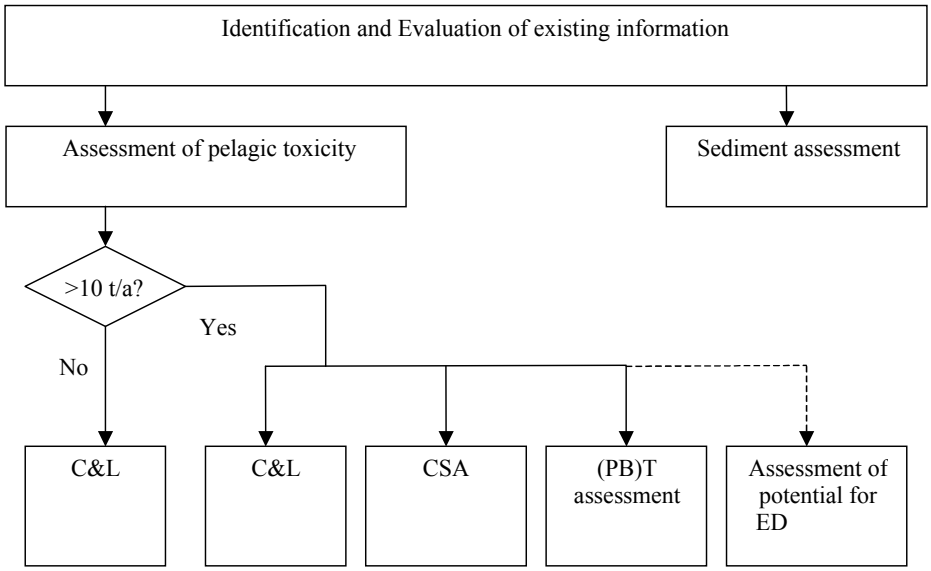
Depending on the intrinsic properties of the substance and available exposure information, examination of additional possible adverse effects relevant for the aquatic ecosystem could be necessary:

- Substances that are potentially capable of depositing on or sorbing to sediments to a significant extent have to be assessed for *toxicity to sediment-dwelling organisms*. In addition, marine sediment effects assessment is necessary for substances that are known to be persistent in marine waters and may accumulate in sediments over time. Guidance for the assessment of toxic effects on sediment organisms is provided in Section R.7.8.7.

- In addition, if, in the course of evaluation of available information, it is confirmed or indicated that a substance displays an *endocrine mode of action* in aquatic organisms, this may constitute a concern that requires further investigation regarding potential adverse effects on development or reproduction. If a clear link between serious adverse effects and an endocrine mode of action can be established, the substance may fall under the provisions of Article 57(f), which specifies that *substances - such as those having endocrine disrupting properties (…) – for which there is scientific evidence of probable serious effects to human health or the environment which give rise to an equivalent level of concern* to those of CMR, PBT or vPvB substances may be included in Annex XIV of substances subject to the authorisation procedure. The inclusion will be decided on a case-by-case basis following the preparation of an Annex XV dossier by the Competent Authorities. As this kind of information is not part of the standard information requirements set out in REACH Annexes VII-X (see below), this part of the guidance is based on the evaluation of available information. Guidance for the evaluation of available information on endocrine activity is provided in Section R.7.8.11.

Figure R. 7.8-1 summarises the general regulatory steps that are relevant for aquatic toxicity. It starts with the evaluation of existing information and, based on this information a conclusion whether evaluation of waterborne exposure is sufficient or evaluation of toxicity to sediment dwelling organisms should be included. As a second step in the hazard assessment has to be performed the classification and labelling (C&L) (for substances manufactured/imported at less than 10 tonnes per year and more than 10 tonnes per year) and the determination of the. $PNEC_{water}$ in the frame of the Chemical Safety Assessment (CSA) (for substances manufactures/imported at ≥10 t/y) as well as for PBT assessment. Guidance for gathering of and evaluation of information for these steps is provided in this document. The guidance for the evaluation of sediment toxicity is provided in a separate document. If, based on available information, a substance is suspected to exhibit endocrine activity, it might be necessary to assess the endocrine disruption potential of the substance. Guidance for this step is provided in Section R.7.8.11 of this document.

Figure R. 7.8-1: Regulatory steps relevant for aquatic toxicity



R.7.8.2 Information requirements for aquatic pelagic toxicity

As described in **Annex VI** all available existing information should be collected and considered in the hazard assessment, regardless whether testing for a given endpoint is required or not at a specific tonnage level. Minimum information requirements are set out in Annex VII- X. If information required in Annex VII- X is not available, testing is required unless modification according to general rules described in Annex XI is possible. If the test needed concerns Annex IX or X a testing proposal has to be prepared and submitted to the Agency. Further information on general rules described in Annex XI is provided in Chapter R.5 and Section R.7.8.3. The following paragraphs summarise requirements according to Annex VII–X.

For substances covered by **Annex VII** short-term toxicity testing on invertebrates (preferably Daphnia) and growth inhibition study on aquatic plants (preferably algae) are required. However, these short-term studies do not need to be conducted if there are mitigating factors indicating that aquatic toxicity is unlikely to occur (e.g. the substance is highly insoluble in water or the substance is unlikely to cross biological membranes).

In addition, the short-term testing on invertebrates does not need to be conducted if a long-term aquatic toxicity study on invertebrates is available or if adequate information on environmental classification and labeling is available.

If the substance is poorly water soluble the long-term toxicity testing (according to Annex IX) shall be considered (For more detailed description of potentially mitigating factors see Section R.7.8.7, for interpretation Section R.7.8.5).

For substances covered by **Annex VIII** short-term toxicity testing on fish is additionally required. In analogy to the tests required on Annex VII, this test does not need to be conducted if there are mitigating factors indicating that aquatic toxicity is unlikely to occur (e.g. the substance is highly insoluble in water or the substance is unlikely to cross biological membranes).

However, if the chemical safety assessment according to Annex I indicates the need to investigate further effects on aquatic organisms, long-term testing as described in Annex IX shall be considered. Long-term testing should also be considered if the substance is poorly water soluble. For explanation and interpretation see Section R.7.8.4.3 on exposure considerations.

For substances covered by **Annex IX** long-term toxicity testing on invertebrates (preferably Daphnia) and fish is required, if the chemical safety assessment according to Annex I indicates the need to investigate further the effects on aquatic organisms.

In case of the long-term toxicity testing on fish, information on one of the following studies shall be provided: (for explanation see Section R.7.8.5 on suitability of data on CSA).

- Fish Early Life Stage (FELS) toxicity test

- Fish short-term toxicity test on embryo and sac-fry stages

- Fish, juvenile growth test

For substances covered by **Annex X** there are no additional information requirements for pelagic aquatic toxicity.

As stated above the data are generated for environmental hazard assessment of substances (i.e. classification, derivation of PNEC) and (PB)T assessment (see Section R.7.8.5 on conclusion on the endpoint).

R.7.8.3 Information on aquatic pelagic toxicity and its sources

Below different types of information relevant for assessing aquatic toxicity are presented. This includes available testing (*in vitro* and *in vivo*) and non-testing methods ((Q)SAR, read-across and categories) that generate information on aquatic toxicity relevant for regulatory purposes.

R.7.8.3.1 Data on aquatic pelagic toxicity

Testing data on aquatic pelagic toxicity

IN VITRO DATA

At present, there are no EU / OECD guidelines for *in vitro* tests of relevance to aquatic toxicity.

There are ongoing efforts to develop and validate *in vitro* methods, which in future might be useful in a testing strategy for acute aquatic toxicity (e.g. ECVAM study on optimisation of cytotoxicity tests and CEFIC LRi study ECO 8 aiming to replacing the acute fish toxicity test using fish cell lines and fish embryos).

The use of fish cells in environmental toxicology was recently reviewed at the ECVAM workshop (Castano et al 2003, ECVAM workshop report 47) and ECETOC (2005).

Primary cells: Primary cells are freshly isolated cells from various tissues: liver, gill epithelia, gonads, kidney macrophages, skin epithelia, endocrine tissues, muscle cells and white blood cells. Primary cells require the use of living animals. They express many of the differentiated cellular structures and functions of their source tissue and are particularly suitable for mechanistically oriented studies on cell-specific toxicant fate and action.

Fish cell lines: More than 150 permanent fish cell lines are available, most of them are fibroblast or epithelia-like and derive from tissue of salmonids and cyprinids. Most of the tests with permanent cell lines (monolayers or suspension cultures) measure the basal cytotoxic effects of chemical substances.

Results from *in vitro* studies based on mammalian systems may be of interest for the assessment of endocrine activity (see R.7.8.11).

IN VIVO DATA (SINGLE SPECIES)

Information on aquatic toxicity may be acquired from studies performed according to existing national and international guidelines as well as from scientific literature, where different aspects of aquatic toxicity are examined. The available guidelines are focused on measuring of adverse effects of substances due to waterborne exposure. Since there are no internationally harmonised guidelines for feeding studies in pelagic species, tests employed in assessment of oral exposure are designed on case-by-case basis.

In general, the majority of the test guidelines for pelagic system are exclusively developed for testing of either freshwater or saltwater species. There are, however, guidelines providing procedures that are suitable for testing of species from both water systems (see Tables in Section R.7.8.8).

EU/OECD Test guidelines

The EU/OECD test guidelines comprise internationally agreed testing methods for environmental effects. Tests undertaken using these guidelines are useful for both risk assessment and classification purposes. Data obtained from a test carried out in accordance with an OECD test guideline are covered by the principle of mutual acceptance of data (MAD), thereby reducing the number of tests that needs to be conducted saving both animals and money.

There are a number of the tests guidelines available. They provide information on short-term and long-term toxicity to aquatic species (both freshwater and marine) due to waterborne exposure. Several new test methods, including potential alternative methods to vertebrate animal testing, are currently under development and validation. Both the available tests guidelines and these under development are presented in Section R.7.8.8.

The information requirements of REACH are, in principle, met by studies carried out according to the currently adopted OECD test guidelines. However, if required by further evaluation, additional (more adequate) tests (e.g. on organisms not included in OECD test guidelines) may be selected from the lists of guidelines developed by other regulatory bodies (see Section R.7.8.8[2]).

Other test guidelines

Acceptable alternatives to the OECD test guidelines are published by the OPPTS, US-EPA, various EU countries (national standard methods) and organisations such as ASTM, ISO (for detailed list of available guidelines see R.7.8.8).

[2] Following development in the field of eco-toxicology new test guidelines are developed and available test methods undergo changes. Their procedures may be revised or some of the guidelines may even be exchanged by other, better tests. Therefore every table that aims at compiling all available test guidelines will soon become obsolete. The table in Appendix III gives the status from 1998 (OECD 1998) Therefore, the user is advised to consult the organisation that has issued the selected guidelines for its current status (addresses to the organisations are also presented in chapter R.0).

Non-guideline studies

In addition to results from guideline studies, also results from non-guideline non-GLP studies may be available. The studies may vary in duration, endpoints measured; species exposed etc. compared to the standard test guidelines. Despite the variability in the test performance the results may be useful for hazard assessment (e.g. direct in calculation of PNEC or indirect in application of *Weight of Evidence*). However, these data should be particularly assessed for their adequacy (reliability and relevance) and completeness (for details see Section R.7.8.4.1 on criteria for the evaluation of *in vivo* testing data).

Information sources

Data from different tests measuring toxicity to aquatic species (results from tests performed according to the test guidelines and to non-standard procedures) may be gathered in different databases. Not all databases routinely make a quality check of the data before their inclusion in the database. Unless the data quality is known user is recommended to consult original scientific paper where these data were derived. Aquatic toxicity data may also be reviewed in scientific reports. References to these databases and documents are presented in Section R.7.8.8.

IN VIVO – MULTIPLE SPECIES (FIELD DATA)

Experimental ecosystem studies are aiming at understanding both fate and effects at higher tiers of ecological integration. The design of any study is dependent on the objectives and includes:

- to gain more knowledge about ecosystem structure and function (and thus help to develop better ecosystem models);
- to develop and validate predictive models for chemical effect; with enough information about the chemical fate in the particular experimental ecosystem to be able to define NOECs, ECx or effect levels at different loading rates;
- to evaluate environmental quality standards derived from laboratory toxicity data through extrapolation (improvement and refinement of extrapolation models);
- to study the resilience of ecosystems in terms of time required for restoration after chemical disturbance; and,
- to obtain data required for regulatory purposes of assessing fate and/or effects in natural ecosystems (Crossland et al 1992).

Because different objectives exist for conducting model ecosystem tests, not all test results may be equally useful, especially with respect to regulatory purposes.

Numerous expert meetings concerning the development and design of experimental ecosystem studies involving all stakeholders have been held over the past 20 years. An OECD guidance for the conduct of simulated freshwater lentic (standing water) tests in the form of outdoor microcosms and mesocosms is available (OECD 2006a).

The choice of endpoints to measure during an experimental ecosystem study should not be exhaustive and preferably targeted based on knowledge developed from lower tiers of fate and effects assessment.

However, because experimental ecosystems offer the advantage of addressing ecological properties that cannot be considered in lower tiers (and inherently addressed in subsequent PNEC extrapolation), such as species diversity, trophic structure, species interactions and so on, these may be useful to consider when designing, conducting and interpreting a study (OECD 2006a).

Non-testing data on aquatic pelagic toxicity

A general guidance on the use of (Q)SAR results and chemical grouping approaches is given in Sections R.6.1 and R.6.2. The following section provides an overview of different information sources for (Q)SAR predictions and grouping approaches specific for the assessment of aquatic toxicity. Additional, more generic sources of information are summarised in Chapter R.4. Guidance for the evaluation of the results of these approaches is provided in Section R.7.8.4.1.

(Q)SAR

General guidance on QSAR is given in Section R.6.1 and a more specific guidance on QSAR for estimating for toxicity to the environment is given in Chapter R.10.

Available (Q)SAR methods can be summarised using the following categories:

- Schemes for the prediction of the mode of action/structural class of a compound (baseline toxicity, excess toxicity)
- Qualitative information from structural alerts
- QSARs predictions from individual models (e.g. narcosis, other modes of action, QICARs and QCARs for metals and inorganic metal compounds)
- QSARs predictions from expert systems
- Databases of (Q)SAR predictions
- Activity-activity relationships (QAARs) predictions

GROUPING APPROACHES

General guidance on grouping approaches is given in Section R.6.2 and a more specific guidance on QSAR for estimating for toxicity to the environment is given in Chapter R.10.

R.7.8.4 Evaluation of available information on aquatic pelagic toxicity

Below criteria for evaluation of the gathered information are presented. Integration of the gathered information should lead to an understanding of the toxic profile of the substance, its potential exposure routes, its mechanism of action and its potential for distribution in the environment.

Toxic effects of substances in the aquatic environment are among others related to (i) intrinsic physical and chemical properties of substances and (ii) physical and chemical properties of the aquatic (tests) systems. These two information have to be taken into account when evaluating the available information on aquatic pelagic toxicity.

Properties of substances and of test systems

For most organic chemicals uptake from water is believed to be the predominant route of uptake (for very hydrophobic or very sorptive substances does uptake from food become important). It is believed that substances dissolved in water and taken up by organisms may accumulate to a certain internal concentration, which may then cause adverse effects. Therefore factors that influence bioconcentration influence also toxicity to aquatic species. Molecular weight, water solubility and log K_{ow} of substances are such factors. They are described in detail in Section R.7.8.7. In addition other substance related factors like degradation are described in this chapter.

In the context of toxicity, properties of aquatic (test) systems may or may not create optimal conditions for recording possible adverse effects. Therefore they are important quality parameters to be taken into account while evaluating toxicity studies. The water quality parameters that influence toxicity testing are also described in Section R.7.8.7.

For metals and inorganic metal compounds exposure through the water is also the predominant route. For many metals bioavailability and detoxification mechanisms is known to modulate both accumulation and toxicity (McGeer et al, 2002).

The criteria for evaluation of information on the physico-chemical properties of substances are provided in Section R.7.1.Furthermore consideration should be given to whether the substance being assessed can be degraded, biotically or abiotically, to give stable and/or toxic degradation products. Where such degradation can occur, the assessment should give due consideration to the properties (including toxic effects) of the products that might arise.

Other considerations

Information on exposure must also be taken into account when deciding on the aquatic pelagic tests to perform. Before their use the exposure data should be validated in respect to their representativeness, completeness, relevance and reliability.

For existing data evaluation it is common that the full study information will not be available to fully assess in detail all of the considerations above. The study may be of good quality, however, and the study result can still be considered for use as part of a *Weight of Evidence*. Under these circumstances, key information should be available to give some confidence that the underlying data are of good quality. Where such circumstances exist it is critical to know that the test has been carried out to standardised test guidelines. The study method should be reported. In addition key study information should also be provided in the technical dossier (further guidance is given in the Section 8 of the guidance on registration). These are 1) test substance identification, 2) sample purity, 3) test species and 4) test duration. Without this information and in the absence of other key study information or other studies for the same endpoint it is extremely difficult to justify use of that particular study result on its own. The study may be used in combination with other data as part of a *Weight of Evidence* approach (see Section R.4.4)

Other programmes/ secondary sources of data

There are also circumstances where reported values have already been through a screening process such as the SIDS program or through an EU existing substances risk assessment (http://ecb.jrc.it/esis/). In such circumstance the data may be considered sufficiently reviewed as to not require further evaluation assuming that the problems have been highlighted with the study(ies) of interest. Data reported as part of other equivalent peer reviewed risk assessment programs (e.g. HERA (http://www.heraproject.com/); US-EPA HPVC Challenge Programme) may also be considered in this way although a level of expert judgement is required to evaluate the quality of these programmes and further justification in the use of such a programme data may be required.

R.7.8.4.1 Data on aquatic pelagic toxicity

Testing data on aquatic pelagic toxicity

In vitro data

Although the extrapolation of *in vitro* data to *in vivo* data is discussed in literature further research in this area is needed (ECETOC, 2005) and there is currently not enough information available to give guidance for the extrapolation from *in vitro* data to *in vivo* data. Various publications show that, for the correlation with *in vivo* results the *in vitro* bioavailability of the substances tested should be considered (Guelden and Seibert 2005; Bernard and Dyer 2005; Schirmer 2006).

Currently, there are no validated fish cell systems available. Nevertheless, information from *in vitro* studies might be considered in a *Weight of Evidence* approach provided that they fulfil certain data quality aspects and comply with the Annex XI criteria.

Annex XI states that *suitable in vitro* methods should be well developed and fulfil certain criteria, e.g. the ECVAM criteria to enter a pre-validation study (Curren et al, 1995). Based on these, the following information on the study/method would be useful:

- the source of data should be named (e.g. publication, study report, in-house data, interlaboratory study)
- fish cell system:
  - primary cells (tissue used for isolation)
  - fish cell line and if available passage number
  - for both, culture conditions (e.g. medium, serum, serum-free)
- protocol used (e.g. incubation temperature, exposure time, replicants, endpoint measured, positive and negative controls, data analysis and interpretation, limitations, etc)
- status of standardisation of protocol
  - in house validated (evidence of repeatability)
  - used in other labs (evidence of reproducibility)

- nominal or measured concentration
- comparison to other *in vitro* / *in vivo* tests
- data on other substances tested with the method

Primary cells are more suitable to evaluate specific toxic effect, e.g. isolated hepatocytes for liver toxicity, metabolism or isolated gill epithelia for effects on the gill barrier function, toxicant uptake and metabolism. However they require the use of living animals. Cytotoxicity tests using fish cell lines are more likely to indicate acute toxic effects although it is necessary to consider that they might lack of realistic toxicokinetics including metabolism

The ongoing standardisation and validation efforts might provide validated methods which will then be included into testing strategies.

In vivo data (single species)

INITIAL RELIABILITY SCREENING

An initial review of the reliability of data should be made in order to filter out the most reliable values for consideration. For many existing substances the test data available will have been generated prior to the establishment of standard protocols and GLP. To address the potential variability in data quality in older data collections, there are various possible approaches. These include methods such as those employed by the OECD (2000a), U.S. EPA (2002), Hobbs et al. (2005) or the recommendations of Klimisch et al. (1997) which are introduced and described in Chapter R.4 of this guidance document. Further data on structurally similar substances may be available and these may add to the toxicity or ecotoxicity profile of the substance under investigation.

Klimisch et al. (1997) describe the parameters that need to be considered to evaluate the quality of a non-standard test. However, the authors do not describe the expert judgement process by which the strengths and weaknesses in the reporting of these different parameters are integrated to determine an overall quality assessment. To address this limitation, the following set of quality criteria, which are a development of Klimisch et al (1997), should be considered (see below for further details):

- Description of the test substance.
- Description of the test procedure including exposure period.
- Data on the test species and the number of individuals tested.
- Description of measured parameters, observations, endpoints.
- Control data available and acceptable according to guidelines. For some species used in environmental toxicity tests, guidelines are not available and in this instance, the guideline for the taxonomically closest equivalent species should be used.
- A concentration-response has been established, except in the case of limit tests determining a NOEC/ECx.
- Achieved exposure concentrations were measured in the test medium or vehicle. For aquatic toxicity tests, measurements should be made at least at t_0 and t_{end} and exposure should be calculated in terms of geometric mean measured concentrations unless measured concentrations were within 20% of the nominal concentration, in which case the nominal concentrations may be used.

If available data do not conform to the quality standards, the data should be reconsidered, to determine whether any of them are acceptable under current circumstances, and in particular, that they will not underestimate toxicity. For example, in an environmental toxicity test the data could have been rejected due to an absence of measured concentrations in the test media, but for a test substance whose physical/chemical properties suggest a low potential for biodegradation / volatilisation / sorption, the data may be acceptable.

Irrespective of whether or not data meet the full set of quality criteria, consideration should be given as to whether the data:

- are outliers in a large data-set for a particular substance;
- fit with what is known of the toxicity of other related substances.

CHECKLIST

After an initial screen, a number of studies will be screened out on which to focus and a second stage of screening is likely to be necessary. In an ideal world this considers what is essentially a minimum set of criteria which should be met. The following considerations relate to the aquatic toxicity testing at this second screening:

Test substance/ test substance identification

It is important to be able to accurately identify the substance tested. This should include an adequate description of the test substance. Ideally this should include an internationally recognised identifier such as the CAS number. However, the CAS number is not always unique to a substance and so a chemical description may be sufficient as long as the description is sufficiently detailed to allow clear identification. For example, positioning of particular moieties around a ring structure can be important from an (eco)toxicity point of view so a description of dichloro- should be more clearly identified as 1,3-dichlor etc. A further example can be where the term alkyl is used when an exact chain length should be described.

It is critical to ensure that the test material which has been tested is actually consistent with the substance being registered. It may be for example that the material tested is a mixture of homologous chain lengths which are a different distribution to the CAS number being registered. This may be acceptable. However, this information should be clearly described and justified why such data can be used.

Chemical purity should be described and where possible identification of the impurity should be made. The impurity can be important can be responsible for the majority of observed toxicity of a sample even if it is present at low levels. There are cases where studies have been carried out on test materials which have included with them a component which is present intentionally (such as preservatives). In some cases these studies may have been carried out intentionally on this mix in order to replicate more closely the actual material used/ sold. This factor should be considered when assessing the data.

Water solubility should be reported ideally. Results which occur above the limit of water solubility should be considered in further detail – see Section R.7.8.7.

Test Organisms

Details of the taxonomic identity of the organisms used in the study should be described to include the genus and the species. In some cases the genus alone can be sufficient information where it is known that all members of that genus are of similar sensitivity.

Where studies are conducted to standard methodologies such as the OECD guidelines described earlier, often these have listed standard organisms for which the test method is relevant. Non-standard species can also be accepted. However, these should be properly identified and characterised in order to ensure that the test method is suitable.

Test setup
The test system should be adequately described and wherever possible the test should be in accordance with an internationally accepted guideline. Non-standard methods can be accepted but clear description of the methods should be made. If a non-standard method is described or a standard method is followed and a judgement on whether the method has been adhered to, then the following are to be considered:

Test procedures and conditions should be reported to include standard/recognized procedures, appropriate acclimation procedures followed, certain conditions noted (test temperature, dissolved oxygen levels, pH, lighting), and placement of test units to avoid position effects) etc.

Test duration. This is critical information in deciding reliability of a study and must be reported. These do vary by endpoint/ study. Key values have been described previously under Guideline Studies. Deviations from these will make comparison with results from other studies difficult even when these studies are of good quality (e.g. Daphnia sp EC50 results are commonly reported at 24 hours compared to the standard 48hours).

Deviations from standard guidelines. Where deviations are made from the standard guidelines these should be clearly described. Such studies will by default not be scored as reliability 1 under Klimisch. However, with clear documentation the studies may be classified as reliability 2. Without such descriptions the study may be scored as reliability 3 or 4, both of which would indicate less than favourable study results.

Route/Type of exposure. Delivery of the test substance is a critical factor to consider to ensure suitable exposure to the test organisms. For algae, static tests are common. For daphnia studies static or semi-static tests are common and for fish static, semi static and flow-through studies are common. The potential effect of any relevant phys-chem properties of the substance such as solubility, high adsorption, precipitation etc on delivery should also be documented.
In some studies food is added during the exposure period (e.g. green algae are added as food in a Daphnia reproduction test). In such cases exposure may also occur via food for substances that adsorb to the algae.
A description of the test medium and dilution water should be included to ensure that it is for example correctly made, of specified hardness and salinity range etc. Other relevant quality criteria should be included also as appropriate such as total organic carbon, un-ionized ammonia. Besides ensuring that all abiotic factors fall within the tolerance limits of the test organisms a proper description of other abiotic parameters, e.g. dissolved organic carbon concentration (DOC), cations and anions etc., that govern the speciation (i.e. availability) and subsequently may influence the uptake of certain chemicals. In particular influence of abiotic factors on the bioavailability of some metals and inorganic metal compounds have been studied and for certain of these chemicals correction for bioavailability is possible and relevant. The term bioavailability[3] is in the context of environmental risk assessment of metals used to describe both the availability of metals due to speciation phenomena (a part which is independent of the organism and where chemical speciation

[3] Bioavailability of metals: A metal is considered bioavailable when it is free for uptake by an organism and when it result in a toxicity response (Newman and Jagoe, 1994; Campbell et al., 1988). The main idea behind the concept of "bioavailability", is that the toxic effect of a metal does not only depend on the total (or dissolved) concentration of that metal in the surrounding environment, but also on the complex interaction between physico-chemical factors, the free metal ion considered and the biological ligand on which the metal binds and result in a toxic response of the exposed organism. In other words, the same total metal concentration does not result in the same degree of toxic effect on an organism under all environmental conditions.

models could be used as a first tier to reduce variability) and the real bioaccessibility part influenced by biological/physiological factors (e.g. competition effects as captured in Biotic Ligand Models).

Furthermore, in the case of testing essential metals and metal components a proper description of the culture conditions, specifically related to the level of essential metals and inorganic metal compounds added or already present in the culture media could give valuable insight on issues such as acclimation. The way how bioavailability can be taken account of in aquatic effects assessment for metals and inorganic metal compounds is further elaborated in the guidance on metals.

Test concentrations/dose levels and number of concentrations should be known and where possible evidence provided that concentrations have been maintained throughout the duration of the test. Therefore, measured concentrations are preferred over nominal (non-measured) concentrations. If measured concentration are <80% of nominal concentrations, effect values should be related to mean measured concentrations. For flow-trough studies the arithmetic mean of measured concentrations should be calculated, for static or semi-static tests the geometric mean of measured concentrations (see Section R.7.8.7). In some cases where only nominal concentrations are provided, expert judgement may be required to decide whether test concentrations are likely to have been maintained. Such circumstances may occur if:

- It is known that the material is abiotically and biotically stable (from e.g. stability in water/ biodegradation studies etc such as OECD 111, OECD 113, OECD 301A-F, OECD 310, OECD 302A-C) to conclude that the concentrations are likely to have been maintained during the study.
- The test substance is soluble, well below its limit of solubility,
- Is non volatile
- Has low adsorbance to either delivery apparatus or the exposure vessels

For metals and inorganic metal compounds there is a strong preference for using measured data because potential issues related to natural background, to analytical errors and to the limited solubility of some metals and inorganic metal compounds. If it is not mentioned whether the reported toxicity values are based on measured concentrations, they should be considered as nominal concentrations. In cases where no measured data are available the use of nominal concentrations could be considered. In artificial media, where the metal background concentration is often very low compared to the effects levels, nominal concentrations could usually be used as long as the tests are based on soluble metal salts. When natural waters are used instead of artificial test media there could be a concern with the use of nominal values when the derived NOEC/EC_{10} values are close to the reported background values of the natural water used as these concentrations could potentially contribute to the observed toxicity in a significant way and as result the use of a nominal values would overestimate toxicity.

However, it must be emphasized that most often information on metal background values in natural waters is not readily available Furthermore natural background concentrations for metals can vary substantially and can not easily be distinguished from anthropogenic metal concentrations. For sparingly soluble metals measured data on the dissolved fraction[4] are always required for getting reliable toxicity test data. If the solubility is exceeded the test result has to be considered as unreliable. Results from tests where a visual precipitation is observed should be discarded. The absence of a visual precipitation does not exclude that colloids may be present that could affect the test results. For more specific guidance see section on difficult substances in Section R.7.8.7.

In some cases studies will have been carried out with the use of solubilisers. In these circumstances

[4] Different definitions for the dissolved fraction exist. Most often the dissolved fraction in ecotoxicity tests refers to the fraction that passes through a filter of 0.45 µm. It should be noted, however, that this definition may not necessarily refer to the metals in solution. In the range of 0.01-0.45 µm colloid inert particles that remain suspended may exist.

it is important to consider the change in bioavailability of the test substance and also the potential impact of the solubiliser. Studies performed without solvents/solubilizers are preferred over studies with solvents. Solvent concentrations should be the same in all treatments and controls. Further guidance on the interpretation of studies performed with the use of solubilisers is given in OECD (2000c).
Where a reasonable estimation of the exposure concentration cannot be determined then the test result should be considered with caution unless as part of a *Weight of Evidence* approach.

Controls: All studies must have controls. If a solvent is used, also solvent controls are necessary.
Test endpoints and reported data. Confidence in the reliability of a study can be increased if dose-response or concentration-response is evident and some measure of data quality such as Good Laboratory Practices (GLP) is reported to have been followed. Where a test result is reported as a *less than* (<) value this cannot be used. Results reported as *greater than* (>) can be used as additional information and may in some cases be considered directly instead of a fully defined result. However, this result should be justified with considerations of the test set up and phys-chem properties etc which may influence the result.

Statistical analyses. Statistical methods for derivation of LC_{50}, EC_{50}, IC_{50}, NOEC values etc should be reported. Where possible these should be presented with relevant reliability criteria. However, in the absence of these a description of the method could be considered acceptable.

Test design: Studies should be designed to enable sufficient statistical differences to be established between controls and test ingredient solutions. Further guidance on number of replicates, number of test organisms per replicate, number of concentrations necessary for a reliable ECx and/or NOEC/LOEC determination can be found in the different OECD test guidelines.

Hormesis effect: Hormesis has been observed for metal as well as organic substances and has been related to enhanced performance at low levels of induced stress (=at lower test concentrations). In such cases it is indeed important to use the neutral control data as a reference or to use specific models designed to model hormesis phenomenons (Brain and Cousens, 1989, Van Ewijk and Hoekstra, 1993; Schabenberger et al., 1999; Cedergreen et al, 2005). The need to take the activating part into account when deriving an ECx should be considered when appropriate.
For metals and especially, essential metals, the observation of hormesis may however also indicate a metal deficiency of the control medium and this needs to be avoided (see - description of the test medium). The possibility of a hormesis effects, observed for essential nutrients, needs to be considered when evaluating the calculation of EC_{10} values beyond the lowest tested concentration.

GUIDANCE OF SPECIFIC TEST TYPES FOR FRESHWATER SPECIES

In the following practical guidance is given for the evaluation of data from non-standard ecotoxicity tests.

Evaluation of data from growth inhibition testing on algae, aquatic plants (OECD 201 (2006c), 221 (2006d) and other standard and non-standard tests):
Commonly used and favoured tested species are *Pseudokirchneriella subcapitata* (previously named *Selenastrum capricornutum*) *Scenedesmus subspicatus* and *Chlorella vulgaris*. All can be considered as equally accepted preferred species.
The algal test is a short-term test although it provides both acute and chronic endpoints. The preferred observational endpoint in this study is algal growth rate inhibition because it is not dependent on the test design, whereas biomass depends both on growth rate of the test species as well as test duration and other elements of test design.
Often both acute growth rate EC_{50} (ErC_{50}) and biomass (EbC_{50}) endpoints are reported however the latter should not be used. The reason is that direct use of the biomass concentration without logarithmic transformation cannot be applied to an analysis of results from a system in exponential growth. Where only the EbC_{50} is reported, but primary data are available, a re-analysis of the data should therefore be carried out to determine the ErC_{50}. Where other supporting data exist as part of a *Weight of Evidence* approach it may be possible to consider an EbC_{50} value if only this value is reported. However, if only an EbC_{50} is reported and no primary data are available, it should be considered to perform a new algae study to obtain a valid ErC_{50} and NOEC or ErC_{10} especially if algae are the most relevant species for the effects assessment.
The typical test duration for this study is 72 hours. However, 96 hours is also commonly reported. This should be used as an equally acceptable value. For existing substances often algae tests with a duration of >96 h are available. As it cannot be assumed that the algae are in the exponential growth phase during the whole exposure period, the result from such tests cannot be used, unless the available raw data show monotone exponential growth of the controls. This also applies to reported chronic NOEC values. Common examples of this are 7-day and 14-day reported values.
It is sometimes seen also when test was done according to standard test guidelines, that the exponential growth ceased in the control before the end of the test period. Likewise it may be seen that the validity criteria of the test were not fulfilled (pH increase etc.) or growth of the algae in the exposed concentrations was increased (due to e.g. loss of test substance from the test system) at the end of the test. In such cases only data from the part of the test where exponential growth occurs and the validity criteria for the controls are fulfilled, should be used. In many such cases this may be achieved by excluding data from the last test day from the calculation of ErC_{50} and NOEC or ErC_{10}.

Common problems associated with algal study measurements result from coloured test materials and those with particular particle size (see Section R.7.8.7).
The most commonly used vascular plants for aquatic toxicity tests are duckweeds (*Lemna gibba* and *Lemna minor*). The Lemna test is a short-term test although it provides both acute and sub-chronic endpoints. The tests last for up to 14 days and are performed in nutrient enriched media similar to that used for algae, but may be increased in strength. Test design can be static, semi-static or flow-through. Frond number is the primary measurement variable. Other additional measurement parameters are total frond area, dry weight/fresh weight. The ECx/NOEC should be related to growth rate.
Evaluation of data from short-term toxicity testing on invertebrates (OECD 202 (2004b) and other standard and non-standard tests):
In addition to *Daphnia magna, Daphnia pulex*, *Ceriodaphnia affinis* and *C. dubia* are commonly tested species. Overall, there is no significant difference in sensitivity of *D. magna* and *D pulex*. Good correlation has been reported between acute toxicities of all three species (ECETOC 2003c).

All these can be considered as equally accepted preferred species.
Acute tests with crustacea generally begin with first instar <24 hours old juveniles. If the test organisms used are >24 h old, their sensitivity might be lower and the test can be accepted only in conjunction with other available data.
For daphnids, a test duration of 48 hours is standard. However, 24 hour LC_{50} or EC_{50} values are often reported for this study. 24 hour values can have considerable variability in the repeatability of results and should not be compared to 48 hour values. The standard 48 hour reported values are favoured over 24 hour values for these reasons. 24 hour values should be considered only in the absence of good quality 48 hour values and in conjunction with other available date (non-testing, read-across, information on time-dependence of effects etc). For other crustacea, such as mysids or others, a duration of 96 hours is typical.
The observational endpoint for short-term invertebrate tests is immobilization (EC_{50}) as a surrogate to mortality as it is quite difficult to make a clear judgement on mortality. Immobilisation is defined as unresponsive to gentle prodding.
Studies are often conducted under semi-static conditions where test solutions are renewed at periods (usually after 24 hours) during the study. This helps to maintain test concentration during the duration of the study. These studies are preferable over those studies conducted under static conditions, when the test material is known to degrade rapidly (either biotically or abiotically) or where known test material properties could lead to reduced test solution concentration due to adsorption processes for example. Results from flow-through studies can also be used as long as test duration is as already described.
Often a NOEC is reported for this acute study. This value cannot be used as surrogate value for a chronic NOEC as reported from OECD guideline 211.

Evaluation of data from long-term toxicity testing on invertebrates (OECD 211 (1998b) and other standard and non-standard tests):
Chronic tests with crustacea also generally begin with first instar juveniles and continue through maturation and reproduction. At least 3 broods should be produced during the exposure period. For daphnids, 21 days is sufficient for maturation and the production of 3 broods. For mysids, 28 days is necessary while *Ceriodaphnia dubia* produces 3 broods within 7 d. Observational endpoints include time to first brood, number of offspring produced per female (reproduction), growth, and survival (lethality). Reproduction and lethality are the most sensitive endpoints. Where uncertainly arises from which endpoint to consider, the lowest reported value should be used. Due to the test duration there is higher potential for loss of test material concentration over the test period. Studies with analytical support are thus preferable where available. Where such data are not available, consideration of other properties which may lead to doubt over test material concentration should be made, where these data are available. In addition to solubility these would include biotic and abiotic degradation and adsorption potential of the test material (resulting in loss to test glassware/ feed etc).
Typically the 21 day study may report ECx/NOEC values for survival or reproductive endpoints. The lowest value should be used for establishing ECx/NOEC for reproduction although in practice the two endpoints results tend to be close to each other.

Evaluation of data from short-term toxicity testing on fish (OECD 203 (1992a) and other standard and non-standard tests):
A number of species are recommended for use across several OECD Test Guidelines. Section R.7.8.8 indicates commonly used recommended species from OECD Test guidelines 203: Fish, Acute Toxicity Test; 204 Fish, Prolonged Toxicity Test: 14-Day Study; 210: Fish, Early-life Stage Toxicity Test; 212: Fish, Short-term Toxicity Test on Embryo and Sac-fry Stages and 305:

bioconcentration: Flow-through Fish Test. These can be considered as equally accepted preferred species.

The differences in fish species sensitivity sometimes can be substantial. This can often be due to differences in toxicity of the test material rather than inherent differences in species sensitivity. Often substances with the highest toxicity also have the largest variation in toxicity to different species. Acute tests are generally performed with young juveniles 0.1-5 g in size for a period of 96 hours. Fish larger than this range are generally less sensitive.
Where values are reported with shorter test duration, these should be treated with caution and should be used only in conjunction with other data (non-testing), read-across etc. as exposure phases shorter than 96 h generally lead to higher effect values.
Care should be taken also when considering studies carried out where the test material is readily biodegradable and where the nominal test concentration is low (<10mg/l). In these cases there is high likelihood that test concentrations will be lower than nominal.
The observational endpoint in these tests is mortality (LC_{50}).
Studies are often conducted under semi-static or flow-through conditions where test solutions are renewed at periods (usually after 24 hours) or continuously during the study. This helps to maintain test concentration during the duration of the study. These studies are preferable over those studies conducted under static conditions, when the test material is known to degrade rapidly (either biotically or abiotically) or where known test material properties could lead to reduced test solution concentration due to adsorption processes for example.

Evaluation of data from long-term toxicity testing on fish (OECD 210, 212, 215 and other standard and non-standard tests):
Only such studies can be regarded as long-term fish test, in which sensitive life-stages (juveniles, eggs, larvae) are exposed. Thus, tests performed according to OECD 204 (Fish, Prolonged Toxicity Test: 14-Day Study (OECD 1984)) or similar guidelines cannot be considered suitable long-term tests. They are, in effect, prolonged acute studies with fish mortality as the major endpoint examined. The most relevant long-term fish tests are described below.

OECD Test Guideline 210 (1992b) Fish, Early-Life Stage (FELS) Toxicity Test:
For the test the following freshwater species are recommended Brachydanio rerio, Pimephales promelas, Oryzias latipes, and Oncorhynchus mykiss as well as saltwater Cypridon variegatus. Among the currently available standardised test methods, the FELS toxicity test is considered as the most sensitive of the fish tests. It covers several life stages of the fish from the newly fertilised egg, through hatch to early stages of growth and is also the only suitable test currently available for examining the potential toxic effects of bioaccumulation. The required test duration is species-dependent: 60 days post-hatch for rainbow trout or approximately 30 days for warm water fish. Observational endpoints include hatching success, survival and growth.
OECD Test Guideline 212 (1998a) Fish, Short-term Toxicity Test on Embryo and Sac-Fry Stages:

For the test the following freshwater species are recommended Danio rerio, Pimephales promelas, Cyprinus carpio, Oryzias latipes,, and Oncorhynchus mykiss. This test measures the sensitive early life stages from the newly fertilised egg to the end of the sac-fry stage. It is considerably shorter, and hence less expensive, than the FELS toxicity test but it is also considered less sensitive. The method offers an alternative to the FELS toxicity test for substances with log K_{ow} less than 4.
OECD Test Guideline 215 (2000b) Fish, Juvenile Growth test:
Oncorhynchus mykiss is recommended freshwater specie for the test, however also *Danio rerio* and

Oryzias latipes may be used. This test measures the growth of juvenile fish over a fixed period, and it is considered a sensitive indicator of toxicity. Although it is considered to be of insufficient duration to examine all the sensitive points in the fish life-cycle, it provides a shorter and less expensive option to the FELS test for substances of log K_{ow}<5.
Non-standard tests using similar methods can be accepted if the studies are well documented and comply with the guidelines in critical points (exposure duration, endpoints studied). Studies should be performed preferably under flow-through conditions or under appropriate semi-static conditions.

MARINE SPECIES

There are few standardised marine species protocols available (see Section R.7.8.8).
In general the same criteria as described for freshwater tests should be applied for the evaluation of the tests for marine species. Additional attention should be paid to the fact that the solubility of the substance might be influenced by the salinity (see Section R.7.8.7 for further detail).

DIFFICULT SUBSTANCES

A significant number of chemicals are described as 'difficult substances', which the OECD (2000c) class as difficult to test for the purpose of determining their aquatic toxicity. Typical characteristics of difficult substances include:

- Difficulty in maintaining substance concentration during the test, for example degradation in the test medium or loss of substance from media (e.g. absorption or evaporation)
- Difficulty in dissolving the substance, either due to poor solubility in test medium or a multi-component substance of varying solubility
- Difficulty in being able to measure substance concentration, due to problems in developing an analytical method or again multi-component substances

Such properties and the problems these cause for carrying out valid tests and their interpretation are described in Section R.7.8.7, and more fully in publications issued by the OECD and ECETOC (ECETOC 2003a). These also describe practical ways to deal with such issues. The possibility of a substance being difficult to test can often be determined from its physico-chemical properties such as water solubility, volatility, biodegradability, hydrolysis and photodegradability. This re-emphasises how important it is to know these parameters prior to new test being carried out, or before reviewing a test report.

In vivo – multiple species (field data)

Model ecosystems represent the highest experimental tier in the hazard and fate assessment processes. When tests are well-designed, the exposure of chemicals to environmental organisms can be directly related to the route applied in model ecosystem tests. The diversity of organisms and their interactions cannot be adequately modelled in simpler laboratory single species tests, therefore valuable information on fate and effect responses of biota can be gained. Test systems should contain sufficiently complex assemblages to address the objectives. In order to be useful for environmental protection, results should be statistically reliable and capable of identifying response patterns.

CONCEPTS OF DATA INTEGRATION AND STATISTICS

Conclusions developed from model ecosystem tests are based on expert judgment using a combination of univariate and multivariate statistical analyses of measured endpoints.

Explicit evaluation of model ecosystem data should be systematic. Combinations of both univariate and multivariate analyses are preferred if the measurements collected during the test are amenable to both. Effects observed through time, whether or not the effects are permanent or transitory, and the nature of the exposure-response relationship for important endpoints should be explored. OECD (2006a) provides reporting needs for standing water studies, but similar considerations exist for flowing water studies. These include information on the test substance, thorough description of the test system, experimental design and measured data, and how data were evaluated. As described in Section R.7.8.3.1, the actual reporting of a study will largely depend on the objectives of the work.

EVALUATION OF DATA

Mesocosms are not commonly employed for general chemicals partly because the dosing methods employed may not be representative of the way that these chemicals reach the environment (unlike pesticides which may reach ponds, ditches or rivers via drift or run-off). Another reason is without doubt that only for few industrial chemicals resources were available to conduct such higher tier expensive tests. In certain exceptional cases (notably down the drain chemicals) lotic mesocosm data may be most useful. However, if water concentrations can be maintained adequately and the mesocosm can be maintained long enough that sediments reach equilibrium concentrations, the results may be highly relevant in addition to laboratory tests on individual species.

Within the Existing Substance Regulation only for few substances results from mesocosm studies were available (e.g. metals such as zinc and cadmium, acrylamide, nonylphenol).

In summary, the main conclusions seem to have been that mesocosm data suffer from some of the following drawbacks:

- Observation intervals may be too long
- There can be overlap with other pollutants (e.g. metals) which makes interpretation difficult.
- Analytical inconsistencies may occur.
- There may be difficulties in maintaining exposure concentrations over prolonged periods and in confirming concentration (e.g. in relation to river flow rates).
- Some potentially sensitive life stages (e.g. larval stages), endpoints or species might not be included.
- Given the natural variation inherent in such test systems, very large changes in population abundance may have to occur for them to be statistically significant when compared to the variation in control populations.
- The number of endpoints measured may be insufficient to draw reliable conclusions, or a clear concentration-effect relationship may be lacking.

Non-testing data on aquatic pelagic toxicity

General guidance for the evaluation of non-testing data is provided in Chapter R.6 (cross-cutting guidance QSAR). The following section includes information specific for the evaluation of the reliability of non-testing data in aquatic toxicity.

EVALUATION OF QSAR RESULTS

As outlined in Section R.6.1, the evaluation of the reliability of a non-testing result includes two steps:

1. Evaluation of the validity of the model or expert system

The validity of a model should be assessed according to the OECD validation principles for QSARs (OECD 2004a). They can be used for the evaluation of expert systems respectively. An in depth interpretation of the OECD principles can be found in Worth et al. (2005) and in Chapter R.6 (cross-cutting guidance QSAR). Table R. 7.8-1 summarizes specific aspects for the assessment of aquatic toxicity endpoints.

Table R. 7.8-1: specific aquatic toxicity aspects of the OECD validity criteria

OECD Principle	Specific considerations for aquatic toxicity assessment
Principle 1: a defined endpoint	A defined endpoint is assumed if the QSAR model is based on experimental data with a) a single measured biological endpoint (eg. mortality of a specific fish species) b) comparable exposure conditions (e.g. exposure duration, same age of test organisms) and c) a single statistically derived endpoint (e.g. LC_{50})
Principle 2: an unambiguous algorithm	No specific considerations. Models based on linear regressions using $logK_{ow}$ as sole descriptor are considered to have an unambiguous algorithm. General considerations for the scientific validation of (Q)SAR models are described in Section R.6.1.3.
Principle 3: a defined domain of applicability	A defined domain of applicability can be based on a) definition of the descriptor domain of the model (i.e. range of log K_{ow} of the training set) b) definition of the structural domain of the model (e.g. description of fragments and functional groups covered by the model) c) definition of the mechanistic domain of the model
Principle 4: appropriate measures of goodness-of-fit, robustness and predictivity	No specific considerations for aquatic toxicity assessment. General considerations for the scientific validation of (Q)SAR models are described in Section R.6.1.3.
Principle 5: a mechanistic interpretation (if possible)	A mechanistic interpretation is possible if the QSAR model is based on chemicals assumed to have the same mode of action (e.g. models for polar or non-polar narcosis) or on chemical classes with a known mode of action (e.g. carbamates).

The outcome of the analysis might not be a simple yes/no answer and it might be impossible to conclude on the validity of the model without considering the regulatory context of the decision. However results of the analysis should be reported in a transparent way. Templates, so called QSAR model reporting formats (QMRFs) are provided in Section R.6.1.9.

2. Evaluation of the reliability of the outcome of a prediction

General guidance for the evaluation of model predictions is provided in Section R.6.1.3. The outcome of the assessment should be reported in detail. Templates, so called QSAR prediction reporting formats (QPRFs) are provided in Section R.6.1.10.

Evaluation of the outcome of schemes for the identification of modes of actions

Assessing the result of a prediction of a mode of action is mainly connected with an analysis of the possible short comes of the prediction with respect to the background (mechanistic domain) of the scheme. Some of the schemes include rules that focus on the identification of possible structural alerts/structural classes, while other focus on the active identification of chemicals acting via narcosis (e.g. Verhaar et al, 1992). Some information about the background of the different schemes is provided in Chapter R.10 (Appendix 1).

In general the following issues should be considered:

- Is the characterisation based on the identification of specific structural properties?
  E.g. was a substance identified as being narcotic because of its chemical structure or just because it does not fit to any of the classes described by the scheme?

- Is the chemical within the applicability domain of the characterisation scheme?
  E.g. does the chemical include substructures that are unknown by the schemes? This becomes increasingly important if the scheme is based on the identification of substructures that might be responsible for excess toxicity. If a substructure of the chemical is not known by the scheme, the scheme might not be able to assess if this substructure will create excess toxicity.

Evaluation of the outcome of a research for structural alerts

Structural alerts as described in Section R.7.8.3 and Section R10.2.2.2, indicate the presence of substructures that might increase the aquatic toxicity of the substance. Thus, if a structural alert was identified for a given substance, it can be assumed that the substance exhibits excess toxicity. On the other hand, the absence of a structural alert does not necessarily indicate the absence of excess toxicity since lists of structural alerts are not exhaustive. Thus results from a structural alert research can be used as a confirmation or evidence of excess toxicity only. It can not rule out other information if no alerts are identified. In order to assess the reliability of the structural alert research the same criteria as described above should be applied.

Evaluation of the outcome of a QSAR/QAAR prediction

Assessing the reliability of a QSAR/QAAR prediction for aquatic toxicity endpoints is mainly connected with the question whether the substance is within the predictive space of the model or not. Guidance for the assessment is provided in Section R.6.1. Additional information about the reliability can be achieved by comparing the mechanistic domain of the model with the assumed mode of action of the substance.

Evaluation of information derived by the grouping approach

The reliability of results obtained by grouping approaches highly depends on the selection of appropriate analogues and chemical classes. General guidance for the assessment of the reliability an applicability of grouping approaches is provided in Section R.6.2. With respect to aquatic toxicity the following additional aspect should be considered:

- Are substances used for the grouping approach that are comparable with respect to substructures (e.g. do they all contain/ not contain structural alerts)?
- Can a similar mode of action/structural class be assumed for all substances?
- Are the substances comparable with respect to physico-chemical properties that influence aquatic toxicity (e.g. comparable lipophilicity)
- Is the metabolic pathway of the substances comparable? E.g. specific attention should be paid to substances with methyl groups as the metabolic activation might differ from similar compounds that do not include methyl groups.

The selection of chemicals for read-across and chemical categories should be combined with a reliable documentation. Reporting formats are provided in Section R.6.2.6.

R.7.8.4.2 Remaining uncertainty for aquatic pelagic toxicity

For the pelagic compartment generally there are more tests available than for other environmental compartments. However, even for effect assessment on pelagic organisms there will nevertheless normally often remain substantial uncertainty in relation to estimating a concentration which will not affect structure and function of the pelagic ecosystem (PNEC).

Often a few monospecies laboratory tests on pelagic organisms are extrapolated to a PNEC value for the pelagic compartment which introduces uncertainty as it does not take more complex interactions in the ecosystem into account. When only acute tests have been performed, extrapolation of acute effect concentrations to chronic no effect concentrations also implies uncertainty because short term data have only limited predictive value for long term no effect concentrations (Ahlers et al., 2006).

The more chronic studies are available the more likely sensitive species are represented and hence the remaining is less. When the PEC/PNEC ratio is close to 1, it is preferable to have a robust database with as many as possible chronic data on pelagic species available, ideally including life cycle exposure.

The remaining uncertainty may in many cases be reduced when in an integrated assessment is being made taking all available information into account (e.g. including toxicity information on pelagic organisms from standard and non-standard tests, and taking into account results from alternative test methods and non-testing information).

R.7.8.4.3 Exposure considerations for aquatic pelagic toxicity requirements.

The information requirements for a substance as proposed by REACH may be modified based on information on exposure (i.e. triggering or waiving of further testing). This section considers triggering of further data requirements only (according to rules for adaptation of the standard information requirements, Column 2). For waiving the specific guidance on exposure based waiving should be consulted (Section R.5.1).In general, further testing is proposed if the CSA indicates the need to investigate further the effects on aquatic organisms, which implies long-term testing on fish and Daphnia for substances covered by Annex VIII and Annex IX. The need to conduct further testing may be triggered by the following cases, e.g.:

i. Results from a quantitative assessment, where PEC/PNEC>1;

ii. Results from a qualitative assessment, where a possible risk should be confirmed/rejected, e.g. when due to low water solubility of a substance, short term toxicity tests do not reveal any toxicity, long-term tests are performed;

iii. Information on a specific mode of action and unexpected sensitivity of a group of organisms to the substance under investigation;

iv. Monitoring data showing occurrence of a substance in the aquatic compartment.

If further tests are required, considerations provided in Section R.7.8.5 regarding the alternatives for vertebrate tests should be taken into account

R.7.8.5 Conclusions for aquatic pelagic toxicity and integrated testing strategy (ITS)

Section R.7.8.3 (information sources) presents an overview about the possibilities to collect available or generate new information of different kinds (*in vivo* testing, *in vitro* testing, non-testing). Section R.7.8.4 gives guidance how the adequacy, i.e. reliability and relevance, of every single piece of information from these different sources can be judged and ranked. Section R.7.8.5 is supposed to guide through the assessment of the toxicity of the substance in cases where the total amount of available information is suitable for regulatory decisions and in cases, where there are data gaps which have to be filled.

The overall purpose of REACH is to provide a high level of protection for man and the environment. To achieve this, the potential hazards associated with chemical substances must be evaluated and to this end, information about the intrinsic properties of each chemical is needed. At the same time, also according to the REACH regulation, vertebrate animal testing must be restricted to the necessary minimum. Column 1 of REACH Annexes VII–X specifies what is regarded as minimum information requirements. Column 2 of Annexes VII–X as well as Annex XI specify possibilities to modify these requirements. The prerequisite is the availability of other information that is a) equivalent to the results that would be obtained by standard testing and b) adequate for the three regulatory endpoints: Classification and Labelling, PBT assessment and Chemical Safety Assessment. The equivalence and adequacy will have to be substantiated by a *Weight of Evidence* approach, making best use of all existing information.

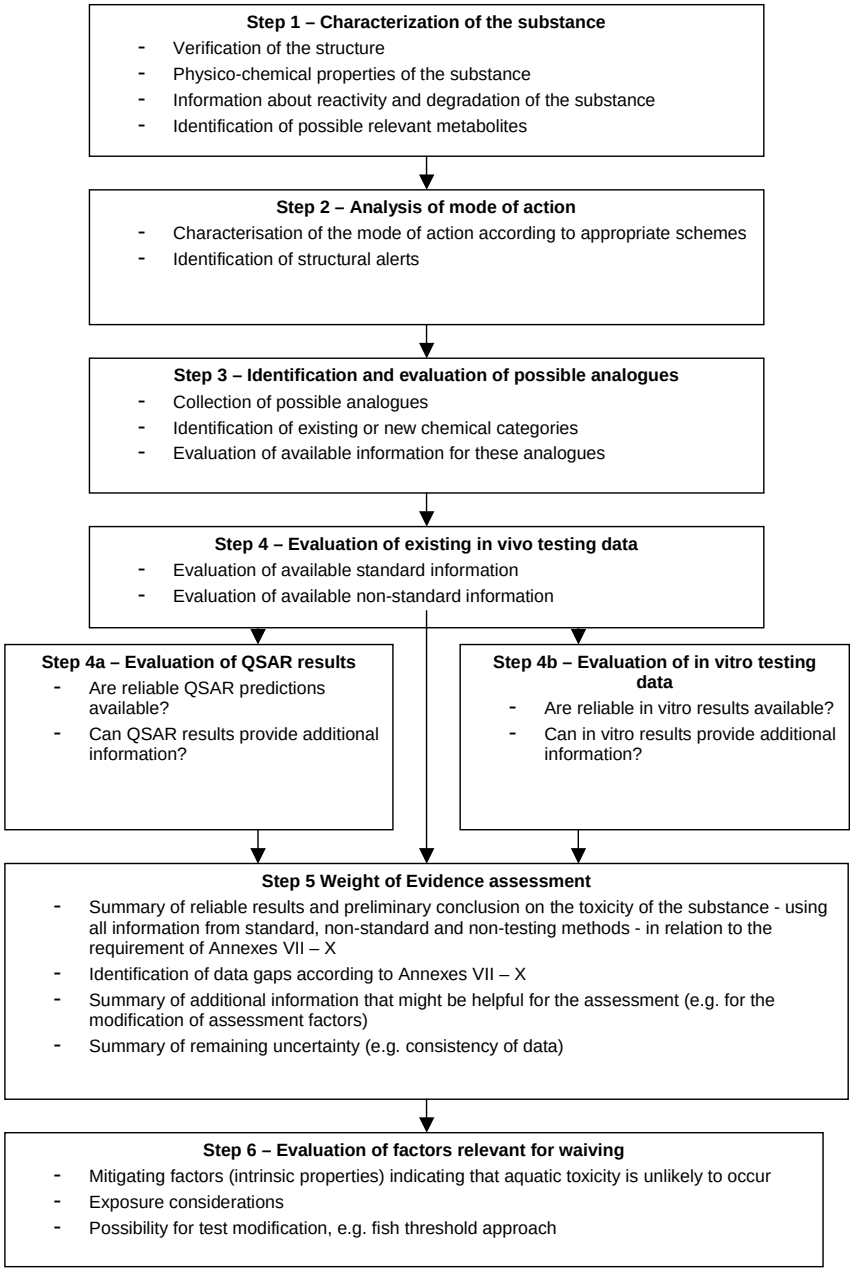
Weight of Evidence is closely linked to *Integrated Testing Strategies* (ITS,), in that the available evidence can help to determine the subsequent testing steps. Results from these subsequent tests affect the *Weight of Evidence*, which leads to a new decision on whether there is any need of further testing, and so on. ITS are particularly characterised by flexibility and case specificity. No general ITS can be developed but a case-by-case decision will always be necessary. Guidance on how to develop an individual ITS has to focus on decision making criteria and underlying considerations rather than on ready-to-use procedures.

Figure R. 7.8-2 outlines a systematic approach how to use all available data on a *Weight of Evidence* decision. It provides a step-wise procedure for the assessment of different types of information, which might be helpful to come to an overall conclusion. The scheme proposes a flexible sequence of steps, the order of which depends on the quality and quantity of data and might be changed, e.g. for a substance with available *in vivo* data of adequate quality, performance of steps 2, 3 and 4a and 4b might not be necessary. On the other hand, steps 2 and 3 might be particularly helpful in cases of varying data quality, and steps 4a and 4b in cases where not enough data are available. Step 1, which is a collection of information on physico-chemical properties

rather than an assessment of available information, is a prerequisite for the further assessment of other information. All steps are associated with three distinct activities: (i) the gathering of information (see detailed guidance in Section R.7.8.3), (ii) the evaluation of the quality of a distinct piece of information, e.g. a test report or a QSAR result (see detailed guidance in Section R.7.8.4), and finally (iii) the overall assessment of all available information, which will be the focus of this chapter. Additional guidance on generic aspects of a *Weight of Evidence* approach is provided in Chapter R.4.

Weight of Evidence is a decision making activity aiming at concluding on toxicity of a substance based on integration of information from different sources and various aspects of uncertainty. It will often require expert judgement. To make this expert judgement transparent and comprehensible it is essential that all information used, all steps carried out in the evaluation process and all conclusions drawn are fully documented and justified.

Figure R. 7.8-2: Suggestion for a *Weight of Evidence* approach



*****The scheme proposes a flexible sequence of steps, the order of which depends on the quality and quantity of data and might be changed.

STEP 1:

This step includes consideration of the following issues:

- Selection of the representative structure for the assessment (see Section R.6.1.7.3)

This step is essential for the assessment of the mode of action of a substance and for the potential use of non-testing techniques, e.g. QSAR models. In the case of multi-constituent substances (mixtures), it may be necessary to regard two or more structures, if a single representative structure is not considered sufficient.

- .

  - Preliminary analysis of uptake and fate

A preliminary assessment of expected uptake, toxicity, and fate is performed on the basis of the information collected so far, i.e. analysis of the chemical structure, chemical and physical properties, degradation pattern, abiotic and biotic reactions involving the parent compound and other information as available.

It is important to evavaluate at this stage the molecular structure and stability of the substance as well as identify the relevant metabolites. This is essential for the overall hazard assessment of a substance and especially for the evaluation of available in vivo tests (e.g. for the assessment if the test concentration was maintained during the test duration in cases where no analytical data are available) as well as for the use of QSAR results (in order to decide if the QSAR models should be used for a metabolite rather than the parent compound).

Further guidance is provided in Section R.6.1.7.4.

STEP 2:

As described in Section R.7.8.3 several schemes and programmes are available to derive information about the possible acute mode of action of a substance and to identify structural alerts. In Section R.7.8.4 some help for the evaluation of the outcome of these methods is provided. For the overall assessment of the mode of action, results are available in terms of QSAR prediction reporting formats (QPRFs). In addition, information about the existence of structural alerts will be available (for more guidance see Section R.7.8.4).

The overall assessment of the acute mode of action should take the following questions into account:

- Does the chemical contain structural alerts?
- Is the characterisation of different tools consistent with respect to the mode of action?
- If the results of different classification schemes differ, is there a reasonable explanation?
- Can additional information be derived from the results?

In many cases it will be difficult to detect a specific mode of action such as inhibition of photosynthesis. Therefore the evaluation should focus on the question whether the substance is likely to show baseline toxicity or if it is likely that it will exceed baseline toxicity. The answer to this question will be helpful for the evaluation of QSAR predictions as well as for the assessment of the reliability of experimental data and for the assessment of the relative species sensitivity. For the assessment the following considerations might be helpful:

STRUCTURAL ALERTS

The presence of a structural alert gives a strong indication, that the toxicity of the substance under investigation exceeds baseline toxicity with respect to the acute endpoint under investigation (e.g. acute fish toxicity). On the other hand the absence of a structural alert does not mean that the substance can be classified as baseline toxic.

CONSISTENCE OF DIFFERENT SCHEMES FOR THE CHARACTERISATION OF THE MODE OF ACTION

As outlined in Sections R.7.8.3 and R.7.8.4, the algorithm of different characterisation schemes and the outcome (identification of specific mode of actions or identification of excess toxicity) differs. Some advantages and disadvantages of the different schemes are outlined in Section R.7.8.4. With respect to the question if the substance shows baseline toxicity, different tools should be combined.

It can be assumed that the characterisation of a substance as being baseline toxic is reliable if different tools, based on different algorithms characterise the substance as baseline toxic and if no structural alerts could be identified. For a high reliability it is important that characterisation tools were included that are able to actively identify baseline toxicity (e.g according to Verhaar, 1992). However it should be carefully assessed if the overall assessment considers all parts of the molecule or if substructures are present that were not evaluated.

EXPLANATION OF DIFFERENCES

If the reliability of the outcome of the assessment is low because the outcome of the different schemes differs, the following considerations might be helpful:

- Can the difference be explained by different algorithms of the tools?
  E.g. if the characterisation as baseline toxic is based on tools that do not actively identify baseline toxicity a higher uncertainty can be assumed because of the possibility that the substance simply can not be characterised by the scheme (e.g. ECOSAR).
- Can the difference be explained because different parts of the molecule were considered for the assessment?
  In this case, the characterisation should generally be based on the most conservative result (e.g. excess toxicity rather than baseline toxicity).

ADDITIONAL INFORMATION

Results of step 2 may help for the decision on choosing the appropriate test conditions for a new test. E.g. If the substance is classified as reactive, it might be reasonable to perform a semi-static or flow-through test rather than a static test.

Attention should be paid to the fact, that, at the current state of the art not enough information is available for a characterisation of chemicals according to their chronic mode of action. If tools become available and will be used for the assessment, it should be clearly identified if the characterisation is valid for acute or chronic mode of actions.

THE REPORT OF THE OUTCOME OF THE ASSESSMENT SHOULD IDEALLY INCLUDE THE FOLLOWING INFORMATION

- Description of the mode of action if possible, or description if the substance can be characterised as baseline toxic or excess toxic.
- Reliability of the result
- Possible outliers and reasons for the outliers.

STEP 3:

This step includes the following issues:

IDENTIFICATION OF ANALOGUES FOR THE VERIFICATION OF EXPERIMENTAL AND NON-TESTING DATA

As the identification of possible analogues is a helpful tool for the assessment of the reliability of existing data, the identification of analogues and categories might be particularly helpful in cases of varying data quality.

In Section R.6.2.3 and in Section R.10.2.2.2 tools that might be helpful for identification of analogues are described. Guidance how to conclude on possible analogues and categories is provided in Section R.7.8.4.

ANALYSIS OF SUBSTITUTES FOR NEW TESTS

In certain cases, when information on a group is available it may be possible to extrapolate results for studies that would otherwise be technically very difficult to perform. I.e. for a substance where the hydrophobicity is just too high or solubility just too low to maintain or measure a test concentration, studies on more soluble members of the group could be used to predict the likely endpoint value.

STEP 4 – EVALUATION OF *IN VIVO* DATA:

Guidance on how to evaluate the quality of information from individual *in vivo* tests is given in Section R.7.8.4. The following paragraphs describe approaches for the overall assessment of all available information from *in vivo* testing. This may include consideration of the following issues:

HOW TO DEAL WITH CONFLICTING DATA?

When there is more than one set of data on the same species, (strain if known), endpoint, duration, life stage and testing condition the greatest weight is attached to the most reliable and relevant one. When there is more than one set of data with the same reliability rating, it might be necessary to look into more detail at the study reports to see whether a specific reason could explain the difference. If no explanation can be found and the results are not more than one order of magnitude apart, they can be harmonised by a geometric mean. If they are more than one order of magnitude apart, this may be questionable. If the endpoint is critical for the outcome of the regulatory decision,., a repetition of the study may sometimes be the easiest and most efficient solution, especially for non-vertebrate tests. A decision might also be possible on the basis of additional available data, e.g. from studies of a lower reliability rating or from non-testing methods, if these show a distinct tendency in support of a certain result.

ONLY SECONDARY DATA SOURCES AVAILABLE

Normally, data from a secondary source will lack several of the criteria required for a sufficient reliability rating and can therefore not be considered for use in regulatory conclusions. An

exception to this can be made when these data have previously been considered under widely accepted/ justified programmes which themselves contain adequate review processes for data reliability.

CAN AVAILABLE DATA, WHICH ARE NOT ADEQUATE IN THEMSELVES, PROVIDE SUFFICIENT INFORMATION WHEN USED IN COMBINATION?

Some generic guidance on this issue is provided in Chapter R.4. This also mentions the technique of *meta-analysis*, a statistical tool used for analysing the combined data from multiple studies. Such pooling of data may increase the statistical power of certain findings. It requires, however, that the studies from which data are pooled are sufficiently similar with regard to critical parameters of test conditions, set-up, endpoints, reporting etc.

There may be several studies available for the same test substance for the same endpoint, which are deemed to not be fully reliable. However, when used collectively the study results may indicate an effect at approximately the same concentration and time. In these cases there could be justification for using all the studies collectively to conclude on a specific endpoint.

Examples:

- Valid fish toxicity data are only available for a short exposure regime (e.g. 24h).Tests over 96h might be available, which cannot be judged as reliable (e.g. because of poor documentation), but which provide information that the main effect occurs within the first 24h. In this case the 24h value might be used.
- Toxicity data are available for several time points from a 72h test. In this case, the time-effect curve may allow extrapolation of the 96h value.

DO AVAILABLE DATA ALLOW THE DERIVATION OF A SEMI-QUANTITATIVE RESULT?

This consideration applies in relation to given effect values, for example:

- an LC50 value cannot be calculated from an available acute fish tests because no mortality was observed but the tested concentrations are above the EC50 value determined for algae or Daphnia (retrospective threshold approach).
- an EC/LC50 value cannot be derived, because test concentrations were either too high or too low, but it can be stated that the LC50 is either above or below a specific regulatory relevant trigger value, such as C&L criteria or the T criterion in PBT assessment.

THE SUMMARY OF THE GATHERED INFORMATION FROM THE AVAILABLE *IN VIVO* STUDIES SHOULD CONTAIN THE FOLLOWING:

- Results of standard tests available for all trophic levels?
- Reliable results of non-standard tests available for all trophic levels?
- Reliable results from aggregation of different studies available?
- Reliable half-quantitative results available?

- Description of additional information available, of the reliability of this information and of its intended use?

STEP 4A:

The overall assessment of QSAR results highly depends on the availability of additional data such as information about the mode of action and experimental results for analogues. Therefore if this step is used, information generated by step 2 and 3 should ideally be available.

As described in Section R.7.8.3, several QSAR models and programs including models and expert systems are available in order to derive non-testing data. For the overall assessment of the results, the outcome of the analysis of different QSAR models (provided as QSAR prediction formats (QPRFs)) should be considered.

Step 4a aims at answering the following questions:

- Are reliable QSAR results available that can be used instead of experimental data if data gaps are present?
- Can additional information provide a rational for the waiving of tests?
- Can additional information provide a rational for the performance of specific additional tests?

RELIABLE QSAR RESULTS

In general, due to development of regulatory experience in use of non-testing data, guidance at this point is rather tentative. The conclusion on the use of non-testing data alone or in combination with experimental data on decision making will benefit from a case-by-case discussion. It is foreseen to develop a manual of experience which could continuously be updated, revised and improved by a suitable mechanism. This manual will turn practical experience in the validity and acceptance of using (Q)SARs under REACH into a continuously growing REACH QSAR guidance.

However the following considerations might be helpful for the conclusion:

- At the present (2006) higher confidence is based on QSAR models for acute effects compared to QSAR models for chronic effects. Thus QSAR predictions should focus on acute effects, while QSAR results for chronic effects will be in most cases highly unreliable.
- In general higher confidence is provided by QSAR predictions based on baseline toxicity compared to QSAR predictions based on specific modes of action or chemical classes that show more than baseline toxicity. Thus if for a substance a highly reliable classification as baseline toxic according to step 2 and a valid QSAR model where the substance fits into the applicability domain is available the confidence in the prediction might be high.
- Reliability of the result may increase if a close analogue is available and experimental results for this analogues fit to the QSAR prediction.

WAIVING OF TESTS

In general for most substances with a log K_{ow} between 1 and 6 a reliable QSAR model for acute baseline toxicity will be available. Thus in most cases it will be possible to calculate the baseline

toxicity of the substance. If the acute effect concentration calculated for baseline toxicity already triggers a regulatory decision (e.g. baseline toxicity <1 mg/L for classification and labelling) this result might be used. But attention should be paid to the fact that the real toxicity of the substance might be much higher due to a more specific mode of action.

In addition, there could be cases where a substance was classified as having a specific mode of action and a valid model for this specific mode of action is available. Although the result of the prediction may not be reliable enough for a definitive risk assessment, it might be possible to base the decision on the results as a worst case decision (see step 5).

THE SUMMARY OF THE GATHERED INFORMATION FROM THE AVAILABLE QSAR MODELS SHOULD CONTAIN THE FOLLOWING:

- Reliable results of QSAR predictions available?
- Other half-quantitative information available?
- Description of additional information available?
- Description of the reliability of the information and of its intended use?

STEP 4B:

Available *in vitro* tests and their use for regulatory decision are described in Chapters R.3 and R.4. At the present (2006) no *in vitro* tests are available that can substitute *in vivo* data. However *in vitro* data might be helpful to get further insight into the mode of action of a substance:

Some permanent cell lines might express specific characteristics/functions of their source tissue/organ. Their use for more specific modes of action has to be evaluated. Specific modes of action are more likely to be detected with primary cell cultures. For example, primary hepatocytes are used for studying metabolism, hepatotoxicity, genotoxicity and vitogellin induction and isolated gill cells for studying the effect on the branchial epithelium. Transfected permanent fish cell lines were used to detect estrogenic effects of substances.

STEP 5:

In step 5 all available data from the different steps should be integrated in the assessment of the toxicity of the substance in order to understand the toxicity pattern of the substance:

- Experimental data (especially of standard tests) have the highest priority when conclusions on the various endpoints (C&L, PBT assessment, PNEC derivation) have to be drawn. Non-standard or *in-vitro* as well as non-testing data are important in cases where standard experimental data are missing, are not reliable or inconsistent in order to verify experimental data and avoid an assessment on the basis of invalid data (e.g. if two acute fish toxicity tests give two different LC_{50} values (e.g. 10 and 100 mg/L) and the chemical under concern shows non-polar narcosis with an appropriate QSAR result of LC_{50} = 120 mg/L, more confidence might be given to the 100mg/L LC_{50} value). Non-testing data can be considered also as additional information to experimental data in a *Weight of Evidence* approach even if experimental data exist. Moreover, they can be used for elaboration of a test-design for higher-tier-tests or for a decision to perform chronic tests instead of acute ones.

- Ideally, at the end all available information (test data and non-testing information) should be used for a comprehensive conclusion on the endpoint (multi task assessment). This conclusion has to be substantiated and described in the text. The amount of information necessary to draw such conclusions will definitely be different dependent on the regulatory endpoint. For C&L, in certain cases limit tests may be sufficient as only a decision has to be drawn whether the toxicity is below a certain trigger value, whereas for derivation of the PNEC a quantitative figure has to be given. In the latter case it is of particular importance to use all available information, as PNEC derivation means to extrapolate from a few monospecies laboratory tests to maintenance of structure and function of ecosystems. Especially the extrapolation from acute to chronic toxicity is hardly possible. Analysis of a large number of validated data on new and existing chemicals revealed that acute data have only limited predictive value for long-term effects in aquatic ecosystems. The acute/chronic ratio correlates neither with acute toxicity nor with baseline toxicity as modelled through log Kow and no acute/chronic ratio correlation is found across trophic levels, meaning that it is generally not possible to conclude e.g. from daphnia or algal ACR on fish ACR (Ahlers et al 2006).
- In contrast to C+L and PBT assessment, which solely base on intrinsic properties, for PNEC derivation also exposure-based decisions (PEC/PNEC ratio) have to be considered. E.g. EC_{50} values for alga and daphnia are available. In addition QSAR calculations for fish have been performed. From these data a high or low PEC/PNEC ratio has been derived. In the first case a chronic fish test has to be considered. In the second case no additional data are necessary.

STEP 6:

INTRINSIC PHYSICO-CHEMICAL PROPERTIES

Column 2 of REACH Annexes VII and VIII contains the provision that acute studies do not need to be conducted if there are mitigating factors indicating that aquatic toxicity is unlike to occur for instance if the substance is highly insoluble in water or the substance is unlikely to cross biological membranes. On the other hand, REACH asks registrant to consider long-term study when substance is poorly water soluble.

There is no scientific basis to define a cut off limit value for solubility below which no toxicity could occur. There may be technical difficulties to perform the test, e.g. sensitivity of the analytical method used for the determination of test concentration. Such difficulties and proposed solutions should be clearly documented. Results from tests above the limit of solubility should not be interpreted as pelagic toxicity, but as confounded by physical effects. For further details see testing of difficult substances in Section R.7.8.7.

Equally, there is no scientific basis to define molecular characteristics that would render a substance unlikely to cross biological membranes.

Thus no scientifically based cut off criteria for these mitigation factors can be provided at the moment. Nonetheless, it might be possible to decide on a case-by-case basis, that aquatic toxicity is unlikely to occur due to very low water solubility and unlikelihood to cross biological membranes. Issues which may be considered in this regard are the indicators used for low likelihood of a high bioaccumulation potential (Chapter R.11). When such indicators are used in the context of triggering derogation from toxicity testing on aquatic organisms however a more cautious approach should be used. The reason is that indications of lack of a high bioaccumulation potential does not necessarily imply lack of toxicity to aquatic organisms.

In any case any proposal to deviate from the standard testing requirements in reference to this clause should be carefully justified. For poorly water soluble substances (e.g. water solubility below 1 mg/L or below the detection limit of the analytical method of the test substance) it should instead of an acute test be considered to perform a long term test (REACH Annex VII and VIII, 9.1) bearing in mind any possibilities for waiving (REACH Annex XI).

THRESHOLD APPROACH FOR TOXICITY TESTING IN FISH

This approach offers a possibility to significantly reduce the number of fish to be used in acute aquatic toxicity testing when a test on fish is required. It takes into consideration that only the lowest value of the acute toxicity in species of three trophic levels is considered for regulatory purposes.

The approach was originally described as threshold/step-down approach by Hutchinson et al. (2003) for pharmaceuticals. Several authors retrospectively evaluated acute aquatic toxicity data of chemical substances (Jeram et al, 2005; Hoekzema et al, 2006) by applying this approach. ECVAM and the ECB further developed the threshold approach taking into account existing guidelines and reflecting the requirements for the limit test (OECD TG 203, Annex V C.1). The ECVAM Scientific Advisory Committee (ESAC) has endorsed the scientific validity of the threshold approach following the advice of the ESAC peer review panel.

The approach is currently part of the rolling workplan for the OECD test guidelines program 2006/2008 (Project 2.23: New Guidance Document on Application of the Step Down Approach (or Upper Threshold Concentration) as a Limit Test for Acute Fish Toxicity Testing).

With the lowest of the two EC_{50} concentrations obtained for algae and *Daphnia*, (the Upper Threshold Concentration, UTC), a limit test according to OECD TG 203, using 7-10 test and 7-10 control fish, is carried out. In case that no mortality is observed, no further tests are carried out and the acute fish toxicity result (LC_{50}) is reported as *greater than* (>) the UTC value. In case that mortality is observed, a full LC_{50} test should be performed.

The same principle could also be applied when instead of fish, fish embryos or early life stages are used for toxicity testing.

FROM INTEGRATED TESTING TO INTEGRATED ASSESSMENT

When the *Weight of Evidence* approach has been finalised as described above, the amount of validated information may in some cases largely exceed the minimum information requirements of the Annexes of REACH and thus reduce the uncertainties when extrapolating from monospecies laboratory tests to the structure and function of ecosystems. As for PNEC derivation these uncertainties are to be covered by the assessment factors it may be considered to use these factors in a more flexible way according to the altered degree of uncertainty; (*it has to be mentioned that such flexibilities on assessment factors are already foreseen, when the assessment is based on Species Sensitivity Distribution (SSD) and on mesocosm as well as field studies and also use of QSAR for narcotic mode of action, to be confirmed*).

Such a *multi-criteria assessment* should cover - beside the information mentioned above – also:

- The number and representativity of species tested
- The quality of non-standard tests
- the time-dependence of the toxicity
- the steepness of concentration/effect curves
- Information from mammalian toxicity normally not used in standard assessments. Specific guidance on this approach with regard to potential reproductive or developmental toxicity via endocrine modes of action is provided in Section R.7.8.11.

At the end the derivation from the degree of uncertainty defined in the standard situations and represented by certain assessment factors given by the Section R.10.3 has to be substantiated fully.

The proposal presented here is an optimal possibility to use *all available information* in order to protect human health and the environment from hazardous chemicals

R.7.8.5.1 Concluding on suitability for Classification and Labelling

Environmental classification and labelling of a substance is generally based on data from short-term tests for fish, invertebrates and algae. Information from other tests may be used under the *safety net* provisions, i.e. in cases where substances do not fall under the *core set of criteria* but on the basis of the available evidence concerning their toxicity may nevertheless present a danger to the structure and/or functioning of aquatic ecosystems. There are no defined criteria for this classification; its possible application to substances that cause adverse effects on development or reproduction is discussed in Section R.7.8.11.

Classification and labelling has to be performed for all substances registered in REACH. The following strategy gives guidance how to classify a substance for the environment based on its toxicity, if different levels of information are available (see also Figure R. 7.8-3).

As a first step all available information on substance has to be collected and evaluated as described in Section R.7.8.5 and Chapter R.3.

1. If acute effect values for all three trophic levels are available, classify based on the lowest effect value available and derive specific concentration limits (M-factor) if relevant, i.e. toxicity <0.1 mg/l.
2. For substances with tonnages between 1 and 10 t/y, Annex VII requires acute toxicity tests with invertebrates and algae/aquatic plants:
    a) If EC_{50} for invertebrates and algae/aquatic plants are available according to Annex VII, classify the substance based on the lowest effect value; if, according to step 4a of Section R.7.8.5, a reliable QSAR result for fish is available or if additional information e.g. using read-across can be provided, consider this value for the classification. Specific concentration limits (SCLs) (M-factor) should be derived, if relevant (GHS and RIP 3.6. Guidance).

b) If no acute data are available for invertebrates and/or algae/aquatic plants, it should first be checked, if mitigating factors (water solubility, molecular size) are justifiable:

- if this is the case, no acute tests have to be performed for the substance. *Safety net* classification based on fate data (degradation and bioaccumulation) should nevertheless be considered.
- if the mitigation factors are not applicable, it is necessary to perform an acute Daphnia and an acute algae test to fulfill the requirements of Annex VII. If a reliable QSAR prediction for fish can be made, consider this value for classification. SCLs (M-factor) should be derived, if relevant.

3. For substances with tonnages >10 t/y, Annex VIII requires in addition an acute fish test. However derogations from the standard information requirements may be made if the provisions of REACH for this are fulfilled. In the following, guidance is given to use available aquatic toxicity data on classification and labelling:

a) Acute toxicity data on invertebrates and algae/aquatic plants are available and the EC_{50} for at least on of these species is <1 mg/l. In this case, no acute fish study is necessary for substances that are not used in preparations, as the available effect value(s) already trigger the classification as N, R50. However, for substances used in preparations, an acute fish test might nevertheless be a prerequisite for setting specific concentration limits (SCL, M-factor) for the classification of preparations containing the substance.

b) Acute toxicity data on invertebrates and algae/aquatic plants are available and EC_{50} for both species is >1 mg/l. In this case, information on acute toxicity to fish is necessary for the judging whether the aquatic toxicity to fish may warrant classification. Thus it should be checked whether the calculation of an LC_{50} for fish with a reliable QSAR is possible or whether estimation is possible that fish may be less sensitive than invertebrates and/or algae/aquatic plants (see to Section R.7.8.5). Derive SCLs (M-factor) if necessary.

- if this is possible, this information can be used together with the available effect data on invertebrates and algae/aquatic plants for the purpose of classification.
- if this is not possible, an acute toxicity test with fish would provide data which may be used for classification purposes. However if alternative and adequate test methods are available for acute fish toxicity they may be considered to be used instead for classification (see Figure R. 7.8-3). E.g. a proposal to use the fish embryo test (FET) as an alternative to the acute fish toxicity test has been made and is currently under evaluation in the OECD Guideline program (see Sections R.7.8.3.1 and R.7.8.8). There is currently no specific guidance in GHS on whether or not data from a FET can be used instead of acute toxicity test data on fish for hazard classification purposes (to be elaborated later in RIP 3.6).
- if data from suitable alternative test methods are not available, a fish limit test following OECD TG 203 using as exposure concentration the lowest EC_{50} from acute tests on invertebrates and algae/aquatic plants may be performed. If no mortality is observed, this is an indication of fish not being more acutely sensitive than invertebrates and algae/aquatic plants. Hence classification can then be based on the lowest available EC_{50}-value (for invertebrates and algae/aquatic plants). If mortality occurs in the fish limit test, data from an acute fish toxicity test according OECD TG 203 should be made available according to the needs of the chemicals safety assessment and the LC_{50} (fish) can then be used together with the EC_{50}-values for invertebrates and algae/ higher plants as basis for classification (GHS & RIP 3.6. Guidance).

In the following, guidance is given for the specific cases, that instead of acute invertebrate/fish tests long-term invertebrate/fish tests are available (Column 2 of Annex VII and VIII). It is very likely that such cases do not commonly occur, and therefore guidance is only given in the text and, not in the flow chart:

1. Substances with tonnages between 1 and 10 t/y (Annex VII): EC_{50} algae/aquatic plants and *long-term* invertebrate instead of acute invertebrate test are available.

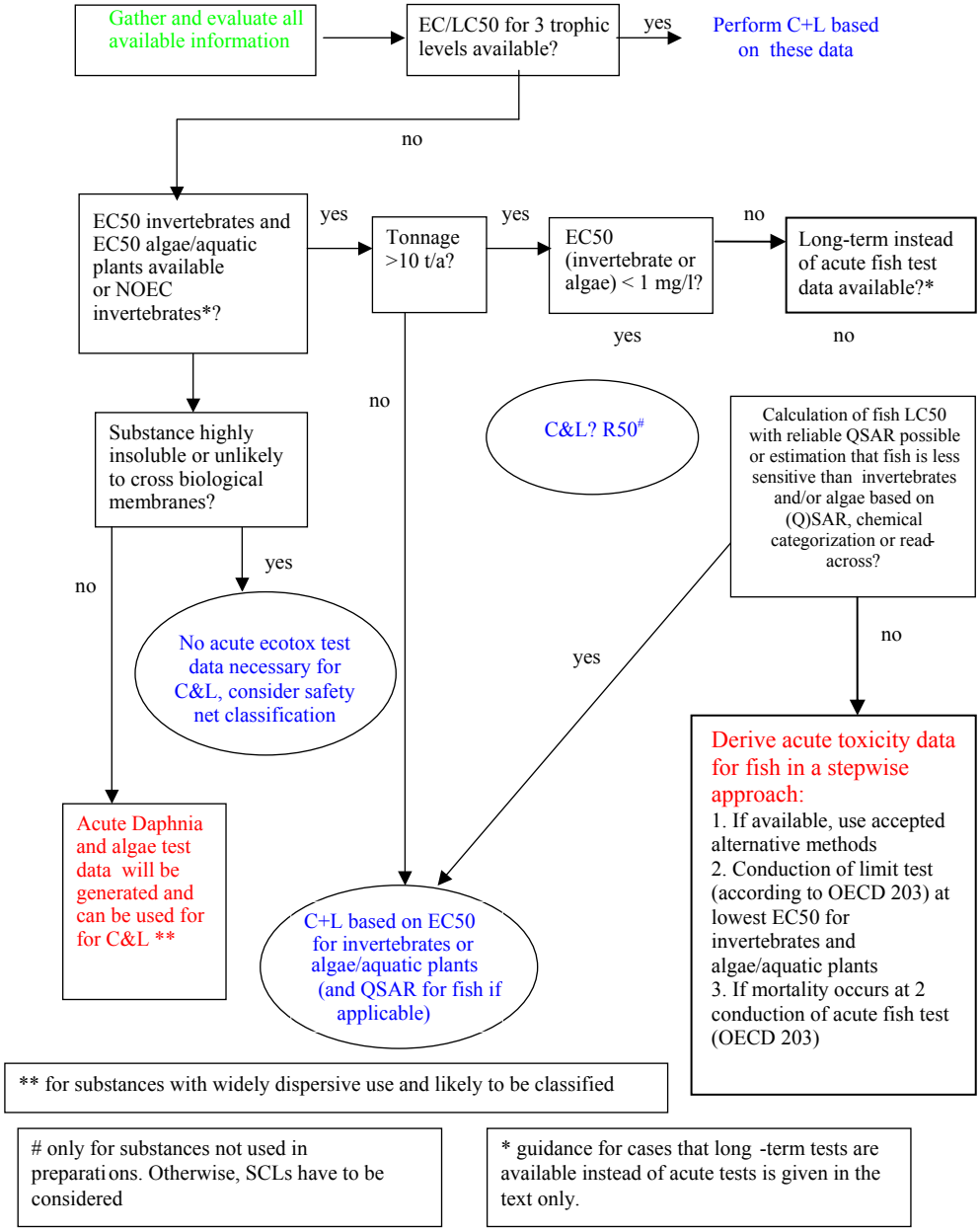
2. Substances with tonnages ≥10 t/y (Annex VIII): EC_{50} invertebrates and algae/aquatic plant and *long-term* fish instead of acute fish are available.

   For both points above:

   a) At least one available EC_{50} is <1 mg/l: In this case no further acute data are necessary for the classification for substances that are not used in preparations, as this value triggers already the classification as N, R50. However, for substances used in preparations, further information on acute toxicity might nevertheless be useful for classification purposes of substances. The reason is that particular high acute toxicity may imply that a specific concentration limit (SCL, M-factor) should then be set for the substance.

   b) Available EC_{50} >1 mg/l: In this case it should be checked whether the derivation of an acute EC_{50} from the long-term studies is possible (e.g. if raw data of the study are available and at the tested concentration range included immobilisation of parent invertebrates (OECD TG 202, part 2) resp. mortality of fish (OECD 215) of >50 % the test parental animals). This effect value can then directly be used for classification purposes together with available EC_{50} values.
   If this is not possible, it should be checked whether reliable predictions of EC_{50} for invertebrates resp. fish with valid QSAR models are possible that can be used for the classification of the substance. An additional option is to check whether classification can be considered based on a grouping approach relating to the species for which data are missing regarding acute toxicity. If no estimation is possible of the acute toxicity for the aquatic organism with no acute toxicity test data , then classification have to be considered based on the available data on other aquatic organisms.

Figure R. 7.8-3: Decision Scheme for Classification and Labelling



R.7.8.5.2 Concluding on suitability for PBT/vPvB assessment

DEFINITIVE CRITERIA

According to the REACH legislation (Annex XIII), a substance is considered to fulfil the toxicity criterion (T) with regard to aquatic toxicity when:

- the long-term no-observed effect concentration (NOEC) for marine or freshwater organisms is less than 0.01 mg/l.

Due to animal welfare concerns, the general scheme of testing P, B and then T (cf. section on PBT assessment being developed by in Chapter R.11) should be applied and also vertebrate-animal testing should be minimised by first testing non-vertebrate species. For determination of definitive criteria for T, chronic tests must be performed. Under normal circumstances, these criteria are applied based on the methodologies listed in the tonnage triggered information requirements (REACH Annexes VII to X). The suitability of information obtained with other test methods must be assessed on a case by case basis based on the principles outlined in Chapter R.4.

Use of non-testing data

Only a few QSAR models predicting chronic aquatic toxicity are available but further research on the QSAR prediction of chronic toxicity may increase our predictive capacities. Therefore at the current state of the art, QSAR models seem not to be applicable for the definitive assessment of the T criteria.

Screening criteria

A substance is considered to potentially meet the criteria for T when an acute L/EC50 value from a standard E/LC50 toxicity test (REACH Annexes VII to X) is less than 0.1 mg/l. In addition to data from standard toxicity tests, data from reliable non-standard tests and non-testing methods (including valid QSAR predictions) may also be used if available. These data should be particularly assessed for their reliability, adequacy, relevance and completeness.

The toxicity criterion (T) for PBT assessment cannot be decided on the basis of acute studies alone. If the screening criterion is met, the substance is referred to definitive T testing and chronic studies are required regardless of the tonnage band unless the acute L/EC50 < 0.01 mg/l. In the latter case, it is automatically confirmed that the substance fulfills the definitive T criterion. Normally, the testing order for conclusion on T based on chronic data is Daphnia and then fish[5]. If the T-criterion is fulfilled by the chronic algae or Daphnia data, a chronic fish test is not necessary.

5 Algae are not mentioned here because chronic algae data (i.e. 72h NOEC) normally will be available, as it can be obtained from the same 72h standard test from which the acute endpoint (72h EC50) is derived. However an exception may be when the original data on algae toxicity from existing reports are not obtainable.

For certain lipophilic substances, with a log Kow >5, the acute toxicity value may be greater than the water solubility of the substance under review. In such situations, chronic toxicity with a NOEC <0.01 mg/l cannot be excluded, as these substances may not have had sufficient time in the acute test to reach steady-state. In the absence of definitive information on T, for substances with very high lipophilicity, a weight of evidence or group approach for long term toxicity may be used to predict whether long term effects are likely to occur. If convincing evidence is available that aquatic toxicity is not expected to occur at <0.01 mg/l, chronic testing may not be required. Such evidence should be based on expert judgement and weight of evidence of data including reliable QSAR predictions/read-across/grouping approach indicating narcotic mode of action together with measured low chronic fish toxicity from a related compound. Supporting information could be chronic data on aquatic species such as daphnids, algae or sediment dwelling species.

In choosing the appropriate test organism, the data from the available base set of toxicity tests for algae (acute / chronic), Daphnia (acute) and fish (acute) should be evaluated under consideration of the possible hydrophobic properties of the test substance, and hence the expected time to steady-state. Any indication for a specific mode of action of the test substance also needs to be considered. If it can be concluded that one taxonomic group is significantly more sensitive than the others, e.g. because there is evidence for a specific mode of action, this sensitive group should be chosen for chronic testing and conclusion on the T-properties[6]. If no conclusive evidence for significant differences in sensitivity between the groups can be found the testing order as mentioned above shall apply.

If the relevant test species is selected in accordance with the suggested approach in the paragraph above, lack of toxicity at or below the definitive T criterion for the tested species is evidence that further studies on T are not necessary. If however a long term test on Daphnia or algae provides a NOEC close to but above 0.01mg/l, a long-term fish study is likely to be needed on substances confirmed to fulfill the P and B criteria to confirm “not T” unless, taking into consideration the above-mentioned approach, convincing evidence exists that the fish NOEC will be higher than 0.01 mg/l. Supporting evidence in such considerations could be an acute fish value a factor of 10 or more greater than that of the other two trophic levels under the provision that the acute daphnid test showed toxicity at least one order of magnitude lower than the limit of solubility.

Certain chemical characteristics (such as high adsorption or extremely low solubility) are likely to make any toxicity testing extremely laborious if not technically impossible. Some examples together with recommendations to overcome the technical difficulties are provided in the section on assessment of difficult substances (see Table 7.8-3 on physicochemical properties - summary of difficult testing issues).

Use of non-testing data

At preliminary stages in the assessment, in cases where no acute or chronic toxicity data are available, the assessment of the T criterion at a screening level can be performed using data obtained from quantitative structure activity relationships (QSARs) for acute aquatic toxicity. In order to be suitable, the QSAR prediction should comply with the general principles described above. Long-term testing is required if QSAR estimations indicate that the substance fulfils the screening criteria for T (EC50 or LC50 < 0.1 mg/l).

6 This could for example mean that no further testing is necessary if it is concluded that algae are significantly (i.e. more than 10 times) more sensitive than daphnids or fish and the available chronic algae data are well above a NOEC of 0.01 mg/l.

It may on a case by case basis be decided whether confirmatory chronic testing on fish is necessary if valid QSAR prediction indicate that the acute LC50 is < 0.01 mg/l. Alternatively either first an acute fish toxicity limit test could be performed to check whether the acute toxicity is below 0.1 mg/l or the QSAR-prediction could be accepted as providing sufficient evidence of the T criterion to be fulfilled.

Further Guidance

Further guidance on the test strategy including how to avoid aquatic toxicity testing due to lack of persistence and/or lack of bioaccumulative properties is given in Chapter R.11.

R.7.8.5.3 Conclusions on Chemical Safety Asessment (PNEC Derivation)

The Chemical Safety Assessment (CSA) is based on all available toxicity information. The information should at least cover species of three trophic levels: algae/aquatic plants, invertebrates (*Daphnia* preferred), and fish. The following strategy gives guidance how to assess the pelagic toxicity of a substance for chemical safety assessement, if different levels of information are available (see also Figure R. 7.8-4)

A first sequence of considerations is primarily based on the availability of short-term toxicity data as specified in REACH Annexes VII and VIII (combined). If results from the hazard assessment or the risk characterisation indicate the need for further investigations, long-term toxicity data will be considered in subsequent considerations.

Short-term toxicity data

1. Check available data from standard testing:

   Algae: If a 72 hour ErC_{50} value from a growth inhibition study according to OECD 201 or a 96 hour ErC_{50} value from a growth inhibition study is available this can be used directly for PNEC assessment. If possible, it is recommended to calculate the 72 h growth rate based on data from the test report of 96 h tests.

   Invertebrates: If a 48 hour EC_{50} value from short-term toxicity testing on *Daphnia sp.* according to OECD 202 or a NOEC/EC_X value from long-term toxicity testing on *Daphnia sp.* according to OECD 211 or results from tests using equivalent test guidelines are available, these can be used directly for PNEC assessment.

   Fish: If an LC_{50} value from short-term toxicity testing on fish according to OECD 203 or a NOEC value from long-term toxicity testing on fish e.g according to OECD 215 (fish juvenile growth test) or 210 (fish early life stage test) or OECD 212 (egg and sac-fry test) or results from tests using equivalent test guidelines are available these can be used directly for PNEC assessment.

2. Check other available data - standard testing data might be substituted by one of the following:

   Algae: The ErC_{50} is the preferable and more meaningful value from a standard growth inhibition (OECD 201) study. Where this is not available/ reported but an EbC_{50} is available/reported it should be considered to perform a new algae study, especially if algae are the most relevant species for the effects assessment. If possible it is recommended to calculate, the 72 h value based on data from the test report of 96 h tests.

Invertebrates: A 24 hour EC_{50} value from short-term toxicity testing on *Daphnia sp.* according to OECD 202 but this should only be used in conjunction with other data (e.g. on time-dependence of toxicity) as part of a *Weight of Evidence* approach.

Other reliable experimental data on algae/aquatic plants, invertebrates or fish (e.g. data from non-standard studies or for non-standard organisms).

Reliable QSAR results (see Section R.7.8.4.1 for evaluation of QSAR results).

Reliable read-across from available experimental data on a structurally related substance.

An adequate value for growth inhibition of algae/aquatic plants or for short-term toxicty in invertebrates or fish from any of the sources listed above may be used directly for PNEC assessment.

3. Check possibilities for the prediction of relative species sensitivities:

   The sensitivity of fish relative to invertebrates and algae might be predicted from one of the following:

   Experimental data from standard studies

   Other experimental data (e.g. data from non-standard studies or for non-standard organisms)

   Data generated with QSAR models

   Read-across from available experimental data on a structurally related substance.

   If there is compelling evidence, using these methods, to suggest that the fish value is likely to be at least a factor of about 10 less sensitive than invertebrates or algae there are no further requirements for acute fish testing. There may be other considerations for testing, e.g. if a test result would help to build or improve a data base for a chemical category. Consideration should also be given to needs for chronic testing e.g. whether range finding data is needed to determine test concentrations etc.

4. Check possibilities for adaptation of the standard information requirements:

   If there are mitigating factors, such as those specified in Section R.7.8.5, indicating that aquatic toxicity is unlikely to occur, studies on the growth inhibition of algae/aquatic plants or the short-term toxicity in invertebrates or fish do not need to be conducted (column 2, Annex VII and VIII).

5. If no adequate data is available and there are no mitigating factors indicating that aquatic toxicity is unlikely to occur perform a growth inhibition study on algae according to OECD 201 and a short-term toxicity study on *Daphnia sp.* according to OECD 202 or a long-term toxicity study according to OECD 211 (According to column 2, Annex VII, a long-term study shall be considered if the substance is poorly water soluble, i.e. solubility <1 mg/L, TGD 2003). Alternatively risk management measures reducing exposure and hence risk sufficiently might be considered.

6. Fish: Check availability of accepted alternative methods

   If there is a need to generate new data on the toxicity in fish and an accepted alternative method is available instead of *in vivo* fish testing perform the alternative test. At the time of writing (2006) no alternative methods have been accepted as an alternative to the *in vivo* fish study. A possible alternative, the fish embryo toxicity test, is currently under evaluation in the OECD Guideline program (see Sections R.7.8.3.1 and R.7.8.8).

7. Fish: Determine relative sensitivity

   If there is no alternative to generating new toxicity data from *in vivo* fish testing a limit test should be performed as described in OECD 203 using the lowest EC_{50} from invertebrates or algae. If no mortality occurs in the limit test that indicates that fish are less sensitive than invertebrates or algae there are no further requirements for short-term fish testing.

8. Fish: If mortality occurs in the limit test, perform a short-term toxicity study in fish according to OECD 203 or a long-term toxicity study as appropriate (for detailed guidance see below long-term toxicity testing) (according to column 2, Annex VIII, a long-term study shall be considered if the substance is poorly water soluble, i.e. solubility <1 mg/L, TGD 2003). Alternatively risk management measures reducing exposure and hence risk sufficiently might be considered.

   Normally a Fish Early Life Stage test (OECD 210) would be considered appropriate for examining long-term fish toxicity. However, the fish, juvenile growth test (OECD 215) (for substances with log K_{ow} <5) or egg and sac-fry stage test (EU Annex V C., OECD 212) (for substances with log K_{ow} <4) may also be considered. Specific guidance on the consideration of available data on developmental or reproductive effects from non-standard tests is provided in Chapter R.7.

   Using the data specified in the preceding steps, the PNEC value can be derived considering the results from all three trophic levels. If the substance meets the criteria for classification as dangerous or is assessed to be a PBT or vPvB, an exposure assessment should be carried out and the risk should be characterised.

   If the CSA indicates no risk, there are no further requirements for aquatic toxicity testing. If the CSA indicates a need to investigate further effects on aquatic organisms long-term toxicity testing shall be considered. These considerations apply in the same way to all substances in quantities >10 t.

   A risk from CSA is indicated

   - If PEC/PNEC >1
   - For substances with log K_{ow} >3 (or BCF >100) and a PEC_{local} or $PEC_{regional}$ >1/100th of the water solubility.

Long Term Testing

1. Check available data from standard long-term testing:

   Invertebrates: If a NOEC value from long-term toxicity testing on *Daphnia sp.* according to OECD 211 or results from tests using equivalent test guidelines are available these can be used directly for the refinement of the PNEC value.

   Fish: If a NOEC value from long-term toxicity testing on fish according to OECD 215 or 210 or 212 or results from tests using equivalent test guidelines are available these can be used directly for the refinement of the PNEC value.

2. Check other available data:

   Standard testing data might be substituted by one of the following:

   Other reliable experimental data on aquatic invertebrates or fish (e.g. data from non-standard studies or for non-standard organisms)

   Reliable QSAR results[7]

   Reliable read-across from available experimental data on a structurally related substance

   An adequate value for long-term toxicity in invertebrates or fish from any of the sources listed above may be used directly for the refinement of the PNEC value.

3. Check possibilities for the prediction of relative species sensitivities:

   The sensitivity of fish relative to algae and invertebrates might be predicted from one of the following:

   Experimental data from standard studies

   Other experimental data (e.g. data from non-standard studies or for non-standard organisms)

   Data generated with QSAR models

   Read-across from available experimental data on a structurally related substance.

   If there is compelling evidence, using these methods, to suggest that the fish value is likely to be at least a factor of about 10 less sensitive than invertebrates or algae there are no further requirements for fish testing. There may be other considerations for testing, e.g. if a test result would help to build or improve a data base for a chemical category.

   The same considerations as detailed above apply to the sensitivity of invertebrates relative to algae and fish, i.e. if there is compelling evidence to suggest that the invertebrate value is likely to be at least a factor of about 10 less sensitive than algae or fish there are no further requirements for invertebrate testing.

4. If invertebrates are likely to be more sensitive than fish and algae or the relative sensitivity of invertebrates cannot be predicted prepare a testing proposal for a long-term toxicity study on *Daphnia sp.* according to OECD 211 for submission to the Agency. Alternatively risk management measures might be considered.

5. If fish are likely to be more sensitive than invertebrates and algae or the relative sensitivity of fish cannot be predicted prepare a testing proposal for a long-term toxicity study on fish according to one of the below listed OECD testing guidelines for submission to the Agency. Alternatively risk management measures reducing exposure and hence risk sufficiently might be considered. Normally a Fish Early Life Stage test (OECD 210) would be considered appropriate for examining fish toxicity. However, the fish, juvenile growth test (OECD 215) (for substances with log Kow <5) or egg and sac-fry stage test (EU Annex V C.) (for substances with log Kow <4) may also be considered. Specific guidance on the consideration of available data on developmental or reproductive effects from non-standard tests is provided in Chapter R.7.

[7] Currently reliable QSAR models for chronic toxicity are rare and thus reliable QSAR results will be seldom available. However if QSAR models for chronic toxicity will be available in future they need to be evaluated equivalent to acute toxicity QSAR models as described in chapter R.R.7.8.4.1.

Further possible methods for the refinement of the risk assessment, e.g. the use of Species Sensitivity Distributions may be considered.

R.7.8.5.4 Overall conclusion

In Section R.7.8.5 guidance is given, how to combine all gathered information in order to understand the toxicity pattern of the substance and how to draw overall conclusions on the different regulatory endpoints, Classification and Labelling, PBT Assessment as well as PNEC derivation. A major feature of these assessments will be flexibility and expert judgement. The results have to be substantiated thoroughly and communicated.

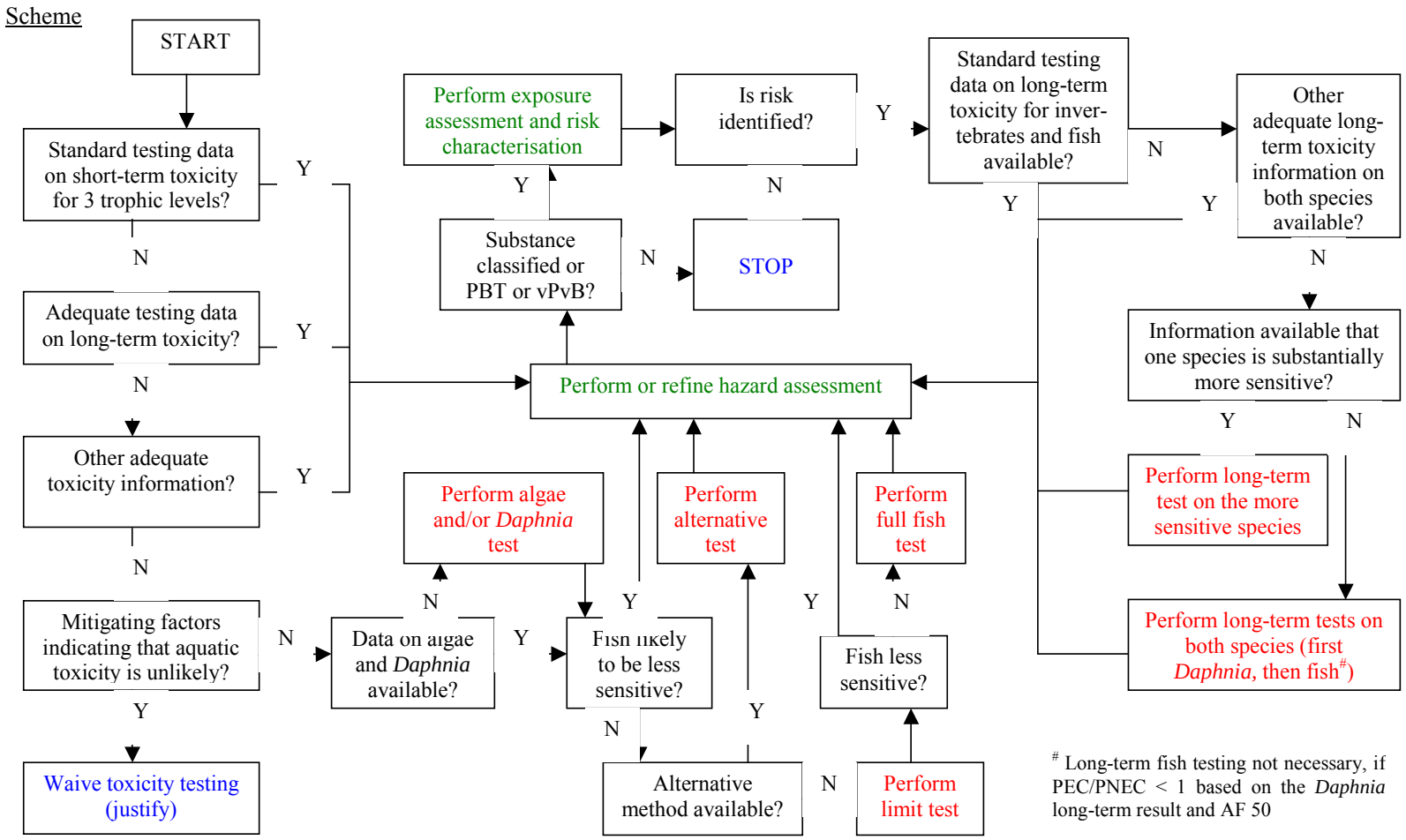


Figure R. 7.8-4: Decision scheme for the conclusion on chemical safety assessment (PNEC Derivation)

For the conclusions on the different endpoints often variable amounts of information are required with the consequence that the testing strategies proposed may differ accordingly; e.g. for classification and labelling a limit test may be sufficient, whereas the CSA assessment for the same substance requires a chronic fish test.

Therefore, to avoid unnecessary testing the different strategies should be compared critically at the end of the exercise. Moreover, a few rules have to be followed:

PBT assessment: chronic fish toxicity testing is generally only necessary, when the P and B criteria are fulfilled (cf. further in RIP 3.2. PBT Guidance on strategies to minimize testing with vertebrate animals).

Priorities for future research: To perform substantiated conclusions on the different endpoints the available tools have to be developed further. The following items among others should be considered for further research:

1. Mechanistic approaches
    a. Develop knowledge of modes of action so that future CSAs can be adapted to technical progress.
    b. Sub-lethal acute endpoints as predictors. Better use of information from chronic toxicity tests as well as toxicokinetics to make predictions of Mode of Action. Use data acquired to increase knowledge of structural alerts.
2. Development, including validation and applicability domain description, of QSAR models for chronic toxicity to pelagic and sediment organisms
3. Develop validated Test Guidelines for feeding studies on pelagic organisms
4. Improve knowledge of critical body burdens and compile databases and establish and improve links to various classes of modes of action.
5. Improve read-across for freshwater to marine organism toxicity and increase database for marine Phyla.
6. Improve understanding of how to read-across from Human Health and, if possible, biodegradation data to environmental risk assessment (e.g. to increase understanding of biotransformation and identification of relevant metabolites).
7. Improve predictive techniques for extrapolating from laboratory to field studies.
8. Consider how population dynamics can be included into ecotoxicology.
9. Develop & validate *in vitro* tests and based on this develop guidance how to use *in-vitro* tests.
10. Develop Guidance how to use genomic information (“omics”)
11. Develop guidance for *multi-criteria assessment*, that means how to use all available information on derivation of a PNEC, including flexibility of assessment factors.

R.7.8.6 References on aquatic pelagic toxicity

Ahlers J, Riedhammer C, Vogliano M, Kühne R-U, Schüürmann G. 2006.Acute to Chronic Ratios in Aquatic Toxicity – Variation across Throphic Levels and Relationship with Chemical Structure. Environmental Toxicology and Chemistry 25: 2937-2945.

Bernard MJ, Dyer S. 2005. Fish critical cellular residues for surfactants and surfactant mixtures. Environ Toxicol Chem. 24: 1738-1744.

Brain P, Cousens R. 1989. An equation to describe dose responses where there is a stimulation of growth at low doses. Weed Research 29: 93-96.

Campbell PGC, Lewis AG, Chapman PM, Crowder AA, Fletcher WK, Imber B, Luoma SN, Stokes PM, Winfrey M. 1988. Biologically available metals in sediments (NRCC No. 27694), National Research Council Canada, Ottawa.

Castaño A, Bols N, Braunbeck T, Dierick P, Halder M, Isomaa B, Kawahara K, Lee LEJ, Mothersill C, Pärt P, Repetto G, Riego Sintes J, Rufli H, Smith R, Wood C, Segner H. 2003. The use of fish cells in ecotoxicology. Altern Lab Anim 31:317-351.

CEC. 2003. 2nd Edition of the Technical Guidance Document (TGD) on Risk Assessment of Chemical Substances following European Regulation and Directives. April 2003. European Chemicals Bureau (ECB). JRC-Ispra, Italy. http://ecb.jrc.it/tgdoc.

Cedergreen N, Ritz C, Streibig JC. 2005. Improved empirical models describing hormesis. Environmental Toxicology and Chemistry, 24 (12): 3166-3172.

Crossland NO, Heimbach F, Hill IR, Boudou A, Leeuwangh P, Matthiessen P, Persoone G. 1992. Summary and recommendations of the European Workshop On Freshwater Field Tests (EWOFFT). Potsdam, Germany, 25-26 June 1992. 37p.

Curren et al. 1995. The Role of Prevalidation in the Development, Validation and Acceptance of Alternative Methods. ECVAM Prevalidation Task Force Report 1. ATLA 23, 211-217

ECETOC. 1997. The Value of Aquatic Model Ecosystem Studies in Ecotoxicology. Technical Report No. 73. 123p. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

ECETOC. 2000. Comparative Ecotoxicity of Chemicals to Freshwater and Saltwater Organisms. Background paper provided to the TGD working group on Marine Risk Assessment. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

ECETOC. 2003a. Environmental Risk Assessment of Difficult Substances. Technical report No 88. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

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

ECETOC. 2003c. Aquatic Hazard Assessment II. Technical report No 91. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2005. Alternative Testing Approaches in Environment Safety Assessment. Technical Report No. 97. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

Escher BI, Hermens JLM. 2002. Modes of toxic action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. Environ. Sci. Technol. 36: 4201-4217.

Giddings JM, Solomon KR, Maund SJ. 2002. Probabilistic risk assessment of cotton pyrethroids: II. Aquatic mesocosm and field studies. Environmental Toxicology and Chemistry 20:660-668.

Gombar VK, Enslein K. 1995. Use of predictive toxicology in the design of new chemicals. ACS Symp. Ser. 589: 236-249.

Gulden M, Seibert H. 2005. Impact of bioavailability on the correlation between in vitro cytotoxic and in vivo acute fish toxic concentrations of chemicals. Aquat Toxicol. 2005 72, 327-37.

Hobbs DA, Warne MSJ, Markich SJ. 2005. Evaluation of Criteria Used to Assess the Quality of Aquatic Toxicity Data. Integrated Environmental Assessment and Management 1(3): 174–180

Hoekstra JA, Van Ewijk PH. 1993. Alternatives for the no-observed-effect level. Environ. Toxicol. Chem. **12**, 187-194.

Hoekzema, C, Murk, J.A., van de Waart, B.J., van der Hoeven, J.C.M., de Roode, D.F. (2006) Alternative approaches can greatly reduce the number of fish used for acute toxicity testing. Environ. Toxicol. Chem. 22, 1322-1325.

Hutchinson, T.H., Barrett, S., Buzby, M., Constable, D., Hartmann, A., Hayes, E., Huggett, D., Länge, R., Lillicrap, A.D., Straub, J.O., Thompson, R.S., 2003. A strategy to reduce the numbers of fish used in acute ecotoxicity testing of pharmaceuticals. Environ. Toxicol. Chem. 22, 3031-3036.

Jaworska JS, Hunter RS, Schultz TW. 1998. Quantitative structure-toxicity relationships and volume fraction analyses for selected esters. Arch. Environ. Contam. Toxicol. 29: 86-93.

Jeram, S., Riego Sintes, J.M., Halder, M., Baraibar Fentanes, J., Sokull-Klüttgen, B., Hutchinson, T.H. (2005) A strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances notified in the European Union. Regulatory Toxicology and Pharmacology 42, 218-224.

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

Lipnick RL. 1991. Outliers: their origin and use in the classification of molecular mechanisms of toxicity. Sci. Total. Environ. 109-110:131153.

Mc Geer JC, Brix KV, Skeaff JM, DeForest DK. 2002. The use of bioaccumulation criteria for hazard identification of metals. Fact sheet on environmental risk assessment. Published by the International Council on Mining and Metals (ICMM)

Mekenyan OG, Dimitrov S, Pavlov TS, Veith GD. 2004. A systems approach to simulating metabolism in computational toxicology. I. The TIMES heuristic modelling framework. Current Pharm. Design 10: 1273-1293.

MERAG. 2006. Metal Environmental Risk Assessment Guidance Document. www.euras.be/merag

Moore DRJ, Breton RL, MacDonald DB. 2003. A comparison of model performance for six quantitative structure-activity relationship packages that predict acute toxicity to fish. Environ. Toxicol. Chem. 22: 1799-1809.

Newman MC, Jagoe CH. 1994. Ligands and the Bioavailibility of Metals in Aquatic Environments. In: Hamelink JL, Landrum PF, Bergman HL, Benson WH (eds.) Bioavailability: Physical, Chemical and Biological Interactions. Lewis Publishers, Boca Raton, U.S.A.

Newsome LD, Johnson DE, Nabholz JV. 1993. Validation and upgrade of a QSAR study of the toxicity of amines to freshwater fish. ASTM Special Technical Publication, STP 1179 (Environmental Toxicology and Risk Assessment), 413-426.

OECD. 1984. Guideline for the testing of chemicals No. 204. Fish, prolonged toxicity test; 14-day study. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 1992a. Guideline for the testing of chemicals No. 203. Fish, acute toxicity test. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 1992b. Guideline for the testing of chemicals No. 210. Fish, early life stage toxicity test. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 1995. Environment Monographs No. 92. Guidance documents for aquatic effects assessment. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 1996. OECD guidelines for testing of chemicals. 305. Bioconcentration: Flow-through fish test. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 1998a. Guideline for the testing of chemicals No. 212. Fish, short-term toxicity test on embryo and sac-fry stages. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 1998b. Guideline for the testing of chemicals No. 211. *Daphnia magna* reproduction test. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2000a Guidance for Determining the Quality of Data on the SIDS Dossiers (Reliability, Relevance and Adequacy), Interim Manual, June 2000.

OECD. 2000b. Guideline for the testing of chemicals No. 215. Fish, juvenile growth test. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2000c. Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures, ENV/JM/MONO (2000)6. OECD Environmental Health and Safety Publications Series on Testing and Assessment No. 23, Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2004a. The Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs. ENV/JM/TG(2004)27/REV. Organisation for Economic Cooperation and Development, Paris, 17pp.

OECD. 2004b. Guideline for the testing of chemicals No. 202. *Daphnia* sp. acute immobilisation test. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2006a. Draft Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms). 26 p.

OECD. 2006b. Guidance Document on Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. OECD Environmental Health and Safety Publications Series on Testing and Assessment No. 54, Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2006c. Draft Guideline for the testing of chemicals No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Organisation for Economic Co-operation and Development, Paris, France

OECD. 2006d: Draft Guideline for the testing of chemicals No. 221: *Lemna sp.* Inhibition Test. Organisation for Economic Co-operation and Development, Paris, France

Ownby DR, Newman MC. 2003. Advances in Quantitative Ion Character-Activity Relationships (QICARs) using metal-ligand binding characteristic to predict metal toxicity.

Pack S. 1993. A review of statistical data analysis and experimental design in OECD aquatic toxicology test guidelines. Shell Research, Sittingbourne, UK, prepared for OECD.

Roberts DW, Costello JF. 2003. Mechanisms of action for general and polar narcosis: a difference in dimensions. QSAR Comb. Sci. 22: 226-233.

Russom CL, Anderson EB, Greenwood BE, Pilli A. 1991. ASTER: An integration of the AQUIRE database and the QSAR system for use in ecological risk assessments. Sci. Total Environ. 109/110: 667-670.

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

Schabenberger O, Tharp BE, Kells JJ, Penner D. 1999. Statistical tests for hormesis and effective dosages in herbicide dose response. Agronomy Journal 91: 713-721.

Schirmer K. 2006. Proposal to improve vertebrate cell cultures to establish them as substitutes for the regulatory testing of chemicals and effluents using fish. Toxicology 224, 163-83

Schüürmann G, Kühne R, Kleint F, Ebert R-U, Rothenbacher C, Herth P. 1997. A software system for automatic chemical property estimation from molecular structure. In Chen F, Schüürmann G, eds, *Quantitative Structure-Activity Relationships in Environmental Sciences – VII.* SETAC Press, Pensacola, FL. pp. 93-114.

Seth et al. 1999. ES&T 33:2390

U.S. EPA. 1994. ECOSAR: A computer program for estimating the ecotoxicity of industrial chemicals based on structure-activity relationships. EPA-748- R-002. Office of Pollution Prevention and Toxics, National Center for Environmental Publications and Information, Cincinnati, OH.

U.S. EPA. 2002. ECOTOX user guide: ECOTOXicology database, version 3.0. www.epa.gov/ecotox..

Urrestarazu RE, Vaal MA, Hermens JLM. 2002. Interspecies sensitivity in the aquatic toxicity of aromatic amines Environ. Toxicol. Pharmacol. 11: 149-158.

Vaes WHJ, Urrestarazu RE, Verhaar HJM, Hermens JLM. 1998. Acute toxicity of nonpolar versus polar narcosis: is there a difference? Environ. Toxicol. Chem. 17: 1380-1384.

Van Ewijk PH, Hoekstra JA. 1993. Calculation of the EC50 and its confidence interval. Ecotoxicology and Environmental Safety 25: 25-32.

Veith GD, Call DJ, Brooke LT. 1983. Structure-toxicity relationships for the fathead minnow, Pimephales promelas: narcotic industrial chemicals. Can. J. Fish. Aquat. Sci. 40: 743-748.

Verhaar HJM, Van Leeuwen CJ, Hermens JLM. 1992. Classifying environmental pollutants. 1: structure-activity relationships for prediction of aquatic toxicity. Chemosphere 25: 471-491.

Von der Ohe PC, Kühne R, Ebert R-U, Altenburger R, Liess M, Schüürmann G. 2005. Structural alerts – a new classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute Daphnid assay. Chem. Res. Toxicol. 18: 536-555.

Walker et al. 2003. Quantitative Cationic-Activity relationships for predicting toxicity of metals. Environmental Toxicology and Chemistry 22 (8): 1916-1935.

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

Appendices to Section R.7.8

Appendix 7.8-1 Critical parameters for aquatic toxicity testing

Appendix 7.8-2 Information an dits sources*: in vivo*

Appendix 7.8-4 Methodology for body burden approaches in aquatic effects assessment

Appendix 7.8-5 Assessment of available information on endocrine and other related effects

Appendix 7.8-1: Critical parameters for aquatic toxicity testing

(Properties of substances and (tests) systems and other factors influencing evaluation of aquatic toxicity

The following table summarizes the critical parameters that influence toxicity testing and potentially testing strategy in the aquatic environment. The table is divided into two main headings, Test related parameters, and Substance related parameters. Both are useful for evaluating the validity of existing studies however, the Substance related parameters also concern information that should be acquired prior to initiating new studies. For more detailed information the reader is referred to OECD (2000) and (ECETOC, 2003). This document gives some first guidance for inorganic compounds and metals. More extensive guidance can be found in Van Gheluwe 2006.

Table R. 7.8-2: Critical parameters for aquatic toxicity testing

Parameter	Sub-parameter	Issue	Recommendation
Test related parameters			
General		Water quality	All ecotoxicological tests should include information on key parameters influencing general water quality, indicating the fitness of the medium to support the organisms being tested and the likelihood that the exposure of the test substances occurred in a way that resembles the conditions in the environment. Frequency of measurement should also be indicated. Any single parameter which was out of the range indicated by the test method should trigger an in depth inquiry into the validity of the study and careful consideration of the relevance of the results.
Oxygen			Oxygen requirements depend on the organism with e.g. rainbow trout requiring very high levels (less than 50% could result in mortality) and certain benthic dwelling organisms capable of survival with almost negligible oxygen availability. However, in sediment tests, oxygen should always be measured close to the sediment as there may be much lower concentrations in the peribenthic layer than in the water column. In certain cases, (e.g. if biodegradation of the test substance or tertiary solvent is high) with non-volatile chemicals, aeration may be provided directly in the test system to increase oxygen concentration but for some species, (e.g. daphnids) this may lead to physical damage of the organisms and significant stress and should be avoided.
pH			*Pelagic* – pH is generally acceptable within the range of 6.5 – 9 but this depends on the organism. Algae tests, for example, may reach a pH of 10 without any notable effect on the growth rate. However, in certain cases, notably ionising organics and metals, pH has an impact on speciation and thus toxicity. In such cases a decision needs to be made on the test strategy to be employed and the acceptable range of pH in the tests. Use of buffers or modified test strategies (e.g. reduction of initial cell numbers) can help to prevent major modifications of pH during the test. *Sediment* – The pH of sediments may vary during the study. This may have an impact on the sediment dwelling organisms but also, for ionising substances, may change the ion exchange capacity of the substrate, increasing or decreasing bioavailability of the test substance and the pore water concentrations. Such changes should be monitored and controlled if possible.
Temperature			Most Guidelines include temperature as a standard physical parameter as the organisms may be stressed or the validity of the results may not be achievable outside of the recommended limits (e.g. at less than 18ºC it may be difficult to achieve the validity criterion of >60 juvenile daphnids per surviving adult within 21 days recommended in OECD 211). However, the change in temperature during the test is just as important. Fish are particularly sensitive to temporal temperature variations which can lead to temperature shock. In any test, spatial variation in temperature is also critical, and as climate rooms are often inconsistent, comprising

Parameter	Sub-parameter	Issue	Recommendation
			both hot and cold spots, ideally oxygen should at least be measured in test systems with the greatest spatial separation. Any suggestion that systematic differences in temperature occurred between groups should lead to consideration of the validity of the study.
Hardness/ Conductivity			The optimal ion requirement and composition varies from species to species and these are generally indicated in the test method. Hardness may influence the bioavailability of certain test substances (such as metals and metal compounds) and in these cases measurement of this parameter is relevent. For example, hardness is used in bioavailability models such as Biotic Ligand models (BLM) to describe competition effects for metals.
Alkalinity			Carbonate ions may alter speciation of metals. Hence for a proper understanding of metal speciation in the test medium knowledge on the alkalinity may improve our understanding of the test results.
Chlorides/salinity			Salt effects may have a pronounced influence on test results. Most organisms tolerate chloride levels up to 500 mg/L. Above this threshold, depending on the organisms tested, osmotic stress may occur and bias the test results. For some metals like Ag the formation of chloride complexes may also influence the bioavailability.
NH_3/NH_4			Ammonia is highly toxic and in dynamic equilibrium with the less toxic ammonium ion, is thus influenced by pH and to some extent temperature. Many species, including fish, directly excrete ammonia via the gills and faeces into the water and in static systems, or in high stock density tests, the ammonia concentration is likely to increase during the study. This may be a particular problem for sediment based systems which may be static for long periods of time. In studies where ammonia can cause a problem, measurements are generally included in the methodology, however for less validated methods it is worth considering whether the ammonia concentration is likely to have influenced the results.
DOC			Dissolved organic carbon may be present in some studies, particularly those where natural water has been used. In such cases, DOC measurement is needed. Many adsorbing substances bind to DOC either ionically or hydrophobically and this may increase or decrease the bioavailability of the test substance. DOC is also a key parameter which is incorporated for most bioavailability models for metals. E.g. Biotic Ligand models using speciation models like WHAM VI.
TOC			*Sediment:* Total Organic Carbon (TOC) of sediments will vary depending on the type of sediment used in the study. This may have an impact on the sediment dwelling organisms but influence the bioavailability of both organic substances and metals/metal compounds..
AVS			*Sediment-metals*: Acid Volatile Sulfides (AVS) may influence the bioavailability of metals and metal compounds. AVS concentrations in artificial sediments are very low and quite often below detection limit. However, when field sediments are used AVS concentrations can be measured in order to allow a proper interpretation of test results of metal sediment toxicity data.

Parameter	Sub-parameter	Issue	Recommendation
Substance related parameters			
Molecular weight and size			Molecular weight and size might influence the bioavailability and the uptake of the substance
Water solubility		General	Water solubility is an essential parameter in ecotoxicological testing and data should be available prior to any aquatic effects testing. Failure to do so could result in testing above the solubility limit leading to misinterpretation of the results. Poorly soluble substances are defined by OECD (2000) as substances with a limit of solubility <100 mg/l although technical problems are more likely to occur at <1mg/l as defined in TGD (1996). Very low water solubility (i.e. in the low µg/l range) could be used as a reason to significantly modify a standard test or to test non-pelagic organisms preferentially (see Table 7.8-3 for more information). Whenever possible pelagic tests should be performed at or below the water solubility of the test substance in that medium. Tertiary solvents are often used in order to prepare stock solutions so that they can be further diluted to provide test solutions. Solvents used at the maximum allowed concentration (100 mg/l) will rarely increase the solubility of the test substance significantly but may lead to emulsion formation which could cause physical effects. Solvents should be avoided when possible for pelagic tests and if employed, care should be taken that they do not lead to an increase in BOD due to their (in some cases) rapid degradation. They are also employed to spike sediment and in such cases they are generally removed by air drying prior to use. However, traces of contaminants they contain may remain and furthermore, the organic solvent may have a negative effect on the sediment being used by redistributing or changing the organic carbon fraction. Typically solvents distribute the test substance onto the substrate in a way that does not occur in the environment and therefore the technique should be used with care. Dispersants have been employed in a similar way to solvents but are used more to achieve a stable dispersion than to dissolve the substance in the stock solution. OECD (2000) does not generally advocate the testing of dispersants unless they are natural properties of the substances under scrutiny (e.g. detergents or oil dispersing agents). OECD recommends the use of the column generator method for poorly soluble, solids which do not contain impurities with higher solubility than the test substance itself.
		Multi-component substances (UVCBs)	Multi-component substances are mixtures comprising a complex mix of individual substances with different solubilities and physico-chemical properties. In most cases, they can be characterised as a homologous series of substances with a certain range of carbon chain length/number or degree of substitution. Typically it is difficult to test and evaluate these substances. For further information see Table 7.8-3
	Freshwater		Natural freshwater contains inorganic ions and DOC as well as suspended matter. Synthetic media contain many of the compounds found in natural freshwater but sometimes also other substances are employed to help buffer or

Parameter	Sub-parameter	Issue	Recommendation
			maintain bioavailability of certain micronutrients. Standard solubility tests on the other hand are usually performed in deionised water. It is not unusual for measured values at maximum solubility in aquatic tests to differ from the solubility test result. Usually, the maximum solubility of a substance in synthetic medium is lower than the solubility test result indicates but this is not always the case. This should be taken into account generally when testing is proposed close to the limit of solubility of the test substance but may be exacerbated for certain groups of chemicals e.g. chelates. For strongly adsorbing chemicals adsorption to suspended solids (SS) and for ionised organics such as surfactants, also binding to DOC may occur and the truly dissolved fraction may be difficult to evaluate. In such cases total load may be reported or used as a more applicable endpoint. In such cases it is important that the DOC and SS concentrations are known. For more information see Table 7.8-3
	Marine		In the marine environment the salinity is so high that the solubility of most substances decreases and precipitation may occur by a process known a salting out. The decrease in solubility has been calculated as approximately 10-50% for neutral non-polar substances. A simple correlation for the salting out factor in seawater as a function of organic solute molar volume is to consider a reduction in solubility by a factor of 1.36 (ECETOC, 2001). For ionising substances, pH dependency should be known when the pH of seawater (approximately 8) is close to the pKa value. Testing considerations should be taken into account as above (freshwater).
	Poorly soluble	Physical effects	These usually apply only to difficult substances with very low solubilities. Certain substances may form mycelles when mixed with water even at very low concentrations (100 µg/l or less) or form a surface film covering aquatic organisms and potentially smothering them. Signs of these effects can be considered likely when daphnids are trapped at the surface in the test solutions (not always reported) or when there is a great variation in effect between replicates of the same concentration
Coloured substances			See Table 7.8-3 difficult substances
Sorption	General		Sorption/desorption tests provide information on Koc (organic carbon normalised adsorption coefficient) and Kd (distribution coefficient) to the appropriate compartment. For many chemicals, such studies (or values of Koc derived from Kow QSARs) provide useful information on their likely partitioning behaviour in aquatic studies although it should be noted that for certain chemicals (notably surfactants and metals) the standard Freundlich isotherms derived from such studies are inappropriate.
	Neutral (hydrophobic) (expressed as log Kow)	Loss of substance from the test system	Highly lipophylic substances (log Kow >4, OECD 2000) are likely to pose problems during testing due to their expected low water solubility and tendency to stick to hydrophobic surfaces such as glassware, tubing, food and test organisms binding by van der Waals forces. Loss from the test solution may also be expected due to bioconcentration in the test organisms. For these reasons the organism stocking density should be low enough and the test system volume should be high enough so that the concentration of the test substance can be maintained throughout the studies. Naturally, static systems tend not to be appropriate for such substances. Flow-through is preferred when possible but achieving an adequate stock solution under such circumstances may be a challenge.
	Ionic	Loss of	May be positively or negatively charged organic or inorganic chemicals which bind to substrates of opposite

Parameter	Sub-parameter	Issue	Recommendation
		substance from the test system	charge e.g. cationically charged substances bind to negatively charged humic acids, clay, glassware, microorganisms etc; anionic compounds bind to positively charged Si, Al or Fe oxides). Adsorption mainly becomes an issue when test concentrations are below 1 mg/l. Attempts should be made to minimise binding sites and to saturate them if possible by pre-exposing them to similar concentrations of test chemical as those to be used in the study.
Surface active		Loss of substance from the test system	Surface active substances are a sub-set of the ionic substances mentioned previously and may be cationic, anionic, non-ionic or amphoteric. In all cases supplementary difficulties in estimating Koc arise and the Kow method cannot be used.
Ionising		Change of bioavailability with pH	Knowledge of the PKa will allow prediction of the extent of ionisation of such substances in test water. As unionised organic species tend to be more hydrophobic than the ionised forms, the solubility and bioavailability of the substance may vary dramatically even between environmental extremes in pH. Consideration should be given to appropriate pHs (to be) used in the test as, solubility may be lower but toxicity may be higher in the unionised form than in the ionized form.
Degradation			OECD recommends testing parent compound for DT50 >3 days, breakdown products for DT50 <1h and case-by-case basis for anything in between. A flow-through test is recommended for substances with a DT50 of 4 h as 50% of the nominal parent substance concentration can be maintained with 6 volume renewals per day. ECETOC (2003) and the TGD recommend to test parent substance with a DT50 as low as 12 h, as based on maximum half life allowing 80% maintenance of parent compound in flow-through system and >1% in short term test. However, this should be considered on a case-by-case basis depending on the technical feasibility of performing such a study.
	Photodegradation		Photodegradation is the reaction of a chemical after absorption of light leads to an electronically excited state with increased reactivity and subsequent transformation. Photodegradation may be either direct (transformation of the substance by direct excitation) or indirect (transformation of another chemical due to transfer of energy from another photosensitive molecule. Kinetic photodegradation is determined experimentally.
	Hydrolysis		Hydrolysis is a common degradation route in the environment, where reaction of a substance with water with a net exchange of the X group with an OH at the reaction centre such that $RX + H_2O \rightarrow ROH + HX$. Hydrolysis is often dependent upon pH as the reactio is commonly catalysed by hydrogen or hydroxide ions. Hydrolysis kinetics are usually determined experimentally and should be used to consider the test type and whether parent or degradation product should be tested.
	Biodegradation		In the cases of readily biodegradable substances, biodegradation may be so fast that it is difficult to maintain test concentrations throughout the study. If such situations are likely then consideration should be given to regular cleaning or replacement of the test vessels during testing and preparation of stock solutions under sterile or near sterile conditions.

Parameter	Sub-parameter	Issue	Recommendation
Volatility			Vapour pressure is a measure of the equilibrium between the condensed and vapour phases of a substance. The Henry's law constant (*H*) for a substance is a measure of its equilibrium between an ideal solution phase and the vapour phase. As such it is a measure of the potential for a substance to be lost from solution by evaporation. As an approximation, if *H* is greater than 100 Pa.m3/mol, more than 50% of the substance could be lost from the water phase-in 3-4 hours (Mackay, 1992).If there is evidence that the substance may volatilise from the test solution during the study, steps should be taken to reduce the loss by using closed systems or reducing headspace.

References:

OECD, 2000a. Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures, ENV/JM/MONO (2000)6. OECD Environmental Health and Safety Publications Series on Testing and Assessment No. 23, Organisation for Economic Co-operation and Development, Paris, France.

Ecetoc 2003. Aquatic Hazard Assessment II. Technical report No 91. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

Difficult Substances

Valid aquatic toxicity tests require the test substance to be dissolved in the water medium under the conditions recommended by the guideline, and the maintenance of a bioavailable exposure concentration for the duration of the test. One or both of these requirements may be difficult to achieve or measure in practice for some types of substance – collectively referred to as *difficult substances*. This can affect both the performance and interpretation of tests, and can be especially problematic when considering existing data from older studies. Such data typically require expert judgement to determine whether there is sufficient information in a test report for a decision to be made on its validity, and also whether the result is suitable for regulatory use.

Figure R. 7.8-5 indicates the thought processes that must be followed when considering a difficult substance. In general, it is important that the composition of the substance is as well-defined as possible. In some cases, it may be relatively straightforward to make a decision on the use of the data. It should be remembered, however, that a substance may be 'difficult' in several ways (e.g., it might be both a multi-component mixture and unstable), and each property can present complex challenges, even for experts. It is therefore impossible to provide simple advice that can apply in every situation. Nevertheless, the OECD has produced detailed guidance on how to adjust standard methods for such substances (OECD, 2000) and guidance on data interpretation for classification (OECD, 2001). Table R.7.8-3 presents a summary of the main issues identified in these important sources, which should be consulted for more detailed information.

One of the key issues for difficult substances is the ability to quantify actual exposure of the test organisms to the test substance. In general, test results should be expressed in terms of mean measured concentrations as far as possible (though it is often useful to quote both the measured and nominal effect concentrations). The following general principles apply:

- For static, semi-static and flow-through tests, where the concentrations remain within 80-120% of nominal, the effect concentrations can be expressed relative to nominal or measured concentrations.
- For static tests, where the concentrations do not remain within 80-120% of nominal, the effect concentrations should be expressed relative to the geometric mean of the measured concentrations at the start and end of the test.
- For semi-static tests, where the concentrations do not remain within 80-120% of nominal, the effect concentrations should be expressed relative to the mean concentration over the whole exposure period, calculated from the geometric mean of the measured concentrations at the start and end of each media renewal period.
- For flow-through tests, where the concentrations do not remain within 80-120% of nominal, the effect concentrations should be determined and expressed relative to the arithmetic mean concentration.

- For tests with chemicals that cannot be quantified by analytical methods at the concentrations causing effects, the effect concentration can be expressed based on the nominal concentrations. However this might result in an underestimation of the toxicity and it should be justified why no quantification by analytical methods is possible.

  Where loss processes are very fast, the median of the concentrations that are measured after the decline would be more appropriate as a surrogate for the mean exposure concentration. In the absence of a suitable analytical method, a semi-static renewal or flow-through regime may be necessary to ensure that exposure concentrations are in line with target values.

  Where a measured concentration at the end of the exposure period is absent or where it indicates that the substance is not detected, the validity of the test should be reconfirmed. In order to calculate a mean exposure concentration, the final concentration may be taken as the limit of detection for the method if the substance is not detected. When the substance is detected but not quantified, it is good practice to use half of the limit of quantification. Since there may be various methods for determining that, the method selected to determine mean measured concentrations should be made explicit in the reporting of test results.

  NOTE:

  a) Polymers are not considered either, because they do not require registration in the initial phases of REACH implementation.
  b) Finally, some substances can contain impurities that can change in proportion and/or chemical nature between production batches. Interpretational problems can arise where either or both the toxicity and water solubility of the impurities are greater than the parent substance. This is not currently considered in this document, but is closely linked with the identity of the registered substance.

Figure R. 7.8-5: Considerations for difficult substances

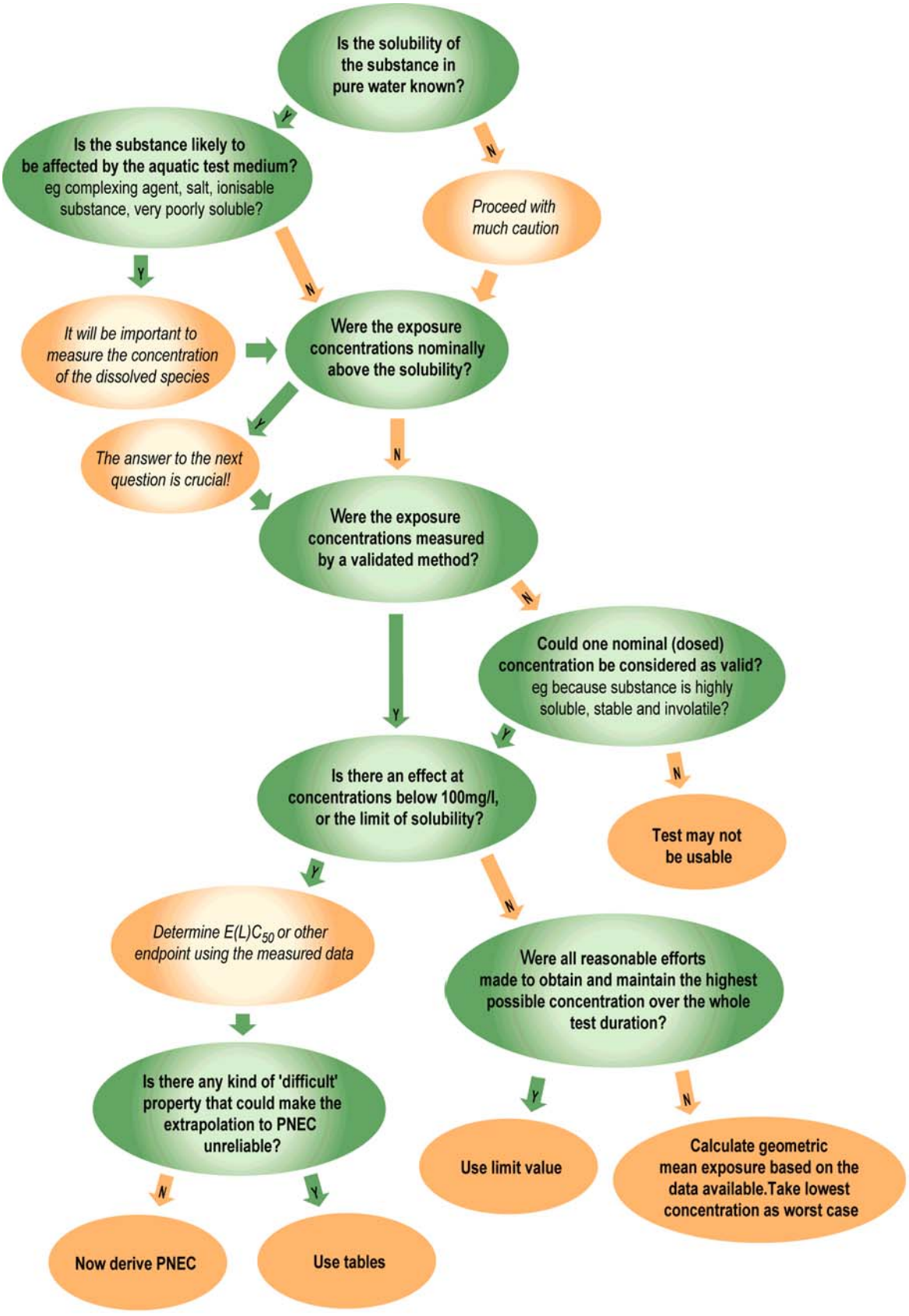


Table R. 7.8-3: Summary of difficult substance testing issues

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
The substance contains many components	Multiple components may make analytical monitoring impossible. Differences in partitioning behaviour and water solubility between components can make it difficult to achieve a homogeneous solution by direct addition to the test medium (e.g. if some components are highly insoluble). This can also present interpretational problems. For example, it might not be possible to know which components have caused any observed adverse effects.	Figure 7.8-5 presents a general pathway for considering such substances. If all the components of the substance are fully soluble in the medium across the range of test concentrations, standard test methods are appropriate. Some components may have individual properties (e.g. degradability, volatility, etc.) that require steps to be taken to control losses (see below). If the substance is only partially soluble, the components should be identified and the toxicity estimated using available information on them. For example, components that have structural and physico-chemical similarities should be grouped and treated as if the whole 'block' were one single compound. This approach has been developed for petroleum hydrocarbons in particular, and is known as the 'hydrocarbon block method'. (see draft ESR risk assessment for gasoline, and guidance from CONCAWE) Each 'block' is assembled on the basis of those properties that will influence the outcome of the PEC and PNEC calculations, i.e. usually octanol-water partition coefficient, Henry's Law constant, biodegradability and toxicity. The properties of each block may be estimated using a combination of non-testing methods for representative structures and the available measured data. If this is not possible, tests using water-accommodated fractions (WAFs) are appropriate. The method used to prepare the WAF should be fully described in the test report, with evidence provided of attainment of equilibrium and its compositional stability over time if possible. WAFs are prepared individually and not by serial dilution of a single stock WAF. Solvents should also be avoided, and generator systems are not appropriate. Test data obtained with WAFs apply to the multi-component substance	It maybe possible to analyse for one of the components during the test This approach was used in the UK CCRMP assessment of tetrapropenyl phenol, for one of the long-term aquatic studies.

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
		as an entity. The exposure is generally expressed as the 'loading rate' (mass to volume ratio of the mixture to medium) used to prepare the WAF. The measured mass of test substance in the WAF can also be used (as a concentration). For test data obtained with WAFs the following apply if the mixture contains components with a large range in water solubility: actute test data will correspond to the toxicity of the more soluble components, whereas chronic tests will reflect toxicity of the less soluble components. The acute lethal loading level (typically expressed as the E(L)L50) is comparable to L(E)C50 values determined for pure substances tested within their solubility range. It may therefore be used directly for classification. However, it cannot be used to derive a PNEC, since partitioning in the environment will make the comparison with a PEC meaningless. No Observable Effect Loading Rate (NOELR) values from chronic tests may be sufficiently low to be of the same order as the level at which most components are dissolved (or the PEC value), in which case they can be used for PNEC derivation. If direct dosing of the medium can be achieved, e.g. by use of solvents within the limits allowed by the test guideline, the data will represent the hazard of the sum of the components and the E(L)C50 can be used to obtain a PNEC (though it will still not be known which components caused the effects).	

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
The substance is poorly soluble in the test medium (water solubility typically <1 mg/L) [similar problems can apply if the substance is simply difficult to analyse in the test medium]	Solubility may be difficult to determine and is frequently recorded as less than the analytical detection limit. It may be difficult to dissolve the substance in a test solution, and to maintain and verify concentrations. Toxicity may be observed at concentrations below the lowest measurable concentration. Results may be expressed in terms of nominal concentration, which might exceed the true dissolved concentration of the substance in the test medium. This is a particular problem for older studies. Physical effects (e.g. entrapment) may occur if the test concentration is significantly above water solubility. Interpretation of partitioning behaviour can also be problematic where poor solubility in water and octanol may be compounded by insufficient sensitivity in the analytical method.	Ideally, tests using appropriate dissolution techniques and with accurately measured concentrations within the range of water solubility should be used. Where such test data are available, they should be used in preference to other data. However, some techniques may present certain drawbacks, which must be taken into account. For example, the effect of any solvent needs to be determined, and solvents are not appropriate for mixtures where the use of the solvent can give preferential dissolution of one or more components (this may also apply to impurities). OECD (2000) provides more examples. The study report should be read carefully for indications of the presence of undissolved test material (e.g. droplets or surface layer). If this is the case and effects are observed, the results should be treated as invalid. Toxicity may be observed at concentrations nominally in excess of water solubility, or below the detection limit of the analytical method. Such data are not automatically invalid since the original solubility estimate may be uncertain, and the solution may have been prepared appropriately (e.g. provided any undissolved substance is removed prior to testing). If physical effects are not obvious, then as a realistic worst case, the lowest effect concentration may be based on either the water solubility limit or detection limit of the analytical method, whichever is the lower. If no toxicity is expressed at concentrations up to the water solubility limit, judgement must be applied as to whether the result can be considered valid. The hazard should not be underestimated, and interpretation should stress the side of caution. Due account should be taken of the techniques used to achieve the maximum dissolved concentration. Where these are inadequate, the test should be considered invalid.	If the PNEC represents an upper limit, further testing may be required following risk assessment. This may require a more appropriate method or sensitive analysis (e.g. using radio-labelled test compound). For substances that are not acutely toxic at their limit of water solubility, the need for chronic testing has to be addressed if required by the risk assessment (provided the solubility is less than 100 mg/L). Substances that are not chronically toxic to aquatic organisms at their limit of solubility rarely need further consideration. If the substance to be tested is a member of a chemical category or if there are analogue substances, a possibility is to test the analogue substance that has a higher solubility and to extrapolate the results from this test to the substance in question. See ESR on Decabromodiphenylether and MCCP.

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
The substance is ionisable or is a salt	The extent of ionisation may vary according to pH or the level of counter ions in the media, 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. This in turn may cause the expression of different toxicities in freshwater and marine environments. For salts, both the anionic and cationic parts need to be considered. Solubility measurements for regulatory purposes are usually made in distilled water (pH 6-9), whereas the pH of test media is usually 7-8. This may significantly affect solubility, especially for substances with a pKa between 5 and 9.	For hazard and risk assessment, the data must be obtained under environmentally relevant conditions. If the relevant dissociation constant (pKa value) for the ionisation process is available (required for substances supplied at 100 t/y), it should be compared with the pH reported in the test report to determine which chemical species were present. It may also be important to check which chemical species are monitored by any analytical method used. The absence of this information may make it impossible to interpret the results. The definitive test should be conducted at a pH consistent with the more toxic form of the substance whilst remaining within the range required to maintain the health of the control organisms. A stable pH is important to ensure that the balance between dissociated and non-dissociated forms of the substance is maintained. If no data are available on a salt, effects may be read-across from the anion or cation, whichever has the most toxic effect. If the effect is related to only one of the ions, the classification of the salt should use the effect concentration multiplied by the salt:ion molecular weight ratio. Where a substance causes a change in pH of the test medium (e.g. strong acids and bases), the pH should be adjusted to lie within the specified range for the test using a suitable technique. Care should be taken that this does not lead to removal of the substance (e.g. via sedimentation and/or degradation). The use of buffers can affect the test result, particularly for algae. Growth of algal test cultures can cause an increase of pH due to consumption of bicarbonate ions. Strategies for maintaining the concentration of these ions and therefore reducing pH shifts are discussed in OECD (2000).	If the test substance ionises to a significant extent, it may be necessary to determine the toxicity of both anionic and cationic species. The solubility at different relavent pH should be determined, and pH and substance concentration should be analysed during the test. An example where this issue has been considered is in the ESR assessment of tetrabromo-bisphenol A.

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
The substance is a complexing agent	Speciation may change in the presence of cations (e.g. Ca, Mg) and anions (e.g. SO4, PO4), co-complexing agents and other properties of the medium such as pH. This can influence solubility, bioavailability and toxicity of the substance. It may also reduce the availability of essential nutrients (which is only a secondary effect, not direct chemical toxicity). Adsorption to sediments is not easily predicted – adsorption is often strong for these types of substance.	This issue is generally of most significance for aquatic plant growth tests. It is important to distinguish between chelated and non-chelated fractions in the test medium if possible, and the extent to which effects are a direct consequence of chemical toxicity (based on the bioavailable fraction). Speciation models may be helpful for this purpose. Data from tests in which complexation is judged to have had a significant bearing on the result are likely to be of questionable value for regulatory use. Compensatory adjustment to water quality parameters (e.g. the concentration of the essential ions) or the testing of an appropriate salt of the test substance may help to achieve a valid test result but protocols incorporating modifications to standard procedures should be validated and approved for use by the regulatory authority. The issue has arisen in the ESR assessment of EDTA, as well as for other complexing agents for the interpretation of algal studies. One approach used has been to run additional tests using enriched nutrient media, reduced substance concentration or addition of extra nutrients at test completion, and then extending the study. This is described in a paper presented at the 24th North American SETAC meeting: PW070 Effects of Iron amd Micronutrient Metals on Algal Growth in the Presence of Chelators	If toxic effects are believed to be due to complexation, then this could be substantiated by measuring the complexation stability constants. Tests with provision of additional nutrient (to compensate for the complexed fraction) may be helpful in some cases. OECD (2000) suggests testing the substance in both standard algal growth medium and in a modified medium with a higher hardness, as well as the calcium salt. See UBA guidance too.

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
The substance is surface active	Surfactants and detergents can form dispersions or emulsions in which the bioavailablity is difficult to ascertain, even with careful solution preparation. Micelle formation can result in an overestimation of the bioavailable fraction even when “solutions” are apparently formed. This presents significant problems of interpretation. QSAR modelling is potentially very difficult since the Kow cannot usually be measured.	Toxic effect concentrations for dispersions and emulsions should be compared with the dispersibility limit (i.e., the limit at which phase separation takes place) or the critical micelle concentration (CMC) for a substance in water rather than with its water solubility limit. The bioavailable concentration does not change above the CMC, even at higher dosing levels. The highest test concentration should either be 1000 mg active ingredient/litre or the dispersibility limit/CMC, whichever is lower. In the ESR programme, a number of surfactants have been assessed - DODMAC and the alkylamines. For these, one of the main difficult properties was the strong tendency to adsorb on surfaces such as test vessels or organic material. If the E(L)C50 or NOEC(L) is below the CMC then the data can be treated in the usual way for classification and to derive a PNEC. If the substance is not toxic at the CMC, the CMC may be used as a NOEC to derive a precautionary PNEC. If a test has been conducted at concentrations above the CMC and shows effects, the effect concentration should be set as the CMC as a precautionary worst case, unless it is clear that physical effects have occurred. For sediments, it is very important to know the adsorption coefficient, preferably by measurement. An estimated Kow value, though of low reliability for surfactants, may be helpful. Guidance for the selection of appropriate methods of Kow measurement is provided in Chapter R.9 (Guidance from RIP 3.2 for physico-chemical properties) .	Techniques for physically separating the test organisms from non-dissolved material, whilst maintaining contact with the water column, should be considered where physical effects are likely to be significant.

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
The substance is coloured	Absorption of light at relevant wavelengths may cause an indirect effect on aquatic plant growth by inhibiting photosynthesis. Strongly coloured solutions might make it difficult to observe effects in animals.	Since the amount of light absorbed will vary with solution concentration, effects seen at high concentration are not necessarily environmentally relevant. The endpoint for regulatory use should therefore be based on direct toxic effects. If the test has not been designed to indicate whether any observed effects are caused by light limitation, then the results cannot be used. Early algal studies may not have considered the effect of light absorption, and therefore all observed inhibition was assumed to be inherent toxicity. In the late 90s an approach known as the ETAD method was used. This attempted to compare direct and indirect contact of the test substance with the algae, with the indirect contact used to evaluate light inhibition only. If the results of each experiment comparable, it was interpreted that effects were only due to light inhibition. Such a result could be used to justify not using the algae results for classification or PNEC derivation. More recently, the ETAD method has been thought to be too simplistic for this evaluation, and instead the Manual of Decisions has been updated with the modified algae / Lemna approach as detailed below: The following adjustments to the standard algae growth inhibition test, Annex V method C.3 (or OECD guideline 201) have to be applied: • The irradiation (light intensity) should be in the highest end of the range prescribed in the method C.3 (or (draft revised) OECD guideline 201): 120µE m-2 s-1 or higher. • The light path should be shortened by reduction of the volume of the test solutions (in the range of 5 - 25 ml). • Sufficient agitation (for example by moderate shaking) should be performed in order to obtain a high frequency of exposure of the algae to high irradiation at the surface of the culture.	OECD (2000) provides a number of options for performing algal tests with coloured substances. See latest MoD decision, left. The 7-d Lemna growth test avoids the problem since the fronds grow at the water surface.
The substance is likely to be lost from the water column	The substance is volatile; losses may be particularly significant if the test is conducted using an open system. Vapour pressure, and more specifically the Henry's Law constant (H), are indicative of potential problems. If H is > 100 Pa.m3/mol, > 50% of the substance could be lost from the water phase-in 3-4 hours. Other factors in the test system may affect the rate of loss (e.g. vessel shape, aeration rate, etc). Volatilisation losses may also be significant for substances with H in the range of 1-10 Pa.m3/mol under vigorous mixing conditions. As a general rule vessels should be sealed during preparation and exposure and the headspace kept to a minimum. Problems with using sealed vessels are outlined in OECD (2000).). Within the ESR programme, two volatile substances styrene and 1,3 butadiene have been assessed. For the latter a combination of QSARs and read-across were used		

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
	to provide environmental data; 1,3 butadiene was also a known CMR, so avoiding exposure of the substance to laboratory workers was an additional consideration. For styrene, due to it being readily biodegradable, an additional problem was degradation in ecotoxicity test media lowering oxygen levels for test organisms. Normally this could be mitigated providing additional oxidation, however due to the volatility this was likely to increase substance loss. In the studies steps were taken to minimise degradation (e.g. vessel sterilisation), as well using a flow-through system supported by analysis throughout the test. QSARs were also used to support the test results. The substance is adsorptive to glassware, food and/or test organisms. This property often accompanies low water solubility, since hydrophobic chemicals usually prefer to partition to organic phases (i.e. substances with a log Kow >4 or bioconcentration factor >500). Where this occurs, the loss of concentration is usually rapid and exposure may best be characterised by the concentration at the end of the test. Other reasons for adsorption may be formation of ionic or hydrogen bonds negatively charged surfaces of the test vessel or the biological material. . The ESR assessments of tetrapropenylphenol and tris[2-chloro-1-(chloromethyl)ethyl] phosphate (TDCP) provide good examples where substance absorption was considered. The substance is unstable (i.e. degrades - abiotically, biotically or photolytically - or reacts) over the test duration. The loss may be so rapid that the substance itself cannot be tested, and/or specific degradation products may be formed that need consideration. See notes below on interpretation of exposure concentrations. The substance precipitates (e.g. because it has not truly dissolved despite the apparent absence of particulates, and agglomeration occurs during the test). In these circumstances, the L(E)C50 may be considered to be based on the concentration at the end of the test for classification purposes. Precipitation may occur as a result of degradation, e.g. an insoluble hydrolysis product or oxidation of test substance, other causes include complexation with media salts, pH change, oxidation. Note some substance may form an emulsion/dispersion, which can be tested as such – see surfactants discussion above. The substance bioaccumulates in the test organisms. This may be particularly important where the water solubility is low. The L(E)C50 may be calculated based on the geometric mean of the start- and end-of-test concentrations for classification purposes. It is necessary to determine whether appropriate methodology has been used (OECD (2000) describes a variety of methods to minimise the impact of these properties). In general, if test concentrations fall below 80% of nominal, measures should have been taken to reduce the decline for the test to be considered valid. This may require exposure regimes that provide for renewal of the test material (semi-static or flow-through conditions are preferred), and it is desirable that test concentrations are measured analytically at suitable time points throughout the test (for volatile , adsorptive unstable substances the latter is essential). These factors should be taken into account in deciding on the test data validity. It should be noted that semi-static and flow-through regimes may lead to an accumulation of organic debris and the development of excessive microbial populations. Test organisms may be stressed by cleaning. Special problems arise with respect to algal tests, which are generally static tests. Data providing an indication of the stability of the test substance under the test conditions may be derived from a review of existing data on the physical and chemical properties of the substance, or from a preliminary stability study (see OECD (2000) for further details). In the absence of analytically measured concentrations at least at the start and end of the test, no valid interpretation can be made and the test should be considered as invalid. Classification should account for the loss of the substance during the test, if relevant and possible. For example, if degradation occurs, it is necessary to determine whether it is the substance or the degradate that has been tested, and whether the data produced are relevant to the classification of the parent substance. Measured concentrations of the parent material and all significant toxic degradates are desirable. Where degradation is rapid (e.g. half-life < 1 hour), the available test data will frequently define the hazard of the degradation products since it will		

Difficult property	Potential problems with standard test procedures	Advice on interpretation	Possible refinements
	be these that have been tested. These data may be used to classify the parent substance in the normal way. Where degradation is slower (e.g. half-life > 3 days), it may be possible to test the parent substance and thus generate hazard data in the normal manner using a suitable renewal regime. The subsequent degradation may then be considered in determining whether an acute or chronic hazard class should apply. Where degradation rates fall between these two, testing of either parent and/or degradates should be considered on a case-by-case basis. There may be occasions when a substance may degrade to give rise to a more hazardous or persistent product (this may be determined from preliminary tests or non-testing methods). Leaving a stock or test solution of the parent substance for a period equal to 6 half-lives of the substance will generally be sufficient to ensure that the medium contains only degradation products, which can then be used for toxicity testing. In these circumstances, the classification of the parent should take due account of the hazard of the degradation product, and the rate at which it can be formed under normal environmental conditions. For risk assessment, PECs and PNECs should relate to the same compound(s). For example, the degradation half-life should be compared with the duration of the emission and the time taken for the emission to reach the receiving water. If degradation is rapid, only the degradation product(s) are important. If the substance degrades slowly, the degradation products may be irrelevant for the risk assessment if they are less hazardous than the parent. Between these two extremes, the substance effectively becomes a multi-component mixture. Interpretation of the available data will need to carefully assign effects and properties between the original substance and the degradation products. Non-testing approaches may help this decision, especially where the properties of the products have not been measured separately. In some cases, two risk assessments might be needed to explore the significance of the possible extremes (i.e. 'no degradation' and 'complete degradation'). Such analysis can guide which further measurements are needed to understand the significance of the properties and the extent of risk. Some substances adsorb to organic matter more strongly than might be expected from Kow (e.g. aniline reacts irreversibly with sediment components). In addition, adsorption to inorganic matter (which is the major soil and sediment component) is important for several substance types, including metals, dyestuffs, cationic substances, complexing agents and surfactants.		

Figure R. 7.8-6: Considerations for multi-component mixtures

***i.e. add PEC/PNEC values**

References

OECD (2000). Environmental Health and Safety Publications, Series on Testing and Assessment No. 23, Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures, Environment Directorate, Organisation for Economic Co-operation and Development, Paris, September 2000.

OECD (2001). Environmental Health and Safety Publications, Series on Testing and Assessment No. 27, Guidance Document on the Use of the Harmonised System for the Classification of Chemicals Which are Hazardous for the Aquatic Environment, Environment Directorate, Organisation for Economic Co-operation and Development, Paris, March 2001

Appendix 7.8-2: Information and its sources: *in vivo*

Test guidelines

a) Adopted OECD test guidelines for aquatic pelagic toxicity

Organism	F/S	Type of test	Test guideline (Year)	Exposure
Algae	F	Growth inhibition	201 (2006)	72 h
Lemna sp	F	Growth inhibition	221 (2006)	Up to 14 days
Daphnia sp.	F	Acute immobilisation	202 (2004)	48 h
Daphnia	F	Reproduction	211 (1998)	21 days
Fish	F	Acute toxicity	203 (1992)	96 h
Fish	F	Prolonged toxicity	204 (1984)	14 days
Fish	F/S	Early-life stage toxicity (FELS)	210 (1992)	30-60 days, species dependent
Fish	F/S	Short-term toxicity test on embryo and sac-fry stages	212 (1998)	Species dependent
Fish	F	Juvenile growth	215 (2000)	28 days

F = Freshwater organism S = Saltwater organism

b) Proposed OECD test guidelines for pelagic aquatic toxicity

Organism	F/S	Type of test	Project nr	Exposure	Additional
Daphnia	F	Enhanced reproduction	2.8	21 days	Endocrine endpoints
Copepod	S	Reproduction and development	2.1	20-26 days	
Mysid	S	Life cycle toxicity	2.13	60 days or longer	Endocrine endpoints
Amphibian	F	Thyroid toxicity	2.19	21 days	Endocrine endpoints
Fish	F	Fish embryo toxicity	2.7	Up to 6 days	
Fish	F/S	Life-cycle toxicity	2.12	Species dependent	Endocrine endpoints
Fish	F	Sexual development	2.14	60-90 days	Endocrine endpoints
Fish	F	Screening	2.18	21 days	Endocrine endpoints

F = Freshwater organism S = Saltwater organism

PROJECT 2.1 COPEPOD REPRODUCTION AND DEVELOPMENT

The test assesses the effect of chemicals on the development and reproduction of the harpacticoid copepods *Nitocra spinipes, Tisbe battagliai, Amphiascus tenuiremis* and the calanoid copepod *Acartia tonsa*. Newly hatched larvae (termed nauplia/metanauplia), are exposed to the test substance added to water at a range of concentrations. The test duration is usually 21 days, which is sufficient time for the control animals to reach adulthood, first egg sac females to be isolated individually and produce 2 or 3 broods of offspring. Effects on copepod development are measured by the time taken for nauplii to attain the first copepodite stage. At the end of the test, the total number of living offspring produced per parent animal alive at the end of the test is assessed. The survival of the

parent animals and time to production of first brood may also be reported. Other substance-related effects on reproduction (e.g. brood size, time interval between successive broods), and possibly intrinsic rate of increase, may also be examined.

PROJECT 2.7 FISH EMBRYO TOXICITY TEST

Newly fertilised eggs of zebra fish (*Danio rerio*), fathead minnow (*Pimephales promelas*) or Japanese medaka (*Oryzias latipes*) are exposed to chemicals for up to 48 hours. In case of any evidence of delayed toxicity, the test duration should be extended to a total of 6 days (for zebra fish), i.e. 2 days post hatch. The test is conducted in 24-well multi-plates, 10 embryos/test concentration and at least 5 concentrations. 2 to 3 independent runs per substance are recommended. After 24 and 48 hours incubation, four apical endpoints are recorded as indicators of acute lethal toxicity: coagulation of fertilised eggs, lack of somite formation, detachment of the tail bud from the yolk and lack of heart beat. Embryos are considered dead, if one of these endpoints is recorded as positive.

A comparable test was standardised (DIN 38415/A1; DIN 2001) in Germany and has replaced the conventional fish test for routine whole effluent testing. An ISO guideline is in the pipeline.

PROJECT 2.8 ENHANCED DAPHNIA MAGNA REPRODUCTION

This is an enhanced version of the "*Daphnia magna* Reproduction Test" (TG 211; OECD 1998). Offspring sex ratio and molt inhibition are evaluated as new endpoints. Sex of neonates can be differentiated under a stereo microscope by the length and morphology of the first antennae. Inhibition of molting can be examined by direct observation under a stereo microscope, as well as by comparing number of molts and/or duration of inter-molt period with that in control group(s).

PROJECT 2.12 FISH LIFE-CYCLE TEST

A comparison of a proposed fish full-life cycle test (FLCT) and a proposed fish two-generation test (TGT) is being conducted. This guideline is intended to be applicable to the fathead minnow (*Pimephales promelas*), medaka (*Oryzias latipes*), sheepshead minnow (*Cyprinodon variegatus*) and zebrafish (*Danio rerio*). The fish FLCT is initiated with fertilized eggs (P generation or F0) and the fish are continuously exposed through reproductive maturity, followed by assessment of the early development of the F1 generation. In contrast, in the fish TGT exposure is initiated with the mature male and female fish (P generation or F0): eggs are collected and the F1 generation is evaluated for embryo fertility, development, sexual maturation and reproduction.

Viability of F2 is also assessed. The main difference between FLCT and TGT is their relative potential for evaluation of the effect of maternal transfer of chemicals, which is evaluated once in FLCT and twice in TGT. Measurements are made of a number of endpoints in both P and F1 generations reflective of the status of the reproductive endocrine system, including the gonadal-somatic index (GSI), gonadal histology and plasma or whole body concentrations of vitellogenin. Additionally, plasma sex steroids (17β-estradiol, testosterone, 11-ketotestosterone) and thyroid hormones (T3/T4) may also be measured.

PROJECT 2.13 MYSID LIFE CYCLE TOXICITY TEST

This test evaluates reproductive fitness in two consecutive generations of mysids (preferably *Americamysis bahia*), starting with newly-released (< 24 h) individuals of the F0 generations and continuing until the first two broods (F2 generation) of the F1 generation. The overall test duration is normally 60 days or longer. Observational endpoints include growth, time to maturity, time to first brood release, interbrood duration, number and sex ratio of offspring.

PROJECT 2.14 FISH SEXUAL DEVELOPMENT TEST

This method is an extension of the existing OECD Test Guideline 210 (1992) Fish, Early-Life Stage (FELS) Toxicity Test, focusing on vitellogenin production and sexual development, i.e. sex ratio as determined via histological examination of the gonads. The test aims at detecting substances acting as estrogens, androgens or aromatase inhibitors in organisms at a very sensitive stage of their life-cycle. The test starts with fertilised eggs and lasts until sexual differentiation is completed (e.g. 60 to 90 days post hatch, depending on the fish species).

PROJECT 2.18 FISH-SCREENING TESTS

Reproductively active male and female fish of fathead minnow (*Pimephales promelas*), medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*) are housed in groups of 5 males and 5 females and exposed to test chemical for 21 days. Core endpoints as indicators of endocrine disrupter activity are gross morphology (i.e., secondary sexual characteristics) in sexually dimorphic species and vitellogenin levels in the serum or liver. Additionally the spawning status is checked daily in all groups, and quantified in some. Examination of gonadal histology is optional but will not be included as validated endpoint in the first draft TG.

PROJECT 2.19 METHODS IN AMPHIBIANS

The primary objective of the Amphibian Metamorphosis Assay is the evaluation of thyroid system disrupting activities of the individual test compound. The post-embryonic development (metamorphosis) of *Xenopus laevis* and the regulatory role played by thyroid hormones (TH) during this process are well characterised. In the assay, exposure of *X. laevis* tadpoles is initiated at developmental stage 51 and is continued for a total of 21 days. A sub-sampling of 5 tadpoles per treatment tank is performed at exposure day 7 for hind-limb length measurement. Tadpoles are exposed to 4 different concentrations of a test substance and a dilution water control. During the exposure period, apical morphological endpoints (developmental stage, hind limb length, whole body length) are assessed for treatment-related deviations from normal development and histological analysis of thyroid gland tissue is conducted with head tissue samples taken from test organisms. Chemical exposure is via the aqueous route achieved using a flow-through exposure regime..

OTHER TEST GUIDELINES - NATIONAL AND INTERNATIONAL STANDARD METHODS AND THEIR PUBLISHERS

Acceptable alternatives to the OECD tests (described above) are also published by the OPPTS, EU (Official Journal), U.S. EPA and organisations such as ISO and ASTM

Standard	Publisher	Web	Address
OECD	Organisation for Economic Co-operation and Development	http://www.oecd.org	OECD 2, rue André Pascal F-75775 Paris Cedex 16, France
EU	Official Journal of the European Communities. Annex V	http://ecb.jrc.it/testing-methods	European Chemicals Bureau TP582 Institute for Health and Consumer Protection Joint Reasearch Centre, Ispra Site European Commission Via fermi 1 I-21020 Ispra (VA), Italy
ISO	International Organization for Standardization.	http://www.iso.org	ISO Central Secretariat: International Organization for Standardization (ISO) 1, rue de Varembé, Case postale 56 CH-1211 Geneva 20, Switzerland
AFNOR	Association Française de Normalisation	http://www.afnor.fr	AFNOR Association Française de Normalisation 11, rue Francis de Pressensé 93571 La Plaine Saint-Denis Cedex,France
ASTM	American Society for Testing and Materials	http://www.astm.org	ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA, 19428-2959 USA
BSI	British Standards Institution	http://www.bsi-global.com	BSI British Standards 389 Chiswick High Road London W4 4AL, United Kingdom

Standard	Publisher	Web	Address
CAN	Environment Canada, Environmental Protection Series	http://www.ec.gc.ca	Environment Canada, Inquiry Centre 70 Crémazie St. Gatineau, Quebec K1A 0H3, Canada
DIN	Deutsches Institut für Normung	http://www.din.de	DIN Deutsches Institut für Normung e.V. Stabsstelle Kommunikation Burggrafenstraße 6 10787 Berlin, Germany
DS	Dansk Standard (Danish Standard Association)	http://www.ds.dk	Dansk Standard Kollegievej 6 2920 Charlottenlund, Denmark
NEN	Nederlands Normalisatie-instituut	http://www.nni.nl	NEN Postbus 5059 2600 GB Delft, The Netherlands
NS	Norges Standardiseringsforbund	http://www.standard.no	Standard Norge Postboks 242 1326 Lysaker, Norway
ÖNORM	Österreichisches Normungsinstitut	http://www.on-norm.at	ON Österreichisches Normungsinstitut Heinestraße 38 1020 Wien, Austria
OPPTS	US-EPA Office of Prevention, Pesticides and Toxic Substances	http://www.epa.gov/oppts/index.htm	US-EPA Office of Prevention, Pesticides, and Toxic Substances MC 7101M 1200 Pennsylvania Avenue, N.W. Washington, DC 20460, USA
SFS	Suomen (Finland) Standardisoimisliitto	http://www.sfs.fi	Suomen Standardisoimisliitto SFS PL 116, 00241 HELSINKI, Finland
SIS	Standardiseringskommissionen i Sverige	http://www.sis.se	SIS, Swedish Standards Institute Sankt Paulsgatan 6 118 80 Stockholm, Sweden

National and international standard methods / Guidelines (OECD, 1998):

Taxonomic group	**Fresh/ Salt**	**Species**	**Exposure time / endpoint**	**Guideline**
Algae	F	Selenastrum capricornutum Scenedesmus subspicatus Chlorella vulgaris	Short-term / Growth rate (Chronic)	US-EPA 1994 (40 CFR 797.1060, 40 CFR 797.1075, 40 CFR 797.1050)
	S	Skeletonema costatum Thallassiosira pseudonana Isochrysis galbana		
	F	Selenastrum capricornutum Scenedesmus subspicatus Chorella vulgaris	Short-term / Growth rate (Chronic)	ASTM (E 1218-90), FIFRA (§122-2), OECD (201), ISO (8692), NF (T90-304), DIN (38412 Teil 33), BS (6068: Section 5.10:1990), NEN (6506), SFS (5072), CAN (1/RM/25, 1992), EU (L 384 A Vol. 35 C.3)
	S	Skeletonema costatum Phaeodactylum tricornutum	Short-term / Growth rate (Chronic)	ISO (10253), BS (91/56211 DC), NEN (6506), SFS (5072)
Macrophytes	S	Champia parvula	Short-term / Reproduction (Chronic)	US-EPA (EPA/600/4-87/028)
Plants	F	Lemna gibba	Short-term / EC50 (Acute)	ASTM (E-1415-91), FIFRA (§123-2), US-EPA (1994)(40 CFR 797.1160)
Crustaceans	S	Mysidopsis bahia	Short-term / LC50 (Acute)	ASTM (E 1463-92), FIFRA (§72-3 c), US-EPA (EPA/600/4-90/027), US-EPA (1994): 40 CFR 797.1930)
	S	Artemia salina	Short-term / LC50 (Acute)	US-EPA (EPA/600/4-90/027)
	S	Penaeus aztecus Penaeus duorarum	Short-term / LC50 (Acute)	US-EPA (1994) 40 CFR Ch. 1 (7-1-92) Part 797.1970)

Taxonomic group	Fresh/ Salt	Species	Exposure time / endpoint	Guideline
		Penaeus setiferus		
	S	Nitocra spinipes	Short-term / LC50 (Acute)	SS (028106), DS (2209), ISO/TC 147/SC 5/WG 2N56
	S	Acartia tonsa	Short-term / LC50 (Acute)	ISO/TC 147/SC 5/WG 2N56
	S	Tisbe battagliai	Short-term / LC50 (Acute)	ISO/TC 147/SC 5WR 2N56
	F	Daphnia magna Daphnia pulex	Short-term / LC50 (Acute)	US-EPA (EPA/600/4-90/027), OECD (202), ASTM (E 729-88a), FIFRA (§72-2), ISO (6341), NF (T90-301), DIN (38412 Teil 11), BS (6068: Section 5,1:1990), NEN (6501), ONORM (M 6264), SFS (5052), SS (028180), DS (ISO 6341), CAN (EPS 1/RM/11, 1990), US-EPA (1994) (40 CFR 797-1300), EU (L 384 A vol. 35 C.2)
	F	Ceriodaphnia dubia	Short-term / LC50 (Acute)	ASTM (E 1295-89), US-EPA (EPA/600/4-90/027)
	S/F	Gammarus fasciatus Gammarus pseudolimnaeus Gammarus lacustris	Short-term/LC50 (Acute)	US-EPA (1994) (40CFR 795.120), CAN (EPS1/RM/26, 1992)
	S	Mysidopsis bahia	Long-term /survival, growth, fecundity (Subchronic)	US-EPA (EPA/600/4-87/028)
	S	Mysidopsis bahia Mysidopsis bigelowi Mysidopsis almyra	Long-term / life cycle (Chronic)	ASTM (E-1191-90), US-EPA (1994) (40 CFR 797.1950)
	F	Daphnia magna	Short-term / reproduction (Subchronic)	US-EPA (1994) (40 CFR 797.1330), OECD (202), NEN (6502)
	F	Daphnia magna	Long-term / life cycle (Chronic)	ASTM (E-1193-87), FIFRA (§72-4 C), US-EPA (1994) (40 CFR 797.1350)
	F	Ceriodaphnia dubia	Short-term / reproduction (Subchronic)	CAN (EPS 1/RM/21, 1992), US-EPA (EPA/600/4-89/001)

Taxonomic group	**Fresh/ Salt**	**Species**	**Exposure time / endpoint**	**Guideline**
Insects (mosquito)	F	Wyemyia Smithii	Short-term / LC50 (Acute)	ASTM (E-1365-90), FIFRA (§142-1)
Rotifers	F	Brachyonus	Short-term / LC50 (Acute)	ASTM (E-1440-91)
Bacteria	S	Photobacterium phosphoreum	Short-term / Light emission (Acute)	NF (T90-320), DIN (38412 Teil 34), ONORM (M 6609), ISO/TC 147/SC 5/WG 1, CAN (EPS/1/RM/24, 1992)
	F	Pseudomonas	Short-term / Growth (Chronic)	DIN (38412 Teil 8), NEN (6509 2e Ont w) ISO (DIS 10712. N133)
	F	Activated sludge	Short-term / respiration inhibition (Acute)	OECD (209), EU (L 133 vol 31 p. 118), ISO 9509
Amphibians	F	Xenopus	Short-term / teratogenesis (Subchronic)	
Fish	F	Brachydanio rerio Oncorhynchus mykiss Pimephals promelas Cyprinus carpio Oryzias latipes Poecilia reticulata Lepomis macrochirus Lepomis cyanellus Salmo gairdneri Oncorhynchus kistutch Salvelinus fontinalis Carassius auratus	Short-term / LC50 (Acute)	ASTM (E-729-88a), FIFRA (§ 72-1), US-EPA (EPA/600/4-90/027 + US-EPA (1994) 40 CFR 797.1440), OECD (203), ISO (7346-1-3), NF (T90-303+305), DIN (38412 Teil 15+20), BS (6068: Section 5,2; 5,3; 5,4:1985), SFS (3035+5073), DS (ISO 7346/1-3), CAN (EPS 1/RM/9), EU (L 383 A vol. 35 C.1)

Taxonomic group	Fresh/ Salt	Species	Exposure time / endpoint	Guideline
		Ictalurus punctatus Leuciscus idus		
	F	Poecilia reticulata	Short-term / LC 50 (Acute)	NEN (6504)
	F	Abassis macleayi	Short-term / LC 50 (Acute	OFR 54
	S	Sheepshead minnow Fundulus heteroclitus Menidia sp. Gasterosteus aculeatus Lagodon rhomboides Leiostomus xanthurus Cymatogaster aggregata Oligocottus maculosus Citharichthys stigmaeus Paralichthys dentatus Paralichthys lethostigma Platichthys stellatus Parophrys vetulus Clupea harengus	Short-term / LC50 (Acute)	ASTM (E729-88a), FIFRA (§72-3 a), US-EPA (EPA/600/4-90/027), SS (028189), CAN (EPS 1/RM/10)
Fish (cont)	F	Brachydanio rerio Pimephals promelas Cyprinus carpio Oryzias latipes	Long-term / growth (Subchronic)	OECD (204), ISO (10229-1), BS (93/500175 DC)

Taxonomic group	Fresh/ Salt	Species	Exposure time / endpoint	Guideline
		Poecilia reticulata Lepomis macrochirus Salmo gairdneri (Oncorhynchus mykiss)		
	F S	Brachydanio rerio Oncorhynchus mykiss Cyrinus carpio Oryzias latipes Carassius auratus Lepomis macrochirus Pimephales promelas Menidia peninsulae Clupea harengus Gadus morhua	Short-term / egg and sac-fry stages (Subchronic)	OECD (212)
	F	Pimphales promelas	Short-term / early life stage test (Subchronic)	CAN (EPS 1/RM/22, 1992, US-EPA (600/4-89/001)
	F	Oncorhynchus mykiss Salmo gairdneri Salvelinus fontinalis Esox lucius Pimephales promelas Catostomus commersoni Ictalurus punctatus	Long-term / early life-stage test (Subchronic)	ASTM (E-1241-92), FIFRA (§72-4 a), US-EPA (1994) (40 CFR 797.1600), SS (SS 028193), NS (4763), SFS (5501), CAN (EPS 1/RM/28, 1992)

Taxonomic group	Fresh/ Salt	Species	Exposure time / endpoint	Guideline
	S	Lepomis macrochirus Morone saxatilis Opsanus beta Cyprinodon variegatus Menidia menidia		
Fish (cont.)	F	Mogunda mogunda	Long-term / early life stage test (Subchronic)	OFR 52
	S	Cyprinodon variegatus	Long-term / survival, teratogenecity (Subchronic)	US-EPA (EPA/600/4-87/028)
	S	Cyprinodum variegatus Menidia beryllina	Long-term / survival, growth (Subchronic)	US-EPA (EPA/600/4-87/028)
	F	Salmo gairdneri Pimephales promelas Brachydanio rerio Oryzias latipes Oncorhynchus kisutch Oncorhynchus tschawytscha Salmo trutta Salvelinus fontinalis Salvelinus namaycush Esox lucius Catostomus commersoni	Long-term / hatching, survival, growth, malformations, behavoiur (Subchronic)	OECD (210)

Taxonomic group	Fresh/ Salt	Species	Exposure time / endpoint	Guideline
	S	Lepomis macrochirus Ictalurus punctatus Jordanella floridae Gasterosteus aculeatus Cyprinodon variegatus Menidia menidia Menidia penisulae		
Echinoderms	S	Arbacia punctulata)	Short-term / fertilization (Subchronic)	US-EPA (EPA/600/4-87/038), CAN (EPS1/RM/27, 1992)
Mussels	S	not specified	Short-term / LC50 (Acute)	ASTM (E-724-89), FIFRA (§72-3 b)
	S	Crassostrea virginica	Short-term / shell growth (Acute)	US-EPA (1994)(40 CFR 797.1800)

* Short-term £ 14 days, Long-term > 14 days

Databases

For the endpoint of aquatic toxicity Ecotoxdatabase, IUCLID, ECETOC database and N-class database may be useful sources of information. Other useful sources of information can be found through existing risk assessment or data evaluation programs such as ESIS, HERA and the OECD HPV program (SIDS). It is recommended that you consult the original scientific paper to ensure an understanding of the context of the data retrieved from the databases.

EAT (EUROPEAN CENTRE FOR ECOTOXICOLOGY AND TOXICOLOGY OF CHEMICALS (ECETOC) AQUATIC TOXICITY DATABASE (HTTP://WWW.ECETOC.ORG)

The ECETOC Aquatic Toxicity (EAT) database (ECETOC, 1993) contains more than 5450 entries on almost 600 chemicals, provides the most comprehensive compilation of highly reliable ecotoxicity data published in the scientific press in the period 1970 - 2000. The EAT 3 database is available as an Excel spreadsheet. For each entry there are 32 fields of information on the substance, test species, test conditions, test description, endpoint, results and source references. All the references are held at ECETOC; ECETOC AISBL, Avenue Edmond Van Nieuwenhuyse 4 Bte 6, B-1160 Brussels, Belgium.

Ecotoxdatabase

(http://www.epa.gov/ecotox/)

The database is maintained by the US-EPA and provides single chemical toxicity information on aquatic and terrestrial life for about 8400 chemicals. Peer-reviewed literature is the primary source of information encoded in the database. Pertinent information on the species, chemical, test methods, and results presented by the author(s) are abstracted and entered into the database. Another source of test results is independently compiled data files provided by various United States and International government agencies. Prior to using ECOTOX, you should visit the "About ECOTOX/Help" section of this Web Site.

ESIS (European chemical Substances Information System) (http://ecb.jrc.it/esis/)

ESIS is an IT System which provides you with information on chemicals, related to:

- EINECS (European Inventory of Existing Commercial chemical Substances),
- ELINCS (European List of Notified Chemical Substances),
- NLP (No-Longer Polymers),
- HPVCs (High Production Volume Chemicals) and LPVCs (Low Production Volume Chemicals), including EU Producers/Importers lists,
- C&L (Classification and Labelling), Risk and Safety Phrases, Danger etc...,
- IUCLID (International Uniform Chemical Database) containing information on approx. 10 500 different substances on the effects on human health and the environment.
- Priority Lists, Risk Assessment process and tracking system in relation to Council Regulation (EEC) 793/93 also known as Existing Substances Regulation (ESR).

HERA (Human and Environmental Risk Assessment) (http://www.heraproject.com)

HERA is a voluntary industry programme initiated by A.I.S.E. and CEFIC to carry out focused risk assessments of the ingredients of household cleaning and detergent products.

HSDB (Hazardous Substances Data Bank)

(*http://toxnet.nlm.nih.gov*)

This is a toxicology data file on the National Library of Medicine's (NLM) Toxicology Data Network (TOXNET®). It focuses on the toxicology of potentially hazardous chemicals. It is enhanced with information on human exposure, industrial hygiene, emergency handling procedures, environmental fate, regulatory requirements, and related areas. All data are referenced and derived from a core set of books, government documents, technical reports and selected primary journal literature. 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. HSDB is organized into over 5000 individual chemical records.

N-class database

(*http://www.kemi.se/nclass*)

The steering group for the Nordic Council of Ministers project on Environmental Hazard Classification is responsible for the continuous updating of the N-Class database. The database contains substances that have been discussed by the EC-Commission on the Classification and Labelling for environmental effects. Substance specific data, gathered from various documents that have been discussed at Commission working group meetings on environmental effects (mainly covering ecotoxicity), may be found in the N-Class database.

OECD Integrated HPV database

(*http://cs3-hq.oecd.org/scripts/hpv/*)

This database tracks all High Production Volume (HPV) chemicals through the process of investigation in the OECD programme on the Investigation of Existing Chemicals. Once agreed in the OECD, it shows the results of assessments as well as the actual reports and background information behind them.The database contains the list of HPV chemicals together with any annotations on each chemical provided to the Secretariat by Member countries, there are links to relevant documents.

When making the first evaluation of an existing chemical, a minimum set of data is necessary to determine its potential hazards. To ensure that such data are available, OECD developed the SIDS (Screening Information Data Set). The SIDS outlines the minimum data elements essential for determining whether or not a chemical requires further investigation

The database has a comprehensive search facility allowing searches to be made in a number of categories: e.g., chemical name, CAS number, sponsoring country, stage of investigation.

Members of the general public have “read only“ access to the database and so can follow the progress of a chemical both through and after its assessment. They can also obtain completed assessments on individual chemicals once these have been agreed in the OECD.

OHMTADS

(*http://www.nisc.com/cis/details/ohm-tads.htm*)

The Oil and Hazardous Materials/Technical Assistance Data System includes 1,402 MSDS-like fact sheets prepared by the US Environmental Protection Agency in the 1970s and 1980s. Each fact sheet deals with one chemical substance. The database is no longer updated, and some material in the database has been rendered incorrect over time by changes in regulatory requirements. However, the database still contains a wealth of still-useful data and references. Consequently, each record is presented with a warning about the age of the database and the need to verify critical information through more current sources. Users can retrieve records by CAS Registry Number (the preferred method), chemical name, and/or subject terms/phrases.

Riskline

(http://www.kemi.se/riskline)

Riskline contains peer reviewed information on both environment and health. The database is produced by the Swedish Chemicals Inspectorate, Sweden. Each reference in Riskline is furnished with a critical evaluation. It represents the unanimous opinion of a group of toxicological experts in the value of the research that is presented in the document. The evaluation might vary depending on the organization that reviewed the literature. All documents center around one chemical element of family of elements. Abstracts from the original documents are added to the unit record. All items are indexed and the chemical substances identified by CAS numbers.

Japanese Ministry of the Environment

(http://www.env.go.jp/en/chemi/)

The Ministry has conducted numerous aquatic toxicity tests in accordance with OECD TGs and GLP for many chemicals. The results from these tests are available on the indicated website.

Literature sources

ENVIRONMENTAL RISK LIMITS IN THE NETHERLANDS, REPORTS 601640001 PART I, II AND III (1999)

This report, produced by the National Institute of Public Health and the Environment (RIVM), documents risk limits, i.e. Maximum Permissible Concentrations (MPCs) and Negligible Concentrations (NCs) for approximately 200 substances in water, soil, sediment and air from the last decade in the framework of the project, 'Setting Integrated Environmental Quality Standards'. The objective was to present the procedures to derive the environmental risk limits to interested parties involved in environmental policy or environmental risk assessment of chemical substances. These risk limits are the none-regulatory standards used in the Dutch environmental policy. The reports include aquatic toxicity data on a number of chemicals. The quality of data has been assessed and ranked.

CANADIAN ENVIRONMENTAL QUALITY GUIDELINES (1999) ISSUED BY CANADIAN COUNCIL OF MINISTERS OF THE ENVIRONMENT.

Canadian Water Quality Guidelines for the Protection of Aquatic Life help to protect all plants and animals that live in lakes, rivers, and oceans by establishing acceptable levels for substances or conditions that affect water quality such as toxic chemicals, temperature and acidity. The guidelines

are based on toxicity data on the most sensitive species of plants and animals found in Canadian waters and act as science-based benchmarks for the protection of 100% of the aquatic life species in Canada, 100% of the time. The guidelines are available on CD-ROM and can be purchased from Canadian Council of Ministers of the Environment (http:/www.ccme.org).

US-EPA WATER QUALITY CRITERIA FOR AQUATIC LIFE

The Aquatic life criteria provide protection for plants and animals that are found in surface waters. The US-EPA develops these criteria as numeric limits on the amounts of chemicals that can be present in river, lake, or stream water without harm to aquatic life. Aquatic life criteria are designed to provide protection for both freshwater and saltwater aquatic organisms from the effects of acute (short term) and chronic (long term) exposure to potentially harmful chemicals. Aquatic life criteria are based on toxicity information and are developed to protect aquatic organisms from death, slower growth, reduced reproduction, and the accumulation of harmful levels of toxic chemicals in their tissues that may adversely affect consumers of such organisms. Developed criteria can be found at http://epa.gov/waterscience/criteria/aqlife.html.

References

OECD, 1998, Detailed Review Paper on Aquatic Toxicity Methods for Pesticides and Industrial Chemicals, OECD SERIES ON TESTING AND ASSESSMENT, Number 11, NV/MC/CHEM(98)19/PART

ECETOC, 1993. Aquatic Toxicity Data Evaluation. ECETOC technical report number 56. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

.

Appendix 7.8-4 Methodology for body burden approaches in aquatic effects assessment

The tests described in the TGD divide data collection into discrete compartments which can be classified as acute and chronic toxicity and bioaccumulation. In practice the data compilations are often obtained from different sources using different species or strains and form different media. The classical approach to risk assessment then compiles these data to arrive at an overall interpretation. In certain cases, there may be benefits in measuring, for example, bioconcentration and toxicity on the same species in the same experiment and in many cases standard tests can be ameliorated by addition of analytical measurement of the internal metric.

The major drawback of relating ecotoxicological effects to *external* concentrations only is in the cases where chemicals do not show (acute) toxic effects at aqueous concentrations below their aqueous solubility, while chronic effects; food-web cascading effects, or aggregate and mixture effects in combination with other non-chemical and chemical stressors may occur. Moreover, measuring external concentrations for low solubility substances is often extremely difficult. For this reason it may be preferable to use an alternative metric for measuring effects: internal body burden. The body burden at which mortality occurs is known as the Lethal Body Burden (LBB) and for sub-lethal endpoints Critical Body Burden (CBB).

This concept of critical body burdens (CBBs) is reasonably well-established, particularly for acute effects ((McCarty and Mackay 1993);(McCarty 1986)) of chemicals that act via a narcosis mode of action. A number of reviews have been made on this concept, (Barron et al. 1997; Barron et al. 2002), (Sijm and Hermens 2000) and Thompson and Stewart (2003). (McCarty 1991) recommended merging acute, chronic and bioaccumulation tests into one to greatly increase the information that could be obtained from a single test. This approach, although having a number of practical difficulties, could provide a more robust method for collating lethal concentration, BCF and chronic effects while adhering to the principle of validated guideline studies rather than performing three standard tests under subtly different conditions and trying to combine the results of the studies.

McCarty and Mackay (1993) were amongst the first to propose that the internal concentration of a chemical that is related to a biological effect is a more accurate and technically correct basis for comparing and ranking toxicity amongst chemicals and this was supported in later publications (Gobas *et al*. 2001) and Mackay, 2001).

The following is the range of body burdens originally tabulated in McCarty and Mackay (1993).

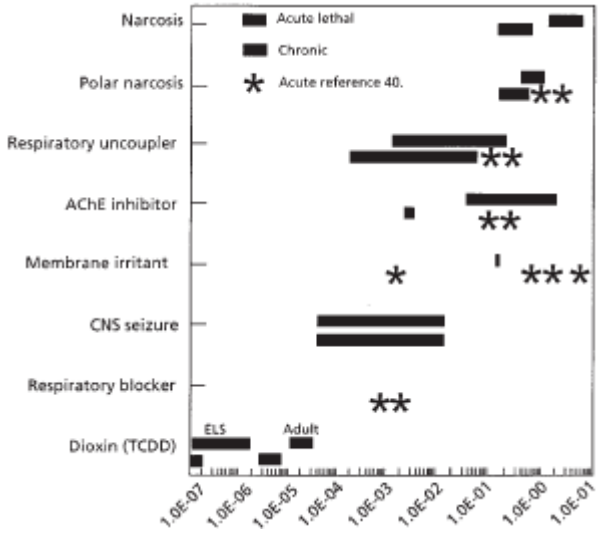


Figure R. 7.8-7 Calculated body burdens (in mmol l^{-1}) associated with different acute and chronic toxicity endpoints for fish exposed to eight categories of organic chemicals.

Similar ranges of L/CBB have also been published (Thompson and Stewart 2003) and shown to be relatively consistent with the Figure:

MoA I (acute = 1 to 10 mmol kg^{-1}, chronic = 0.1 to 1 mmol kg^{-1}) and

MoA II (acute = 0.5 to 2 mmol kg^{-1}, chronic = 0.05 to 0.1 mmol kg^{-1}).

Other MOAs tend to be lower but typically more variable (depending on species and whether LBB or CBB is considered (see Figure R. 7.8-7).

Advantages and disadvantages of the body burden approach

A LBB or CBB can either be measured directly during a study in which biological effects and chemical body burdens are measured in the same test organisms, or estimated indirectly. Indirect estimates can be on the basis of measured bioconcentration and critical external concentrations from different studies, so that LBB = LC50 x BCF and CBB = NOEC x BCF. Alternatively, indirect estimates can be made on the basis of data predicted by QSARs although the domain of applicability of the QSAR should be clearly demonstrated. This approach has been demonstrated for non-polar (Type I) narcotic substances (baseline toxicity) and polar (Type II) narcotic substances (McCarty 1986, McCarty *et al*. 1992, 1993).

The advantages of using the body burden are:

Knowledge of the CBB should reduce uncertainty in risk assessment as CBB can be used as a tool to help classify the known modes of action of chemicals.

Toxic effects should be additive within a MoA class because the CBB is independent of chemical structure, so mixture toxicity can be estimated more readily. Moreover, there is evidence that all

chemicals have narcotic MOA below the level at which their toxic action is exerted (Dyer et al., 2000).

QSARs based on Kow can be used to estimate CBBs for MoA I and II (McCarty 1986). Therefore, CBB can be used as a basis for building category approaches for classes of chemicals.

Data compilations are becoming available that allow theoretical aspects of the body burden approach to be explored and tested empirically, particularly for acute lethal effects caused by chemicals with MoA I and II.

Potentially, body burdens are a technically easier metric to measure than external concentrations for very poorly soluble or highly adsorbing and bioaccumulable substances.

Naturally, the CBB approach currently also has shortcomings however, the following shortcomings are common to both CBB and classical (external concentration) approaches:

1. a value for LBB cannot automatically be used to predict a CBB as the MOA may change from narcotic to non-narcotic for certain chemicals over the long term

2. The critical body burden of a chemical may differ between species, however the use of lipid normalisation may decrease. According to Sijm & Hermens (2000), it can be argued that, on a wet weight basis, fatter individuals may accumulate higher body burdens of toxicants before being affected. Lipid normalisation should, in this case, diminish intraspecies variation but according to the literature only reduces variation by 50%.

3. Other factors may influence CBB such as the sex, life-stage etc.

4. The CBB is usually measured in the whole body of a test organism, although effects may be expected to occur in specific target organs due to high concentrations causing severe damage in particular tissues (e.g., gill). However, this depends on the rate of movement of the chemical in the body.

There are also technical problems associated with precise measurement of CBB:

Body burden data in organisms that die early in a test may be lower than those in organisms that survive to the end of a test.. However, there is a similar issue for classical tests where LC_{10} occurs at an earlier stage than LC_{50} due to inter-individual variability.

Tests on body burden will also include the gut content and, in the case of invertebrates, cuticular adsorption of substance which cannot easily be subtracted to determine true body burden. However, the same applies to standard BCF and BAF tests and while these issues can interfere with the approaches used for CBB determination, they can generally be avoided with careful aforethought.

For classically tested invertebrates (e.g. *Lumbriculus* or *Daphnia*) it may be difficult to provide sufficient biomass to achieve quality analytical results. Biomass is an important consideration to take into account prior to conducting the experiment particularly when bioaccumulation is low.

Use of total radioactivity to measure body burden, without measuring parent compound specifically, does not take into account biotransformation and potential incorporation of the metabolites into the biomass. This can lead to gross overestimations of the body burden.

No normalised studies exist today which take body burdens into account. However, experienced ecotoxicologists should be capable of modifying existing tests to include both bioaccumulation and

toxicity in the same design. While any single study would use more animals than a study not including body burden, collectively there are possibilities for reducing the total number of animals used.

Some data indicate that the body burden technique may not be suitable for substances with a low log Kow (<1). More evidence for this is needed, however, it should be recognised that most applications for the CBB approach really become useful at higher values of log Kow.

Use of body burden data in risk assessment

There are many areas where the generation of body burden data can provide results which can be used in risk assessment: in helping to clarify or form chemical groups and to identify MOA; increasing confidence in data; potential simultaneous provision of BCF and toxicity reducing animal use, for example. Especially, when testing difficult substances it may not even be possible to use standard testing techniques based on aquatic toxicity. In such cases L/CBBs, used in conjunction with QSARs and/or read-across from less difficult substances and quality physico-chemical data, may provide a more reliable data set than standard techniques. The use of such an approach should be reviewed on a case-by-case basis also taking into account the level of technical input required to achieve a suitable result.

Conclusion on body burden techniques

The document provides an overview of the current state of the science for body burden methodology, advantages and disadvantages. There is good experimental evidence to support the hypothesis that Critical Body Burden (CBB), at least for acute lethal toxicity is relatively constant for substances with narcotic mode of action. The CBB approach has been recommended for use in risk assessment (Gobas *et al*. (2001) and Mackay (2001)) for single substances and could help in category approaches. It could also be used to help assess risk of multiple constituent compounds.

If there is information on the critical body burden of a substance in an (aquatic) organism this information could help to identify whether or not the chemical is a baseline narcotic chemical or has a more specific mode of action and thus would provide an indication of its aquatic toxicity.

References related body burden

Barron M, Anderson M, Lipton J, Dixon D. 1997. Evaluation of critical body residue QSARs for predicting organic chemical toxicity to aquatic organisms. SAR QSAR Environ Res 6:47-62.

Barron M, Hansen J, Lipton J. 2002. Association between contaminant tissue residues and effects in aquatic organisms. Reviews of Environmental Contamination and Toxicology 173:1-37.

Dyer SD, White-Hull CE, Shephard BK. 2000. Assessment of chemical mixtures via toxicity reference values overpredict hazard to Ohio fish communities. Environmental Science and Technology, 34, 2518-2524

Gobas F, Purdy R, Granville G, Cowan-Ellsbery C, Gannon J, Lewis M, Socha A, Chapman P. 2001. A proposal for hazard identification of organic chemicals based on inherent toxicity. SETAC Globe 2(2):33-34.

McCarty L. 1986. The relationship between aquatic toxicity QSARs and bioconcentration for some organic chemicals. environmental Toxicology and Chemistry 5:1071-1080.

McCarty L. 1991. Toxicant body residues: implications for aquatic bioassays with some organic chemicals. Mayes M, Barron M, editors. 183-192 p.

McCarty LS, Mackay D, Smith AD, Ozburn GW, Dixon DG. 1992. Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: neutral narcotic organics. Environ Toxicol Chem 11:917-930

McCarty LS, Mackay D, Smith AD, Ozburn GW, Dixon DG. 1993. Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: polar narcotic organics. Ecotoxicol Environ Saf 25:253-270.

McCarty L, Mackay D. 1993. Enhancing ecotoxicological modeling and assessment. Environ Sci Technol 27:1719-1728.

Sijm D, Hermens J. 2000. Internal effect concentrations: Link between bioaccumulation and ecotoxicity for organic chemicals. . Beek B -V-J, editor. Berlin, Germany: Springer-Verlag 167-199 p.

Thompson R, Stewart K. 2003. Critical Body Burdens: a review of the literature and identification of experimental data requirements. . Brixham Environmental Laboratory, AstraZeneca, Brixham, Devon, UK.

Appendix 7.8-5 Assessment of available information on endocrine and other related effects

This chapter is appended to the main guidance document on aquatic toxicity testing. It provides guidance for the evaluation of information relating to (potential) endocrine activity of a substance or long-term adverse effects on development and/or reproduction in aquatic organisms. As this kind of information is not part of the standard information requirements set out in REACH Annexes VII-X (see below), this part of the guidance is based on the evaluation of available information and none of the screening and testing methods discussed has been fully validated or approved as OECD Test Guideline (Status January 2007). Relevant information on the assessment of (potential) endocrine activity in aquatic organisms may also be derived from *in vitro* studies, mammalian screening assays for endocrine activity and other human health endpoints from repeated-dose toxicity, carcinogenicity and reproductive toxicity studies.

Endocrine disrupting guidance

DEFINITION

According to a widely accepted consensus reached at an international workshop in Weybridge, UK, in 1996 (which was later also adopted by OECD expert groups) *"an endocrine disruptor is an exogenous agent that causes adverse health effects in an intact organism, or its progeny, consequent to changes in endocrine function."*

"Endocrine disruption" is not a toxicological endpoint *per se* but a functional change of the endocrine system which may involve a variety of molecular mechanisms and which may result in adverse health effects in an organism or its progeny. This guidance document distinguishes between the identification of an endocrine mode of action and the characterisation of sub-lethal chronic and adverse effects on development and reproduction, which may also arise from other mechanisms of toxicity; the causal link between an endocrine mode of action and an adverse effect should be established to meet the Weybridge/OECD definition of an endocrine disruptor.

OBJECTIVE OF THE GUIDANCE

Endocrine disruption is the occurrence of adverse effects on development or reproduction of aquatic organisms due to a substance's endocrine activity. Such adverse effects, particularly involving reproduction and development, are of high relevance for the assessment of the potential hazards a substance may pose to the aquatic environment.

The guidance in this chapter is supposed to cover the following cases of available information beyond the standard information requirements:

- information indicating potential endocrine activity in aquatic organisms (from human health endpoints, molecular structure, or non-standard *in vitro* assays)
- information on an endocrine mode of action in aquatic organisms
- information on adverse effects on reproduction or development of aquatic organisms

Available information on adverse effects on development or reproduction should be considered for use in classification, the chemical safety assessment, and the PBT assessment in regards to the toxicity properties of a substance.

Furthermore, if a clear link between serious adverse effects and an endocrine mode of action can be established, the substance may fall under the provisions of Article 56 f), which specifies that *substances - such as those having endocrine disrupting properties (…) – for which there is scientific evidence of probable serious effects to human health or the environment which give rise to an equivalent level of concern* to those of CMR, PBT or vPvB substances may be included in Annex XIV of substances subject to the authorisation procedure. The inclusion will be decided on a case-by-case basis following the preparation of an Annex XV dossier by the Competent Authorities.

Information requirements

As indicated above, for registration of a chemical, there is no requirement set out in REACH Annexes VII to X to provide information on the endocrine activity of a substance or on a substance's reproductive or specific developmental toxicity in aquatic organisms.

However, according to Article 12, the information specified in Annexes VII-X is to be seen as a minimum requirement. The technical dossier shall include all physico-chemical, toxicological and ecotoxicological information that is relevant and available to the registrant. This general requirement is confirmed with regard to the chemical safety report and the safety data sheets in REACH Annexes I, II, and VI.

If, in the course of evaluation of available information, it is indicated that a substance displays an endocrine mode of action in aquatic organisms, this may constitute a concern that requires further investigation regarding potential adverse effects on development or reproduction. Such investigations may be requested on a case-by-case basis by a Member State, when performing the substance evaluation of a registration dossier (Article 45). This provision includes the request of specialised studies not covered by the REACH Annexes VII-X, such as the endocrine-specific studies described in this Appendix.

Information and its sources

NON-TESTING DATA

Non-testing data include information derived from SARs, QSARs, read-across and chemical categories. The general principles how to generate information by these methods are explained in the main part of this guidance document. Models are under development under the umbrella of OECD and ECB programmes for specific endocrine-related mechanisms, in particular for estrogen and androgen receptor binding (see Netzeva et al 2006; Saliner et al. 2006; for a recent overview of models see Devillers et al, 2006; for structural requirements specific for ER binding see Fang et al, 2001; for structural requirements specific for AR binding see Fang et al, 2003; Tamura et al, 2006).

Due to availability and quality of experimental data, more SAR and QSAR models are available for mechanism-related endpoints than for endocrine activity in intact organisms and for long-term adverse effects. However, the development of models that can predict *in vivo* effects, in view of their saving potential, may become more important in the future. Among the models (SARs and

QSARs) that predict mechanism-related endpoints, more models were developed for estrogenic activity compared to androgenic activity.

Along with the classical SAR and QSAR models, a number of 3-dimensional QSARs (3D QSARs, derived from Comparative Molecular Field Analysis, CoMFA) and docking studies were published in the literature. There is a good scientific basis for the development of the latter models since most of the endocrine disrupting effects are provoked by binding of chemicals to specific receptors (i.e. interactions, suitable for molecular modelling). However, there are still technical constraints in the transferability of such models for quantitative application unless the result of them is presented in different form (e.g. translated into structural alerts).

There is a large range of computational models that have been successfully applied to model endpoints, related to endocrine disruption. These range from structural features and structural alerts[8] (e.g. the presence of steroid skeleton, diethylstylbestrol skeleton or phenolic ring increase the probability of a chemical to be a binder to the estrogen receptor), to pharmacophore queries, to different discriminant models for assignment to an activity class (e.g. derived from linear discriminant analysis, k-Nearest neighbour modelling, decision tree analysis, biophore-type analysis, common reactivity pattern analysis etc.) to various quantitative models for prediction of potency, derived from local (e.g. congeneric) or global (diverse) data sets. The descriptors in the models also vary from structural fragments, through various hydrophobic, steric and electrostatic descriptors, to steric and electrostatic fields in CoMFA analysis and energies in docking studies. The choice of descriptors and modelling technique is largely dependent on the purpose and data series and no single recommendation can be given but rather critical and realistic evaluation of the models and underlying data is required depending on the problem to be solved.

Testing data

Throughout this Appendix, laboratory (experimental) methods are further divided into *screening assays* and (confirmatory) tests. In this sense, *screening assays* are lower tier *in vitro* or *in vivo* investigations which allow the identification of a potential endocrine mode of action of a substance, while definitive or confirmatory tests are higher tier *in vivo* methods to confirm the screening results and to characterise any adverse effects that may result from such a mode of action. Note should be taken that the term *screening assay*, in this context, does not relate to a blind screening of large numbers of chemicals. All of the methods described below are endocrine-specific studies that will only be relevant for a limited number of substances.

IN VITRO SCREENING DATA

At present, validated *in vitro* assays and internationally accepted Test Guidelines for regulatory purposes are not yet available. However, molecular mechanisms of the endocrine system, especially of the sexual hormone system of vertebrates, are well characterised and a large number of *in vitro* assays are used in scientific research. Although the basic principles have been applied to biological material from a variety of species, including aquatic vertebrates, assays based on mammalian systems are usually in the most advanced stage of development as expressed by their validation

[8] A discrimination between structural feature and structural alert could be done. For example, a tert-butyl moiety and phenol group are structural features associated with high potential for estrogen binding. However, the combination is viewed as a structural alert for estrogenicity only if the two functional groups are in p-position to each other, while, for example, o-position is not linked to a receptor-mediated gene activation.

status. Given the similarity of endocrine systems across vertebrate taxa, these assays may also provide valuable information on the assessment of potential endocrine activity of chemicals in aquatic organisms, in particular fish.

The following *in vitro* assays for the detection of possible endocrine activity of substances were selected for further development with the aim of validation for regulatory use. They are at different stages of development, validation and regulatory acceptance; their status in 2006 is indicated below.

ESTROGEN AND ANDROGEN RECEPTOR BINDING ASSAYS

Principle: Binding of a hormone to its receptor in the cytosol is an early event in the pathway of hormonal regulation. Assays that study the capacity of xenobiotic substances to compete with natural hormones from their binding sites have been developed with estrogen and androgen receptors from several species in different cellular systems. This type of assay cannot predict whether the binding of a substance to a hormone receptor will result in its activation (agonistic activity) or inhibition (antagonistic activity).

Status: Prevalidation of two receptor binding assays within the integrated project ReProTect funded by the 6th Framework Programme of the European Commission is now continuing under the umbrella of the OECD into validation led by the US-EPA and in collaboration with Japan. The US-EPA has completed validation of an assay based on the androgen receptor from rat prostate cytosol and conducted studies on the nature of binding interaction for 50 structurally diverse chemicals with the estrogen receptor from rat uterine cytosol (Laws et al. 2006).

TRANSCRIPTIONAL ACTIVATION (REPORTER GENE) ASSAYS

Principle: The active ligand-receptor complex translocates into the cell nucleus, where it aligns to specific DNA sequences and induces gene transcription. Incorporation of recombinant hormone-responsive gene elements and their promoters together with elements encoding easily detectable proteins into suitable host cells allows the detection of hormone receptor activation by visualising the response at the gene transcription level. As these assays can only show receptor activation, while antagonistic receptor interactions remain undetected, a positive test result does not always mean that exposure to the substance would result in an agonistic effect *in* vivo. The relevance of these genetically engineered systems to *in vivo* dose response of endogenous receptor and target genes has been evaluated in the Japanese Report in peer review at the OECD (see below).

Status: Validation of the Stably Transfected Transcriptional Activation (TA) Assay to Detect Estrogenic Activity was performed in Japan for ER agonists and is at the stage of peer-review within the OECD Test Guidelines programme. Prevalidation of four transcriptional activation assays for ER and AR (anti)agonists detection has been carried out within the integrated project ReProTect funded by the 6th Framework Programme of the European Commission and these are now progressing to validation.

VITELLOGENIN ASSAYS

Principle: Activation of the estrogen receptor in the liver of fish induces the biosynthesis of the egg yolk protein vitellogenin (VTG). Based on this principle, assays have been developed using primary cultured hepatocytes (e.g. from medaka or rainbow trout) to assess the influence of substances on VTG production via estrogenic or anti-estrogenic activity.

Status: This assay has been studied in several common fish species, with most data available for mature male rainbow trout and carp. The sensitivity of the cell cultures and the methods of detection of VTG protein by ELISA are being validated while those measuring VTG mRNA, using RT-PCR, still need to be validated.

STEROIDOGENESIS ASSAYS

Principle: Certain cell cultures express the enzymatic systems to metabolise cholesterol via native biosynthetic pathways into the final active steroid hormones such as androgens and estrogens in sufficient quantities for analytical determination of the rate of steroid synthesis. This provides a basis to develop an *in vitro* assay for stimulators and inhibitors of steroidogenic pathways relevant to vertebrates (see OECD Draft Detailed Review Paper on Steroidogenesis, May 2002). A particular focus of investigations is placed on the enzyme aromatase, which converts androgens into estrogens (see OECD Draft Detailed Review Paper on Aromatase, February 2002).

Status: Pre-validation work within the OECD framework is in progress for an assay based on the H295 human adrenocortical carcinoma cell line that has been shown to express all of the key enzymes necessary for steroidogenesis. The US-EPA is conducting prevalidation studies on human recombinant aromatase.

The latest information on the status of *in vitro* methods that are under development can be obtained from the ECVAM websit (current address: http://ecvam.jrc.it).

IN VIVO SCREENING AND TESTING DATA

Principle: Intact organisms are exposed through the water to the chemical in a range of sub-lethal concentrations for a period of a few weeks at minimum. Males and females are tested and a number of endpoints are measured to either trigger further investigation or conclude on the absence of concern. Biomarker endpoints will play an important role in screening whereas reproductive and developmental landmarks will be assessed in long-term toxicity testing.

Status: At present, there are no validated *in vivo* screening assays for the identification of substances with potential endocrine activity in aquatic organisms or test methods for the investigation whether a substance with endocrine activity has adverse impact in aquatic organisms. However, a number of methods are used in scientific research (see monographs No. 21, 55, and 57 in the OECD Series on Testing and Assessment). The performance of such methods is not included in the minimum requirement by REACH but for some substances relevant information may be available, e.g. from the scientific literature. For these cases, the compilation of available methods is given below as an orientation about the current state of development in the field of endocrine screening and testing and as references for the evaluation of older studies. The following methods were selected for further development with the aim of validation for regulatory use for the detection of endocrine activity or the characterisation of chronic effects on the development and reproduction of aquatic organisms. They are at different stages of development, validation and regulatory acceptance; their status in 2006 is indicated below.

VERTEBRATES

In relation to the sexual hormone system of fish, a range of methods is under development and validation, covering different levels of biological complexity.

Screening Assays

21-Day Fish Screening Assay, draft TG proposal (OECD, 2004)

This assay is proposed for the detection of estrogenic, androgenic or aromatase inhibiting substances in adult organisms which have reached sexual maturity. It can be run with several common fish species: zebrafish, fathead minnow, medaka and possibly the three spined-stickleback. The assay lasts over a period of 21 days. Core endpoints are VTG levels in the serum or liver (medaka), which indicate disturbances of the estrogenic balance, and secondary sex characteristics in sexually dimorphic species (not in zebrafish), which are liable to disturbances of the androgenic balance. The OECD validation studies are completed and the peer-review will take place early 2007 (see monographs No. 47, 60, and 61 in the OECD Series on Testing and Assessment).

Confirmatory Tests

Fish Sexual Development Test, draft TG proposal (OECD, 2006)

This method has been proposed as an extension of the existing OECD Test Guideline 210 (1992) Fish, Early-Life Stage (FELS) Toxicity Test. The enhancements focus on sexual development, i.e. sex ratio as determined via histological examination of the gonads, and on VTG production. The test aims at investigating the impact of substances acting as estrogens, androgens or aromatase inhibitors in organisms at a very sensitive stage of their life to endocrine activity. It can be run with several common test species: zebrafish, fathead minnow, medaka and possibly the three-spined stickleback. The test starts with fertilised eggs and lasts until sexual differentiation is completed (e.g. 60 to 90 days post hatch, depending on the fish species). After test development work in Denmark, the initial OECD validation study for fathead minnow and zebrafish has recently been initiated.

Fathead Minnow Reproduction Test, draft TG proposal (US-EPA, 2001):

A draft proposal for a fathead minnow reproduction test, including vitellogenin, secondary sex characteristics, gonad histopathology, fecundity and fertility assessments, is being validated in the United States. The test duration is 42 days, with 21 days of pre-exposure where fecundity is recorded daily, and 21 days of chemical exposure. The US-EPA validation programme is in progress and guidance documents should be developed for the interpretation of gonad histopathology.

Fish Full Life Cycle / 2-Generation Test

These tests allow an assessment of chronic effects on developmental and reproductive endpoints (see OECD Draft Detailed Review Paper on Fish Two-Generation Toxicity Test and Proposal for a Fish Two-Generation Test Guideline, March 2003). The most complete test design, which allows assessment of trans-generational transfer of effects, begins with exposure of adult, reproducing fish (F0 generation) and continues until in-life biological effects of the F2 generation can be determined. This time point as well as the total test duration may vary considerably depending upon the species used.

Measurements include developmental and reproductive endpoints (hatching, sex ratio, survival, growth, fecundity, fertility and behaviour) as well as biochemical, histological and morphological markers that are indicative of specific mechanism of endocrine disruption. The validation is under preparation. Results from such tests have already been used in risk assessments of specific substances of concern within the EU priority existing substances programme and in the authorisation of pesticides.

21-Day Amphibian Metamorphosis Assay, draft TG proposal (OECD, 2005)

This assay was developed for the detection of chemicals affecting the thyroid hormone system in amphibian species (see monograph No. 46 in the OECD Series on Testing and Assessment). The metamorphosis of amphibians, and in particular *Xenopus laevis,* the test species in this assay, is a well-studied phenomenon under the dependence of thyroid hormone signalling. Development stage, whole body length, hind-limb length and thyroid histology are the endpoints measured during the assay. The assay lasts for 21 days; hind-limb length is measured after 7 days and other endpoints are measured at termination of the assay. The test allows the characterisation of adverse effects on amphibian metamorphosis and growth as well as the identification of a thyroid disruptive mode of action, which may also be of relevance for other vertebrate species. Validation of this test method is ongoing.

INVERTEBRATES

The endocrine systems of aquatic invertebrates differ considerably from those of vertebrates and the knowledge in this field is less advanced. Consequently, consideration of specific endocrine-related endpoints in long-term invertebrate testing is only at the beginning (see also monograph No. 55 in the OECD Series on Testing and Assessment) of its development and its status and implication should be checked carefully:

Confirmatory Tests

Enhanced Test Guideline 211, Daphnia Magna Reproduction Test, (OECD, 2006)

Principle: This method is an enhancement of TG 211 which is intended to detect chemicals interacting with the hormone system of aquatic arthropod species, i.e. chemicals acting like the juvenile hormone or like ecdysteroids. In addition to the traditional endpoints measured in the existing Daphnia reproduction test, the new endpoints are offspring sex ratio and molt inhibition. This enhanced version has the same exposure duration as the existing TG 211, but additional technical efforts and time are required for the microscopic evaluation of the endpoints.

Status: The validation study is on-going in the OECD TG programme with Japan as lead country.

Other Test Guideline projects are currently in progress for marine or estuarine species, where development and reproductive endpoints are assessed. These assays are not intended to specifically identify endocrine modes of action:

Copepod Development and Reproduction Test, draft TG proposal (OECD, 2005)

This test examines the development and reproduction of marine harpacticoid and calanoid copepod species. Eggs or newly hatched larvae (< 24 h) are exposed for 20-26 days. Endpoints are larval mortality, larval development rate and reproductive success. The validation study is in progress in the OECD TG programme with Sweden as lead country.

Mysid 2-Generation Test, draft TG proposal

This test evaluates reproductive fitness in two consecutive generations of mysids (preferably *Americamysis bahia*), starting with newly-released (<24 h) individuals of the F0 generations and continuing until the first two broods (F2 generation) of the F1 generation. The overall test duration is normally 60 days or longer. Observational endpoints include growth, time to maturity, time to first brood release, interbrood duration, number and sex ratio of offspring. The pre-validation is ongoing in the United States under OECD auspices.

Evaluation of information

This section attempts to assist the user (e.g. registrant) in judging and ranking the adequacy (i.e. reliability and relevance) of information related to (potential) endocrine activity of a substance or its reproductive and developmental toxicity towards aquatic organisms. Since information of this kind is not part of the REACH information requirements, the following considerations are supposed to apply to those cases where this information is already available, e.g. from the scientific literature, or where it is specifically requested by a CA, e.g. in the course of substance evaluation. This is a relatively new area of testing and assessment where information needs to be evaluated carefully on a case-by-case basis.

NON-TESTING DATA

The evaluation of QSAR results consists of 1) evaluation of the validity of the model and 2) evaluation of the reliability of the individual model prediction. Guiding principles are explained in the general introduction to the TGD as well as in the main text on aquatic toxicity. Guidance on the application of grouping approaches (read-across and chemical categories) is given in the general introduction.

A special attention deserves the way, in which the activity class is assigned for development of the model, if it is intended to discriminate between active and inactive chemicals. The cut off, if such utilized to obtain binary classification from continuous data, should be clearly described when arguing the validity of the model prediction. Generally, the classification models tend to demonstrate higher accuracy than those predicting continuous values but the borderline predictions will need additional consideration. Nevertheless, both types of models should be evaluated according to the OECD principles and commonly encountered pitfalls (e.g. over-fitted models), described in the cross-cutting guidance on (Q)SAR, should be avoided. The global models, derived on diverse data sets, have generally larger domains of applicability but local models can be preferred if available for a specific chemical of interest. An understanding of structural features that form structural alerts is highly desirable and mechanistic interpretation of models and descriptor combinations should be looked for. Finally, the use of several models is expected to increase the confidence in the prediction but expert judgment might be required in case of contradicting results (e.g. the chemical is predicted active in classification model but with extremely low activity from a potency model, or vice versa).

SCREENING AND TESTING DATA

IN VITRO SCREENING DATA

Guiding principles to judge the adequacy of information obtained from *in vitro* assays are explained in the general introduction to the TGD as well as in the main text on aquatic toxicity (it should be noted that for the assessment of potential endocrine activity, data from mammalian systems may also provide information of relevance to aquatic organisms).

IN VIVO SCREENING DATA

Guiding principles of evaluating the reliability and relevance of *in vivo* data are explained in general parts of this guidance document. In addition, many of the specific considerations for aquatic test systems and organisms detailed in the main text on aquatic toxicity apply.

The purpose of *in vivo* studies for the investigation of endocrine activity of chemicals is to determine 1) whether the chemical is active on the endocrine system of aquatic organisms (e.g. vitellogenin induction as indicator of estrogenic activity), and 2) whether this mechanism induces adverse effects in long-term studies (e.g decrease in the number of offspring, effect on sex ratio in developing organisms).

21-Day Fish Screening Assay, draft TG proposal (OECD, 2004)

For the results to be meaningful, the vitellogenin data in control males and females should be within the range reported in the literature and indicated in the draft test guideline. For test results to be considered positive, significant responses should be observed at sub-lethal concentration (e.g. 0.5 or 0.1 times the LC_{50}; this value would need further discussion and agreement). Importantly, a homologous ELISA method (using standard VTG from the same species and homologous antibodies) should be used. Any loss of biological sample and any deviation from the protocol should be reported. As experience with compounds that are negative for estrogenic modes of action and experience with the rate of false positives for the VTG endpoint is limited, some caution with positive results is currently necessary.

For the evaluation of androgenic substances, a fish species should be used, which possesses the necessary characteristics to determine an endpoint relevant for androgenic stimulation, for instance secondary sex characteristics or an androgen-sensitive biochemical marker such as spiggin induction in the stickleback. In the case of suspected androgen activity fathead minnow, medaka, or stickleback are therefore the only recommended test species in a fish screening assay. Zebrafish is not suitable for the evaluation of androgenic substances in this assay.

No response on the endpoints measured in this assay indicates that the substance does not act as estrogen or androgen agonist or aromatase inhibitor/estrogen antagonist in fish *in vivo*. However, such a test compound may still have endocrine activity mediated through other, non-investigated mechanisms. Together with partial and full-life cycle studies that include developmental and reproductive parameters, these data can be used in a *Weight of Evidence* assessment whether adverse effects may be occurring through the covered endocrine modes of action.

IN VIVO TESTING DATA

Fish Sexual Development Test, draft TG proposal (OECD, 2006)

The current TG210 is suitable for the characterisation of a substance's adverse effects on fish survival, growth and development. The proposed extension, whether an enhanced or separate Test Guideline, focuses on a more detailed evaluation of sexual development, where the sex ratio and the production of vitellogenin are the main core endpoints. The discussion and attention for the evaluation of data should be focused on the statistical analysis and interpretation of the sex ratio endpoint. There may be concerns on the interpretation of results, due to a natural high variability in the sex ratio (i.e. male to female ratio can naturally vary between 35-65%) in control populations. Consequently, the value of "x" in EC_x currently poses question for a regression analysis (i.e. x=10 is not realistic, x=25 may be possible). Alternatively, if the LOEC/NOEC determination is the objective of the assay, a large number of replicate tanks (> 4) is necessary to level off the between-replicate variability and maintain sufficient power of the assay. Solutions to level-off the variability of the sex ratio exist, like the increase of the number of egg clutches (minimum of 5) used at the start of the test. When evaluating data from this test, attention should be paid to such test parameters and adherence to validity criteria specified in the test guideline.

Fathead Minnow Reproduction Test, draft TG proposal (US-EPA, 2001):

Care should be exercised in the evaluation of fecundity and gonad histopathological findings to differentiate toxic response which may not always be indicative of specific reproductive toxicity. An analysis of the data in a *Weight of Evidence* approach is foreseen and should be documented. The data should be transparently reported, especially for gonad histopathology, so that a transparent judgement can be made of the nature and reliability of the responses observed and whether the results are sufficient to conclude on the cause of the effects on reproductive capacity. Guidance documents are in preparation in the US and the OECD to assist pathologists in preparing the samples and evaluation the slides in a standardised fashion.

Fish Full Life Cycle / 2-Generation Test

These tests allow an assessment of apical developmental and reproductive endpoints. Effects observed in these studies are of high relevance for the assessment of chronic toxicity to aquatic vertebrates. The inherent assumption is that effect levels derived from these endpoints are relevant to protect populations. However, the endpoints are not indicative or specific to any particular endocrine mode of action.

21-Day Amphibian Metamorphosis Assay, draft TG proposal (OECD, 2005)

This test allows the detection of interaction of a substance with the thyroid system. This test may be used when there is some indication that the substance may disturb growth and development, essentially for confirming the mode of action (i.e. thyroid). As thyroid is heavily conserved in vertebrates, a negative response in the 21-Day Amphibian Metamorphosis Assay indicates that the substance does not impact the thyroid system in any vertebrate taxa. A positive response may be used in conjunction with chronic tests to conclude on the hazard and the derivation of effect levels.

Invertebrate life cycle tests, including developmental and reproductive endpoints

The life cycle of invertebrates is controlled by distinct and different endocrine systems than vertebrates. In some cases (e.g., mollusks), the hormones may be similar to the steroids found in

vertebrates, while in other cases (e.g., aquatic arthropods) the hormones are specific to certain invertebrate groups, such as juvenile hormone or ecdysteroids.

Test methods for invertebrates, such as life cycle or multi-generation studies, focus on non-specific population-relevant endpoints of reproduction and development, rather than identifying any specific endocrine mode of action for particular invertebrate groups (except for the proposed enhancement to the existing Daphnia reproduction test).

Enhanced OECD TG 211 on Daphnia magna Reproduction Test, draft TG proposal, 2005;

The evaluation of test results is not any different from the existing OECD TG 211. The evaluation of additional endpoints provides a mechanistic insight into the effects observed on development and reproduction. Care should be exercised in the interpretation of changes in the sex ratio in the daphnids as this is not specific for an endocrine mode of action in these parthenogenic organisms where several test conditions (e.g. temperature, food abundance) can affect the sex ratio of the offspring. The regulatory interpretation of changes in the sex ratio endpoint is still new and requires further discussion.

Several new reproductive and developmental assays have been recently proposed for aquatic invertebrates and are listed in Section 3. These proposals are based on endpoints relevant for reproduction and development, and do not include additional markers to indicate any endocrine mode of action. None of these tests have advanced to the stage of regulatory guidelines, and none are currently required by Annexes VII to X in the REACH legislation.

Mammalian toxicity data

Results from mammalian *in vitro* and *in vivo* screening assays should provide both positive and negative indications of endocrine modes of action which are also relevant for aquatic vertebrate species.

Studies on repeated dose toxicity, long-term toxicity and carcinogenicity, reproductive or developmental toxicity in mammals may provide both positive and negative indications of endocrine modes of action which are also relevant for aquatic vertebrate species.

For detailed guidance on the evaluation of such data the relevant sections of the chapter on Human Health Hazard Assessment should be consulted.

Interpretation and use of this data within an integrated assessment of endocrine activity in aquatic organisms is outlined in section 6 of this Appendix.

Conclusions on endocrine activity

The purpose of this section is to give guidance if and how information relating to endocrine activity of a substance and to the adverse effects that may arise from such activity should be considered for conclusions on the regulatory endpoints classification & labelling, PBT assessment and chemical safety assessment and on the assessment of endocrine disrupting properties as referred to in Article 57 f).

SUITABILITY OF INFORMATION ON CLASSIFICATION AND LABELLING

Endocrine activity, which may result in long-term toxicity, are usually not of relevance for classification according to the current EU system, which is based on information from short-term

toxicity testing. A basis for exceptions is provided by the 'safety net' categories for substances, which do not fall under the 'core set of criteria' (N; R50, N; R50-53, N; R51-53 or R52-53).

The risk phrase R52, 'harmful to aquatic organisms', may be assigned to substances *"which on the basis of the available evidence concerning their toxicity may (…) present a danger to the structure and/or functioning of aquatic ecosystems"*. The risk phrase R53, 'may cause long-term effects in the aquatic environment' may be assigned to substances *"which, on the basis of the available evidence concerning their persistence, potential to accumulate, and predicted or observed environmental fate and behaviour may (…) present a long-term and/or delayed danger to the structure and/or functioning of aquatic ecosystems"*. There are no defined criteria for these classifications but both have been proposed and argued for in the course of the classification of bisphenol A, in order to take account of its endocrine disrupting properties. The eventual majority decision in favour of R53 – despite the apparent lack of persistency or accumulation potential - was based on the understanding that this would constitute the "stronger" case. However, from the course of the discussion of that time, it may be conceived that in general both categories might be applied to substances with similar (endocrine disrupting) properties. In any case, such a decision should be based on available information that a substance causes adverse effects on development or reproduction of aquatic organisms which should be derived not from screening assays, but from suitable long-term confirmatory tests, such as those detailed in sections 3 and 4.

SUITABILITY OF INFORMATION ON PBT ASSESSMENT

The assessment whether a substance fulfills the T criterion with respect to freshwater or marine organisms (long-term NOEC < 0.01 mg/l) is usually based on results from standard long-term toxicity testing of the kind that is specified in REACH Annexes VII-X. Standard toxicity testing in fish is based on the assessment of growth and mortality. Some substances, however, may cause sublethal chronic effects in concentrations below those affecting growth or survival, which may also be of serious concern for the aquatic environment, such as an impairment of sexual development or reproductive performance.

Information on reproductive or developmental effects in fish is not part of the requirements of REACH Annexes VII-X but may be available for some substances, e.g. from the scientific literature. Suitable long-term studies are those studies which are designed to investigate specific toxicity on reproduction or sexual development as in the Fish Sexual Development Test, the Reproduction Test or the Full Life-Cycle / Two-Generation Test that are described in sections 3 and 4. Parameters derived from such studies with a widely accepted relevance for reproduction, which may have an impact on population level, are egg numbers, fertilization rate, time to hatch, hatching rate and sex ratio. This information should be considered for use in the assessment of chronic toxicity as part of PBT assessment if it is derived from a suitable long-term study and judged as adequate according to the principles outlined in section 4.

The relevance of changes in fish gonad histology or spermatogenesis and whether these should be considered adverse effects is controversial. Changes to secondary sex characteristics or biochemical parameters such as vitellogenin or spiggin are regarded as evidence that a substance acts via a specific endocrine mode of action, which may or may not result in long-term adverse effects. In itself, information on such parameters is not suitable for use in PBT assessment, but it may be the basis for a CA to request further investigations of potential long-term adverse effect in the course of substance evaluation.

SUITABILITY OF INFORMATION ON CHEMICAL SAFETY ASSESSMENT

The use of information on sub-lethal long-term effects in Chemical Safety Assessment (CSA) should generally be considered according to the same principles as outlined above for PBT assessment.

It is subject to a controversial debate whether the conclusion that an adverse effect is elicited by an endocrine mode of action justifies a modification of the assessment factor used in risk assessment. For the further progress of this debate it might be helpful to bear in mind the provision contained in the TGD 2003 with regard to this issue: *In general, justification for changing the assessment factor could include one or more of the following: (…) Knowledge of the mode of action including endocrine disrupting effects* (p 100).

More guidance on the selection of the appropriate assessment factor is given in guidance provided by RIP 3.2.

SUITABILITY OF INFORMATION ON ASSESSMENT IN RELATION TO ARTICLE 57 (F)

According to Article 57 (f), the list of substances subject to authorisation (Annex XIV), may include "substances – such as those having endocrine disrupting properties (…) – for which there is scientific evidence of probable serious effects to human health and the environment which give rise to an equivalent level of concern to those of other substances listed in points (a) to (e) and which are identified on a case-to-case basis (…)".

While the identification of such substances is a responsibility of the Member States, executed by the preparation of an Annex XV dossier, which should justify the proposal and specify the concern, the evaluation of environmental hazard information will form the basis for it. In accordance with the principles outlined in the previous sections, available information on a accordance with the principles outlined in the previous sections, available information on a substance can be evaluated for its suitability to support a conclusion that

- there is an indication or evidence of *endocrine disrupting properties* (instead of this wording, which is a direct quote from the REACH regulation, the more fitting term *endocrine activity or mode of action* is used throughout this Appendix)
- there is *scientific evidence of probable serious effects* to the aquatic environment due to these properties (i.e. within the terminolgy of this Appendix "adverse effects on development and/or reproduction")

Indication of potential endocrine activity in aquatic organisms may be provided by considerations relating to the molecular structure, available information from endocrine-specific *in vitro* screening assays, such as those outlined in sections 3 and 4, or available information from mammalian toxicity studies. However, structural data *alone* should be regarded as an insufficient basis at this time.

Evidence of an endocrine mode of action in aquatic organisms may be provided by information on biochemical, histological or morphological changes measured in endocrine-specific studies. Generation of this kind of information is not a standard requirement under REACH but may be requested by a CA in specific cases during substance evaluation, e.g. on the basis of available alerts such as those listed above.

Evidence of *probable serious effects* to the aquatic environment due to *endocrine disrupting properties* may encompass information regarding adverse effects on development or reproduction, which can be obtained from suitable long-term studies such as those outlined in sections 3 and 4. However, reproductive or developmental toxicity can also be caused by other toxicological mechanisms and a case-by-case decision must be reached based on *Weight of Evidence* considering all available information on adverse effects in conjunction with information on specific endocrine modes of action. Again, it should be noted that this kind of information is not a standard requirement.

It may be available in some cases, e.g. from the scientific literature, and it may also be requested by a CA under substance evaluation in specific cases, e.g. on the basis of available information that a substance acts via an endocrine mode of action.

The overall conclusion should be on the presence or not of endocrine disrupting properties of the substance and the characterisation of adverse effects, based on existing information or information that is generated on specific request by the Competent Authority under substance evaluation. It is not the responsibility of the registrant to conclude on an *equivalent level of concern*, as specified under Article 54 (f). This task is the responsibility of the Competent Authority or the Agency, who prepare a dossier according to Annex XV for the identification of substances of very high concern and for their eventual inclusion in Annex XIV.

Integrated assessment of potential endocrine activity

In the following, a strategy for an integrated assessment of all available information on potential endocrine activity of a substance is proposed (see scheme). It takes up concepts developed by the OECD in its conceptual framework for endocrine disrupter testing and assessment, which provides a toolbox with methods categorised according to levels of increasing biological complexity (OECD, 2002).

This section is intended to summarise what has been outlined before about how to gather and evaluate existing information on endocrine activity and how this may relate to the purposes and requirements of REACH.

Most of the presently available knowledge, experience and methodology relates to the system of sexual hormones (estrogens/androgens) of vertebrates, with fish as the most extensively studied aquatic species. Progress is also being made with regard to the thyroid system in amphibians. Coverage of invertebrate species and their distinct endocrine systems, such as those of juvenile or ecdysteroid hormones, remains sparse.

In the proposed assessment strategy, three types of information are distinguished: preliminary information that indicates potential endocrine activity in aquatic organisms; information that indicates a specific endocrine mode of action in an intact aquatic organism; information that allows the characterisation of long-term adverse effects, which may be caused by endocrine activity but also by other mechanisms of toxicity.

1) Preliminary indication of potential endocrine activity in aquatic organisms

Preliminary indications of potential endocrine activity that might be of relevance for the aquatic environment but are derived from information sources outside aquatic toxicity testing include considerations of the molecular structure, which will apply to all substances, and results from *in vitro* screening assays, which are not part of the standard information requirements but may be available in certain cases, e.g. from scientific research. Preliminary indications applicable to

vertebrate species may also come from results from mammalian toxicity testing, which may to a certain extent be part of the standard information requirements.

Non-testing information (molecular structure):

The different approaches of generating information by non-testing methods have been outlined in sections 3 and 4. In relation to the steroid sexual hormone system of vertebrates, a number of QSAR models based on experimental data are available resp. under development. Qualitative approaches, such as SAR, read-across or categorisations, may consider similarities with natural hormones or xenobiotic substances of confirmed hormonal activity with regard to all known endocrine systems.

Within the domain of non-testing data, a sensible tiered approach can be applied for screening and prioritization purposes (Tong et al., 2003). Such approach can start with rejection filters (e.g. molecular weight lower than 94 or higher than 1000 is not likely to be associated with estrogen binding affinity), include models for qualitative assignment of activity (e.g. classification as active or inactive compounds) and then applying models for quantitative estimation of the potency in case that the chemical is predicted active as a result of the previous step. The last step includes incorporation of human knowledge and expertise in the evaluation of the results of the previous steps and additional rules for refinement can be applied.

With regard to the endpoint under prediction, a differentiation is to be made between mechanistic endpoints, i.e. mainly interactions with a defined molecular target, endpoints relating to biochemical responses (screening assays) or adverse effects (definitive tests) *in vivo*. Among these, endpoints that derive from methods which are included in this document are to be considered with priority since there is an intensive research ongoing in the field of test methods for endocrine disruption. As is generally the case in the evaluation of the non-testing data, the quality of experimental data they are based on might also be important (e.g. does it come from a single source or it is compilation from different sources).

Information from in vitro screening assays:

Although there are principally *in vitro* systems for the study of all kinds of endocrine systems and mechanisms in use in scientific research, the most relevant methods to date are those related to the sexual steroid hormones, which are described in section 3. Other types of assays, e.g. *in vitro* thyroid receptor binding assays, may become more important in the future.

Given the high degree of conservation of the molecular components of endocrine systems across vertebrate taxa, the ability of a substance to bind to a mammalian hormone receptor, activate transcription of hormone-responsive genes or interfere with steroid hormone biosynthesis in a mammalian cell line may suggest similar activity in aquatic vertebrates.

Regarding the relevance of test results, the usual limitations of *in vitro* methods apply: focus on a single mechanism of action *in vitro vs.* the diversity and complexity of molecular structures and regulatory pathways *in vivo*; lacking or limited metabolic capacity of some test systems; disregard of complex physiological processes, such as the toxicokinetic distribution of a substance, the organ- or tissue-specific expression of its molecular targets, feedback regulations or mechanisms of adaptation.

Information from mammalian toxicity testing:

Standard studies on repeated dose toxicity, long-term toxicity and carcinogenicity, reproductive and developmental toxicity or non-standard studies on specific endocrine mechanisms in mammals can provide indications of endocrine activity that might also be of relevance for aquatic vertebrates.

With respect to the sexual hormone system, this includes changes in endocrine-responsive tissues (gonads, secondary sex organs), reproductive functions (estrous cycling, spermatogenesis, mating behaviour, fertility, gestation, parturition or lactation) or developmental landmarks (e.g. anogenital distance, vaginal opening, preputial separation). All of these changes might be caused by impact on molecular pathways that are also present in aquatic vertebrates such as interactions with steroid hormone receptors or biosynthesis, transport and metabolism of steroid hormones.

Indications of thyroid activity include developmental impairments, histopathological changes of the thyroid gland or (not routinely investigated) thyroid hormone levels.

Weight of evidence:

If there is information available for the same chemical from different sources, the following questions should be considered for the overall conclusion: Is the information consistent or is it in conflict with each other? In the case of conflicting data, the quality of each piece of information should be evaluated in accordance with the principles described in section 4, as should its biological relevance with respect to aquatic organisms, and, finally, the potential impact of such information on the overall regulatory decision.

2) Indication of specific endocrine activity in intact aquatic organisms

Evidence that a substance can operate by a specific endocrine mode of action in aquatic organisms can only be derived from the investigation of specific, endocrine-responsive endpoints. None of these are covered by standard aquatic toxicity testing. Endocrine-specific screening assays are, however, under development and validation for both mammalian rodents (uterotrophic and Hershberger assays) and for aquatic vertebrates (21-day fish screening assay and amphibian metamorphosis assay).

In the endocrine specific aquatic assays, vitellogenin in fish responds to estrogens (induction in males) and aromatase inhibitors (suppression in females), and secondary sexual characteristics in fish respond to androgens (induction in females). Specifically for the stickleback, spiggin may also provide the means to specifically characterise (anti-)androgenic modes of action. Specifity and significance of other endpoints such as other biochemical parameters (e.g. hormone levels) or histopathological changes of the gonads, including impairment of spermatogenesis, are under debate. The specific endpoints which are included in the 21d-Fish Screening Assay can also be assessed in conjunction with higher tier chronic tests. As isolated information, biomarker responses cannot be used for regulatory conclusions. They may raise a strong concern that the substance in question might cause serious long-term adverse effects, in particular if environmental exposure, persistence and/or bioaccumulation are high. Such a concern may lead to a specific request for further investigations by a Competent Authority in the course of dossier or substance evaluation.

Evidence of thyroid activity is provided by histopathological changes to the thyroid gland, which can be observed in the Amphibian Metamorphisis Assay or similar test systems. If a protocol was used in accordance to the current OECD test guideline development, effects information on the progress of metamorphosis will be available from the same study and can be considered for use in regulatory decisions as outlined below. Thyroid histology reported as isolated information may not be suitable for use in regulatory decisions. It may support the interpretation of other toxicity data, also from mammalian toxicity studies. It may also raise a strong concern that the substance in question might cause serious long-term adverse effects, in particular if environmental exposure, persistence and/or bioaccumulation are high. Such a concern may lead to a specific request fo further investigations by a Competent Authority in the course of dossier or substance evaluation.

Evidence of specific endocrine mode of action in invertebrates as isolated information will only be found in very rare cases and no general guidance can be given for its use.

3) Characterisation of long-term adverse effects

The reproductive capacity of fish can be adversely affected by a number of mechanisms of toxicity. Observation of such effects, which can threaten fish populations, can be made during studies that cover a distinct sensitive life stage such as sexual development or active reproduction or studies that cover a complete life-cycle or even two or more consecutive generations. Only the latter allow the identification of delayed reproductive effects through endocrine disruption during early life stages. Information on sublethal adverse effects, if judged as adequate, should be considerd for use in PBT assessment or Chemical Safety Assessment/PNEC derivation. Classification as R52 or R53 according to the safety net criteria might be proposed. A causal link between a reproductive adverse effect and an endocrine mode of action might prompt a proposal for identifying the substance as a substance of very high concern (Annex XV) by a Competent Authority. If the adverse effects information is provided by a reproductive and developmental study similar to those currently under development in the OECD TG programme, information on endocrine-specific endpoints will be available from the same study and assessment of a causal link may be possible based on similar dose responses.

Long-term toxicity caused by chemicals with thyroid activity can be manifest as developmental disturbance, e.g. promotion or inhibition of amphibian metamorphosis. Similar considerations apply as outlined above for adverse effects in fish.

Adverse effects on development or reproduction of invertebrates may be reported from non-standard studies and, if rated adequate, should be considered for use in the assessment of chronic toxicty. A causal link to a specific endocrine mode of action will only be found in rare cases.

Figure R.7.8-8: Integrated assessment of potential endocrine activity in aquatic organisms; based on the evaluation of available information which is not part of the REACH requirements

<table>
<tr><td colspan="3">1) Preliminary indication of potential endocrine activity in aquatic organisms</td></tr>
<tr><td>Estrogen/androgen axis:
- molecular structure
- mammalian toxicity
- in vitro screening</td><td>Thyroid:
- molecular structure
- mammalian toxicity</td><td>Invertebrate systems:
- molecular structure</td></tr>
<tr><td colspan="3">-> determine concern of potential endocrine mode of action of the substance using WoE of all available information, including environmental fate and exposure
-> strong concern may prompt a proposal by the Competent Authority to include the substance in the Community rolling action plan in order to perform a substance evaluation
2) Indication of specific endocrine modes of action in intact aquatic organisms</td></tr>
<tr><td>Estrogen/androgen axis:
- biochemical markers
- morphological changes
(- gonad histopathology)</td><td>Thyroid:
- thyroid histopathology</td><td rowspan="2">Invertebrate systems:
- rare individual cases</td></tr>
<tr><td>Study type:
Fish Screening Assay
Fish Sexual Develpt. Test
Fish Reproduction Test
Fish Full Life-Cycle Test</td><td>Study type:
Amphibian Metamorphosis Assay</td></tr>
<tr><td colspan="3">-> determine concern of potential endocrine mode of action in intact aquatic organisms using WoE of all available information, including environmental fate and exposure
-> strong concern may prompt a proposal by the Competent Authority to include the substance in the Community rolling action plan in order to perform a substance evaluation
3) Characterisation of long-term adverse effects[#]</td></tr>
<tr><td>Estrogen/androgen axis:
- fish (sexual) development
- fish reproduction</td><td>Thyroid:
- amphibian development</td><td>Invertebrate systems:
- development
- reproduction</td></tr>
<tr><td>Study type:
Fish Sexual Develpt. Test
Fish Reproduction Test
Fish Full Life-Cycle Test</td><td>Study type:
Amphibian Metamorphosis Assay</td><td>Study type:
Invertebrate Reproduction or Life-Cycle Tests</td></tr>
<tr><td colspan="3">-> consider use of chronic NOEC for PBT assessment and Chemical Safety Assessment
-> consider classification and labelling according to safety net categories (R52, R53)
-> causal link of adverse effect with an endocrine mode of action may prompt consideration for Annex XV by CA</td></tr>
</table>

[#]It should be noted that the listed adverse effects, which may occur as a result of endocrine activity of a substance, may also be caused by other mechanisms of toxicity

References

Devillers J, Marchand-Geneste N, Carpy A, Porcher JM. SAR and QSAR modeling of endocrine disruptors. *SAR QSAR Environ Res* 2006; 17: 393-412

Fang H, Tong W, Branham W, Moland CL, Dial SL, Hong H, Xie Q, Perkins R, Owens W, Sheehan DM. Study of 202 natural, synthetic, and environmental chemicals for binding to the androgen receptor. *Chem Res Toxicol* 2003;16: 1338-1358

Fang H, Tong W, Shi LM, Blair R, Perkins R, Branham W, Hass BS, Xie Q, Dial SL, Moland CL, Sheehan DM. Structural-activity relationships for a diverse set of natural, syntheetic, and environmental estrogens. *Chem Res Toxicol* 2001; 14: 280-294

Laws SC, Yavanhxay S, Cooper RL, Eldridge JC. Nature of the binding interaction for 50 structurally diverse chemicals with rat estrogen receptors. *Toxicol Sci.*(2006; 94(1):46-56.

Netzeva TI, Gallegos Saliner A, Worth AP. Comparison of the applicability domain of a quantitative structure-activity relationship for estrogenicity with a large chemical inventory. *Environ Toxicol Chem.* 2006 May;25(5):1223-30.

Saliner AG, Netzeva TI, Worth AP. Prediction of estrogenicity: validation of a classification model. *SAR QSAR Environ Res*. 2006 Apr;17(2):195-223.

Tamura H, Ishimoto Y, Fujikawa T, Aoyama H, Yoshikawa H, Akamatsu M. Structural basis for androgen receptor agonists and antagonists: interaction of SPEED 98-listed chemicals and related compounds with the androgen receptor based on an in vitro receptor gene assay and 3D-QSAR. *Bioorg Med Chem* 2006; 14: 7160-7174.

Tong W, Fang H, Hong H, Xie Q, Perkins R, Anson J, Sheehan DM. Regulatory applications of SAR/QSAR for priority setting of endocrine disruptors: A perspective. *Pure Appl Chem* 2003; 75(11-12): 2375-2388.

OECD Series on Testing and Assessment:

No. 21. Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals, 07-Mar-2002, ENV/JM/MONO(2002)8

No. 46. Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances, 22-Oct-2004, ENV/JM/MONO(2004)17

No. 47. Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances, 21-Oct-2004, ENV/JM/MONO(2004)18

No. 55. Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects, 31-Jul-2006, ENV/JM/MONO(2006)22

No. 57. Detailed Review Paper on Thyroid Hormone Disruption Assays, 02-Aug-2006, ENV/JM/MONO(2006)24

No. 60. Report on the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A), 12-Sep-2006, ENV/JM/MONO(2006)27

No. 61. Report of the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1B), 12-Sep-2006, ENV/JM/MONO(2006)29

OECD Draft Guidance and Review Documents:

DRP Draft Detailed Review Paper on Fish Two-Generation Toxicity Test (March 2003 version) and Proposal for a Fish Two-Generation Test Guideline (March 2003 version)

DRP Revised Draft Detailed Review Paper on Aromatase, February 2002

DRP Draft Detailed Review Paper on Steroidogenesis, May 2002

DRP Draft Detailed Review paper on the Use of Metabolising Systems for in vitro Testing of Endocrine Disrupters (version March 2006)

R.7.8.7 Introduction to sediment organisms' toxicity

Substances that are potentially capable of depositing on or sorbing to sediments to a significant extent have to be assessed for toxicity to sediment-dwelling organisms. In addition, marine sediment effects assessment is necessary for substances that are known to be persistent in marine waters and may accumulate in sediments over time. In general substances with a K_{oc} <500 – 1000 l/kg are not likely sorbed to sediment (SETAC 1993). According to this, a log K_{oc} or log K_{ow} of ≥3 is used as a trigger value for sediment effects assessment.

R.7.8.7.1 Definition of toxicity to sediment organisms

Sediments may act as both a sink for chemicals through sorption of contaminants to particulate matter, and a source of chemicals through resuspension. Sediments integrate the effects of surface water contamination over time and space and may thus present a hazard to aquatic communities (both pelagic and benthic) which is not directly predictable from concentrations in the water column.

The sorption or binding behaviour of chemicals to sediment is determined by certain properties. Especially substances with high log K_{ow} or log K_{oc} values adsorb to the organic fraction of the sediment. In addition, substances that bind to components of the sediment via chemical reactions or substances that ionically bind to inorganic as well as organic fractions may accumulate in the sediment.

Effects on benthic organisms are of concern because they constitute an important link in the aquatic food chain and play an important role in the recycling of detritus material. Whole-sediment tests using benthic organisms are most suitable for a risk assessment for the sediment compartment. By using such tests it is possible to adequately address all routes of exposure. Due to the generally long-term exposure of benthic organisms to sediment-bound substances, long-term tests with sublethal endpoints like reproduction, growth or emergence are most relevant.

R.7.8.7.2 Objective of the guidance on toxicity to sediment organisms

The aim of sediment toxicity tests is to find out at which concentrations a substance adsorbed or bound to sediment exhibit toxic effects to benthic organisms. Special attention should be given to the pathways by which the test organisms are exposed to the chemical. In particular spiking methodology should be considered in detail and be performed in the most realistic way possible.

The determination of the concentration-response relationship should lead to the identification of the No Observed Effects Concentration NOEC or EC_{10} from long-term tests or median lethal concentration LC_{50} from acute tests. This NOEC/EC_{10} or LC_{50} is subsequently used for deriving a Predicted No Effect Concentration for the sediment ($PNEC_{sediment}$). This $PNEC_{sediment}$ is compared with the Predicted Environmental Concentration in the sediment ($PEC_{sediment}$) to decide whether there is a risk to sediment organisms from the exposure of the chemical.

R.7.8.8 Information requirements for toxicity to sediment organisms

The information requirements for sediment toxicity are described by REACH Annexes VII to XI, that specify the information that shall be submitted for registration and evaluation purposes.

For this endpoint information requirements are formulated for substances produced or imported in quantities of ≥1000 t/y. (Annex X):

Column 1 **Standard information required**	**Column 2** **Specific rules for adaptation from column 1**
7.5.1 Long-term toxicity to sediment organisms	7.5.1 Long-term toxicity testing shall be proposed by the registrant if the results of the chemical safety assessment indicate the need to investigate further the effects of the substance and/or relevant degradation products on sediment organisms. The choice of the appropriate test(s) depends on the result of the chemical safety assessment.

Further guidance on this issue is given in section.

R.7.8.9 Information on toxicity to sediment organisms and its sources

For most chemicals uptake from water (bioconcentration, defined as the net result of uptake, transformation, and elimination of a substance in an organism due to waterborne exposure) is believed to be the predominant route of exposure for aquatic organisms. For organic substances and metals pore water is one of the primary exposure routes for benthic organisms (Di Toro et al, 1991; Ankley et al, 1991). However, for highly lipophilic compounds or other substances that adsorb to particles (e.g. metals), uptake from food or sediment may contribute to the overall exposure, depending on the living and feeding strategy of the exposed organisms. Therefore factors that influence adsorption and thus distribution between sediment and water influence also toxicity to aquatic (pelagic and benthic) species. A compilation of such factors is given in Section R.7.8.7

R.7.8.9.1 Laboratory data on toxicity to sediment organisms

Non-testing data on toxicity to sediment organisms

For most chemicals the number of toxicity data on sediment organisms is limited. In the absence of such data, a read-across from pelagic effect values is possible as a screening approach (equilibrium partitioning method, EPM) (reference to R16 and R10). It has to be considered that the equilibrium partitioning method may result both in an overestimation or underestimation of the toxicity to benthic organisms (Di Toro et al. 2005). Therefore, this method can only be used as rough screening to decide whether sediment toxicity tests with benthic organisms are required.

A general guidance on how to extrapolate via read-across or chemical categories is given in Section R.6.2.

Up to now there are no QSAR models available for the prediction of toxicity to sediment organisms exposed via a water-sediment system.

Testing data on toxicity to sediment organisms

Only few standardised test methods for sediment tests with benthic organisms are available. An internationally harmonised test guideline exists only for *Chironomus spec.*. In the following an overview of available standardized (short- and long-term) test methods for sediment tests with benthic organisms is given. In Annex 1 the different test species are further characterised in term of taxonomic group, habitat and feeding mode.

OECD TEST GUIDELINES:

Test No 218: Sediment-Water Chironomid Toxicity Using Spiked Sediment

Test No 219: Sediment-Water Chironomid Toxicity Using Spiked Water

Both guidelines are designed for studying long-term toxicity (28d exposure) of chemicals to the sediment-dwelling larvae of the freshwater midge *Chironomus* spec. Measured endpoints are total number of adults emerged and time to emergence. Spiking the sediment (OECD 218) is recommended for continuous and intermittent release of chemicals while spiking the waterphase (OECD 219) was developed for pesticide specific exposure situations.

PROPOSAL FOR NEW OECD TEST GUIDELINE:

Sediment-Water Lumbriculus Toxicity using spiked sediment (OECD 2006)

This guideline is designed for studying long-term toxicity (28d exposure) of chemicals to the endobenthic oligochaete *Lumbriculus variegatus*. Measured endpoints are total number or worms and biomass at end of exposure.

ASTM TEST GUIDELINES

In Table R. 7.8-4 an overview of active ASTM standards for sediment toxicity tests is given. The single test methods cover a selection of different test species that are given in the 2nd column.

Table R. 7.8-4: Overview of active ASTM standards for sediment toxicity tests

Guideline	Species
E1706-05. Standard Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates.	*Chironomus* sp. *Hyalella azteca* *Hexagenia* spp. *Tubifex tubifex* *Diporeia* sp.
E1611-00*. Standard Guide for Conducting sediment toxicity tests with marine and estuarine polychaetous annelids.	*Neanthes arenaceodentata* *Neanthes virens*
E1367-03e1*. Standard Test method for measuring the toxicity of sediment-associated contaminants with marine and estuarine invertebrates.	*Leptocheirus plumulosus* *Ampelisca abdita* *Eohaustorius esturaius* *Rhepoxynius abronius*

***The general procedures described in the above cited standards (ASTM E1611-00 and E1367-03e1) might be useful for conducting tests with other estuarine or marine invertebrates.**

Most of the cited ASTM guidelines are designed as short-term tests (10 d exposure) with mortality as endpoint. However, for some of these species (*Hyalella azteca*, *Chironomus* sp., *Leptocheirus plumulosus, Neanthes arenaceodentata)* also long-term toxicity tests (28d exposure) with sublethal endpoints are recommended by the guidelines.

US-EPA TEST GUIDELINES

EPA 600/R-99/064 Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates.

- 100.1: *Hyalella azteca* 10-d survival and growth test for sediments (short-term)
- 100.2: *Chironomus tentans*: 10-d survival and growth for sediments (short-term)
- 100.4: *Hyalella azteca*: 42-d test for measuring the effects of sediment-associated contaminants on survival, growth and reproduction (long-term)
- 100.5: Life-cycle test for measuring the effects of sediment-associated contaminants to *Chironomus tentans* (long-term): 50 – 65-d test

EPA 600/R-01/020 Method for assessing the chronic toxicity of marine and estuarine sediment-associated contaminants with the amphipod *Leptocheirus plumulosus*. 28-d test with survival, growth and reproduction as endpoint.

OTHER STANDARD TEST METHODS

ISO 16712: Water quality - Determination of acute toxicity of marine or estuarine sediment to amphipods. This specifies a method for the determination of acute toxicity to amphipods exposed over a period of 10 d to (among others) chemicals or preparations spiked into clean sediment.

Proposal for ISO norm: Determination of the toxic effect of sediment and soil samples on growth, fertility and reproduction of *Caenorhabditis elegans* (Nematoda). This test has a duration of 72 h but can be considered as a long-term test as it measures both growth and reproduction endpoints.

OSPAR Guideline (OSPAR 2005): A Sediment Bioassay using an Amphipod *Corophium sp.* – Marine sediment toxicity test. Either *Corophium volutator* or *Corophium arenarium* are considered acceptable for use. In the test adult *Corophium* are exposed to spiked sediments for 10 days. Endpoints are survival and burrowing activity.

Environment Canada. Biological Test Method: Test for Growth and Survival in Sediment Using the Freshwater Amphipod *Hyalella azteca*.

Environment Canada. Biological Test Method: Test for Growth and Survival in Sediment Using Larvae of Freshwater Midges (*Chironomus tentans* or *Chironomus riparius)*.

NON-STANDARD TEST METHODS

There are a lot of non-standard methods for the testing of effects of substances to sediment organisms available. An overview of available non-standard test methods can be found in OECD (1998).

TESTS PERFORMED WITHOUT SEDIMENT

There may be several non-standard tests available in which benthic organisms are exposed in a water-only test system to the chemical in question. Such tests do not take into account the different routes of exposure that may occur under environmental conditions. Therefore, for the derivation of the $PNEC_{sediment}$, these tests can only be used for screening purposes in combination with the equilibrium partitioning method. In addition, such tests may provide information on the importance of sediment ingestion, if compared with tests on the same species in the presence of sediment or may provide evidence of lethal and critical body burden data (*Weight of Evidence* approach).

R.7.8.9.2 Field data on toxicity to sediment organisms

Experimental ecosystem studies examine the effect of chemicals on aquatic model ecosystems. These studies generally study both the effects of chemicals on pelagic organisms via the waterphase and on benthic organisms via the sediment. Therefore, it is referred to the Section R.7.8.4.1.

R.7.8.10 Evaluation of available information on toxicity to sediment organisms

A general overview of the properties of substances and test systems that influences the evaluation of aquatic toxicity tests are described in section. Some of these properties are also related to sediment toxicity.

R.7.8.10.1 Laboratory data on toxicity to sediment organisms

Non-testing data on toxicity to sediment organisms

Equilibrium partitioning method: Several factors have to be considered when using the equilibrium partitioning method for the estimation of the toxicity of chemicals to sediment organisms:

This method considers only uptake via the water phase. However, uptake may also occur via other exposure pathways like ingestion of and direct contact with sediment depending on the organism used for testing. This may become important especially for highly adsorbing chemicals. As uptake via the gut is likely to play an increasingly important role with increasing adsorption, for compounds with a log K_{ow} greater than 5 or with a correspondingly high adsorption or binding behaviour (e.g. aromatic amines forming covalent bound to sediment components, ionisable substances, surface active substances), the equilibrium partitioning method can only be used in a modified way. In order to allow for uptake of substances via ingestion of sediment, an additional factor of 10 is applied to the PEC/PNEC ratio for such substances. It should be borne in mind that this approach is considered only as a screen for assessing the level of risk to sediment dwelling organisms. If with this method a PEC/PNEC ratio >1 is derived, then data improvement is necessary either by refining the exposure assessment or by performing tests with benthic organisms using spiked sediment to support a refined risk assessment for the sediment compartment.

A general guidance on how to extrapolate via read-across or chemical categories is given in Section R.6.2.

Testing data on toxicity to sediment organisms

The effects of sediment-bound substances to benthic organisms can be best assessed by performing long-term whole-sediment tests that take into account all possible routes of exposure that may occur under environmental conditions (overlying water, porewater, ingestion of sediment, direct contact with sediment).

In general, if tests have been performed according to standard test guidelines, the validity criteria or acceptability requirements specified in these guidelines has to be fulfilled for acceptance of the study.

Due to the complex test system, results from whole-sediment tests may be influenced by several parameters (e.g. sediment composition, spiking method, feeding mode).

Critical factors important for evaluating sediment toxicity tests (standard and non-standard tests) are provided, as follows:

TEST ORGANISMS

Only species that act as ecological representatives for the sediment compartment are acceptable as test organism. The available test methods (see Section R.7.8.9) refer mostly to invertebrates of the trophic level *primary consumer* or *decomposer*. Therefore, the concept of covering several trophic levels which has been applied for the pelagic compartment cannot be followed for the sediment. Instead, the test species should cover different habitats and feeding modes in the sediment as well as different taxonomic groups. In general, a distinction is drawn between endobenthic and epibenthic species. Endobenthic species burrow in the sediment and preferably ingest sediment particles below the sediment surface. Epibenthic organisms live on or slightly above the sediment surface and feed mainly on freshly deposited organic material or suspended solids.

Especially for strongly adsorbing or binding substances sediment-dwelling organisms that feed on sediment particles (e.g. *Lumbriculus variegatus*, *Tubifex tubifex*) are most relevant. However, also a specific mode of action that is known for a given substance may influence the choice of the test species (e.g. for substances suspected of having specific effects on arthropods a test with *Chironomus* is more appropriate than tests on other Phyla).

Table R. 7.8-3 gives an overview of different benthic test species in term of taxonomic group, habitat and feeding mode.

ENDPOINTS

Endpoints studied in sediment toxicity tests should be of ecological relevance, i.e. show effects relevant at the population level, where possible. For long-term tests the sub-lethal endpoints reproduction, growth and (insect) emergence are most relevant. Behavioural endpoints like sediment avoidance or burrowing activity have not been standardized. Such endpoints can give indications on toxic effects but should not be interpreted in isolation. For short-term tests survival is the normal endpoint to be considered.

Some endpoints, particularly reproduction endpoints, show a high variability thus making a reliable test evaluation difficult. As a general rule, if any indications for a high variability are found (i.e. control coefficient of variation >20%), the endpoint in question may not be interpretable and should not be used for the assessment.

EXPOSURE PATHWAYS

Sediment organisms can be exposed via their body surfaces to substances in solution in the overlying water and in the porewater and to bound substances by direct contact or via ingestion of contaminated sediment particles. The exposure route that is most important is strongly influenced by species-specific feeding mechanisms, gut retention time and the behaviour of the organisms in or on the sediment. For the evaluation of available sediment tests it has to be assessed which exposure routes are covered by the test design and the test organisms used. For strongly adsorbing or binding substances, preference should be given to test designs and test organisms that cover the exposure via sediment ingestion, as this is the most relevant exposure route for such chemicals.

COMPOSITION OF SEDIMENT, ARTIFICIAL-NATURAL SEDIMENT

Both artificial and natural sediments have advantages and drawbacks. Artificial sediment may separate into layers depending on particle size with the clay particles settling at the surface and this may prevent penetration of certain species into the sediment layer. It is also a step further from natural conditions with no significant microbial flora and thus results may not be the same as those from natural sediment. On the other hand, the constituents of artificial sediment are generally well characterised and the residual contaminants that may be found in natural sediment are generally not present in artificial substrates. Generally characterised natural sediments are not available on the open market. On the whole, due to the level of characterisation and reproducibility possible, artificial sediment is generally considered superior to natural substrate (OECD 2004a and b) unless effects at a specific local site are being considered.

Artificial sediment may be conditioned by continued mixing of the components for days or even weeks prior to spiking to improve the homogeneity, increase the microbial flora and transform the organic matter into a more environmentally realistic form. However, this may dramatically increase the BOD of the sediment-water system requiring supplementary aeration to prevent suffocation of the test organisms.

Artificial sediments used in studies should be characterised (e.g. particle size, organic matter (OM), cation exchange capacity (CEC)/anion exchange capacity (AEC)). If natural sediment is used in the test, it should be characterised preferably by origin, pH and ammonium content of pore water, total organic carbon content and nitrogen content, particle size distribution and percent water content. For metals SEM (= Simultaneously Extracted Metals) and AVS (= Acid Volatile Sulfides) concentrations should be measured.

Grain size of the sediment used in the test may influence bioavailability of the test substance. Sediment grain size can also be an important factor in tests for other reasons. For example, the extent to which bacteria can be adsorbed onto the sediment varies with particle size. Likewise, different species of amphipods prefer sediments with different particle size distributions. One should thus consider the tolerance of a given species with regard to the grain size distribution of the sediments in question.

METHOD OF SPIKING

There are two methods to spike the test system with the test substance: one method is to spike the water phase, the other method is to spike the sediment phase. For both methods an equilibration time without exposure of the test organisms is necessary to enable the distribution of the test substance between water and sediment according to the distribution behaviour of the substance. The time needed for the formation of the equilibration between dissolved and sorbed substance may be rather long and may not be reached during the equilibrium period (dependent on substance properties). In general, spiking of the sediment is preferred over spiking of the waterphase.

If spiking via the waterphase was performed for a study, it must be carefully considered whether an exposure via the sediment has taken place. If possible or relevant (e.g. in the absence of analytical measurements) sediment concentration should be calculated from the water concentration using the equation for the equilibrium partitioning method.

Spiking sediments tends to be problematic for poorly soluble chemicals. The standard approach is to dissolve the test substance in a solvent and then to spike sand, blow-off the solvent and then mix sediment with the remaining sand at various concentrations. The drawback with this technique is that even after hours or sometimes days of mixing, the substance may not be homogeneously mixed to the sediment but still present as solid particles on the original sand. Use of an organic solvent added to wet sediment is not recommended as this may have irreversible effects on the organic matter fraction of the sediment. More appropriate are methods such as use of a generator column. OECD (2000) describes several ways in which generators can be used to spike test solutions. Alternatively use of a circulating system where low concentrations of the substance are added over a long period of time (hours to days) may also be appropriate.

EQUILIBRIUM BETWEEN WATER-PHASE AND SEDIMENT-PHASE

After spiking the water-sediment system with the test substance, an equilibrium period is necessary to ensure partitioning of the substance between water-phase and solid-phase according to the substance-specific distribution characteristic. This partitioning should take place under temperature and aeration used during the exposure phase. Appropriate equilibration time is sediment and chemicals specific and can be in the order of hours to days and in some cases up to several weeks. As this would leave time for degradation of many organic chemicals, equilibrium should normally not be awaited too long and the equilibration period should last between 48 h and 7 days (OECD 2006).

For metals and inorganic metal compounds both short equilibration times and high spiked metal concentrations in sediments will accentuate partitioning of metals to the dissolved phase and increase the probability of exposure and/or toxicity via dissolved metals (Lee et al, 2004, Simpson et al., 2004, Wang et al., 2004). As a consequence, it is in static and semistatic tests recommended that the concentration of the test substances is measured in the overlying water of sediment toxicity tests, and that testing is initiated only when overlying water concentration reaches ambient concentrations. Aging and weathering processes may have an impact on sediment toxicity. However, currently there are no agreed methods available to take these phenomena into account in standard sediment test protocols.

FEEDING

In long-term tests, especially with reproduction or growth as endpoint, feeding of the test organisms is necessary. Supplementary feeding of test organisms during the study should be avoided whenever possible; otherwise the exposure route *sediment ingestion* may be underestimated due to selective feeding of the test organisms on the fresh uncontaminated food. (Åkerblom and Goedkoop 2003). When possible, the tests should be designed in such a way that the food necessary for the test organisms during the study is added to the sediment prior to spiking with the test substance. Thereby, it is ensured that the food taken up by the test organisms is also contaminated with the test substance comparable to environmental conditions. Food types are diverse depending on the study, varying from ground, flaked fish food to plant material (e.g. Urtica powder, ground spaghnum peat or alpha cellulose) to cultured *E. coli* cells at known concentration. If organic matter from other sources is included in the spiking this may not be critical. Examples for sediment studies with existing substances in which the food for the test organisms was added to the sediment prior to spiking with the test substance include tetrabromobishphenol-A, tris[2-chloro-1-(chloromethyl)ethyl] phosphate, aniline or 2,4-toluylendiamine.

It has to be considered that any food added to the test system either periodically or only at test initiation may influence water quality due to degradation (see point water quality below).

DURATION OF EXPOSURE

The duration of exposure in a sediment test should be long enough to ascertain, that test substance is really taken up by the test organisms. Especially for strongly adsorbing substances it may take some time to reach equilibrium between the sediment concentration in the test system and in the test organisms. It is recommended that a sediment test should have a duration of at least 10 days. Most standardized test methods (see Section R.7.8.8) envisage an exposure period of at least 10 days for short-term and 28 days for long-term tests. However, there are (non-standard) methods available in which the exposure period is much shorter (e.g. *Caenorhabditis elegans* 72 h). The short duration of such a test can be regarded as an advantage, as it is cost- and time-saving and as nematodes are commonly found in the sediment compartment, it is a biologically relevant species. However, the mouth aperture is extremely small and therefore it cannot ingest whole sediment and therefore should not be the only sediment species tested.

WATER QUALITY PARAMETER

Water quality parameters like oxygen content, pH, ammonium concentration, temperature, water hardness should be measured in regular intervals during the test and the results of these measurements should be reported in the study report. This is important for the evaluation of sediment studies, as these water quality parameters may have an influence on the result of the toxicity study.

Ideally, oxygen content in the overlying water should not fall below 60% of saturation at test temperature, as limited oxygen may result in adverse effects on the test organisms. This should be measured as close to the sediment layer as possible. However, a temporary shortfall below this value may not automatically mean that a test is not valid. In this case it should be checked that the control response is within the normal range. Many sediment dwelling species are capable of surviving at oxygen concentrations as low as 2 mg/l.

The pH of the overlying water should be in a range between 6 and 9. However, it has to be considered that a pH value above 8 may enhance the formation of toxic NH_3 from NH_4^+. Ammonium may be formed during the study e.g. from the food added to the test system and certain species excrete ammonia directly. As NH_3 that is built up at pH values above 8 is toxic to most aquatic organisms, it has to be verified that toxic effects observed during the study are not caused by high ammonium concentrations.

TEST SYSTEM

Sediment tests can be performed in static, semi-static or flow-through systems concerning the overlying water. Semi-static or flow-through systems may contribute to a good water quality in term of e.g. oxygen content or ammonium concentration thus limiting the influence of such factor on the test results. However, static systems are recommended as regular renewal of overlying water is expected to affect chemical equilibrium resulting in losses of test substance from the system. If semi-static or flow-through systems are used, the maintenance of the test substance concentration in the sediment should be supported with analytical monitoring.

TEST DESIGN

The following guidance should be applied when evaluating non-standard tests. Tests performed according to standard guidelines should follow the guidance given there.

For a proper statistical evaluation of the test results, the number of test concentrations and replicates per concentration are critical factors. If a solvent is used for the application of the test substance, a solvent control is necessary. Estimations of the number of replicates should be based on the statistical power required for the test and therefore the coefficient of variation of the parameter under review. As a general rule the statistical power will be sufficient if the following recommendations are observed:

For the estimation of an ECx, five test concentrations with three replicates per concentration are suggested. For the control (and solvent control) six replicates are recommended. The factor between the concentrations should not exceed a factor of 2.

If NOEC/LOEC estimation is performed, five test concentrations with at least four replicates and six replicates for the control (and solvent control) should be used. The factor between the concentrations should not exceed a factor of 2.

A limit test using only one test concentration and a control (and solvent control) may be performed.

Samples for chemical analysis of the test substance should be taken at least from the control, lowest and highest concentrations. Samples should preferably be taken weekly, but at least at end of equilibration phase (start of exposure) and test end.

At least the sediment and the overlying water should be sampled for analysis. If possible pore water concentrations can be analysed, as this will provide a more accurate determination of the concentration to which the sediment dwelling organisms were actually exposed.

OECD 218 states that effect values should be based on initial measured concentrations. However, this approach should only be followed if analysis shows that over the exposure period >80% of the nominal concentrations are maintained.

If the measured concentrations are below 80% of nominal concentrations, the effect values normally have to be related to the mean measured concentration of the chemical in the test system. The reasons for the decrease in test substance concentration should be investigated. If only a measurement at start and end of the exposure phase is performed, the geometric mean of the measured concentrations has to be used.

For some substances complete recovery of irreversibly bound substance may not be technically possible (e.g. aromatic amines). In this case, nominal concentrations can be used provided that the substance is stable in the test system, i.e. no biotic or abiotic degradation or removal from the test system is expected to occur.

R.7.8.10.2 Field data on toxicity to sediment organisms

It is referred to Section R.7.8.4.1.

R.7.8.10.3 Exposure considerations for toxicity to sediment organisms

c) The rule in Annex X Column 2

According to Annex X long-term toxicity tests for sediment organisms shall be proposed if the result of the Chemical Safety Assessment (CSA) indicates the need to investigate further the effects of the substance and/or relevant degradation products on sediment organisms. The need to conduct testing may be triggered by the following cases, e.g.:

i. PEC/PNEC >1 based on Equilibrium Partitioning Method (EPM)

ii. PEC/PNEC >1 based on available sediment studies (short/long term)

iii. Information on degradation of the parent compound in the water column showing formation of relevant metabolites (see Section R.7.1) that will be distributed to the sediment

iv. Information on degradation of the parent compound in the sediment showing formation of metabolites exclusively in this compartment (i.e. indications of anaerobic/aerobic degradation in the sediment to relevant metabolites)

v. Monitoring data showing occurrence of the substance or relevant metabolites in sediment

General rules in Annex VI and XI

In Annex VI it is stated that, in some cases, the rules set out in Annex VIII to X may require certain tests to be undertaken earlier than or in addition to the tonnage-triggered requirements. For substances that strongly adsorb or bind to sediment, uptake from sediment or food may become more important than uptake from water. Compounds that do not adsorb to particles are covered by the pelagic tests. On the other hand, substances that are highly hydrophobic (log K_{ow} >5) require sediment assessment even at tonnages below 1000 t/y. Therefore, a screening assessment using the equilibrium partitioning method (EPM) has to be performed also for such substances. If this screening assessment results in a PEC/PNEC value above 1, data improvement is necessary independent on the tonnage of the substance either by performing further long-term testing with sediment organisms or by refining the exposure assessment.. The same approach also applies to substances with intermittent release that adsorb to particles and that do not degrade rapidly.

Furthermore, it has to be considered that for substances that do not exhibit a toxic effect when tested in water only test systems because equilibrium was not reached during exposure phase may nevertheless exert significant toxic effects in sediment tests. Therefore, for these substances a read-across from pelagic data to sediment data is not possible. In such cases, it should be considered to perform toxicity test on sediment organisms (whole sediment tests) at lower tonnage levels (in accordance with annex VI).

Bioavailability considerations metals and inorganic metal compounds

Metal bioavailability in freshwater and marine sediments is governed by different ligands/processes (e.g. organic carbon, sulfides, iron and manganese oxy hydroxide and redox potential) and the relative importance of these binding phases may differ depending on the metals binding capacity and general behaviour).

It is recommended to make a clear differentiation between for example metal/inorganic metal compounds that are susceptible for binding with sulfides and those metals that are not sulfide binders, but where the use of partitioning to Fe-Mn (oxy)hydroxides, speciation calculations(reduced forms under anoxic conditions) and organic carbon normalisation may be more appropriate[9].

If it is relevant to take bioavailability of metals/inorganic metal compounds in sediments into account in the CSR, such as SEM/AVS[10] for metals, then it is recommended this correction be performed for both the effect data and exposure data.

Degradation products

For substances that degrade in the environment (but are not readily biodegradable) it might be necessary to test the degradation products, instead of or in addition to the parent substance. Generally, metabolites tend to be less hydrophobic than the parent substance and therefore have a lower adsorption potential, thus the relevance of the metabolites for the sediment compartment is normally lower than for the parent compound. However, there may be cases where the metabolites accumulate in the sediment compartment. In these cases, testing of metabolites might be necessary.

R.7.8.10.4 Remaining uncertainty

Compared to the pelagic compartment, there are only few tests available that examine the effects of industrial chemicals on sediment organisms. Thus, experience with these tests and with the assessment concept is still limited.

Up to now the available standardized test methods only deal with benthic invertebrates. Therefore, specific effects of chemicals on plants (that root in the sediment) or microorganisms are not covered by the available test methods. As these organisms also play an important role for benthic community, there is the necessity to further develop standard test methods and to revise the sediment assessment concept accordingly in future.

In the absence of any sediment tests, the equilibrium partitioning method is used as a screening to decide whether sediment tests are necessary. This is a further uncertainty as the EPM may also underestimate the toxicity of chemicals on sediment organisms. The additional factor of 10 on the PEC/PNEC ratio for highly adsorbing/ binding substances is meant to account for the possibility of uptake via sediment ingestion and so take account of this uncertainty. It should, however, be remembered that this is only a screening approach.

The assessment normally already starts with long-term tests without having information on the relative sensitivity of the test organisms to the chemical under consideration. Thus, there is the uncertainty that if only one long-term test is being performed, the employed species may not be the most sensitive. This uncertainty is only partly covered by the assessment factor of 100 and the result from this approach should therefore be treated with some caution.

[9] It should be noted that strictly spoken the outcome of the use of the SEM-AVS concept and the use of organic carbon and other ligands to normalise the total metal concentrations is a physico-chemical correction and do no not represent the true bioavailable fraction. As for the other compartments the effect of competition with biotic ligands should ideally be taken into account

[10] SEM = Simultaneously Extracted Metals; AVS = Acid Volatile Sulfides

R.7.8.11 Conclusions for toxicity to sediment organisms

R.7.8.11.1 Concluding on suitability for Classification and Labelling

Whole sediment tests with benthic organism are not standard tests for classification and labelling, as only exposure via the waterphase is normally considered for this purpose.

R.7.8.11.2 Concluding on suitability for PBT/vPvB assessment

Guidance on the suitability for PBT/vPvB assessment is given in Chapter R.11.

R.7.8.11.3 Concluding on suitability for use in Chemical Safety Assessment

The available data on sediment toxicity have to be evaluated for their adequacy for use in effect assessment and PNEC derivation according to the criteria described in section 6.8.15. Normally, little if any data will be available for sediment toxicity. In this case the equilibrium partitioning method can be used as a first screening approach to decide whether experimental data on toxicity to sediment organisms are necessary. For substances with a log K_{ow} >5 or substances with a correspondingly high adsorption or binding behaviour (e.g. ionisable substances, surface active substances, substances forming covalent bound to sediment components like e.g. aromatic amines) an additional factor of 10 has to be applied on the PEC/PNEC ratio, to take into account exposure of the benthic organisms via sediment ingestion.

If sediment tests are available in which the test substance was applied to the test system via spiking of the water phase, the effect values given in mg/l have to be converted into a sediment concentration (mg/kg) using the substance-specific partitioning coefficient or if available, measured sediment concentrations can be used.

If only one long-term sediment test is available, it should preferably be for an endobenthic, sediment-ingesting species and the exposure time should be long enough to enable adequate uptake of the sediment-associated substance by the test organism. E.g. if only a 72 h test with the bacterivourus nematode *Caenorhabditis elegans* is available (is considered as long-term test as growth inhibition and egg production are measured), the result from this test cannot be used alone for the derivation of the PNECsediment. However, such a test can be used as 2nd or 3rd test to lower the assessment factor if (a) long-term test(s) with other benthic species like *Lumbriculus* or *Chironomus* are already available.

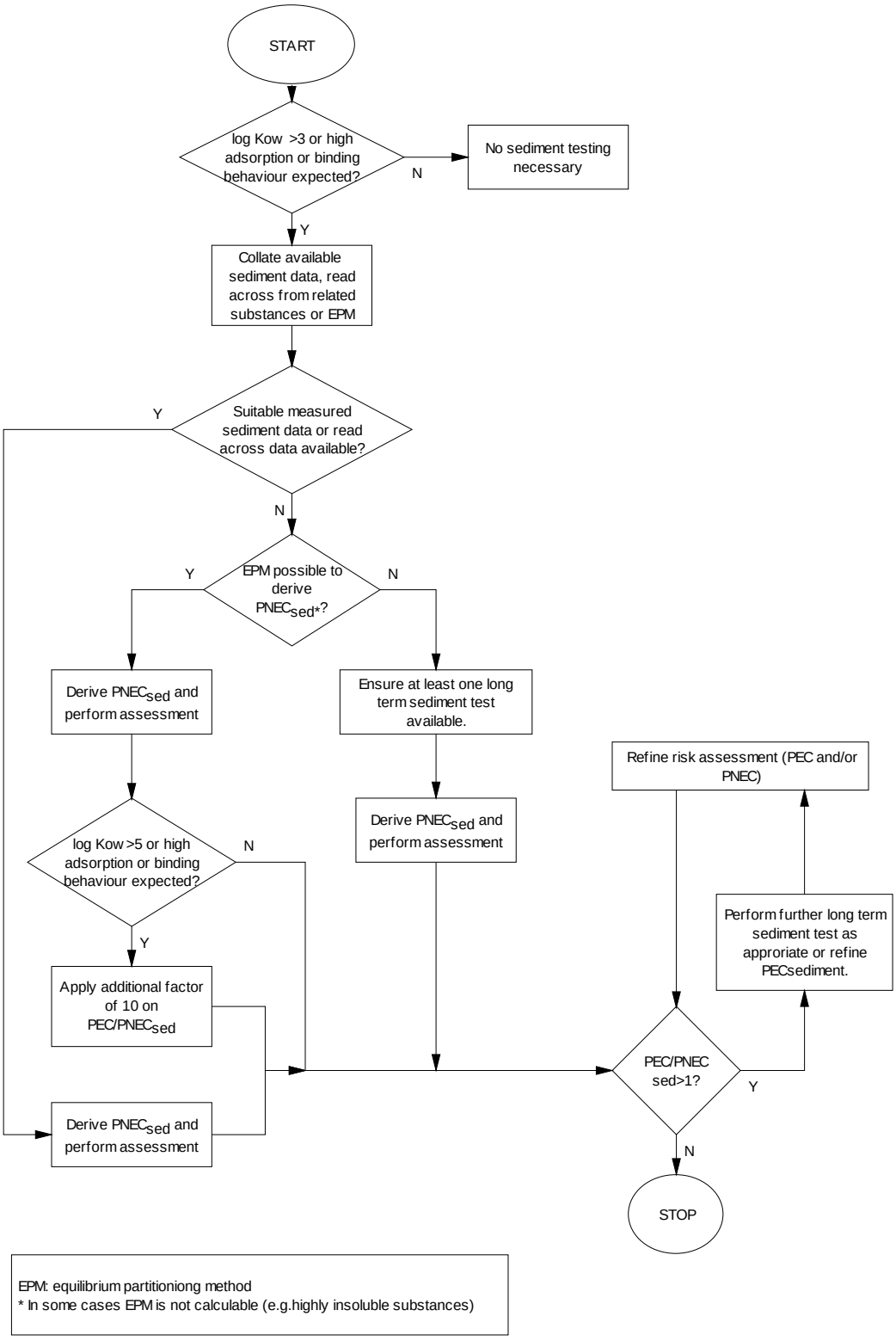
R.7.8.12 Integrated Testing Strategy (ITS) for toxicity to sediment organisms

R.7.8.12.1 Objective / General principles

An integrated testing strategy for the sediment compartment is necessary primarily for the *use in chemical safety assessment*, i.e. for the derivation of a PNECsediment. For C&L sediment tests are not necessary and therefore, no new sediment tests need to be performed to fulfil this regulatory demand. Concerning PBT assessment, long-term sediment toxicity tests may be appropriate to decide whether a substance fulfils the T criterion.

The testing strategy visualised in Figure R. 7.8- described below has the objective to give guidance on a stepwise approach to fulfil the regulatory demand.

Figure R. 7.8-9: Integrated Testing Strategy (ITS) for toxicity to sediment organisms



Note: in case no further risk refinements are possible, then apply appropriate risk reduction measures (e.g. minimizing exposure sufficiently so that RQ<1)

R.7.8.12.2 Testing strategy for toxicity to sediment organisms

The main property of a substance that triggers the assessment for the sediment compartment is the potential to adsorb or bind onto sediment. As trigger value for a sediment assessment a log K_{ow} of 3 is proposed. For substances exceeding this trigger value, at least a screening assessment using the equilibrium partitioning method has to be performed. For substances with a log K_{ow} between 3 and 5 this screening assessment results in the same risk characterisation ratio for sediment as for the pelagic compartment, as both $PEC_{sediment}$ and $PNEC_{sediment}$ are modelled from the corresponding pelagic data. Special attention should be given to substances with a log K_{ow} >5 or a correspondingly high adsorption or binding behaviour (ionising substances, surface active substances, substances that bind chemically with sediment components; for all these substances the adsorption is not triggered by the lipophilicity i.e. log K_{ow} of the substance but by other mechanisms). For these compounds a more comprehensive sediment assessment is needed.

If the need for a sediment assessment is clear, the availability of existing sediment toxicity data should be checked. In the absence of any (acceptable) sediment tests, the equilibrium partitioning method is applied as a first screen. If there is no measured sediment concentration available that is used as $PEC_{sediment}$, the PEC/PNEC ratio derived for the pelagic compartment can be used directly as both $PEC_{sediment}$ and $PNEC_{sediment}$ are derived from the corresponding aquatic values using the same partitioning coefficient. However, to take into account uptake of sediment-bound substance by benthic species, this PEC/PNEC ratio is increased by a factor of 10 for substances with log K_{ow} >5 or correspondingly high adsorption or binding behaviour (ionising substances, surface active substances, substances that bind chemically with sediment components; for all these substances the adsorption is not triggered by the lipophilicity i.e. log K_{ow} of the substance but by other mechanisms), unless scientific evidence can be provided that the extra factor is not applicable for that specific groups of substances. In this case the non-application of this additional factor has to be substantiated in detail. If the PEC/PNEC ratio is below one, no risk for the sediment compartment is indicated for the substance under consideration and further tests are not needed. If the PEC/PNEC ratio is above one, there is a need to perform long-term sediment tests with benthic species.

For substances that are highly insoluble and for which no effects are observed in aquatic studies, the application of the equilibrium partitioning method is not possible. In this case, at least one sediment test has to be performed.

If there is already one or more (acceptable) acute or long-term sediment test(s) available, a $PNEC_{sediment}$ is derived from these tests using an appropriate assessment factor (dependant on the data basis). If sediment tests with more than one benthic species are available, it has to be considered whether these organisms represent different habitats and feeding strategies and are thus exposed via different exposure pathways. Only in this case, a reduction of the assessment factor is possible. If the PEC/PNEC ratio is below one, no risk for the sediment compartment is indicated and further tests are not needed. If the PEC/PNEC ratio is above one, there is a need to perform (further) long-term sediment tests with benthic species.

If there are no adequate long-term sediment tests available, a test with preferably either *Lumbriculus variegatus* or *Chironomus* spec. using spiked sediment should be performed. A $PNEC_{sediment}$ has to be derived from the (lowest available) NOEC/EC10 using an appropriate assessment factor.

If the PEC/PNEC ration is below 1, no risk for the sediment compartment is indicated and there is no need to perform further tests. If the PEC/PNEC ratio is still above 1, the uncertainty can be reduced either by refinement of PEC or by performing another long-term sediment test with species representing different habitats and feeding strategies.

The following benthic species are recommended for testing:

- Long-term test with *Lumbriculus variegatus* using spiked sediment
- Long-term test with *Chironomus* spec. using spiked sediment
- Long-term tests with a further benthic species using spiked sediment. Selection of 3rd species should supplement the first 2 species in terms of habitat, feeding strategy, taxa or life-stage. This could be e.g. *Hyalella azteca*.

However, if there is in addition to the risk for the sediment compartment a risk for the pelagic compartment and the PEC/PNEC for the pelagic compartment is higher than the PEC/PNEC for the sediment compartment, any risk reduction measures applied to reduce the exposure of the aquatic compartment will also influence/cover the sediment compartment. In such a case the need to perform further sediment tests may be postponed to await the outcome of the emission reducing measures.

If the $PNEC_{sediment}$ is derived from the lowest NOEC/EC10 from three long-term sediment tests covering different exposure pathways and taxa and the PEC/PNEC ratio for the sediment compartment is still above one, further action must be taken to reduce the PEC.

In order to reduce testing, group approaches and read-across methods should be considered to partially or completely waive sediment studies. There should be sufficient studies available that further toxicity values can be reasonably predicted.

Examples: if for a certain chemical category clear evidence exists, that the additional factor of 10 significantly overestimates the toxicity to sediment organisms, the EPM can be used without this additional factor. This must be substantiated in detail. In other cases it may be sufficient to perform only one (long-term) sediment test, if for another substance from which read-across is possible, it can be deduced which is the most sensitive test species / test system in order to attain the lowest assessment factor.

Generally, the more substances that can be demonstrably classed into a single group, the less testing is required. A general guidance on how to extrapolate via read-across or chemical categories is given in Section R.6.2.

For the marine compartment, the same testing strategy is followed. However, for this compartment more tests may be necessary due to the higher assessment factor applied.

Table R. 7.8-5: Characterisation of benthic test species

Species	Taxonomic group	Habitat	Feeding mode
Chironomus sp.	insect	freshwater , endobenthic	Suspension and deposit feeder
Lumbriculus variegatus	oligochaete	freshwater, endobenthic	Sediment ingestor
Hyalella azteca	amphipod	Freshwater, Epibenthic	Detrivore, some subsurface deposit feeding
Hexagenia sp.	insect	freshwater, endobenthic	Surface particle collector
Tubifex tubifex	oligochaete	freshwater, endobenthic	Sediment ingestor
Diporeia spec.	amphipod	freshwater, endobenthic	Deposit feeder
Caenorhabiditis elegans	nematode	freshwater, endobenthic	bacterial ingestor
Leptocheirus plumulosus	amphipod	estuarine, endobenthic	Suspension and deposit feeder
Ampelisca abdita	amphipod	marine, endobenthic	Suspension and deposit feeder
Eohaustorius esturaius	amphipod	estuarine, endobenthic	Deposit feeder
Rhepoxynius abronius	amphipod	marine endobenthic	Meiofaunal predator, deposit feeder
Neanthes arenaceodentata *Neanthes virens*	polychaete	marine, endobenthic	Omnivorous deposit feeder
Corophium volutator	amphipod	marine, endobenthic	Suspension and deposit feeder

R.7.8.13 References on toxicity sediment organisms

Åkerblom N, Goedkoop W. 2003. Stable isotopes and fatty acids reveal that *Chironomus riparius* feeds selectively on added food in standardized toxicity tests. Environ. Toxicol. Chem. 22:1473-1480.

Ankley G.T., D.M. Di Toro, D.J. Hansen and W.J. Berry, 1991. Technical basis and proposal for deriving sediment quality criteria for metals.. Environmental toxicology and chemistry, 15 (12), 2056-2066.

ASTM. Standard Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates. ASTM Standard 1706-05, American Society for Testing and Material, Philadelphia

ASTM. Standard Guide for Conducting Sediment Toxicity Tests with Marine and Estuarine Polychaetous Annelids. ASTM Standard E1611-00, American Society for Testing and Material, Philadelphia

ASTM. Standard Test Method for Measuring the Toxicity of Sediment-associated Contaminants with Marine and Estuarine Invertebrates. ASTM Standard E1367-03e1, American Society for Testing and Material, Philadelphia

Di Toro D.M., C. Zarba, D.J. Hansen, R.C. Swartz, C.E. Cowan; H.E. Allen, N.A. Thomas, P.R. Paquin and W.J. Berry, 1991. Technical basis for establishing sediment quality criteria for non-ionic organic chemicals using equilibrium partitioning

Di Toro DM, Berry WJ, Burgess RM, Mount DR, O'Connor TP, Swartz RC. 2005. The Predictive Ability of Sediment Quality Guidelines Derived Using Equilibrium Partitioning. In R. J. Wenning & C. G. Ingersoll (Eds.), Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments Pensacola, FL: SETAC Press.

Environment Canada. 1997. Biological Test Method: Test for Growth and Survival in Sediment Using the Freshwater Amphipod *Hyalella azteca*. Environment Canada, Ottawa, Ontario. Technical report EPS 1/RM/33.

Environment Canada. 1997. Biological Test Method: Test for Growth and Survival in Sediment Using Larvae of Freshwater Midges (*Chironomus tentans* or *Chironomus riparius*). Environment Canada, Ottawa, Ontario, Technical report EPS 1/RM/32.

Hendriks AJ, van der Linde A, Cornelissen G, Sijm DTHM. 2001. Environ. Toxicol. Chem. 20: 1399-1420.

ISO. 2005: Water quality - Determination of acute toxicity of marine or estuarine sediment to amphipods. International Organisation for Standardisation, No. 16712

Lee JS, Lee BG, Luoma SN, Yoo H. 2004. Importance of equilibration time in the partitioning and toxicity of zinc in spiked sediment bioassays. Environmental Toxicology and Chemistry, 23 (1): 65-71.

OECD.1998: Detailed Review Paper on Aquatic Testing Methods for Pesticides and Industrial Chemicals. Organisation for Economic Cooperation and Development (OECD), OECD Environmental Health and Safety Publications. Series on Testing and Assessment No. 11, Paris.

OECD. 2000. Environmental Health and Safety Publications, Series on Testing and Assessment No. 23, Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures, Environment Directorate, Organisation for Economic Co-operation and Development, Paris, September 2000.

OECD. 2004a: OECD Guidelines for testing of chemicals No. 218: "Sediment-water chironomid toxicity test using spiked sediment".

OECD. 2004b: OECD Guidelines for testing of chemicals No. 219: "Sediment-water chironomid toxicity test using spiked water".

OECD. 2006: Rolling Workplan for the Test Guidelines Programme 2006/2008: Project 2.21: New Test Guideline Sediment-Water *Lumbriculus* Toxicity Test Using Spiked Sediment

OSPAR Commission. 2005: Protocols on methods for the testing of chemicals used in the offshore oil industry. A Sediment Bioassay using an Amphipod *Corophium* sp. Reference number 2005-11

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 Envionmental Toxicology and Chemistry – Europe, Brussels.

Simpson, SL, Angel BM, Jolley DF. 2004. Metal equilibration in laboratory contaminated (spiked) sediments used for the development whole-sediment toxicity tests. Chemosphere 54: 597-609.

U.S. EPA. 2000. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. 600/R-99/064 U.S. Environmental Protection Agency, March 2000

U.S. EPA. 2001. Method for assessing the chronic toxicity of marine and estuarine sediment-associated contaminants with the amphipod *Leptocheirus plumulosus*. 600/R-01/020 U.S. Environmental Protection Agency, March 2001

Wang F, Goulet RR, Chapman PM. 2004. Testing sediment biological effects with the freshwater amphipod *Hyalella azteca*: The gap between laboratory and nature. Chemosphere.57(11):1713-24.

R.7.8.14 Introduction to stp microorganisms' toxicity

R.7.8.14.1 Definition of toxicity to STP microorganisms

Adequate functioning of a STP (Sewage Treatment Plant) is essential to protect the downstream aquatic environment and to minimize operational costs. The endpoint of STP toxicity, as part of environmental risk assessment, was also included in the EU TGD (CEC, 2003). The aim of the assessment is the protection of the biodegradation and nutrient removal functions, and process performance in general, of municipal and industrial STPs.

Since chemicals may cause adverse effects on microbial activity in STPs, it is necessary to derive a PNECmicro-organisms (here called $PNEC_{stp}$). The PNECstp will be used as toxicity measure for the calculation of the risk quotient ($PEC_{stp}/PNEC_{stp}$) for microbial activity in STPs.

R.7.8.14.2 Objective of the guidance on toxicity to STP microorganisms

$PNEC_{stp}$ is determined by means of microbial toxicity tests. Currently used test systems for measuring the effect of chemicals on microbial activity have different endpoints and different levels of sensitivity. A number of internationally accepted test systems have been proposed in the past and their recommended use under REACH will be discussed further in this document.

For the engineered environment of a STP, *functional* endpoints (i.e. good and stable functioning) take precedence over *structural* endpoints (i.e. microbial population composition).

If the substance under consideration is released to both industrial- (i.e. production site) and municipal STPs, the toxicity assessment should be conducted separately for both types of STPs, with parameters relevant to the respective systems (see higher)[11].

R.7.8.15 Information requirements for toxicity to STP microorganisms

The assessment of $PNEC_{stp}$ is a requirement as of volumes of 10 tonne/year and above (REACH Annex VIII test requirement 9.1.4.). The type of test specified under 9.1.4 of REACH is an activated sludge respiration test (e.g OECD 209). Respiration inhibition is only one of many possible test approaches for measuring effects on microbes, but it is the most widely accepted indicator of the combined activity of sludge microorganisms. As such, the respiration inhibition test is preferred for the generation of new microbial toxicity data. This test can be substituted by a nitrification inhibition test if there are indications that the substance may be toxic to nitrifying bacteria.

Good quality data obtained with other types of microbial inhibition test methods, degradation- or sewage treatment simulation tests, can be also used to meet the REACH requirements, in particular if these studies were already existing (ITS scheme see Section R.7.8.12).

Column 2 of Annex VIII in REACH indicates that STP toxicity testing is not needed in the following cases:

- no emissions to STP (PEC = 0)
- the compound is readily biodegradable and PEC below test concentration applied
- there are mitigating factors, such as a very low solubility that would limit the exposure.

R.7.8.16 Information on toxicity to STP microorganisms and its sources

R.7.8.16.1 Laboratory data on toxicity to STP microorganisms and its sources

Non-testing data on toxicity to STP microorganisms

The practical use of QSARs for predicting STP toxicity is still limited. Although there are some QSARs for toxicity to microorganisms published (e.g. Blum & Speece 1990; Ren & Frymier 2002b; Redman et al. 2005; Schulz et al. 2005), this is not a very well developed science domain today. The existing microbial toxicity QSARs are mainly developed for baseline toxicity towards individual species of microorganisms, such as the ciliate *Tetrahymena pyriformis* (see work of T. Schulz and colleagues), and the bioluminescent *Vibrio fisher*i, formerly known as *Photobacterium phosphoreum* in the Microtox® test. On top of models for non-polar narcotics, some additional models specific to a particular class of chemicals are available. Since conceptual consistency is to be achieved between the experimental and QSAR approach for protecting microorganisms in STPs, the use of QSAR models developed for ciliates and individual species of bacteria not indigenous to STPs is to be excluded, however.

[11] In practice, many STPs treating domestic sewage also receive a fraction of industrial effluents, and a clear separation can not always be made. Municipal/domestic STPs are defined here as those plants of which the load predominantly consists of domestic waste waters.

Preliminary QSAR models for baseline toxicity to *P. putida* and for activated sludge respiration inhibition are reported in Redman et al. (2005). The reported models are based on a limited number of observations and have not been published yet in the peer reviewed literature. More validation work is needed here.

No QSAR models exist that accurately predict and protect nitrification inhibition. This is a significant outage, since nitrification can be the most sensitive endpoint – as illustrated in the experience of the EU existing chemicals programme.

The ProperEst website developed by the Fraunhofer Institute, to be publicly released, intends to provide a comprehensive compilation and documentation of microbial QSAR models (http://www.ime.fraunhofer.de/fhg/ime/EN/aoe/chp/expo/qsar/propertest.jsp).

In a *Weight of Evidence* context, consideration can be given to the use of read-across instead of testing, in particular for series of close chemical homologues for which there exist experimental data on some of the individual homologues.

Testing data on toxicity to STP microorganisms

INFORMATION FROM SUBCELLULAR MICROBIAL SYSTEMS:

A number of microbial inhibition test approaches exist which are based on subcellular systems, e.g. the Triphenyl Tetrazoliumchloride (TTC) Dehydrogenase assay (Ryssov-Nielsen 1975), β-galactosidase activity (Katayama-Hirayama 1986). Such in-vitro systems based on a single reaction have not been sufficiently validated in the context of STP risk assessment, and their use is therefore not accepted.

INFORMATION FROM MICROBIAL INHIBITION TESTS:

$PNEC_{stp}$ is routinely determined by means of microbial toxicity tests. Section R.7.8.14 provides an overview of the most commonly used microbial toxicity tests and their underlying concept. The toxicological endpoints are: respiration (i.e. O_2 uptake) inhibition, nitrification (i.e. ammonia conversion) inhibition, growth inhibition and bioluminescence. The list in Section R.7.8.14 is not aimed to be exhaustive, as many methodological variations and a suite of different test organisms have been proposed in the literature.

Literature information on the toxicity for microorganisms has to be assessed for its relevance with regard to the endpoint considered, i.e. microbial processes in a STP. In general, short-term measurements in the order of hours are preferred, in accordance with the hydraulic retention time in a STP (e.g. 10 h). Data on microbial toxicity from standard- and non-standard test methods is available for some compounds in the open literature (e.g. Blum & Speece 1991), in handbooks (e.g. Verschueren 2001), and in various databases (e.g. TETRATOX (www.vet.utk.edu), IUCLID).

Data from ciliate growth inhibition tests, preferably with the species *Tetrahymena* (OECD 1998; Pauli & Poka 2005), are also relevant for the risk assessment for STPs[12]. Ciliated protozoa, constituting the most important class of protozoa in STPs are, except for certain industrial plants, important for their functioning (NB: mainly for floc formation and settling properties, rather than for degradation processes). Toxicity data on ciliates are considered to be supplementary to the data on activated sludge or specific bacterial strains, i.e. no correlation exists between activated sludge and ciliate test results, neither are ciliates consistently more sensitive.

Tests using other characteristics (e.g. ciliary motion, cell movement, etc.) should not serve as a basis for the PNEC-derivation. For *Tetrahymena sp.* growth inhibition there exists a very large single endpoint database *TETRATOX* (www.vet.utk.edu). More than 2400 industrial organic compounds - of which more than 1,600 are published - have been tested at the University of Tennessee.

INFORMATION FROM BIODEGRADATION- AND SIMULATION TESTS

Absence of microbial toxicity can often be inferred from biodegradation studies in the laboratory. The information content of ready biodegradability tests (available as of 1 t/y) can under certain conditions also be used to derive a NOEC. This can be used to avoid new testing. The assumption that the substance under investigation is not inhibitory to the micro-organisms when dosed in the test system is implicit in ready biodegradability testing (i.e., EC C.4A-F, OECD 301A-F (OECD, 1992) and OECD 310 (2006)). If a compound degrades well in a ready biodegradability test, or does not inhibit the degradation of a positive control at a certain concentration, this concentration can be used as a NOEC value.

Any Ready Biodegradability Test relying on continuous monitoring, e.g. the MITI I test (EC C.4F; OECD 301C) or the Manometric Respirometry test (EC C.4D; OECD 301F) is considered more reliable for observing the effects of a chemical on the inoculum. A partial or transient toxic effect often results in a delayed mineralisation of the test substance and/or the positive control.

Data from biodegradation/removal studies using either inherent degradability tests (OECD 302A-C), or the laboratory/pilot scale Activated Sludge Simulation test (Continuous Activated Sludge (CAS) – OECD 303A and ISO-11733) may also be acceptable to derive a $PNEC_{stp}$ (OECD 1981; OECD 2001). The latter are laboratory scale models for simulation of activated sludge, representing realistic approximation to actual conditions in full scale STPs. The $PEC_{effluent}$ (or in the absence of that value the $PEC_{influent}$) from well-conducted simulation studies using domestic activated sludge would correspond to the concentration of the chemical substance that does not perturb the proper functioning of the CAS unit with regard to performance parameters such as test substance elimination, BOD/COD removal, nitrification, etc., when compared to a parallel non-dosed control.

[12] Following an international pilot ring test, a growth test with the ciliate *Tetrahymena pyriformis* was recommended for ecotoxicological risk assessment by the German Federal Environmental Agency. A full validation study to establish an internationally recognized Test Guideline has been conducted in the years 2000-2003. The resulting draft for an OECD protozoan test Guideline is currently under review.

R.7.8.16.2 Field data on toxicity to STP microorganisms and its sources

Absence of toxicity of a chemical can in a number of cases also be inferred from observations made at full scale plants. In particular for industrial STPs, the operators may have plant performance data in combination with chemical emission/exposure information, which can potentially be used to justify a $PNEC_{stp}$.

In addition, many full scale STPs are monitored on-line by commercial respirometer apparatus. A variety of commercial respirometers for activated sludge are available on the market (e.g. Strathtox, RODTOX, Oxitop, etc.). These systems monitor the Oxygen Uptake Rate (OUR) of the plant and can be used to derive a NOEC for respiration inhibition similar to laboratory tests and equipment. Some apparatus can also measure nitrification inhibition.

R.7.8.17 Evaluation of available information on toxicity to STP microorganisms

R.7.8.17.1 Laboratory data on toxicity on STP microorganisms

Non-testing data on toxicity on STP microorganisms

Use of non-testing data (QSARs) for STP Toxicity is not generally recommended given the limited availability of validated models relevant to STP organisms, and because an activated sludge respiration inhibition test is not particularly costly, complex or time-consuming to perform. Actual experimental data will typically overwrite calculated data, but QSARs may be useful to provide a preliminary estimate of toxicity for difficult-to-test substances.

In cases where relevant and well validated (Q)SARs for microbial toxicity would be developed in the future, this information could be fitted into the ITS to estimate PNECstp. Sound scientific judgement is needed to evaluate whether this information can replace the need for laboratory testing.

Testing data on toxicity on STP microorganisms

Information derived from sub-cellular microbial test systems (e.g. enzyme activity) as indicator of STP toxicity cannot be used.

The core microbial functions of a STP that need to be protected include carbon (BOD/COD) removal and nitrification. For some installations it is also important to protect other processes such as denitrification and biological P removal. Since there are no standardized test protocols for the latter endpoints, an assessment factor approach is routinely used to provide an adequate level of protection. There exists an anaerobic toxicity test ISO 13641 (2003) based on inhibition of biogas production, but its use to estimate the risk to STPs with biological nutrient removal would require further study.

TOXICITY TESTS WITH BACTERIA

In general, preference is given to tests with a mixed inoculum that assess the functioning of the entire microbial community in an STP, rather than tests based on single species or even microbial sub-systems. Respirometry is generally considered as an approach that will integrate the functioning of all organisms in an STP. The respiration inhibition test is generally positioned as a screening-level test (Painter 1986).

Nitrification inhibition tests, which assess the functioning of the sub-population of nitrifying organisms, are also amongst the preferred tests.

Not all microbial test systems are equally sensitive, however. Umweltbundesamt (UBA 1993) and Reynolds et al. (1987) suggest the following order of increasing sensitivities among particular test systems: respiration inhibition test < inhibition control in base-set tests < growth inhibition test with *P. putida* < inhibition of nitrification. Ren & Frymier (2003b) showed that nitrifying bacteria have a different, and generally higher sensitivity to toxicants, than other test systems. The response of the respiration-, *Tetrahymena*- and Shk1-assay clustered quite closely together in terms of sensitivity.

If activated sludge from an industrial sewage treatment plant is used as inoculum for a respiration or nitrification test, it is assumed that the microorganisms are adapted to the substance. Therefore, the test results cannot be extrapolated to municipal sewage treatment plants, since in municipal plants the bacteria may not be as adapted to the substance as the industrial sludge.

Often inhibition test data on individual bacterial species may be available. Results of the cell multiplication inhibition test with *P. putida* (Bringmann and Kühn 1980) should be used for calculation of the PNECmicro-organisms only in cases where no other test results are available. A similar recommendation is made for the Shk1 assay, which is based on a constructed bioluminescent *Pseudomonas sp.* originally isolated from activated sludge (Kelly et al. 1999; Ren & Frymier 2002a; Ren & Frymier 2003a).

Other single species tests with e.g. *Vibrio fischeri* (used in the MICROTOX® test), *Pseudomonas fluorescens* or *Escherichia coli* should be considered of low relevance for STPs. The tests with *P. fluorescens* and *E. coli* (Bringmann and Kühn 1960) cannot be used for determination of the PNECstp as they use glucose as a substrate (nor is E. coli a bacterium that will tend to multiply in an activated sludge environment). Likewise, *Vibrio fisheri* requires a high salinity environment. The information from such single-species screening tests may eventually be considered together with other existing data in a *Weight of Evidence* approach.

BIODEGRADATION AND SEWAGE TREATMENT SIMULATION TESTS:

The information content of ready or inherent biodegradability tests can also be used to derive a NOEC under the following conditions:

- when in a ready or inherent biodegradability test the compound is found to be respectively readily or inherently biodegradable,
- when in a ready or inherent biodegradability test a toxicity control has been included that shows good degradation of a positive control substance (e.g. glucose, sodium acetate) in the presence of the test substance.

Subject to expert judgement, data from biodegradation/removal studies using the laboratory/pilot scale Activated Sludge Simulation, Continuous Activated Sludge (CAS - OECD303A and ISO-11733) may also be acceptable to derive a PNECstp. In such tests it will be needed to monitor parameters such as BOD/COD removal, N-removal, sludge settling, etc., as compared to a parallel non-dosed control. Measuring chemical removal in such tests is optional, but can provide valuable additional information.

It should be noted that laboratory or field results obtained with an industrial sludge should be seen as plant-specific and cannot be extrapolated. Results for a municipal sludge can be extrapolated to other municipal installations provided that the emission pattern of the chemical is similar.

PROTOZOA TOXICITY TESTS

Ciliate-based test data can be used for deriving a PNECstp in case these are the sole data available, or in multiple-data situations where the ciliates have the lowest NOEC.

SUBSTANCES DIFFICULT TO TEST FOR STP TOXICITY:

Volatile and semi-volatile substances should not be tested in an open test system, e.g. the activated sludge respiration inhibition or nitrification inhibition test, since the chemical may be stripped from the system by the aeration. In such case, the recommended approach is to use a closed system, such as in OECD301F (Manometric Respiratory test) or OECD 310 (CO_2 headspace test).

R.7.8.17.2 Field data on toxicity on STP microorganisms

Also subject to expert judgement, data from full scale domestic or industrial sewage treatment plants that have received a certain chemical for prolonged periods can provide information useful to derive a $PNEC_{stp}$. This information can be used to avoid the need for additional laboratory testing. It would require that the concentrations of the chemical in the effluent or influent are well known, and the stable and efficient operation of the plant in the presence of the chemical has been confirmed (as e.g. indicated by prolonged BOD/COD- and N-removal performance, sludge settling, etc.).

R.7.8.17.3 Exposure considerations for toxicity on STP microorganisms

The paragraph below provides some guidance on exposure considerations for deriving a $PNEC_{stp}$:

Microbial toxicity testing above the solubility limit of a chemical is to be avoided, similar to toxicity test with higher organisms. It is also unrealistic because insoluble chemicals will be removed in the primary settling tank or fat trap of full scale installations, and thus will not reach the activated sludge.

However, data from existing tests where the experimentally derived NOEC is higher than the aqueous solubility can still be used as valid information to derive a $PNEC_{stp}$. This can be justified because it is a conservative estimate unlikely to occur in practice, and because undissolved test substance is found to be less confounding in microbial tests than in tests with higher organisms.

In the case of the respirometric method OECD 209, the test duration is very short; 30 or 180 minutes exposure to the chemical, followed by the measurement of oxygen uptake rate over 5-10 minutes. For chemicals with a low solubility, a contact time of 180 minutes (3 h) is to be used to ensure sufficient exposure. Some authors have proposed even longer exposure in respiration tests to lower the variability of the results (e.g. Gendig et al. 2003).

Keeping exposure constant during microbial toxicity tests: In batch microbial tests, the exposure is often not constant due to degradation, adsorption and other loss processes. It is generally assumed that the microorganisms have been exposed at the maximum level at the onset of the test and that the toxic effect, if any, has taken place at that point. Observation of degradation is further evidence of the detoxification ability of the microbes. For very unstable or sorptive chemicals, the need for a simulation test with continuous dosing such as the OECD 303A test may be considered if a batch test is deemed unreliable. This is not recommended as a routine procedure, however. The reader is also referred to OECD (2000) on testing of difficult substances.

R.7.8.17.4 Remaining uncertainty for toxicity on STP microorganisms

The choice of assessment factors to derive PNEC from microbial tests in the past has been rather empirical/arbitrary, and is not based on the same amount of comparative research as e.g. for the acute/chronic ratio for higher organisms (Table R.10-6 and Section R.10.4). One of the reasons that tests with single species of microorganisms have a lower assessment factor as compared to the recommended activated sludge respiration test, is that the latter is short term screening-type test, while former measure a chronic-type endpoint (growth).

Another aspect which requires consideration is that microbial toxicity results (e.g. respiration inhibition) tend to be proportional to the density of the culture, i.e. the test substance/biomass ratio. In other words, *dose* rather than *concentration* will determine the toxicity. This aspect is often overlooked in STP toxicity testing but can explain part of the differences in sensitivity sometimes noted between microbial inhibition tests (Elnabarawy et al. 1988).

The OECD 209 method operates at 1.6 g SS/l. The SimpleTreat Model version 3 (implemented in EUSES) uses 4 g SS/l in the aeration vessel as a default model value. When comparing microbial inhibition data from different test systems and origins it is good practice to verify if biomass levels are comparable. As a rule of thumb, deviations in biomass larger than a factor 10 are not suitable for direct cross-comparison. Inhibition tests executed at typical SS levels (1–4 g/l) should be considered as more reliable (nb: this guidance does not apply to nitrifying organisms for which levels in sludge are always much lower).

R.7.8.18 Conclusions for toxicity to sewage treatment plant microorganisms

Microbial toxicity tests on STP organisms are not required for Classification & Labelling, nor do they qualify for PBT assessment. Therefore the test data will only find application in Chemical Safety Assessment.

Mainly experimentally-derived microbial inhibition data will be used to derive a $PNEC_{stp}$ in the absence of well-established QSARs. As a general rule, data generated according to international standard guidelines and to GLP are to be preferred over other types of data.

Equally, however, it is important to appreciate that conclusions are to be based on the best available data, and that GLP studies can sometimes be flawed in other aspects. Thus, also available non-standard tests can be used, provided the data are considered scientifically valid.

In case of multiple microbial inhibition data, the $PNEC_{stp}$ is usually derived from results obtained for the most sensitive test system available, regardless of whether this is a test with activated sludge, relevant single bacterial species or ciliated protozoa. If there is considerable uncertainty around individual datapoints or questionable outliers, a *Weight of Evidence* approach can be followed.

R.7.8.19 Integrated Testing Strategy (ITS) for toxicity to STP microorganisms

R.7.8.19.1 Objective / General principles

The main objective of an ITS for STP Toxicity is to ensure that all available relevant exposure and effects information can be used before any new testing is initiated. This way, time and financial investment can be minimized, but without compromising on the quality of the assessment. On the other hand, the ITS should also allow to refine unfavourable screening data by means of higher tier testing. In the case of STP toxicity, the most realistic and highest tier test is a sewage treatment plant simulation test (OECD303A or equivalent).

The proposed scheme is to be followed for both industrial and/or domestic (i.e. municipal) sewage treatment plants, as applicable from the chemical's release pattern.

R.7.8.19.2 Preliminary considerations

In accordance with REACH Annex VI, the preliminary step of the ITS consists of a collection and critical evaluation of all (public) data that may be available for the STP Toxicity endpoint.

It should be noted that based on the test requirements in Annex VII for most substances a Ready Biodegradability test will be available. As such, there may be some relevant –but not necessarily fully conclusive- STP toxicity data available (except for inorganic chemicals which cannot be tested for degradability). The principle followed in the ITS is that existing data from short term tests can be retested/overwritten by more realistic/higher tier data, except if the existing data already come from simulation or field testing.

Step 1 covers calculation of exposure (PEC_{stp}) in both domestic and industrial plants, as applicable; this information will be needed to calculate the PEC/PNEC ratio and decide on need for more data/higher tier testing. Guidance on the PEC_{stp} calculation is provided by Chapter R.16.

Steps 2-4 cover evaluation of existing hazard information and the strategy to make optimal use of existing information, and avoid the need for new testing where possible.

Step 5 covers the execution of an activated sludge respiration test; i.e. first tier of STP toxicity testing (short term test).

Step 5* covers the retesting option for short term tests for industrial plants, based on sludge from that plant. These results are only relevant for this single plant, and cannot be extrapolated to other industrial or domestic plants.

Step 6 covers the execution of a confirmatory, longer term simulation test, i.e. the highest possible tier of STP toxicity testing. This is the test level with the highest real world relevance[13].

R.7.8.19.3 Testing strategy for toxicity to STP microorganisms

Stage 1. Calculation of exposure. Outcome: PEC_{stp} or $PEC_{influent}$ (calculate for both domestic and industrial STP, as applicable).

Stage 2. Assessment of information from existing and quality-assured microbial inhibition tests to derive a $PNEC_{stp}$ (i.e. data from respiration inhibition, nitrification inhibition, ciliate growth, sludge growth inhibition, P. putida, Shk1 assay).

Stage 2.1. IF adequate data are available, THEN derive $PNEC_{stp}$.
IF PEC/PNEC <1, THEN stop.
IF PEC/PNEC >1 for domestic plants, THEN move to stage 6, confirmatory testing

IF PEC/PNEC >1 for industrial plants, THEN move to stage 5* (nb: for industrial plants, there is the possibility to perform an activated sludge respiration test (or nitrification inhibition test) test with sludge from the specific installation)

Stage 2.2. IF no data are available, or the data are considered inadequate, THEN move to stage 3.

Stage 3. Assessment of information from Ready Biodegradation tests to derive a $PNEC_{stp}$.

Stage 3.1. IF the chemical is readily biodegradable, or if there is evidence of good degradation of a positive control in the presence of the test substance, THEN derive $PNEC_{stp}$.
IF PEC/PNEC <1, THEN stop.
IF PEC/PNEC >1, THEN go to stage 5 (nb: a respiration inhibition test can be used, if needed, to refine/overwrite the information inferred from a ready test. The respiration inhibition test may need to be done for both domestic and industrial sludge, as applicable).

Stage 3.2. IF no data are available from a Ready tests, or for all other situations not falling under stage 3.1 (e.g. not readily biodegradable and no information on inhibition), THEN go to stage 4.

Stage 4. Assessment of existing and quality-assured information from inherent biodegradability tests, simulation tests, and/or field data.

Stage 4.1. IF adequate data are available, THEN derive $PNEC_{stp}$.
IF PEC/PNEC <1, THEN stop.
IF PEC/PNEC >1, THEN risk reduction needs to be considered (no further refinement testing possible).

Stage 4.2. IF no data are available, or data are inadequate, THEN move to stage 5.

[13] Based on the experience with the existing high production volume chemicals programme in the EU (ca. 150 chemicals), it is expected that this approach will be seldom needed. For the large majority of chemicals, a lower tier assessment based on a short term tests will suffice.

Stage 5. Execution of an activated sludge respiration inhibition test (OECD 209). (NB: this test can also be substituted by a nitrification inhibition test)

Stage 5.1. IF PEC/PNEC <1, THEN stop.

Stage 5.2. IF PEC/PNEC >1 for domestic and/or industrial plants, THEN move to step 6

Stage 5. * Refinement test for industrial plants only: a test resulting in PEC/PNEC >1 can be repeated with sludge from the industrial plant of interest. This results can not be extrapolated to other plants

Stage 5.1. * If on the basis of a test with nitrifying bacteria (existing data), a PEC/PNEC ratio above 1 is derived for an industrial STP, a revised PNECstp for a specific industrial site can be derived from a nitrification inhibition test using the sludge from this site's STP. (NB: For domestic STPs a revision of the PNEC is not possible in this way, since sludge from one single STP can not be regarded as being representative of all domestic STPs with respect to their nitrifying activity).
IF $PEC/PNEC_{revised}$ <1, THEN stop.
IF $PEC/PNEC_{revised}$ >1, THEN proceed to stage 6 (simulation tests with investigation of nitrification performance)

Stage 5.2. * If on the basis of a standard respiration inhibition-, standardised biodegradation- or an activated sludge growth inhibition test (existing data), a PEC/PNEC ratio above 1 is derived for an industrial STP, a revised $PNEC_{stp}$ for can be derived from a respiration inhibition test using sludge from the site's specific STP.
IF $PEC/PNEC_{revised}$ <1, THEN stop.
IF $PEC/PNEC_{revised}$ >1, THEN move to stage 6.

Stage 5.3. * If on the basis of a single species test with ciliated protozoa a PEC/PNEC ratio above 1 is derived for domestic or industrial sewage treatment plants, a test reflecting the integrity of the native ciliate population is necessary (except if it can be shown that protozoa are not relevant in the system under consideration[14]). It is recommended here to move to stage 6, simulation testing, with investigation of settling performance.

Stage 6. Confirmatory simulation testing: an pilot scale simulation test, using activated sludge from the STP of interest (domestic or industrial) as a source of inoculum can be used as a highly realistic test to refine the $PNEC_{stp}$ derived from any short term microbial inhibition test. The stability and performance of the plant should be monitored over a somewhat longer period (e.g. 2 weeks, following a 2 week start-up period). The test should monitor critical performance parameters such as BOD/COD removal, N-removal (nitrification), and the evolution of the sludge volume index (SVI) parameter, versus an undosed control.

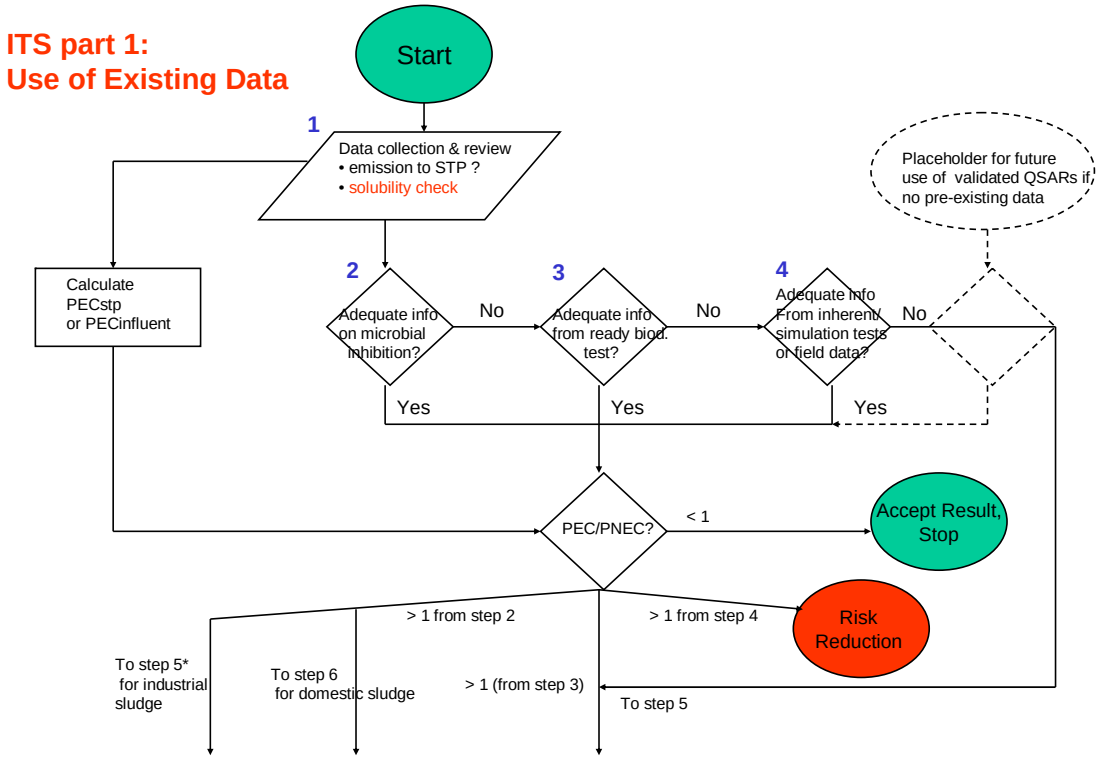
Stage 6.1. IF good and stable reactor performance, THEN stop (i.e. PEC/PNEC <1)

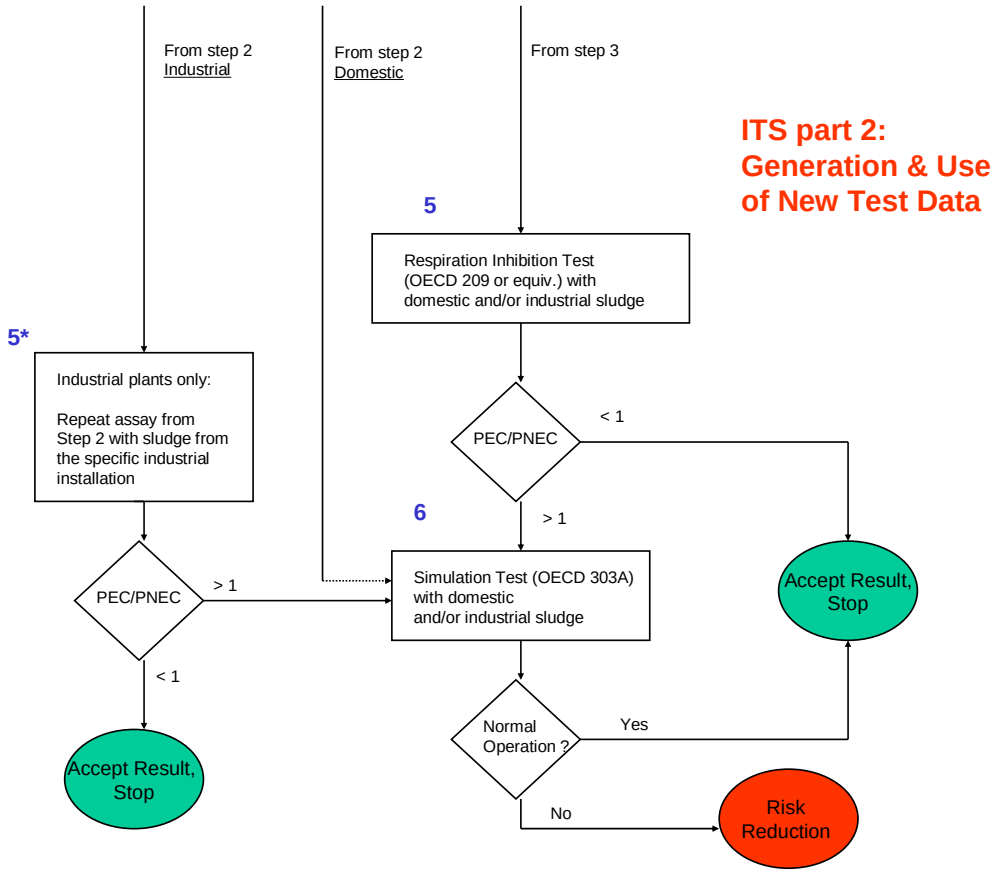
Stage 6.2. IF signs of inhibition or operational issues versus an undosed control unit, THEN PEC/PNEC >1, and risk management (emission reduction at source) is required.

[14] At present a standard protocol for a test on ciliated protozoa which can provide data on revising the PNECstp (based on ciliates) is not available. However, additional research results are underway and will be presented in 2007 by UBA.

(NB: for situations of intermittent release, a simulation test can be more difficult to perform; it would require a realistic dosing regime, which simulates the situation for the emission to the full scale plant).

Figure R. 7.8-10: Integrated Testing Strategy for toxicity on STP microorganisms





R.7.8.20 References on toxicity to STP microorganisms

Blum, D.J.W. and Speece, R.E. (1990). Determining chemical toxicity to aquatic species. Environ. Sci, Technol. 24, 284-293.

Blum, D.J.W. and Speece, R.E. (1991). A database of chemical toxicity to environmental bacteria and its use in interspecies comparisons and correlations. Res. J. Wat. Pollut. Control. Fed. 63, 198-207.

Bringmann, G. and Kühn, R. (1960). Vergleichende toxikologische Befunde an Wasser-Bakterien. Gesundheits-Ingenieur 11, 337-340.

Bringmann, G. and Kühn, R. (1980). Comparison of the toxicity tresholds of water pollutants to bacteria, algae and protozoa in the cell multiplication inhibition test. Wat. Res. 14, 231-241.

CEC (2003). 2nd Edition of the Technical Guidance Document TGD) on Risk Assessment of Chemical Substances following European Regulation and Directives. April 2003. European Chemicals Bureau (ECB), JRC-Ispra, Italy., http://ecb.jrc.it/tgdoc.

Elnabarawy, M.T., Robideau, R.R., and Beach, S.A. (1988). Comparison of three rapid toxicity test procedures: Microtox®, Polytox® and activated sludge respiration inhibition. Toxicity Assessment 3, 361-370.

Gendig, C., Domogala, G., Agnoli, F., Pagga, U. and Strotmann, U.J. (2003). Evaluation and further development of the activated sludge respiration inhibition test. Chemosphere 52, 143-149.

ISO (1986). Method 8192, Water Quality – Test for inhibition of oxygen consumption by activated sludge. Organisation for Standardisation (ISO).

ISO (1989). Method 9509. Water Quality - Method for Assessing the Inhibition of Nitrification of Activated Sludge Microorganisms by Chemicals and Waste Waters. International Organisation for Standardisation (ISO).

ISO (1996). Method 10712. Water quality – *Pseudomonas putida* growth inhibition test. Organisation for Standardisation (ISO).

ISO (1998). Method 11733. Water quality – Determination of the elimination of organic compounds in an aqueous medium. Activated sludge simulation test. Organisation for Standardisation (ISO).

ISO (1999). Method 9408. Water quality – Evaluation of ultimate aerobic biodegradability of organic compounds in aqueous medium by determination of oxygen demand in a closed respirometer. Organisation for Standardisation (ISO).

ISO (1999). Method 15522. Water quality – Determination of the inhibitory effect of water constituents on the growth of activated sludge microorganisms. Organisation for Standardisation (ISO), 15522.

ISO/DIS (2001). Draft guideline No. 14592-1. Evaluation of the Aerobic Biodegradability of Organic Compounds at Low Concentrations – Part 1. International Organisation for Standardisation (ISO).

ISO/DIS (2001). Draft guideline No. 14592-2. Evaluation of the Aerobic Biodegradability of Organic Compounds at Low Concentrations – Part 2. International Organisation for Standardisation (ISO).

ISO (2003). Draft guideline 13641. Water quality – determination of inhibition of gas production of anaerobic bacteria. Part 1: General test and Part 2: test for low biomass concentrations. Organisation for Standardisation (ISO).

Katayama-Hirayama, K. (1986). Inhibition of the activities of β-galactosidase and dehydrogenase of activated sludge by heavy metals. Wat. Res. 20, 491-494.

Kelly, C.J., Lajoie, C.A., Layton, A.C. and Sayler, G.S. (1999). Bioluminescent reporter bacterium for toxicity monitoring in biological wastewater treatment systems. Wat. Environ. Res. 71, 31- 35.

OECD (1981) OECD Guideline for the Testing of Chemicals: 302A, Inherent Biodegradability, Modified SCAS Test, OECD, Paris.

OECD (1984) OECD Guideline for the Testing of Chemicals: 209, Activated Sludge Respiration Inhibition Test, OECD, Paris.

OECD (1992) OECD Guideline for the Testing of Chemicals: 310A-F, Ready Biodegradability, OECD, Paris.

OECD (2000). Series in testing and assessement number 23. Guidance document on aquatic toxicity testing of difficult substances and mixtures. OECD, Paris.

OECD (2001). OECD Guideline for the Testing of Chemicals: 303A, Simulation Test: Aerobic Sewage Treatment - Activated Sludge Units, OECD, Paris.

OECD (2006). OECD Guideline for the Testing of Chemicals: 310 Ready Biodegradability –CO_2 in sealed vessels (Headspace Test), OECD, Paris.

Painter, H. (1986). Testing the toxicity of chemicals by the inhibition of respiration of activated sludge. Toxicity Assessment 1, 515-524.

Pauli, W. and Poka, V. (2005). Entwicklung und Standardisierung eines Testsystems zur Überprüfung der Wirkung von Chemikalien auf zentrale Funktionen der Protozoen-Biozoenose in Belebtschlämmen kommunaler Kläranlagen. Umweltforschungsplan des Bundesministers für Umwelt, Naturschutz und Reaktorsicherheit, Umweltbundesamt-Forschungsbericht 201 67 402.

Redman, A., McGrath, J., Di Toro, D. and Parkerton, T. (2005). Application of the narcosis target lipid model to waste water treatment organisms. Poster presented at SETAC Congress Baltimore, USA, 2005.

Ren, S. and Frymier, P.D. (2002a). Estimating the toxicities of organic chemicals to bioluminiscent bacteria and activated sludge. Wat. Res. 36, 4406-4414.

Ren, S. and Frymier, P.D. (2002b). Modelling the toxicity of polar and nonpolar narcotic compounds to luminescent bacterium SHK1. Environ. Toxicol. Chem. 21, 2649-2653.

Ren, S. and Frymier, P.D. (2003a). Comparative study of two bioassays for applications in influent wastewater toxicity monitoring. J. Environ. Eng. 129, 216-221.

Ren, S. and Frymier, P.D. (2003b). Use of multidimensional scaling in the selection of wastewater toxicity test battery components. Wat. Res. 37, 1655-1661.

Reynolds L., Blok J., de Morsier A., Gerike P., Wellens, H., and Bontinck, W.J. (1987). Evaluation of the toxicity of substances to be assessed for biodegradability. Chemosphere 16, 2259-2277.

Ryssov-Nielsen, H. (1975). Measurement of the inhibition of respiration in activated sludge by a modified determination of the TTC dehydrogenase assay. Wat. Res. 9, 1179-1185.

Schultz T.W., Netzeva, T.I., Roberst, D.W. and Cronin M.T. (2005). Structure-toxicity relationships for the effects to Tetrahymena pyriformis on aliphatic, carbonyl-containing, alpha, beta-unsaturated chemicals. Chem. Res. Technol. 18, 330-341.

Strotmann, U.J., Eglsäer, H. and Pagga, U. (1994). Development and evaluation of a growth inhibition test with sewage bacteria for assessing bacterial toxicity of chemical compounds. Chemosphere 28, 755-766.

Struys, J. Stoltenkamp, J. and van De Meent, D. (1991). A spreadsheet-based model to predict the fate of xenobiotics in a municipal wastewater treatment plant. Wat. Res. 25, 891-900.

UBA (1993). Entwurf zur Bewertung von Bodenbelastungen, Fachgebiet I 3.7. Umweltbundesamt (UBA), Berlin.

Verschueren, K. (2001). Handbook on Environmental Data on Organic Chemicals. John Wiley & Sons.

R.7.9 Degradation/biodegradation

R.7.9.1 Introduction

Degradation is an important process that can result in the loss or transformation of a chemical substance in the environment. Degradation of organic chemicals in the environment influences exposure and, hence, it is a key parameter for estimating the risk of long-term adverse effects on biota. Degradation rates, or half-lives, are determined in, or default rates assigned from, laboratory-based degradation tests. These tests can be simple screening tests (e.g. the OECD 301 ready biodegradability tests and the OECD 111 hydrolysis as a function of pH test), or relatively complex higher tiered simulation types of tests (e.g. the OECD 308 aerobic and anaerobic transformation in aquatic sediment systems, OECD 309 aerobic and anaerobic transformation in surface water and the OECD 303 aerobic sewage treatment).

Information on the degradability of chemicals may be used for hazard assessment (e.g. for classification and labelling), risk assessment (for chemical safety assessment) and persistency assessments (for PBT/vPvB assessment). Hazard and persistency assessments, or risk in general, and aquatic hazard classification in particular, are normally based on data obtained in standardised tests for ready biodegradability and hydrolysis. Results of tests simulating the biodegradation in water, aquatic sediment and soil may also be used for these purposes. Other types of test data that may be considered in an assessment of the potential environmental hazard or risk include sewage treatment plant (STP) simulation data, inherent biodegradability, anaerobic biodegradability, biodegradability in seawater and abiotic transformation (OECD, 2006). In determining which higher tiered or simulation degradation data are required consideration should be given to the partitioning behaviour of the chemical and its release or emission pattern. This may be useful for prioritising testing requirements to those environmental compartments that are the most relevant. Consideration should be given to whether the substance being assessed can be degraded to give stable and/or toxic degradation products. Where such degradation can occur, the assessment should give due consideration to the properties (including toxic effects and bioaccumulation potential) of the products that might arise.

R.7.9.1.1 Definition of degradation/biodegradation

Degradation can result in the loss or transformation of a chemical substance in the environment. Degradation processes can be abiotic or biotic. Abiotic or non-biological degradation can occur by physico-chemical processes such as hydrolysis, oxidation and photolysis. Removal due to biotic or biological degradation is commonly known as biodegradation. Biodegradation can proceed in the presence of oxygen (aerobic biodegradation) or in the absence of oxygen (anaerobic biodegradation).

Biodegradation is often preceded by the terms primary or ultimate. Primary biodegradation describes the initial transformation of a chemical by microorganisms to another organic chemical, a transformation product or metabolite; ultimate biodegradation describes the (multistep) degradation process leading to inorganic endproducts and biomass.

There are numerous terms and phrases associated with assessing degradation. Some of the commonly used terms are defined in Table R. 7.9-1.

Table R. 7.9-1: Glossary of terms associated with degradation

Term	Definition
Fate	Distribution of a chemical in various environmental compartments (e.g. soil or sediment, water, air, biota) as a result of transport, partitioning, transformation, and degradation.
Biodegradation	The biologically mediated degradation or transformation of chemicals usually carried out by microorganisms.
Primary biodegradation	The structural change (transformation) of a chemical substance by microorganisms resulting in the loss of the original chemical identity.
Ultimate aerobic biodegradation	The breakdown of a chemical by microorganisms in the presence of oxygen resulting in the formation of carbon dioxide, sulphate, nitrate and new biomass
Ultimate anaerobic biodegradation	The breakdown of a chemical in absence of oxygen resulting in the formation of carbon dioxide and final reduction products like methane, H2S, or NH3, mineral salts and new biomass.
Ready biodegradability tests	Stringent screening tests, conducted under aerobic conditions, in which a high concentration of the test substance (in the range of 2 to 100 mg/L) is used and ultimate biodegradation is measured by non-specific parameters like Dissolved Organic Carbon (DOC), Biochemical Oxygen Demand (BOD) and CO2 production. Small amounts of domestic sewage, activated sludge or secondary effluent form the microbial inoculum in tests for ready biodegradability. The inoculum should not have been artificially pre-adapted to the test substance through previous exposure to either the test substance or structurally related chemicals. The test substance is provided as the sole source of carbon for energy and growth. A positive result in a test for ready biodegradability can be considered as indicative of rapid and ultimate degradation in most environments including biological STPs
Inherent biodegradability tests	Tests inoculated with a high concentration of microorganisms carried out under aerobic conditions in which biodegradation rate and/ or extent are measured. The test procedures offer a higher chance of detecting biodegradation compared to tests for ready biodegradability and therefore if an inherent test is negative this could indicate the potential for environmental persistence.
Simulation tests	Aerobic and anaerobic tests that provide data on biodegradation under specified environmentally relevant conditions. These tests attempt to simulate degradation in a specific environment by use of indigenous biomass, media, relevant solids (i.e. soil, sediment, activated sludge or other surfaces) to allow sorption of the chemical, and a typical temperature that represents the particular environment. A representative and low concentration of test substance is used in tests designed to determine the biodegradation rate constant whereas higher concentrations for analytical reasons are normally used for identification and quantification of major transformation products.
Persistence	A chemical that resists degradation processes and is present in the environment for a long time. Specific criteria have been established in Persistent Organic Pollutant (POP) protocols, in the TGD and in REACH (PBT/vPvB). In the latter persistent (P) and very persistent (vP) refers to chemicals that have degradation half-lives above certain trigger values in surface water, sediment or soil.
Abiotic degradation	Degradation mediated through processes other than biodegradation such as hydrolysis, photolysis and interactions with other chemicals (e.g. oxidation). Abiotic degradation studies typically provide a

	measure of primary degradation.
Hydrolysis	Decomposition or degradation of a chemical by reaction with water
Photolysis	Chemical decomposition or degradation induced by light or other radiant energy. Direct photolysis in natural water involves the transformation of a chemical resulting from the direct absorption of a solar photon. Indirect photolysis in natural water sometimes involves the transformation of a chemical due to energy transfer from naturally occurring photosensitizers in electronically excited triplet states. However, indirect photolysis more often involves the transformation of a chemical due to reactions with transient oxidants such as hydroxyl radicals, molecular oxygen in a singlet electronic state, and peroxy radicals. Indirect photolysis is an important transformation process for chemicals in the gaseous state in air.
Oxidation	A substance may undergo oxidation/reduction or other transformation reactions (under storage, use etc.). These reactions may be slow and initiated for instance by the atmospheric oxygen or the presence of other oxidising agents.
Degradation rate constant	Typically a first order or pseudo first order kinetic rate constant, k (d-1), which indicates the rate of the degradation processes. However, depending upon the ratio of the chemical to degrader biomass, the rate constants may be Monod constants reflecting growth processes.
Half-life, t1/2	Term used to characterise the rate of a first or pseudo-first order reaction. It is the time interval that corresponds to a concentration decrease by a factor 2. The half-life and the degradation rate constant are related by the equation t1/2 = -ln2/k. Half-lives are usually expressed in hours or days and can be assigned to either primary degradation or ultimate biodegradation (mineralisation).
DT50	(Disappearance Time 50) is the empirically measured time within which the initial concentration of the test substance is reduced by 50%. It should be stated whether the DT50 refers to primary degradation or mineralisation (ultimate biodegradation)
DT90	(Disappearance Time 90) is the time within which the initial concentration of the test substance is reduced by 90%. In the case of a first-order reaction, this time would be slightly longer than 3 half-lives
Degradation product(s)	The chemicals produced as a result of degradation processes. For aerobic ultimate degradation, or mineralisation, these are carbon dioxide, water and mineral salts.
Field Data	Measured concentrations of a chemical in an environmental compartment, which can be related to loading, partitioning, dilution and degradation.

R.7.9.1.2 Objective of the guidance on degradation/biodegradation

The purpose of this report is to define an integrated testing strategy (ITS) that would help collect information on substances, within the context of REACH, i.e. to enable the hazard and risk assessment of substances to be performed. This information should form the basis for classification, PBT- and vPvB-assessment, and exposure assessment for use in risk characterisation. To do this all degradation data sources, including non-testing data, simulation testing data, field data, and exposure data will be taken into account.

Degradation is an important endpoint against the following *regulatory* needs:

- Identifying whether a chemical has PBT or vPvB properties and determining whether a chemical has the potential to cause long-term adverse effects in the environment in Environmental hazard classification
- Determining the Predicted Environment Concentration (PEC) of a chemical in environmental exposure assessment for use in risk characterisation

The general process of information collection will be a step-wise process. The following four processes are foreseen for collection of information on substance properties by a potential registrant according to the Guidance Note in Annex IV on the information requirements referred to in Article 9:

- Gather and share existing information
- Consider information needs
- Identify information gaps
- Generate new data/propose testing strategy

Within the report the proposed general ITS will be tested against selected substances. For exploration of elements of the strategy, fractions of the data of data-rich substances will be used to test the strategy i.e. different tonnage levels, different levels of available data etc.

R.7.9.2 Information requirements for degradation/biodegradation

Article 10 of REACH presents the information that should be submitted for registration and evaluation of substances. In Article 12 the dependence of the information requirements on production volume (tonnage) is established in a tiered system, reflecting that potential exposure increases with volume. Referring to article 10, Annexes VI to XI set out the requirements for generating information on the substance to be registered. However, for existing substances all available information should be used independently from the tonnage trigger.

R.7.9.2.1 Annex VII (Registration tonnage >1 t/y -<10 t/y)

Current text regarding degradation in Annex VII of the REACH proposal: This information is required if the substance meets the criteria laid down in Annex III:

- substances for which it is predicted (i.e.; by the application of (Q)SARs or other evidence) that they are likely to meet the criteria for category 1 or 2 classification for carcinogenicity, mutagenicity or reproductive toxicity or the criteria in Annex XIII.
- substances:
  - with dispersive or diffuse use(s) particularly where such substances are used in consumer preparations or incorporated into consumer articles; and
  - for which it is predicted (i.e. by application of (Q)SARs or other evidence) that they are likely to meet the classification criteria for any human health or environmental effects endpoints under Directive 67/548/EEC.

Column 1 Standard Information Required	Column 2 Specific rules for adaptation from Column 1
7.2. Degradation **7.2.1. Biotic** 7.2.1.1. Ready biodegradability	7.2.1.1The study does not need to be conducted if the substance is inorganic

Ready Biodegradation Test:

The waiving of the requirements for the following tests should be considered in the following circumstance:

Column 2: "The study does not need to be conducted if the substance is inorganic."

Inorganic substances cannot be tested for ready biodegradability.

R.7.9.2.2 Annex VIII (Registration tonnage ≥ 10 t/y)

Current text regarding degradation in Annex VIII of the REACH proposal:

Column 1 Standard Information Required	Column 2 Specific rules for adaptation from Column 1
9.2. Degradation	9.2. Further degradation testing shall be considered if the chemical safety assessment according to Annex I indicates the need to investigate further the degradation of the substance. The choice of the appropriate test(s) will depend on the results of the chemical safety assessment.
9.2.2. Abiotic 9.2.2.1. Hydrolysis as a function of pH	9.2.2.1. The study does not need to be conducted if: – the substance is readily biodegradable; or the substance is highly insoluble in water;

The requirements at this supply tonnage are for data on ready biodegradation (as defined in Annex VII) and for hydrolysis data at pHs 4, 7 and 9. Normally, a test for ready biodegradability would be required, although it may be possible to provide a valid QSAR as described in Section R.6.1.

HYDROLYSIS TEST

This test is designed to provide information on abiotic degradation that can help in classification, persistence testing and in determining the fate of a substance in environmental surface waters. The test may be waived under the following circumstances.

Column 2: "The substance is readily biodegradable."

In these circumstances, the hydrolysis test will provide little additional information since rapid mineralisation in the environment is already assumed.

Column 2: "The substance is highly insoluble in water"

In these circumstances, the test will be practically very difficult to conduct without special analytical techniques. In addition, it is likely that the aqueous environment may not be the principal environmental compartment of concern (see Section R.7.9.6). The test may still be important in certain circumstances however, for example where hydrolysis occurs at the surface of particles of the undissolved substance leading to more soluble products, but may be considered on a case-by-case basis if needed for risk assessment purposes.

R.7.9.2.3 Annex IX (Registration tonnage ≥ 100 t/y)

Current text regarding degradation in Annex IX of the REACH proposal:

Column 1 **Standard Information Required**	**Column 2** **Specific rules for adaptation from Column 1**
9.2. Degradation	9.2. Further biotic degradation testing shall be proposed by the registrant if the chemical safety assessment according to Annex I indicates the need to investigate further the degradation of the substance and its degradation products. The choice of the appropriate test(s) depends on the results of the chemical safety assessment and may include simulation testing in appropriate media (e.g. water, sediment or soil).
9.2.1. Biotic	
9.2.1.2. Simulation testing on ultimate degradation in surface water	9.2.1.2. The study need not be conducted if: – the substance is highly insoluble in water; – the substance is readily biodegradable.
9.2.1.3. Soil simulation testing (for substances with a high potential for adsorption to soil)	9.2.1.3. The study need not be conducted: – if the substance is readily biodegradable; or – if direct and indirect exposure of soils is unlikely.
9.2.1.4. Sediment simulation testing (for substances with a high potential for adsorption to sediment)	9.2.1.4. The study need not be conducted: – if the substance is readily biodegradable; or – if direct and indirect exposure of sediment is unlikely.
9.2.3. Identification of degradation products	9.2.3. Unless the substance is readily biodegradable

Additional biodegradation testing may be required at this tonnage depending on the relevant environmental exposure considerations.

Further biotic degradation testing shall be proposed by the registrant if the chemical safety assessment according to Annex I indicates the need to investigate further the degradation of the substance and its degradation products. The choice of the appropriate test(s) depends on the results of the chemical safety assessment and may include simulation testing in appropriate media (e.g. water, sediment or soil).

This may be taken as providing a general framework by which the exclusion of certain testing may be justified by the need to clarify or revise the conclusions of the CSA.

SIMULATION TESTING OF ULTIMATE DEGRADATION IN SURFACE WATER

Column 2: "The substance is readily degradable."

In these circumstances, the simulation test will provide little additional information since rapid mineralisation in the environment is already assumed. This will be so unless a refinement of the estimated environmental half-life is needed to aid the risk characterisation at regional scale.

Column 2: "The substance is highly insoluble in water"

The solubility in water may be so low that the test may be practically difficult or impossible to conduct at concentrations below the water solubility of the substance. It is also likely that the surface water environment will not be the principal environment of concern and, if a simulation test is required, consideration should be given to a test in a different environmental media. If the substance is considered a PBT/vPvB candidate, e.g. by fulfilling screening criteria on persistence, then it may be necessary to consider additional information.

SIMULATION TESTING ON ULTIMATE DEGRADATION IN SOIL

Column 2: "The substance is readily degradable."

In these circumstances, the simulation test will provide little additional information since rapid mineralisation in the environment is already assumed. This will be so unless a refinement of the estimated soil half-life is needed to aid the risk characterisation at regional scale.

Column 2: "If direct and indirect exposure of soil is unlikely."

If there is no exposure of the soil, or the exposure is so low that no refinement of the $PEC_{regional}$ is required, then this test may not be necessary. If the substance is considered a PBT/vPvB candidate, then it may be necessary to consider this test if soil is environmental compartment of concern (see Section R.7.9.6).

SIMULATION TESTING ON ULTIMATE DEGRADATION IN SEDIMENT

Column 2: "The substance is readily degradable"

In these circumstances, the simulation test will provide little additional information since rapid mineralisation in the environment is already assumed. This will be so unless a refinement of the estimated sediment half-life is needed to aid the risk characterisation at regional scale.

Column 2: "If direct and indirect exposure of sediment is unlikely"

If there is no exposure of sediment, or the exposure is so low that no refinement of the $PEC_{regional}$ is required, then this test may not be necessary. If the substance is considered a PBT/vPvB candidate, then it may be necessary to consider this test if sediment is environmental compartment of concern (see Chapter R.11).

IDENTIFICATION AND/OR ASSESSMENT OF DEGRADATION PRODUCTS

These data are only required if information on the degradation products following primary degradation is required in order to complete the CSA. This is considered further in Section R.7.9.4.

Column 2: "The substance is readily degradable"

In these circumstances, it may be considered that any degradation products formed during such degradation would themselves be sufficiently rapidly degraded as not to require further assessment.

R.7.9.2.4 Annex X (Registration tonnage ≥ 1000 t/y)

Current text regarding degradation in Annex VIII of the REACH proposal:

Column 1 **Standard Information Required**	**Column 2** **Specific rules for adaptation from Column 1**
9.2. Degradation **9.2.1. Biotic**	9.2. Further biotic degradation testing shall be proposed if the chemical safety assessment according to Annex I indicates the need to investigate further the degradation of the substance and its degradation products. The choice of the appropriate test(s) depends on the results of the chemical safety assessment and may include simulation testing in appropriate media (e.g. water, sediment or soil).

These data concerns further confirmatory testing on biodegradation and are only required if information on the degradation products following primary degradation is required in order to complete the CSA including the PBT-assessment or if felt necessary by the registrant because of implications for the hazard classification.

R.7.9.3 Information on degradation/biodegradation and its sources

This section identifies sources of information, including non-testing and testing data, which are important in the assessment of degradation. An inventory of officially adopted EU and OECD test guidelines and their application domain will be provided.

Other information such as the chemical physico-chemical properties are also important in identifying appropriate studies to conduct, for example certain biodegradation tests are not applicable for volatile and poorly water-soluble chemicals. These data can also assist in identifying environmental compartments of concern in order to prioritise higher tiered testing data accordingly.

R.7.9.3.1 Laboratory data on degradation/biodegradation

<u>Non-testing data on degradation/biodegradation</u>

DATABASES

Qualitative information is available for a number of biodegradation pathways, most notable the University of Minnesota Biocatalysis/Biodegradation Database (http://umbbd.msi.umn.edu/). This database collates known biodegradation pathways that have been published in the open literature. Many of these experimental studies were designed to determine pathways of biodegradation using pure cultures of microorganisms. Therefore these data can aid in the identification of potential degradation products where analysis of metabolites will be needed.

The suitability of this data on use in hazard, persistence and risk assessment needs careful consideration and may only contribute as part of the *Weight of Evidence* assessment if other data are available.

Two other major sources of empirical information are the Syracuse Research Corporations Environmental Fate Data Base (EFDB) (http://www.syrres.com/eSc/efdb.htm) that collates biodegradation, photooxidation and hydrolysis data and the Japanese Ministry of International Trade and Industry (MITI) database.

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS

A variety of models have been developed to predict biodegradation. These include structure biodegradability relationships (SBRs) and quantitative structure biodegradability relationships (QSBRs). SBRs provide qualitative endpoints such a passing or failing a ready biodegradation test. QSBRs provide an estimation of rate or half-life. Examples of such models include:

- Syracuse Research Corporations Estimation software that includes packages to determine log octanol-water partition coefficients, Henry's Law constant, indirect photolysis in the atmosphere (by reaction with OH and NO_3), biodegradation and hydrolysis (http://www.syrres.com/esc/est_soft.htm). The Biodegradation Probability Program for Windows (BIOWIN) calculates the probability that a chemical under aerobic conditions with mixed cultures of microorganisms will biodegrade rapidly or slowly (http://www.syrres.com/esc/biowin.htm). In help files of the programme the training set chemicals used for development of the BIOWIN models are presented. Recently HCBIOWIN, a model that predicts the primary degradation half-lives of hydrocarbons in water has also been developed by Syracuse. Description of the model and its development is given in Howard et al. (Howard et al., 2005) (see also http://www.epa.gov/opptintr/exposure/pubs/episuite.htm).

- CATABOL is a mechanistic modeling approach for quantitative assessment of biodegradability in biodegradation pathways of chemicals. It attempts to predict microbial biodegradation (BOD, and CO_2 production) in two ready biodegradation tests (the modified Sturm CO_2 evolution test (OECD 301B) and the MITI I test (OECD 301C). It also displays the metabolic pathways and generates most plausible biodegradation products and provides quantitative estimation of their physico-chemical properties and acute toxicity to aquatic organisms (http://www.oasis-lmc.org/?section=software).

- TOPKAT has an aerobic biodegradability module. This module comprises a statistically significant and cross-validated quantitative structure-toxicity relationship (QSTR) model applicable to a specific class of chemicals, and the data from which these models were derived. A single study that reported the biodegradability of 894 compounds, as assessed by the Japanese Ministry of International Trade and Industry (MITI) I test protocol, was used to develop these models. Molecular structure is the only input required to conduct an assessment of aerobic biodegradability (http://www.accelrys.com/products/topkat/modules/aero_degrade.html).

- Multicase has a META program to create metabolic breakdown pathways of chemicals. All rules have been determined based on reliable literature sources. A tree of products can be saved and analysed for mammal metabolism, aerobic biodegradation, anaerobic biodegradation and photodegradation (http://www.multicase.com/products/prod05.htm).

- ECB has made the Danish QSAR database available (http://ecbqsar.jrc.it), which contains QSAR predictions on degradability from the EPIWIN models and from a MCASE QSAR model developed by the Danish EPA. The database contains predictions on almost all discrete organic EINECs and TSCA substances and require only the CAS number as an input, but in addition allows complex searches to be made (combined search algorithms concerning the predictions for all endpoints included in the database by use of the following conditional options to be fulfilled by specific searches: "OR", AND" and "NOT" and conditions such as ">", "<", =, "contains" plus option for choice of freely selected sub-structures and in relation to recorded EU production tonnage level: 1-10 t/y; 10-1000 t/y & > 1000 t/y.

It is noted that the various QSAR models for biodegradation estimation with the exception of BIOWIN 1, 2, 3 & 4 have been developed based on training set data consisting of results from ready biodegradability tests, in particular MITI I data, which uses a uniquely derived inoculum. The training set for BIOWIN 1, 2, 3 (ultimate degradation time frame) and 4 (primary degradation timeframe) on the contrary, was based on the overall conclusions of a panel of USEPA experts for rapid or slow environmental degradation and based on various types of degradation information on the training set substances. Nevertheless also the BIOWIN 1, 2 and 3 model has been tested (*validated*) in literature for its predictability concerning not ready and ready biodegradability.

For prediction of hydrolysis there are also some freely available models. The Syracuse Research Corporations Estimation software (EPIWIN) includes also a HYDROWIN program to estimate hydrolysis half-life. (http://www.syrres.com/esc/est_soft.htm). Another useful program for estimation of hydrolysis is SPARC (http://ibmlc2.chem.uga.edu/sparc/). The program is available freely on the internet for single substance calculations by use of CAS no or SMILES input of the chemical identity (http://ibmlc2.chem.uga.edu/sparc/).

For prediction of photolysis the Syracuse Research Corporations Estimation software (http://www.syrres.com/esc/; EPIWIN) includes the AOPWIN program, which calculates the indirect photolysis half-life in the atmosphere by reactions with OH^- and NO_3^- radicals. Also a Multicase photodegradation program exists.

The Danish QSAR database also contains EPIWIN predictions for photodegradability and hydrolysis for the chemicals included in the database.

Testing data on degradation/biodegradation

PHYSICO-CHEMICAL DATA

The interaction of a chemical with the environment is an important consideration. The fate and behaviour of a chemical is largely governed by its inherent physico-chemical properties. Knowledge regarding the physico-chemical properties of the substance enables the most appropriate abiotic degradation and biodegradation tests to be identified. These data together with multimedia fate and transport models will also enable higher tiered tests to be prioritized accordingly. Information on the following physico-chemical properties determined using the relevant OECD technical guidelines identified is desirable: vapour pressure, water solubility, absorption - desorption using a batch equilibrium method, partition coefficient (n-octanol/water), dissociation constants in water, partition coefficient (n-octanol/water) - HPLC method, and Estimation of the Adsorption Coefficient (K_{oc}) on Soil and on Sewage Sludge using High Performance Liquid Chromatography (HPLC). Additional information is provided in chapter R.7.1.

For many chemicals measurements on partition coefficients (log K_{ow}, log K_{oa} and log K_{aw}) are not available and estimation methods based on fragment methods are often used. If a substance has properties that do not allow for the reliable estimation of partition coefficients or environmental rate constants, the models may fail to predict realistic environmental concentrations.

ABIOTIC DEGRADATION DATA

Abiotic processes such as hydrolysis, oxidation and photolysis may transform chemicals in aquatic environments, soil and air. Abiotic transformation can be an important step in the pathway for degradation of chemicals in the environment (OECD, 2006). The following guideline exists to assess abiotic degradability:

OECD 111: Hydrolysis as a Function of pH

There are also two draft OECD guidelines considering photolysis. These are (1) Phototransformation of chemicals on soil surfaces and (2) Phototransformation of chemicals in water-direct and indirect photolysis.

For many chemicals measurements of abiotic degradation may not be available and QSARs derived rates or estimates of degradation may be available (see above).

BIODEGRADATION DATA

In general, the assessment of degradation processes should be based on data, which reflect the environmental conditions as realistically as possible. Data from studies where degradation rates are measured under conditions that simulate the conditions in various environmental compartments are preferred. The applicability of such data should, however, be judged in the light of any other degradation data including results from screening tests. Most emphasis is put on the simulation test results but in the absence of simulation test data, degradation rates and half-lives have to be estimated from screening test data. Listed below are the OECD guidelines to assess biodegradability:

- OECD TG 301: Ready Biodegradability
  A: DOC Die-Away Test
  B: CO2 Evolution Test
  C: Modified MITI Test (I)
  D: Closed Bottle Test
  E: Modified OECD Screening Test
  F: Manometric Respirometry Test
- OECD TG 302: Inherent Biodegradability:
- A: Modified SCAS Test
- B: Inherent Biodegradability: Zahn-Wellens/EMPA Test
- C: Inherent Biodegradability: Modified MITI Test (II)
- OECD TG 303: Simulation Test - Aerobic Sewage Treatment
  A: Activated Sludge Units
  B: Biofilms

- OECD TG 304A: Inherent Biodegradability in Soil
- OECD TG 306: Biodegradability in Seawater
- OECD TG 307: Aerobic and Anaerobic Transformation in Soil
- OECD TG 308: Aerobic and Anaerobic Transformation in Aquatic Sediment Systems
- OECD TG 309: Aerobic Mineralisation in Surface Water - Simulation Biodegradation Test
- OECD TG 310: Ready Biodegradability - CO_2 in sealed vessels (Headspace Test)
- OECD TG 311: Anaerobic Biodegradation of Organic Compounds in Digested Sludge - Method by Measurement of Gas Production

Section R.7.8.1 contains a list of the ISO and OPPTS tests that are equivalent to the OECD guidelines listed above. This chapter also lists some of the important attributes of each test.

The existing methods for testing ready biodegradability (OECD 301 series and OECD 310) and the endpoints evaluated are compiled in Section R.7.8.1. It is important to recognise that the guidelines are not applicable to all substances, especially difficult substances with low water solubility, volatile or adsorbing properties. The applicability of the ready biodegradability tests for poorly water soluble, volatile and adsorbing chemicals has been identified by OECD (2006).

Proposed new guidelines currently being reviewed within OECD include a series of simulation tests, which have been designed to assess the primary and ultimate biodegradability of chemicals discharged to wastewater. These tests consider biodegradation in:

- Wastewater
- Activated Sludge
- Anaerobic Digester Sludge
- Mixing Zone for Treated Effluent and Surface Water
- Mixing Zone for Untreated Wastewater and Surface Water

The applicability of these new proposed guidelines for environmental hazard and risk assessment requires further discussion. However, they are not likely to be relevant for classification & labelling, but have their greatest relevance for quantitative risk assessment.

NON-STANDARD PUBLISHED BIODEGRADATION STUDIES

In addition to the standardised data described above there is a vast amount of non-standardised biodegradation data that has been published in the scientific literature. Many of these studies share some common principles with the ready biodegradability tests, for example the test chemical is usually introduced to the microorganism or microbial community as the sole source of carbon for growth and energy. There is a general reluctance to use these types of data on regulatory purposes. However, they may be valuable, as part of a *Weight of Evidence* assessment, and attempts should be made to gather, evaluate and when appropriate use these types of information.

R.7.9.3.2 Field data on degradation/biodegradation

The ultimate verification for an environmental risk assessment is to measure chemical concentrations or removal in the environment (e.g. Fox et al., 2000). Monitoring data can be used directly in the risk assessment and it can also be used to refine the exposure models or the biodegradation rates.

When monitoring data are considered in the risk assessment of substances, the data are often obtained from existing monitoring programmes. In that case the field or monitoring study has not specifically been designed to fulfil regulatory needs. In such cases extra care should be given to the selection of relevant data. When field studies or monitoring campaigns are specially designed to fulfil regulatory needs of REACH the monitoring studies can be designed and implemented accordingly. It must be noted that monitoring data can be required under REACH only as a result of a substance evaluation. For the use of existing and the generation of new field data attention should be given to following aspects:

- reliable and representative data should be selected by evaluation of the sampling and analytical methods employed and the geographic and time scales of the monitoring campaigns. As sampling and measurements are usually performed at a local geographical a justification is required to demonstrate that measured chemical concentrations are representative for the risk assessment, particularly if the data are to be used in regional exposure models.
- the data should be assigned to local or regional scenarios by taking into account the sources of exposure and the environmental fate of the substance.
- the measured data should be compared to the corresponding calculated PEC. For naturally occurring substances background concentrations have to be taken into account. For risk characterisation a representative PEC should be decided upon based on measured data and a calculated PEC.

In the risk assessment of chemicals a cautious approach is followed. This means that PECs are computed for a relevant scenario that describes usually the worst-case (but still realistic) situation. A common quantification of a vulnerable situation is a combination of geochemical scale and parameters, time scale and climate that results in the 90^{th} percentile PEC. An example of this approach for surfactants in surface water is described by Feijtel et al. (1999). This approach is also used in environmental risk assessment for pesticide registrations (FOCUS, 2000).

SEWAGE TREATMENT PLANTS

Monitoring in sewage treatment plants can be very useful. The endpoint usually is a percentage of removal during the residence time in the sewage treatment plant. Also for the determination of metabolites monitoring the sewage treatment plant (STP) is a good tool. Monitoring in STP's is usually not expressed as a biodegradation rate as removal due to degradation and/ or sorption to sludge solids is usually not resolved. Recent publications on monitoring in STP's include Morral et al. (2006), Eadsforth et al. (2006) and Belanger et al. (2006).

SURFACE WATER MESOCOSMS.

A mesocosm is a controlled field experiment. Although the primary endpoints of this study are the effects on aquatic organisms, it is possible to obtain information on the fate of substances at the same time. The system is usually closed, and spiked with the substance under realistic outdoor conditions, with representative flora and fauna included. OECD (2006) provides guidance for the set-up of microcosm and mesocosm experiments.

For the marine environment no such guidance document exists, but the IOCCP (International Oceans Carbon Coordination Project) noted that there was an immediate need to develop guidelines and protocols for mesocosm experiments, and is pulling together appropriate scientists from different research programs to develop these. http://ioc.unesco.org/ioccp/ MesoCosmGuides.htm. The TGD (2003) indicates that the same rules as for fresh surface water should apply for seawater. Relevant literature includes Grice & Reeve (1982), Lauth et al. (1996) and Culp et al. (2000).

Large-scale monitoring studies have been performed for surfactants. These monitoring studies are generally focussing on improvement of PNEC's or better estimates of PEC's instead of better estimates of biodegradation rates. An overview of methods, fate and risk assessment for surfactants is given in Knepper et al. (2003).

SOIL AND GROUNDWATER

Three types of field data can be distinguished for soil and groundwater.

- Lysimeter studies
- Field studies
- Monitoring studies

Lysimeter studies can be compared with mesocosm studies. They are closed, controlled, outdoor systems, making it possible to use radiolabelled substances and to study the mass-balance. Field studies are semi-controlled, because the system is not closed, the mass-balance can not be checked, so loss of substance is more undefined as compared to lysimeter studies. In monitoring studies are even more uncertainties arise, because the exposure of the compartment is not under control and the system is not closed.

Especially for pesticides many lysimeter, field and monitoring experiments have been performed. Guidance for the performance and evaluation of these studies, aiming at risk assessment in soil and groundwater is given by OECD (2000), Verchoor et al (2001) and Cornelese et al (2003).

R.7.9.4 Evaluation of available information on degradation/ biodegradation

R.7.9.4.1 Laboratory data on degradation/biodegradation

Non-testing data on degradation/biodegradation

QSAR CALCULATION

The development and validation of (Q)SARs is an intense and continuous research area. The future European Chemicals Agency in collaboration with the Commission, Member States and the interested parties will develop and provide guidance in assessing which (Q)SARs may be suitable for regulatory purposes and to provide reporting formats for a transparent reporting the extent of validation of the models (QSAR Model Reporting Formats (QMRF)) as well as reporting information relevant for judging the reliability of predictions for individual substances (QSAR Prediction Reporting Formats (QPRF)). QMRF displays a description of the QSAR model relative to the five OECD QSAR validation principles in a systematic and summarised way (OECD 2006c). The QPRF shows how a prediction of an individual chemical and endpoint relates to the applicability domain of the used QSAR model. It furthermore contains test data information on the endpoint on close structural analogs to the chemical that the prediction is made for. In that case it also describes how closely related the analogs are to the chemical that the prediction is made for. Development of QMRFs and QPRFs has already started in the framework of the TCNES QSAR working group and draft reporting formats on for example biodegradability using the BIOWIN models and CATABOL models have been developed and discussed. It is foreseen that the draft formats will be further refined and that the work on development of QRFs will continue. It is also foreseen that further guidance on use of QSAR models/model predictions including the use of *Weight of Evidence* approaches for specific regulatory purposes will continue and that further guidance will be prepared and issued by ECB.

In a recent draft review an overview of existing validations of a range of the most frequently used QSAR models for prediction of ready/ not ready biodegradability have been given (Pavan & Worth, 2006).

One example on the use of QSAR models for predicting ready biodegradability is the BIOWIN model (v4.02), which estimates biodegradation of organic chemicals. It has the following estimation summary line:

READY BIODEGRADABILITY PREDICTION: YES OR NO

A recently proposed criterion (USEPA 2004) for an *overall* YES or NO prediction are as follows: If Biowin3 (ultimate survey model) result is “weeks” or faster (e.g. days or days to weeks) ≥2.75 AND Biowin5 (MITI linear model) ≥0.5, then the prediction is YES (readily biodegradable). If this condition is not satisfied, the prediction is NO (not readily biodegradable) according to this proposal for drawing an overall *Weight of Evidence* -based conclusion (EPIWinn ver. 3.12, 2004). The acceptability of this generic *Weight of Evidence* -based criterion has until now not been considered in the EU working groups dealing with hazard and risk assessment.

Another example of a *Weight of Evidence* procedure that has been used is the TGD (2003) QSAR based screening criterion for identifying substances for persistency (P and vP). BIOWIN 2 <0.5 or BIOWIN 6 <0.5 and BIOWIN 3 <2.2. (- 2.7), i.e. for substances fulfilling this but BIOWIN indicates a value between 2.2 and 2.7 more degradation relevant information is generally warranted in relation to the PBT testing strategy according to the working practices of the EU PBT working group cf. TGD (2003) & EU WG on Substances of very High Concern: Working document: SHC/TS 2-3/029 2002.

In general the following freely available BIOWIN models may be used when predicting the ready biodegradability of chemicals BIOWIN1, 2, 3, 5 and 6. It is noted that according to various validation studies performance of the models seem to differ, but in general not ready biodegradability predictions seems to be more certain than ready biodegradability predictions (GHS 2004 & OECD 2004: (ENV/JM/TG/2004)26Rev1 and references therein). However, in some particular cases arguments may be provided for using also ready biodegradability predictions for regulatory decisions (e.g. when many valid individual QSAR model predictions supported by read-across considerations indicate ready biodegradability, cf. the example with toluene in the chapter with case studies). The prediction value cut off points between ready and not ready biodegradability predictions relative to the particular BIOWIN model is indicated in the table. These cut off points were used in a comparison of 177 high production volume (HPV) chemicals in relation to biodegradation test data compared with model predictions by the shown QSAR models (OECD 2004: (ENV/JM/TG/2004)26Rev1) but the same cut off points have been used before in a range of validations studies in the past (Table R. 7.9-2).

Table R. 7.9-2: QSAR Cut off Points between Ready and Non-Ready Biodegradability

QSAR model:	**Probability cut off point:**	**Reference:**
BPP1 (BIOWIN1, linear)	0.5	Howard et al. (1992); Boethling et al., (1994); and TemaNord (1995)
BPP2 (BIOWIN2, non-linear)		
BPP3 (BIOWIN3)	2.75	Boethling et al. (2004)
BPP5 (BIOWIN5, linear)	0.5	Roije et al. (1999); Tunkel et al. (1999); and Boethling et al. (2003)
BPP6 (BIOWIN6, non-linear)		
DK DEG (MCASE)	yes/no	Report on the Advisory list for self-classification of dangerous substances (http://www.mst.dk/chemi/01050000.htm)

Generally it is only recommendable to use the single QSAR model predictions when these are clearly within the applicability domain of the model. Whether this is the case may not always be easy to conclude. For BIOWIN models the structural domain can be checked *manually* by checking whether or not a prediction on the individual chemical was exclusively based on substructures known to the model or whether the chemical also contained sub-structures unknown to the model. It is noted that the BIOWIN models can give predictions for chemicals which only contain substructures that are unknown to the particular BIOWIN model (i.e. not represented in the training set of chemicals for the model). This is due to the fact that the BIOWIN models then revert to an assumption of the probability of biodegradability which is solely related to the molecular mass of the substance (i.e. the greater the molecular mass the less probability for a high probability score implying rapid biodegradation).

This implies that checking of the applicability of whether predictions are within the applicability domain of BIOWIN models may be important. Contrary to this both Multicase and CATABOL models includes more automated features for checking whether the individual predictions they make are within the applicability domain of the models. For Multicase models the program contains possibilities to pre-define the structural domain by use of statistically defined criteria. However, different possibilities exist for defining the stringency of such definitions of the applicability domain. Also the CATABOL program contains possibilities to check whether predictions are inside the applicability domain of the model. (cf. further in Pavan & Worth, 2006).

When using *Weight of Evidence* and model predications from various QSARS related to the same regulatory endpoint such as not ready/ ready biodegradability in order to increase the confidence associated overall with a general conclusion based on all model predictions, it is important to consider the performance of the individual models based on known validation information (about the sensitivity, specificity and positive and negative predictive values of the individual models). Another factor to consider is the extent to which the training sets of the models do or do not *overlap* (cf. further in OECD 2004, ENV/JM/TG/2004)26Rev1 where different types of *Weight of Evidence* approaches referring to BIOWIN 1, 2, 5, & 6 model predictions have been exemplified and discussed).

When using both individual as well as multiple QSAR model predictions for ready / not ready biodegradability it is relevant is to consider dropping use of predictions which are close to the borderline cut off between ready and not ready biodegradability. It has for example been proposed not using BioWIN 1, 2, 5, 4 or 6 model predictions with a biodegradability probability score between 0.4 and 0.6. (because the cut off point is 0.5). Such a strategy seems according to an analysis done by RIVM on the SIDS data set included in OECD 2004, ENV/JM/TG/2004)26Rev1 to increase the level of predictability (Rorije, 2005).

In relation to abiotic degradation several models are relevant to consider using. For hydrolysis it is the HYDROWIN model (v.1.67), which estimates aqueous hydrolysis rate constants for the following chemical classes: esters, carbamates, epoxides, halomethanes and selected alkyl halides (US-EPA 2004). This QSAR has only a limited coverage of the existing substances e.g. listed in EINECS Another possibility is using the hydrolysis module of SPARC for estimating a hydrolysis half-life.

Finally included in the EPISuite is also a programme for estimation of indirect photo-oxidation. This Atmospheric Oxidation Program for Microsoft Windows (AOPWIN) estimates the rate constant for the atmospheric, gas-phase reaction between photochemically produced hydroxyl radicals and organic chemicals. It also estimates the rate constant for the gas-phase reaction between ozone and olefinic/acetylenic compounds. The rate constants estimated by the program are then used to calculate atmospheric half-lives for organic compounds based upon average atmospheric concentrations of hydroxyl radicals and ozone. The estimation methods used by the Atmospheric Oxidation Program are based upon the structure-activity relationship (SAR) methods developed by Dr. Roger Atkinson and co-workers. Predictions of this programme has recently been evaluated and found reasonable reliable in general (Muller, 2005). Generation of estimates for atmospheric degradation half-life of chemicals in the gaseous phase may be useful when making initial assessment by multi-media modelling of the potential for long-range environmental air transport and its possible implication for the selection of a simulation study of degradation in the open sea (see Section R.7.9.6).

SAR EVALUATION

Various approaches comprised under the heading Structure-Activity Relationship (SAR) may be used for giving an indication of the degradation potential of a substance. Characteristics of a substance may give a first indication of the likely degradation possibilities.

A large number of chemical substances are not completely stable, but have certain reactivity potential. By time or by influence of environmental factors, the substance may undergo transformations, which lead to structural changes. In collecting and reviewing existing information on degradation characteristic of the substance, information on possible transformation properties is important.

Even if biological processes accelerate the decomposition of some simple inorganic substances they may not normally degrade biotically and consequently biodegradability testing is not worth doing. The inorganic substances may dissociate in the environment (like water soluble salts) or undergo other transformation reactions (atmospheric oxidation, photo-oxidation, hydrolysis, slow biomethylation etc.) that may change the character or magnitude of environmental hazards or risks. The rate of these transformations may be fast, indicating remarkable instability of the substance under certain conditions. For unstable substances, the character of instability and the rate of transformation and transformation products (to other substances) need to be described to estimate hazards and fate of the chemical properly. If no test data are available, the rate of transformation needs to be described to some extent, i.e. the expected order of magnitude of rate of transformation at specified conditions ($t_{½}$ = minutes, days or weeks?). In addition, one of the key issues is how relevant the qualitative and temporal conditions, in which the substance is unstable, are for typical use and/or emission scenario situations.

Organic substances may contain structures that indicate a rapid biotic degradation or on the other hand that the substance is recalcitrant. Some organics that are not structurally defined may be of a natural origin, and they may often be degradable (e.g. fatty acids), while other types of organics often are recalcitrant (e.g. multi-branched alkyl structures). Cf. further in OECD (1993).

The two main approaches used in regulatory settings are:

- Read-across from analogues and
- Read-across within a chemical category.

The two approaches generally have a good regulatory acceptance and can be applied to any endpoints, whether physico-chemical property, environmental fate and pathways, ecotoxicity or toxicity.

In principle these two approaches can therefore be applied for most types of degradation tests reviewed in this report and for any type or regulatory purpose (Environmental hazard classification, PBT and vPvB assessment and Exposure assessment) provided that the estimation is sufficiently robust in accordance with the currently available guidance documents as reviewed in the TAPIR report (IWG 3; ECB, 2005) “Non-Testing considerations”. Another way of assessing the robustness of the read-across and categorisation approaches in relation to ready biodegradability of un-tested chemicals will be to make comparison and make an overall evaluation relative to predictions made by QSAR models.

In practice, most experience on the use of non-testing methods for estimation of the potential for biodegradation is available using the approaches for screening, i.e. tests on ready biodegradation, estimation of hydrolysis and atmospheric degradation rate time frame. For other types of tests, specifically those giving kinetic results e.g. simulation tests for an environmental compartment or determination of degradation products, the applicability of these approaches are currently limited as not much experience is available.

Testing data on degradation/biodegradation

ABIOTIC DEGRADATION

HYDROLYSIS

Abiotic hydrolytic transformation of chemicals in aquatic systems may be examined at pH values normally found in the environment (pH 4-9) by use of the guideline: Hydrolysis as a Function of pH (OECD 111). This method is generally applicable to chemical substances (^{14}C-labelled or non-labelled) for which an analytical method with sufficient accuracy and sensitivity is available. The results of a test of hydrolysis may include (OECD, 2006):

- Repeatability and sensitivity of the analytical methods;
- Recoveries;
- Mass balance during and at the end of the study (when ^{14}C-labelled test substance is used);
- Half-life or DT_{50}.

Most hydrolysis reactions follow apparent first order reaction rates and, therefore, half-lives are independent of the concentration. This usually permits the extrapolation of laboratory results determined at high concentrations to low environmentally realistic concentrations. The specific reporting requirements for the hydrolysis test are described in Section R.7.9.9.

TEMPERATURE DEPENDENCE OF HYDROLYSIS

In general, the hydrolysis reactions are relatively sensitive to temperature. Reliable extrapolation of hydrolysis rates from higher to lower temperature (e.g. from 25°C to 10°C) may contain remarkable uncertainties (OECD 2004; Lyman et al. 1990). Temperature dependence of hydrolysis reactions can be reliably determined only by testing the reaction rate at a number of temperatures. The OECD 111 test guideline on hydrolysis points out that higher tier (tier 2) hydrolysis tests should be carried out with a minimum of three temperatures and preferably at least one temperature below the standard reporting temperature of 25°C. If the hydrolysis test conditions do not meet this requirement, the hydrolysis half-life should be compared just to the screening criteria and not to the definitive Annex XIII criteria.

The temperature dependence of hydrolysis reactions reflects to the intrinsic activation energy of the reaction that is taking place. The higher the activation energy is, the slower is the relative rate of hydrolysis at reduced temperature. In practice, temperature dependence of the activation energy is specific for each chemical and reaction, leading to moderate variability in reaction rates between substances at reduced temperature compared to standard reporting test temperature (25°C).

High extrapolation uncertainties can be best avoided by selecting appropriate testing temperatures. For the PBT assessment purposes, the 10°C testing temperature is a good choice for tier 2 testing purposes.

However, a rough hydrolysis temperature correction estimate may be done by using equation: t ½ (X°C) = t ½ e $^{(0.08\ (T\ -\ X)}$. This equation uses "fixed" activation energy (ca. 70 kJ/mol) for all hydrolytic reactions and for all substances. This equation results to fixed 1.5 fold change in hydrolysis rate per each 5°C change in temperature.

MODIFICATIONS TO THE HYDROLYSIS TEST CONDITIONS

At screening level, priority should be given to test results applying standard test methods. However, quite often modifications to standard methods are needed to overcome testing difficulties, but basically these test modifications should not have influence on the observed degradation rates. For instance in highly modified test systems, surface-controlled reactions can predominate over bulk solution hydrolysis (reflecting rather soil than aquatic environment). The highly modified systems may result in different, poorly comparable degradation rates than would be predicted from standard guideline based rates in homogeneous solutions.

Typically very dilute solutions and relatively low temperature are the prevailing environmental conditions. Attention is needed to interpret whether these test conditions, e.g. test temperature and test substance concentration have had such influence on the test results that reliable extrapolation to environmental conditions is possible. If the abiotic transformation is likely to be reversible in the environmental conditions, the relevance of the transformation observed must be carefully interpreted whether results can be used in persistence assessment.

For example, unnecessarily high concentrations of test substances and buffers should be avoided since reaction mechanism may be heavily influenced by high concentrations as well highly elevated temperature.

PHOTOTRANSFORMATION

The potential effects of solar irradiation on the fate of chemicals in surface water and soil may be examined by use of the draft guidelines: Phototransformation of Chemicals in Water – Direct and Indirect Photolysis and Phototransformation of Chemicals on Soil Surfaces, respectively (OECD, 2006).

The direct and indirect phototransformation of chemicals in natural water bodies is a complex process that depends on a number of factors such as:

- the chemical structure and electronic absorption spectrum of the chemical;
- the concentration, composition, and absorption spectra of chromophoric dissolved organic matter (CDOM; from which photosensitizers and singlet oxygen arise);
- the concentration of nitrate (the primary source of hydroxyl radicals); and
- the solar photon flux spectrum to which the chemical, CDOM and nitrate are exposed.

Any data on half-lives or DT_{50}, DT_{75} and DT_{90} values should be reported along with calculations associated with these data, and the results of any outdoor experiments, if the latter have been conducted. Where possible, information on transformation products should be provided as well (OECD, 2006).

The level of information required in the test report depends on the complexity and purpose of the study. Consequently, OECD has identified a number of tiers for direct and indirect photolysis in water (see the relevant guidelines for details, OECD, 2006).

Phototransformation data may be of use for assessing direct photolysis in air. It may also be of use for assessing photolysis in water when factors such as water depth, suspended matter and latitude are taken into account.

BIODEGRADATION

READY BIODEGRADABILITY

Ready biodegradability tests must be designed so that positive results are unequivocal. Given a positive result in a test of ready biodegradability, it may be assumed that the chemical will undergo rapid and ultimate biodegradation under most environmental conditions. In such cases, no further investigation of the biodegradability of the chemical, or of the possible environmental effects of transformation products, is normally required. However, the fact that the chemical is found to be readily biodegradable does not exclude a possible need for further information about biodegradation rate constants and the transformation products in cases of high influx into a receiving environment. Realising that ready biodegradability tests may sometime fail because of the stringent test conditions, in general, and the differences among the individual tests in terms of their stringency, consistent positive test results from test(s) should generally supersede negative test results. However, when conflicting test results are reported, it is recommended to consider such differences in stringency and to check the origin of the inoculum in order to check whether or not differences in the adaptation of the inoculum may be the reason (OECD, 2006).

When faced with conflicting results using different ready protocols, it is also important to consider the following.

- Test material concentration
    - Very high concentrations (100 mg/L) used for some 301 tests increases the probability of inhibition or mass transfer issues for low solubility materials.
    - Very low concentrations (2-5 mg/L) used for the closed bottle test can sometimes overestimate degradation given the poor signal to noise (theoretical vs. background) in the test.
- Inoculum
  The pre-treatment of the inoculum such as in the MITI test (OECD 301C) seriously impact the diversity of the microbes (Forney et al., 2001).

- The Analyte
  O_2 uptake tests result in problems due to difficulties in estimating theoretical O_2 production when chemical structure is not defined and elemental analyses are complicated and the chemicals are resistant to oxidation in a COD analysis. Greater confidence should be given to CO_2 based tests because of better certainty around the theoretical values.

A negative result in a test for ready biodegradability does not necessarily mean that the chemical will not be degraded under relevant environmental conditions and persist in the environment. A failed ready biodegradability test indicates that further testing under less stringent test conditions should be considered at the next level.

The OECD tests which can be used to determine the ready biodegradability of organic chemicals include the six test methods described in the OECD 301 test guidelines. The following pass levels of biodegradation, obtained within 28 days, may be regarded as evidence of ready biodegradability: 70% DOC removal (TG 301 A and TG 301 E); 60% theoretical carbon dioxide (ThCO2; TG 301 B); 60% theoretical oxygen demand (ThOD; TG 301 C, TG 301 D and TG 301 F).

These pass levels have to be reached in a 10-day window within the 28-day period of the test. The 10-day window does not apply to TG 301 C or if the test substance represents a mixture of homologous compounds e.g. technical surfactants. The 10-day window begins when the degree of biodegradation has reached 10% DOC removal, ThOD or $ThCO_2$ and must end before or at day 28 of the test. The pass levels of either 60% ThOD (or ThCO2) or 70% DOC removal practically represent complete ultimate degradation of the test substance as the remaining fraction of 30-40% of the test substance is assumed to be assimilated by the biomass or present as products of biosynthesis (OECD, 2006).

Another test for ready biodegradability, which represents an alternative to the CO_2 Evolution Test (OECD 301 B), is the Headspace Test (Ready Biodegradability – CO_2 in sealed vessels; OECD 310). This test is especially suitable for volatile compounds. In this test, the CO_2 evolution resulting from the ultimate aerobic biodegradation of the test substance is determined by measuring the inorganic carbon (IC) produced in sealed test bottles, and the pass level has been defined as 60% of theoretical maximum IC production (ThIC).

Ready biodegradability tests usually last for 28 days. However, biodegradability tests may be ended before 28 days, i.e. as soon as the biodegradation curve has reached a plateau for at least three determinations. Alternatively, they may be prolonged beyond 28 days when the curve shows that biodegradation has started but that the plateau has not been reached by day 28 (OECD, 1992). Where chemicals have not achieved the pass criterion for ready biodegradability in the 28-day test duration the substances are not considered to be readily biodegradable by OECD (1992). Substances where mass transfer or substance availability is limited fall into this category e.g. poorly-water soluble substances.

Tests should be conducted in accordance with the OECD principles for Good Laboratory Practice, and the test report should include the information identified in Section R.7.9.9.

MARINE BIODEGRADABILITY

The OECD TG 306 on Biodegradability in Seawater includes seawater variants of the Closed Bottle Test (OECD 301 D) and of the Modified OECD Screening Test (OECD 301 E). Degradation of chemicals in seawater has generally been found to be slower than that in freshwater tests inoculated with activated sludge and sewage effluent, and, therefore, a positive result obtained during 28 (Closed Bottle Method) or 60 days (Shake Flask Method) in the biodegradability in Seawater test can be regarded as evidence of a chemical's potential for biodegradation in the marine environment. A result of >20% ThOD or DOC removal is indicative of potential for primary biodegradation in the marine environment, whereas a result of >60% ThOD or 70% DOC removals is indicative of potential for ultimate biodegradation in the marine environment (OECD, 2006). When a chemical attains >60% ThOD or >70% DOC removal in a Biodegradability in Seawater test (OECD 306), it can also be expected to fulfil the criteria for ready biodegradability.

MODIFIED READY BIODEGRADABILITY TESTS

Two modifications to the standard ready biodegradability and marine biodegradability tests have been identified below. These consider biodegradability testing at low test substance concentrations and assessing the biodegradation of poorly water soluble chemicals. Provided that all other conditions in the Ready Biodegradability Tests are fulfilled, these tests are regarded as Ready Biodegradability Tests and the results can be used directly in classification.

TESTING AT LOW TEST SUBSTANCE CONCENTRATIONS DUE TO INOCULUM TOXICITY

For chemicals that are known or expected to exert toxicity to the microbial inoculum a lower test substance concentration should be used. The toxicity of the test substance to microorganisms can be determined using one of a number of microbial toxicity tests e.g. the activated sludge respiration inhibition test (OECD 209). Where possible the lower test substance concentration should still allow the assessment of biodegradability to be determined through the measurement of carbon dioxide evolution, oxygen demand or dissolved organic carbon removal. Reduction in the toxicity in the ready biodegradability tests may also be achieved by the introduction of carriers allowing the 'slow-release' of the test substance during the test period.

Conducting studies at low concentrations may only be possible if the test chemical is available containing an appropriate radioisotope. If this is not possible then the primary biodegradability of the test chemical should be measured using specific chemical analysis. If primary degradation is being measured then an attempt should be made to identify any major degradation products.

GUIDANCE ON BIODEGRADABILITY ASSESSMENTS OF POORLY WATER-SOLUBLE SUBSTANCES

The standardised ready biodegradation test methods adopted by the OECD that are listed above were initially developed to evaluate the biodegradability of test substances which are soluble in water to at least 100 mg l^{-1} provided they are non-volatile and non-adsorbing. For substances that are poorly soluble in water, volatile or adsorbing OECD concluded that only a subset of the ready biodegradability test guidelines were applicable (Table R. 7.9-4).

For poorly-water soluble substances these are the OECD 301B, 301C, 301D and 301F tests and the OECD 310 test. For volatile substances these are the OECD 301C, 301D and 301F tests and the OECD 310 test. For adsorptive substances these are the OECD 301B, 301C, 301D and 301F tests and the OECD 310 test.

Tests using DOC analysis cannot be used to assess the biodegradability of poorly water-soluble substances unless it is measured in addition to another parameter. Specific chemical analysis can also be used to assess primary degradation of the test substance and to determine the concentration of any intermediate substances formed. Specific chemical analysis is obligatory in the MITI method (OECD 301C; OECD, 1992). Strategies to determine the biodegradability of poorly water-soluble chemicals are described in section R.7.9.10.

ENHANCED BIODEGRADATION SCREENING TESTS

A number of potential enhancements to the ready biodegradation test have been identified. These enhancements have been identified to assist in persistency assessments and are not to be used in Classification and Labelling. The enhancements are designed to help improve the environmental relevance of biodegradability assessments without the immediate requirement for simulation level testing. The potential enhancements described below have been published and they would benefit from being ring-tested by appropriate international standards bodies. Test substances that degrade in these enhanced biodegradation screening tests will not be considered as readily biodegradable.

With the exception of the MITI I test (OECD 301C), the current ready biodegradation tests the inoculum can be obtained from a number of sources as long as it has not been pre-exposed to the test chemical or it is not from a site with a high level of exposure to industrial chemicals. The current ready biodegradability testing approach includes use of inoculum from e.g. municipal STP pre-exposed to chemicals which are normally emitted to STPs.

For both ready biodegradability and simulation degradation tests biodegradation depends upon one or more competent micro-organism(s) being introduced into the test flask and these microorganisms being able to establish themselves under the conditions of the test. For many substances the use of replicate flasks may give rise to high levels of variability and several studies for an identical substance can give conflicting results. These variable results are largely due to differences in the composition of the microbial inoculum introduced into the test flask on day zero. Therefore strategies are required to ensure that a representative microbial diversity is introduced into the test system. In simulation tests it is essential to have a representative diversity present in the inoculum source to ensure environmental realism. This is especially true for biodegradation tests that use small test vessels.

Therefore test strategies are required that can maximise the diversity and adaptation of microbes in the test system without compromising environmental realism or the philosophy that the innate ability of the environmental degradation potential is being assessed. It must be reiterated that the purpose of using enhanced biodegradation screening tests is to confirm a potential for degradation, which can be considered in persistency assessment (e.g. PBT and vPvB assessment). These tests, however, do not provide information on ready biodegradability. Test approaches in enhanced biodegradation screening tests could include:

- Test duration - the test duration for poorly soluble substances and substances with extended lag phases is important. Where biodegradation is still occurring in a ready biodegradability test weekly determinations could be continued and made up to day 60.

In accordance with OECD guidance the test should be stopped when degradation has ceased i.e. three time points give the same result.

- Testing in larger vessels – the drive to generate tests that allow rapid and small-scale chemical assessments does not work for biodegradability assessments. At very small test volumes the total number of and the number of different types of microorganisms introduced into the test flask decreases. Conducting biodegradation tests using larger volumes of environmental waters increases the total number of microorganisms introduced into the test, and the number of different types, without changing the density of microorganisms introduced (Ingerslev et al., 2001). This will increase the probability of introducing a competent microorganism into the test vessel.

- Increasing the biomass concentration - Testing at a number biomass concentrations, using tangential flow filtration to concentrate the microbes in environmental waters, as advocated by Thouand et al (1996) and ECETOC (2004) may enable a most probable number (MPN) approach to biodegradation testing to be developed i.e. it may be possible to identify that a competent microorganism was present in x litres of river water etc. This approach recognises that when conducting biodegradability assessments with less than one litre of an environmental water sample it will not reflect the total number and types of microorganism that a chemical will routinely encounter once released to an environmental water course.

- Low-level pre-adaptation test systems – adaptation or enrichment of environmental microorganisms that can degrade particular chemical substances is a natural phenomenon. Low-level pre-adaptation test could include conducting a second ready biodegradability test using the inoculum derived from the initial study. This should reduce the lag period preceding the onset of biodegradation.

- Semi-continuous assessments - conducting a ready biodegradability study using an inoculum derived from test systems fed with the test substance at environmentally realistic concentrations on a semi-continuous basis. Semi-continuous test systems help maintain the diversity, viability and nutrient status of the biodegradability tests whilst allowing the potential for adaptation to be determined over time (such as the semi-static version of OECD TG 309, Toräng L. et al 2005).

INHERENT BIODEGRADABILITY

The tests that can be used to determine the inherent biodegradability of organic chemicals include three methods described in the OECD test guidelines 302 A-C: Modified SCAS Test (OECD 302 A), Zahn-Wellens/EMPA Test (OECD 302 B) and Modified MITI Test (II) (OECD 302 C).

Biodegradation above 20% of theoretical (measured as BOD, DOC removal or COD) may be regarded as evidence of inherent, primary biodegradability, whereas biodegradation above 70% of theoretical (measured as BOD, DOC removal or COD) may be regarded as evidence of inherent, ultimate biodegradability. Care must be taken when using DOC removal to ensure that elimination did not occur through adsorption or volatilization. The shape of the degradation curve should give an indication whether or not a biological degradation process occurred. When results of ready biodegradability tests indicate that the pass level criterion is almost fulfilled (i.e. ThOD or DOC slightly below 60% or 70% respectively) such results can be used to prove inherent biodegradability. This is also the case when the pass level criterion is fulfilled but the 10-day window criterion is not.

Such application of ready biodegradability tests, which may include their incubation beyond 28 days, may in some cases eliminate the need for additional testing of biodegradability in inherent or simulation tests (OECD, 2006).

Inherent biodegradability data may be used for extrapolation to a rate constant in models for estimation of the elimination of chemicals in STP. However, this extrapolation is only allowed, if the pass level of 70% degradation in the Zahn-Wellens/EMPA Test is reached within seven days, including the lag-phase and the log-phase, the log-phase should be no longer than three days, and the percentage removal in the test before biodegradation occurs should be below 15%. The pass level of 70% in the Modified MITI Test (II) must be reached within 14 days, including the lag-phase and the log-phase, and the log-phase should be no longer than three days.

SIMULATION TESTS

Simulation tests aim at assessing the rate and extent of biodegradation in a laboratory system designed to represent either the aerobic treatment stage of STP or environmental compartments, such as fresh or marine surface water (OECD, 2006).

SEWAGE TREATMENT

The fate of chemicals in STPs can be studied in the laboratory by using the Simulation Test.

Aerobic Sewage Treatment: Activated Sludge Units (OECD 303 A) and Biofilms (OECD 303 B). The removal of the test substance is determined by monitoring the concentration of DOC and/or Chemical Oxygen Demand (COD) in the influent and effluent. The test recommends addition of the test substance at a concentration of DOC between 10 mg/L and 20 mg/L. However, many chemicals are normally present at very low concentrations, even in waste water, and procedures for testing the biodegradation at suitably low concentrations (<100 µg/L) are presented in Annex 7 to the TG 303 A (OECD, 2006).

Biodegradation in a DOC based Continuous Activated Sludge (CAS) test can only be determined when the material is non-sorptive since biodegradation is the only relevant removal mechanism assuming the test material is non-volatile. If a radiolabelled CAS is performed and a mass balance is done on the effluent and solids, it is possible to determine biodegradation for any type of non-volatile compound. The value of a CAS for estimating biodegradation increases when off-gases are trapped for CO_2 and other organic volatiles.

No specific pass levels have been defined for the elimination of chemicals in aerobic sewage treatment simulation tests. The test results may be used to estimate the removal in STPs and the resulting effluent concentrations for prediction of the concentration in the treatment plant and the receiving aquatic environment.

The assessment of biodegradability and/or removal in sewage treatment plants should preferably be based on results from tests simulating the conditions in treatment plants. Such a test may be the OECD 303 A test. Data from non-standardised tests and/or tests not performed according to the principles of GLP may be used if expert judgement has confirmed them to be equivalent to results from the standardised degradation tests on which the calculation models, e.g. SimpleTreat, are based. The same applies to STP monitoring data, i.e. in-situ influent/effluent measurements.

There is separate endpoint specific guidance for toxic effects of substances on STPs (see Section R.7.8.18).

SOIL, SEDIMENT AND WATER

The following tests can be used to simulate the biodegradation of organic chemicals under environmentally realistic conditions in soil, sediment or surface water: Aerobic and Anaerobic Transformation in Soil (OECD 307); Aerobic and Anaerobic Transformation in Aquatic Sediment Systems (OECD 308); and Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test (OECD 309).

Aerated soils are aerobic, whereas water-saturated or water-logged soils are frequently dominated by anaerobic conditions. The surface layer of aquatic sediments can be either aerobic or anaerobic, whereas the deeper sediment is usually anaerobic. These conditions in soil or sediment may be simulated by using aerobic or anaerobic tests described in the test guidelines (OECD 307 and OECD 308).

Generally, a low concentration of the test substance is used in tests designed to determine biodegradation. A low concentration in these types of tests means a concentration (e.g. from 1 µg/L to 100 µg/L in TG 309), which is low enough to ensure that the biodegradation kinetics (first order or pseudo-first order) obtained in the test reflect those expected in the environment.

Where possible simulation studies should be conducted at environmentally relevant temperatures e.g. the temperature that the environmental media was collected. However, it is recognized that these higher tiered studies take up a large laboratory footprint and it may not be practically possible to conduct the test at the environmental temperature. In such cases attempts should be made to reduce the temperature as far as practically possible.

When using radiolabelled chemicals, the label should be located in the most recalcitrant part of the molecule when total mineralisation is assessed. Measuring disappearance of the parent compound by chemical analysis does not imply mineralisation. Simulation tests are especially useful if it is known from other tests that the test substance can be mineralised and that the degradation, which is measured, covers the rate determining process.

The results of simulation tests may include:

- First order or pseudo-first order rate constant;
- Degradation half-life or DT50
- Length of the lag phase
- Half-saturation constant;
- Maximum specific growth rate;
- Fraction of mineralised label, and, if specific analyses are used, the final level of primary degradation;
- The fraction of bound residue;
- Mass balance during and at the end of the study;
- Identification and concentration of major transformation products, where appropriate;
- A proposed pathway of transformation, where appropriate;
- Rate of elimination (e.g. for risk assessment purposes)

NON-STANDARD PUBLISHED BIODEGRADATION STUDIES

When judging poorly reported or non-standard data then the following minimum information needs to be available in order to make any use of the data:

- The source and density of the inoculum, this should not be taken from an industrial site and the density should be equivalent to that of a ready biodegradation test
- Any pre-treatment of inoculum including pre-exposure to the test chemical
- The test chemical, its purity and the concentration that is used in the test
- The motivation for the study
- The analyte being measured (parent compound, DOC, BOD or CO_2 evolution)
- Details regarding the biochemical pathway for degradation if available
- Either a removal percentage or a degradation rate

REPORTING BIODEGRADATION STUDIES

FOCUS (2006) makes a distinction between biodegradation endpoints used as a trigger for higher tier studies (trigger endpoint) and biodegradation endpoints used in risk assessment (modelling endpoint). The main difference in approach is that for trigger higher tier studies the best fitting kinetic model is applied, for instance a biphasic kinetic model or a lag-phase model, while for modelling endpoint and use of data on risk assessment the choice of the kinetic model should be in agreement with the kinetics used in the environmental fate model used in the risk assessment. Until now, the environmental fate models are based on first-order kinetics. So in practice modelling endpoints should be derived with first-order kinetics.

The principle of reporting biodegradation studies is that enough information should be provided to allow independent reproduction of the results and verification with alternative software packages. The following aspects of kinetic analysis should be reported:

- Software package(s) and version. To facilitate independent duplication of results it is preferred that the kinetic analyses are performed with publicly available software packages, commonly used for such analyses.
- A listing of all original values to be used in the analysis. When datapoints are discarded as part of the kinetic analyses, the rationale for discarding datapoints should be included in the report
- Analyses. Exact description of kinetic models used in the regressions. Software options like range limits, initial values, restrictions in optimization should be described.
- Visual and statistical assessment of the results. Figures of predicted and observed values (i.e. concentrations) as a function of time and residual plots. Other statistical endpoints that support the decision-making process should be reported.
- Uncertainty (standard deviation or confidence interval) of degradation rate constant and formation fractions of metabolites.
- If the DT_{50} is extrapolated beyond the experimental period this should be clearly stated in the report.

TEMPERATURE CORRECTION

Incubation temperature is one of many factors that need to be considered when conducting higher tiered biodegradation studies. Others include the substance concentration, test volume and geometry, airflow rate and cometabolism.

Temperature is an issue within Europe due to the wide range of environmental temperatures that a chemical may experience in the field. Where the competent degrader is a mesophile, rates of degradation in a test conducted in the laboratory at 20°C may be higher than those measured in the field. However, where the competent degrader is a pyschrophile the rates of degradation in the environment may be higher than those observed at 20°C in the laboratory. Consequently, there can be no systematic or universal correction factor for temperature that should be applied to higher tiered biodegradation studies. However, for persistence assessments where the B and T criterion have been met, and simulation data exist for degradation at 20°C, consideration should be given whether temperature correction should be applied. This will be particularly important where the measured half-life is close to the persistence criteria,. This correction, if applied, should be based on the Arhenius equation and extrapolate from 20°C to the temperature of the environmental media at the point of sampling. No temperature correction is required for sewage treatment plants simulations (OECD 303).

DETERMINATION OF DEGRADATION PRODUCTS

By measuring parent material, bio-transformation products or metabolites and bound material as a function of time, it is possible to assess the fate of the test substance in the specified environmental compartment. When a substance is not fully degraded or mineralised, degradation products may be determined by chemical analysis. The methods will have to be substance specific and consequently no guidance on choice of method can be given. For some substances, radio-labelled chemicals and specific chemical analyses may allow reasonable fate assessment by measuring subsequent metabolite formation and decay.

Where analytically possible, identification, stability, behaviour, molar quantity of metabolites relative to the parent compound should be evaluated. Additionally, the predicted degradation rate of the parent material, log K_{ow} of the metabolites relative to the parent compound, and the potential toxicity of metabolites may be investigated. The first step in a PBT assessment for metabolites should be their degradation half-life. If the metabolites are long-lived or persistent, they should then be assessed for bioaccumulation and toxicity. The following statement from the TGD is relevant in this regard: *"In principle the persistence in the marine environment should be assessed in simulation test systems that determine the half-life under relevant environmental conditions. The determination of the half-life should include assessment of metabolites with PBT-characteristics. The half-life should be used as the first and main criterion in order to determine whether substances should be regarded as persistent"*.

Where the potential toxicity of significant metabolites is concerned, microbial degradation processes usually lead to more polar compounds than the parent, but in some cases to less polar compounds. This can be seen in the HPLC-RAD chromatographs routinely produced during simulation tests. Reduced lipophilicity may be one indication that the metabolites are less harmful than the parent material. Preliminary information on toxicity can be obtained with the help of measured K_{ow} values and QSAR predictions for postulated and identified metabolites.

Knowledge of bound residues and incorporation into biomass also needs to be considered and should be seen as a potential removal pathway. The OECD 308 (2002) Guideline advises as follows: *"Bound residues represent compounds in soil, plant or animal that persists in the matrix in the form of the parent substance or its metabolite(s) after extractions. The extraction method must not substantially change the compounds themselves or the structure of the matrix… In general, the formation of bound residues reduces the bioaccessibility and the bioavailability significantly (1) [modified from IUPAC 1984 (2)]."* Extraction of the sample, often with a suitable organic solvent is generally repeated 3 or 4 times until no further yield is achieved. Typically a range of solvents are used of increasing polarity (e.g. methanol, acetone, acetonitrile and hexane etc.) under ambient conditions. If the entire residual radioactivity cannot be recovered then appropriate solvent may be mixed with weak acids or bases or coupled to ultrasonic extraction. This aims to provide different conditions that may lead to the chemical or metabolite being released back into solution. Finally, the use of strong acids, bases or refluxing could undoubtedly extract the sample more thoroughly but could alter both the compounds of interest and the matrices. Such severe extraction techniques are rarely if employed in e.g. routine soil or sediment/water testing. The extraction methods and efficiencies as well as analytical methods and detection limits should always be reported.

These considerations should aid in determining the following environmental assessments for classification, PBT/vPvB and potential exposure.

ENVIRONMENTAL HAZARD CLASSIFICATION:

When a substance is not fully mineralised, but rapidly degraded to less degradable degradation products, the environmental hazard of these should be considered before a final judgement of whether a substance is readily or rapidly degradable.

PBT AND VPVB ASSESSMENT:

When a substance is not fully mineralised, but degraded to more persistent degradation products, the PBT/vPvB properties of these should be evaluated before a final judgement of whether a substance fulfils the persistence criteria. More guidance is given chapter R.11.

EXPOSURE ASSESSMENT:

When a substance is not fully mineralised, but degraded to more persistent degradation products, the environmental exposure concentrations should be determined for these products. Consequently, the safety assessment should also consider the degradation products.

R.7.9.4.2 Field data on degradation/biodegradation

In higher tier studies biodegradation is not always visible as a separate process. Other processes like transport, adsorption, volatilization, uptake in plants or organisms, hydrolysis also contribute to the fate of the substance simultaneously. In order to derive biodegradation rate inverse modelling can be applied to quantitatively separate biodegradation from other processes.

Measured concentrations in the mesocosm, lysimeter, or field experiments are compared with simulated concentrations in an environmental model, and de biodegradation rate constant is computed by a parameter estimation procedure (manually by trial and error or automated by a software package for example PEST) until the modelled concentration fit to the measured data. Procedures are described in FOCUS (2006), an example is published by Dubus et al (2004).

R.7.9.4.3 Exposure considerations for degradation/biodegradation

The major factors that are related to exposure within the context of degradation relate to:

- the use of the chemical;
- the chemicals emission pattern (continuous or intermittent release);
- the compartment to which the chemical is released (this can be more than one compartment);
- the amount per time unit or rate of chemical released;
- the rate of degradation; and
- the physico-chemical properties of the chemical.

The physico-chemical properties of the chemical and the compartment to which the chemical is released will have a large influence on where the chemical will be transported to and distributed to within the environment (see Chapter R.16). The emission pattern (continuous or intermittent) will influence the ability of competent microorganisms to establish themselves and for biodegradation to occur. The amount of chemical released will also influence the kinetics of biodegradation.

The identification of the environmental compartment(s) is of primary importance for a PBT, vPvB or /and risk/exposure assessments. A simulation test will normally not be required for all environmental compartments. The compartments of highest exposure and risk should be tested first if testing is required for refinement of risk assessment:

- If testing is triggered for PBT assessment different types of considerations should be made: The selection of most appropriate suitable simulation test(s), should consider the intrinsic properties of the chemical (e.g. water solubility, vapour pressure, log K_{ow}, K_p), its use and emission pattern (including the primary receiving compartment(s).
- The K_p (sediment) may be used as an indicator of whether testing in a water-sediment system may be warranted. Although for substances with K_p >2000 an aquatic sediment simulation test might be relevant in addition to a pelagic simulation test, a good test of this type does not exist yet.
- Results from multi-media modelling (e.g. Mackay level 3 models) could also be explored in order to evaluate the environmental compartment(s) of primary concern. It is noted that the results of such models should be used with care, as the results (distribution of mass fraction in the different environmental compartments) are strongly dependent of the relative size of the environmental compartments, and the emission parameters employed in the modelling. Contrary to the result of Mackay level 1 modelling, Mackay level 3 modelling is also dependent of the release pattern (fraction of emission between air, water, soil) and thus also on the use of the substance.

- Nevertheless a case-by-case evaluation of the results of such models may be useful and may even indicate whether or not chemicals may expose pristine environmental compartments (e.g. open sea) to a significant extent (i.e. indicate a significant potential for long range environmental transport via the atmosphere).

One of the key aspects for consideration is the volatility of the compound. By affecting the partitioning to other media and compartments from the source compartment(s) and the kinetics of that transfer, volatility is a key physico-chemical properties that greatly influence the overall persistence of a chemical in the environment, as defined by the mean time that a molecule resides in the system taking into account all intra-media and transfer processes (OECD, 2002).

Webster *et al.* (1998) have pointed out the inconsistencies which result when using only specific degradation half-lives for determining the environmental persistence and ignoring the mode/compartment of entry and the effects of partitioning to other media.

Usually, intra-media and transfer processes are ignored in the assessment of persistence, whereas it should be considered that:

- compartment specific degradation half-lives might be overly conservative when a chemical does not partition significantly into that compartment;
- compartment specific degradation half-lives are not independent of each other;
- the amount lost by degradation in a specific compartment is determined both by the compartment specific degradation rate constant and the amount of substance present in that compartment (Wania & Mackay, 2000).

There are several parameters that impact on the volatility of a chemical and its inter-compartmental partitioning, including aqueous solubility and vapour pressure (VP). There are also a number of parameters that may be useful for assessing volatility and inter-compartmental transport, including octanol-air partitioning constant and the Henry's law constant. When assessing the persistence of a chemical with high volatility, it is therefore recommended not to rely only on specific-medium degradation half-lives but to also consider on a case-by-case basis if these half-lives will cover the overall persistence of a chemical in the environment. This might be achieved by the use of multimedia fate models.

R.7.9.4.4 Remaining uncertainty for degradation/biodegradation

Chemicals that fulfil the criteria for ready biodegradability are likely to undergo rapid degradation in the environment under most conditions (OECD, 2006). However, it must be recognised that these tests are very stringent and most chemicals will not fulfil the pass criteria for ready biodegradability. For chemicals that exhibit between 40 and 60% mineralisation in ready biodegradability test, extensive primary biodegradation would have occurred even though the use of non-specific endpoints such as DOC and BOD do not directly measure this. Therefore there will remain a large degree of uncertainty about the biodegradability of many chemicals and testing at higher levels or tiers will be required.

At present the data set for biodegradation of general chemicals in higher tiered studies such as the OECD 308 test is relatively small. These tests were originally designed for plant protection products and have not been routinely applied to general chemicals. Even though such tests constitute the highest tier testing of biodegradation there are uncertainties connected with their use.

One example is that degradation half-lives may vary between different sites from where the environmental compartments inoculum and test media are sampled. Another example is, that it is uncertain what the value of conducting the strict anaerobic test part of the OECD 308 test is, and how these data can be used in CSA.

Identifying the compartments of concern can also be problematic in the absence of accurate use and emission data or data concerning the potential for environmental long-range transport. Confidence can be improved if such data are comprehensive and accurate.

R.7.9.5 Conclusions for degradation/biodegradation

R.7.9.5.1 Concluding on suitability for Classification and Labelling

Environmental hazard classification requires information on aquatic toxicity, degradation and bioaccumulation. In the current EU classification system (Council Directive 67/548/EEC) and in the "Globally Harmonised System of classification and labelling of chemicals (GHS)" (United Nations GHS (Rev.1) 2005), the determination of the appropriate risk phrases or hazard statements are often based on an integration of this information. However, this integrated approach is not considered here, as the ITS is concerning degradation aspects alone.

Under the degradation part of the EU and GHS classification criteria two aspects need to be evaluated:

Current EU system:

- Whether "the substance is readily degradable or not"
- Whether "additional scientific evidence concerning degradation" is available, i.e. whether there is "a proven potential to degrade rapidly in the environment"

GHS:

- Whether there is a "lack of rapid degradability"
- Whether there is "other evidence of rapid degradation"

Some guidance on interpretation of information on degradation is available given in Annex VI of Directive 67/548/EEC and this has been further developed in part 4 and Annex 9 to the GHS criteria (United Nations GHS (Rev.1) 2005). This latter guidance, which has been internationally agreed by OECD, forms the principal basis for this guidance on the suitability of degradation data on classification. For the purposes of decisions on classification and testing strategies, the two terms 'not readily degradable' and 'lack of rapid degradation' may be considered as synonymous.

The decision criteria for evaluating the suitability of available information on use in a decision on environmental hazard classification should consequently be focused on these aspects. At each step of the ITS, the available information will need to be evaluated against the aspects described above. The definition of ready (or rapid) degradability covers both biotic and abiotic degradation. Under most environmental conditions hydrolysis will be the major abiotic removal process. Data on either or both biotic or abiotic degradation would be sufficient to make a decision on rapid degradation.

Degradation can be monitored by either measuring the complete breakdown of the chemical to carbon dioxide and water (ultimate degradation), or simply the measuring the disappearance of the parent substance, primary degradation. While ultimate degradation is preferred, primary degradation can be used to define the pass levels in each of the degradation tests provided certain conditions are met. Data on primary biodegradability may be used for demonstrating rapid degradability only when it can be satisfactorily demonstrated that the degradation products formed do not fulfil the criteria for classification as hazardous to the aquatic environment.

In general, where experimental data are not available, and there are no additional data from structurally similar substances, a substance must be considered as *not rapidly degraded*. The following types of non-test data may be considered, however, as contributing to a decision on *ready or rapid* degradation for classification purposes.

QSAR Data

In the absence of experimental or environmental data, the predictions from QSARs models described in Section R.7.9.3.1 may be considered. No formal decision has been taken on how to use (Q)SAR derived information on biodegradability for classification purposes in the EU. In relation to the development of the GHS, the usefulness of (Q)SARs for predicting ready biodegradability is considered (United Nations GHS (Rev.1) 2005). It is stated that (Q)SARs for predicting ready biodegradation are normally not yet sufficiently accurate to predict rapid degradation. However, it is a general rule that when no useful information on degradability is available - either experimentally derived or estimated - the substance should be regarded as not readily or not rapidly degradable and (Q)SAR prediction can be used as supporting evidence of this.

The reason for this discrimination on usability of different outcomes of (Q)SAR predictions is that currently conducted validations and comparisons between test data and (Q)SAR predictions often seem to suggest that the probability of a correct prediction of a slow biodegradation is high, while the probability of a correct prediction of a fast biodegradation is significantly lower (e.g. OECD 2004). This is however according to validation studies where (Q)SAR predictions have been compared with ready biodegradability test data and the sensitivity and specificity of not ready biodegradability predictions seem to be dependent on the particular (Q)SAR model in question (cf. OECD 2004:ENV/JM/TG(2004)26Rev1 and references therein). Generally however when a substance is estimated to be *slowly* biodegradable, sufficient information is normally considered available on biodegradability for hazard classification purposes, when no test data are available. When a substance is estimated to biodegrade *fast*, further information gathering is normally necessary (United Nations GHS (Rev.1) 2005).

Structurally related substances

When no experimental data are available, the potential for rapid degradation in the aquatic environment may also be assessed by examining available data on structurally related substances. There will always need to be an element of expert judgement in such an evaluation, but this approach may be particularly relevant where the QSAR prediction described above suggests rapid degradation. If such a prediction is supported by experimental evidence from structurally similar substances, then this can be considered as convincing evidence for rapid degradation for classification purposes. Equally, of course, such data on similar structures may provide evidence of a lack of rapid degradation. In general, expert judgement should be used in a conservative way.

Degradation data suitable for use in classification

READY BIODEGRADATION

Ready biodegradability is defined in the OECD Test Guidelines No. 301 (OECD 1992). All organic substances that degrade to a level higher than the pass level in a standard OECD ready biodegradability test or in a similar test should be considered readily biodegradable and consequently also rapidly degradable. Many literature test data, however, do not specify all of the conditions that should be evaluated to demonstrate whether or not the test fulfils the requirements of a ready biodegradability test. However, provided a test is conducted within the constraints and quality criteria defined in Section R.7.9.4, it may be considered as a ready biodegradability test for the purposes of classification. In the context of classification, the individual test *pass* levels are considered an important part of the criteria.

When conflicting results in ready biodegradability tests are obtained the positive results could be considered valid irrespective of negative results, when the scientific quality of the former is good and the positive test results are well documented, including assurance of the use of non-pre-exposed (non-adapted) inoculum. (United Nations GHS (Rev.1) 2005). Before a decision is made on the appropriate result to use, however, the data should be carefully examined to determine whether there is a simple or clear explanation for the differences in result. Not all of the various screening tests are suitable for the testing of all types of substances, and results obtained by the use of a test procedure which is not suitable for the specific substance should be evaluated carefully before a decision on the use is taken (see Section R.7.9.2). Equally, where possible, the inoculum source should be checked to ensure a positive result is not the result of artificially pre-adapted inoculum.

Nevertheless, where a positive result has been obtained using a standard and valid methodology, this will be used to indicate rapid degradation for classification, irrespective of other negative results. This will hold true unless there are strong *Weight of Evidence* or structural reasons to question this result.

MODIFIED READY BIODEGRADATION TESTS

There are circumstances when it may be necessary to modify the standard guidelines in order to test a particular substance. This is particularly true for poorly water soluble substances, and also those that show toxicity to micro-organisms at the concentrations of the test. These modifications are described in Section R.7.9.4 . Such tests are regarded as ready biodegradation tests and can be used directly in classification.

BOD5/COD

Information on the 5-day biochemical oxygen demand (BOD5) can be used for classification purposes only when no other measured degradability data are available. Thus, priority is given to data from ready biodegradability tests and from simulation studies regarding degradability in the aquatic environment. The BOD5 test is a traditional biodegradation test that is now replaced by the ready biodegradability tests. Therefore, this test should not be performed today for assessment of the ready biodegradability of substances. Older test data may, however, be used when no other degradability data are available. For substances where the chemical structure is known, the theoretical oxygen demand (ThOD) can be calculated and this value should be used instead of the chemical oxygen demand (COD).

TEST DURATION LESS THAN 28 DAYS

Sometimes degradation is reported for tests terminated before the 28 days period specified in the standards (e.g. the MITI (1992) test data). These data are of course directly applicable when degradation greater than or equal to the pass level is obtained. When a lower degradation level is reached, the results need to be interpreted with caution. One possibility is that the duration of the test was too short and that the chemical structure would probably have been degraded in a 28-day biodegradability test. If substantial degradation occurs within a short time period, the situation may be compared with the criterion BOD5/COD ≥0.5 or with the requirements on degradation within the 10-days time window (OECD 301A,C,D,E and F) or 14-days time window (OECD 301B). In these cases, a substance may be considered readily degradable (and hence rapidly degradable), if:

- the ultimate biodegradability exceeds 50% within 5 days and
- the ultimate degradation rate constant in the test system in this period is greater than 0.1 day^{-1} corresponding to a half-life of 7 days in the test system (see Section R.7.9.11).

OTHER CONVINCING SCIENTIFIC EVIDENCE

Rapid degradation in the aquatic environment may be demonstrated by other data than referred to using the standard assessment methods covered above. This may be data on biotic and/or abiotic degradation. Data on primary degradation can only be used where it is demonstrated that the degradation products shall not be classified as hazardous to the aquatic environment, i.e. that they do not fulfil the classification criteria.

Scientific evidence must be provided that the substance is degraded in the aquatic environment to a level of >70% within a 28-day period. If first-order kinetics is assumed, which is reasonable at the low substance concentrations prevailing in most aquatic environments, the degradation rate will be relatively constant for the 28-day period. Thus, the degradation requirement will be fulfilled with an average degradation rate constant, k >0.043 day^{-1} which corresponds to a degradation half-life of 16 days. In determining whether this half-life criterion is met, care should be taken to ensure that an appropriate account has been taken of the temperature of the study.

The evaluation of data on fulfilment of this criterion should be conducted on a case-by-case basis by expert judgement. However, guidance on the interpretation of various types of data that may be used for demonstrating a rapid degradation in the aquatic environment is given below. In general, only data from aquatic simulation tests are considered directly applicable. However simulation test data from other environmental compartments could be considered as well, but such data require in general more scientific judgement before use.

HYDROLYSIS

Data on hydrolysis (cf. OECD 111) might be considered for classification purposes only when the longest half-life $t_{½}$ determined within the pH range 4-9 is shorter than 16 days. However, hydrolysis is not an ultimate degradation and various intermediate degradation products may be formed, some of which may be only slowly degradable. Only when it can be satisfactorily demonstrated that the hydrolysis products formed do not fulfil the criteria for classification as hazardous for the aquatic environment, data from hydrolysis studies could be considered.

When a substance is quickly hydrolysed (e.g. with $t_{½}$ < a few days), this process is a part of the degradation determined in biodegradation tests. Often, hydrolysis is the initial transformation process in biodegradation.

AQUATIC SIMULATION TESTS

Aquatic simulation tests are tests conducted in laboratory, but simulating environmental conditions and employing natural samples as inoculum. It should be noted that the OECD 303 test is not simulating conditions in the aquatic environment but in sewage treatment plants and consequently, results from this test are not valid for classification. Results of aquatic simulation tests (mineralisation rate, degradation half-life) may be used directly for classification purposes when realistic environmental conditions in surface waters are simulated. Such tests are described in Section R.7.9.3 .

SOIL AND SEDIMENT DEGRADATION DATA

It has been argued that for many non-sorptive (non-lipophilic) substances more or less the same degradation rates are found in soil and in surface water (see Section R.7.9.10). For adsorptive substances, a lower degradation rate is generally expected in soil than in the water-phase due to partly immobilization caused by sorption. Thus, when an adsorptive substance has been shown to be degraded rapidly in a soil simulation study, it is most likely also rapidly degradable in the aquatic environment. It is therefore proposed that an experimentally determined degradation in soil is sufficient documentation for a rapid degradation in surface waters. Such tests are described in Section R.7.9.3.

FIELD INVESTIGATIONS

Parallels to laboratory simulation tests are field investigations or mesocosm experiments. In such studies, fate and/or effects of chemicals in environments or environmental enclosures may be investigated. Fate data from such experiments might be used for assessing the potential for a rapid degradation. This may, however, often be difficult, as it requires that an ultimate degradation can be demonstrated. This may be documented by preparing mass balances showing that no non-degradable intermediates are formed, and which take the fractions into account that are removed from the aqueous system due to other processes as e.g. sorption to sediment or volatilisation from the water environment. In general, mesocosms and field studies are not useful for classification and labelling purposes.

MONITORING DATA

Representative monitoring data may demonstrate the removal of contaminants from the aquatic environment. Such data are, however, very difficult to use for classification purposes. The following aspects should be considered before use:

- is the removal a result of degradation, or is it a result of other processes as e.g. dilution or distribution between compartments (sorption, volatilisation)?
- is formation of non-degradable intermediates excluded?

Only when it can be demonstrated that removal as a result of ultimate degradation fulfils the criteria for rapid degradability, such data might be used directly for classification purposes. In general, monitoring data can only be used as supporting evidence for demonstration of either persistence in the aquatic environment or a rapid degradation.

Degradation data not suitable for use in classification

INHERENT BIODEGRADABILITY TESTS

Substances that are degraded more than 70% in tests for inherent biodegradability have the potential for ultimate biodegradation (OECD Test Guidelines). However, because of the optimum conditions in these tests, the rapid biodegradability of inherently biodegradable substances in the environment cannot be assumed. The optimum conditions in inherent biodegradability tests stimulate adaptation of the microorganisms thus increasing the biodegradation potential, compared to natural environments. Therefore, positive results in these tests should not be interpreted as evidence for rapid degradation in the environment.

STP SIMULATION TESTS

Results from tests simulating the conditions in a sewage treatment plant (STP) (e.g. the OECD 303) cannot be used for assessing the degradation in the aquatic environment.

PHOTOCHEMICAL DEGRADATION

Information on photochemical degradation (cf. OECD GD(97)21) is difficult to use for classification purposes. The actual degree of photochemical degradation in the aquatic environment depends on local conditions (water depth, suspended solids, turbidity, etc.) and the hazard of the degradation products is usually not known. Probably only seldom will enough information be available for a thorough evaluation based on photochemical degradation.

VOLATILISATION

Chemicals may be removed from some aquatic environments by volatilisation. In general these data do not represent degradation and are not used in classification. The reason is that the degree of volatilisation from the aquatic environment is highly dependent on the environmental conditions of the specific water body in question, such as the depth and the gas exchange coefficients (depending on wind speed and water flow). In general, therefore, the Henry's Law constant cannot be used for assessment of the degradation (here removal of a chemical from the water phase) in relation to aquatic hazard classification of substances. However, substances that are gases at ambient temperature may be exempted from this general recommendation.

R.7.9.5.2 Concluding on suitability for PBT/vPvB assessment

Annex XIII of the REACH Regulation lays down specific criteria by which the terms Persistent and very Persistent are defined. These are:

Persistent: a half-life in the freshwater environment >40 days, or freshwater sediment >120 days, or marine water >60 days or marine sediment >180 days, or soil >120 days

very Persistent: a half-life in water (freshwater or marine) >60 days or sediment >180 days, or soil >180 days

While the criteria are specific in terms of the defined half-lives, it is recognised that the terms freshwater, marine, sediment and soil cover a range of different environments with different degrading potential, and thus the application of the criteria is by no means straight forward. In general, all available degradation and physico-chemical data should be evaluated and the potential of these data to influence the final decision considered. As a minimum, information on the vapour pressure, water solubility, octanol/water partition coefficient and Henry's Law Constant must be available (see Section R.7.1.22), and the impact of these data on the test design and data interpretation should be considered, as well as appropriate degradation data.

The half-lives described are considered to represent degradation half-lives; it is insufficient to consider removal alone where this may simply represent the transfer of a substance from one environmental compartment to another. Degradation may be biotic or abiotic, e.g. hydrolysis, and result in complete mineralisation, or simply the removal of the parent substance (primary degradation). Where only primary degradation is observed, it may be necessary to identify the degradation products. This will be considered further in Section R.7.9.6.

Degradation Test Data

SIMULATION TESTS

In principle Environmental Simulation Studies in an appropriate environmental media at environmentally realistic conditions are the only tests that can provide a definitive degradation half-life that can be compared directly to the persistence criteria as defined in Annex XIII. Such tests allow both biotic and abiotic degradation processes to operate. A correctly conducted study using either the OECD Guidelines 307 (soil), 308 (water/sediment) or 309 (water), as described in Section R.7.9.4 , with the degradation half-life calculated for the appropriate compartment either by direct substance analysis or some other suitable method such as radiolabel analysis, would allow direct comparison to the criteria. Even with a correctly conducted study, however, results can be difficult to interpret, particular where partitioning between phases and/or aerobic/anaerobic conditions can arise. Tests should report the degradation rate in each media determined through mineralisation, e.g. volatile ^{14}C, and/or direct substance analysis. Where mineralisation is measured a full mass balance of the substance and any degradation products/metabolites should be determined, and in water-sediment or soil tests they should include determination of the level of bound residues present. Where primary degradation is observed, the identity of the principal metabolites (section on assessment of metabolites below) or possibly relevant metabolites should also be determined. Where only degradation of the parent substance is monitored, this may not remove all the concerns and further assessment of the degradation products may be required in order to complete the PBT/vPvB and Chemical Safety Assessments.

In general, a single simulation study may be sufficient provided the environmental media at environmentally realistic conditions selected for study are appropriate. Availability or generation of multiple simulation test data may allow more WoE based conclusions to be drawn in relation to environmental half-lives for one or more environmental compartments by expert judgement.

This may allow more robust decisions to be taken when considering the persistency in relation to the PBT criteria. Selection of the appropriate test, and environmental media are described in Section R.7.9.6.

There may also be available non-standard simulation data, i.e. data generated before the standard Guidelines were agreed. Such data may be useful in reaching a decision on persistence provided the conditions of the tests properly simulate an appropriate environment. Such data would normally be considered along with other evidence such as screening test data, QSAR estimations or chemical categorisation or other structural analogy to support a final conclusion.

INHERENT BIODEGRADABILITY

Lack of degradation (<20% degradation) in an inherent biodegradability test equivalent to the OECD 302 series would provide sufficient information to confirm persistence without the need for a further simulation test. The tests provide optimum conditions to stimulate adaptation of the microorganisms thus increasing the biodegradation potential, compared to natural environments. A lack of degradation, therefore provides convincing evidence that degradation in the environment would be slow. When interpreting such tests, it should be realised that the very low solubility of many PBT/vPvB substances may reduce the availability and hence the degradability of the substance in the test

FIELD DATA

A range of field investigation approaches such as mesocosms, lysimeters etc are described in Section R.7.9.4. These are not normally designed to measure just degradation processes and thus cannot be considered to yield a half-life that can be read directly against the criteria. Nevertheless, evidence of degradation (or lack off) may provide evidence as part of a *Weight of Evidence* approach to making a decision.

MONITORING

Monitoring data can also provide evidence to support a conclusion on persistence. Monitoring in itself cannot demonstrate persistence because the presence of a substance in the environment can arise for a range of reasons. Nevertheless, the presence of a substance in the environment in remote regions, or regions not directly exposed suggests sufficient persistence for transport to occur, which also need to be considered.

ASSESSMENT OF THE POTENTIAL PERSISTENCE OF METABOLITES

Where a substance is degraded by abiotic means or partly biodegraded it may be necessary to consider whether there are any breakdown products or metabolites that are formed that could be potential PBTs/vPvBs. Where the original substance forms a breakdown product or metabolite that could be a PBT/vPvB, there will need to be an assessment of how much the breakdown product or metabolite constitutes compared with the parent substance. In relation to degradation testing results, including those from simulation degradation tests which also include investigation of degradation pathways (OECD TG 307, 308 & 309) there are often practical constraints to the analytical identification of transformation products. Biotransformation/ degradation pathways may be complex and many different degradation products may be formed and some only in small amounts.

Practical constraints in relation to analytical methodologies for identification of degradation products may thus limit the possibility for identifying them chemically, when they occur in very small concentrations. In the simulation degradation test guidelines for soil, water-sediment and surface water, transformation products detected at ≥10% of the applied concentration of the parent compound at any sampling time (principal metabolites) should be identified unless reasonably justified otherwise. However transformation products for which concentrations are continuously increasing during the study should also be considered for identification, even if their concentrations do not exceed the limit given above, as this may indicate persistence. The need for quantification and identification of transformation products should be considered on a case-by-case basis with justifications.

Neither a readily biodegradable substance (based on ultimate degradation) nor its metabolites will normally need to be assessed because any metabolites can be assumed to be minimal and transient.

Likewise a rapidly hydrolysable substance, with $t_{1/2}$ <12 hrs will not need to be assessed. However, for such rapidly hydrolysable substances, which will degrade sufficiently rapidly either in a WWTP or the environment, the degradation products themselves need to be considered in addition to, or instead of, the parent substance. For these degradation products it is likely that a CSA/CSR will need to be prepared, which will include an assessment of the PBT/vPvB properties.

To assess whether the breakdown products or metabolites may be potential PBT or vPvB substances, the following approaches may be helpful;

- Based on the structure of the parent molecule, predictions of the structures of the breakdown products/metabolites may be made. These can be based on QSAR models/ expert systems e.g. CATABOL or Multicase and by employment of expert judgement, supported by appropriate documentation.
- At higher tonnages (>100 t/y) there is a requirement to identify breakdown products/metabolites. The registrant shall provide sufficient evidence that either the approach above is sufficient or conduct specific analytical identification.
- Results obtained from valid (Q)SAR models can be used instead of testing or as supporting test results data when the conditions laid down under Annex XI point 1.3 of the Regulation are met.
- Structural alerts or read-across may also be considered, where the structure of the breakdown products/metabolites is sufficiently described that this can be supported.

Screening criteria

The criteria that apply to the definition of persistence result in effect, to a pass or fail, i.e. the measured or estimated half-life is above or below a specific threshold. It is not always necessary to know the exact half-life value, but rather simply that it is above or below the threshold. Screening data can therefore be applied which, based on long experience in application across a wide range of substances, can be used to make judgements regarding the likelihood that a substance will degrade more or less rapidly than the threshold criteria. Screening data will either lead to a decision that no further testing is needed since the substance is expected to degrade sufficiently rapidly that neither the P nor vP thresholds will be exceeded, or lead to the conclusion that further testing might be required in order to apply the definitive criteria. In general, it would not be possible to apply the screening criteria to a definitive judgement that a substance is P or vP, except as part of a Weight of Evidence argument, or when the degradation of a substance does not exceed 20 % in a test on inherent biodegradation.

READY BIODEGRADATION

Any data available that has been used to show a substance is readily degradable for the purposes of applying the hazard classification criteria can be used to define ready degradability with respect to the screening criteria for Persistence. The principal data available will be that from a standard ready biodegradability test, and a pass/fail in this test can be applied to the screening criteria defined (in Chapter R.11). While normally a 10-day window criterion applies in this test, this is considered unnecessary in defining the pass level when considering persistence and the 'pass' criterion apply over the 28-day period. Depending on the test method, a pass criterion of 60 and 70 % degradation as defined in the respective guidelines should be applied. It should be noted that substances being considered as potentially PBT/vPvB are often poorly soluble in water and this may cause significant difficulties in the conduct of the test, and in particular low levels of biodegradation may be observed due to low substance availability. It is possible to modify the standard test to improve this availability using the techniques described in Sections R.7.9.4 and R.7.9.10. This type of testing is acceptable in defining ready biodegradability for the purposes of screening for persistence.

HYDROLYSIS

Data from the hydrolysis test may be used to determine the lack of persistence. Since the intention is that the half-life determined in the testing should reflect the persistence in the real environment, data on hydrolysis rate will generally be required over a range of environmentally relevant pHs from pH 4 to pH 9. Where data are not available over the full range of environmental pHs, justification must be provided for the selected pH, which should be that pH where the slowest degradation would be expected. Normally the longest half-life would be selected.

Any data generated from laboratory testing would also need to be corrected for temperature. (see Section R.7.9.4.10).

ENHANCED BIODEGRADATION SCREENING TESTS

To obtain data from well-documented studies in which the standard conditions of the ready test have been changed in a specified way to better reflect the timescales and degradation processes in the environment is especially relevant for P- and vP-assessment. Such enhancements of some of the standard conditions of the screening tests address time for adaptation and a more environmentally realistic microbial biomass diversity. Generation of data from enhanced screening tests allow P and vP-assessment to be considered in decision making at the screening phase, i.e. without generation of more expensive simulation degradation test data. The enhanced screening tests are restricted to using only natural environmental media as the source of inoculum e.g. marine and freshwater. Enhanced screening studies using inocula derived from sewage treatment works cannot be used in persistence assessments.

For the enhanced screening tests that extend the test duration, or have increased test vessel size or biomass concentration, or are running two RBTs back-to-back, the normal test criteria could be applied without the 10-day window exclusively for the purpose of assessing persistence (60% or 70% depending on analyte). Both respirometric and parent compound analysis should be reported. Where primary degradation is used to establish a level of degradation, metabolites should be considered further.

For the semi-continuous test using natural environmental waters (OECD 309) then the degree of removal or clearance in the semi-continuous vessels needs to be reported. The hydraulic retention time (HRT) should not exceed 28 days. When a sample of liquor is removed from the semi-continuous vessel to seed the ready biodegradation test RBT then the normal RBT pass criteria without the 10-day window apply, but it will need to be stated how many cycles the semi-continuous system has been run.

There is little experience currently available on the use of these approaches as detailed in Section R.7.9.4, but these data can be considered on a case-by-case basis, particularly where clear additional data are available from QSARs or other structural analogues that support the conclusions drawn

INHERENT BIODEGRADATION

Data from inherent biodegradability tests would not normally be used to determine persistence except where a clear lack of degradation (<20 % degradation in an inherent test) can indicate a lack of environmental degradation as described above. Nevertheless, such data can be examined to determine whether the degradation in the test was sufficiently rapid to meet the special criteria detailed in Sections R.7.9.4 and R.7.9.6. If these conditions are met, then the data can be used at the screening stage. In other conditions, further testing will normally be indicated. Where full mineralisation occurs, with non pre-adapted bacteria, in a MITI II study (OECD 302C) (pass level 70%) within the first 14 days, or in a Zahn–Wellens study (OECD 302B) in 7 days, this is can be used to conclude that the substance is not persistent.

R.7.9.5.3 Concluding on suitability for use in chemical safety assessment

Degradation data are used in the chemical safety assessment to:

- determine the level of removal of a substance from waste water in a Sewage Treatment Plant
- determine the initial soil concentration for the purposes of calculating a $PEC_{soil\ local}$
- determine the steady state $PEC_{regional}$ for each environmental compartment.

READY BIODEGRADATION

Data on ready biodegradation can be used, and is a requirement of Annex VII. The data should contain information of the pass or fail status against the appropriate test thresholds, including whether the 10-day window criteria has been met. For poorly soluble substances, adjustments to the test protocol as described in Section R.7.9.4 are acceptable. Equally, test thresholds may be applied on the basis of primary degradation if these data are available, but if primary degradation is considered as the principal degradation route, further information on the degradation products may be required. For readily biodegradable chemicals, regional environmental concentrations in environmental media i.e. surface water, sediment and soil can be calculated by the use of Mackay level 3 models. The default degradation rates for such readily biodegradable chemicals can be used as input values (see Guidance on CSA).

HYDROLYSIS

Data from the hydrolysis test may be used if hydrolysis is a dominant route of degradation. These data may also be used to indicate:

- where problems may arise in generation and interpretation of aquatic toxicity data
- where degradation can occur such that further consideration may need to be given to major degradation products
- where the degradation rate constant may need adjusting in the determination of the $PEC_{regional}$

Rapid hydrolysis, for example, may influence the fate of a substance entering an STP in the same way as primary biodegradation and may require further investigation of potential hydrolysis products. Where data are only available for the screening part of the hydrolysis study, little quantitative information is available and the calculation of an environmental rate constant is not possible. Nevertheless, where the estimated degradation half-life is <24 hours, this will provide clear evidence of environmental degradation, and consideration must be given to the identification and further evaluation of any degradation products.

Hydrolysis data are needed over the range of environmentally relevant pHs from 4 to 9 (See TG 111) and should be corrected for temperature before use in the CSA (see Section R.7.9.4).

INHERENT BIODEGRADATION

Where information on inherent biodegradation is available, particularly from the Zahn-Wellens, or the MITI (II) studies (OECD 302B & C), these data should be examined to determine whether the special criteria detailed in Section R.7.9.4 are met. Where these criteria are met, the information may be used in the CSA to help determine the fate of the substance in an STP and by use of default degradation rates for inherently degradable chemicals in calculating the regional environmental concentrations in surface water, sediment and soil by the use of Mackay level 3 models (see chapter R.16).

A pass level (>70%) degradation in an inherent test may be used in similar manner to a pass in a ready test, where a specific STP may be considered as adapted. This is described further in the CSA Guidance. In other circumstances to those described above, data from inherent biodegradation testing cannot be used in the CSA.

PHOTOCHEMICAL DEGRADATION

Information on direct photolysis is difficult to interpret in the CSA since its significance in the aquatic environment depends on local conditions (water depth, suspended solids, turbidity, etc.). Nevertheless, where a degradation rate constant can be derived for site specific environmentally realistic conditions, these may be used in the assessment on a case-by-case basis where justified by a knowledge of local conditions. Information on indirect photolytic degradation half-life may be used for estimation of generic regional concentrations in air by use of generic assumptions about light intensity (latitude and season, length of day) and concentration of hydroxyl radicals in the air.

Refining a Chemical Safety Assessment

Where it is necessary to develop further the screening assessment, the following information and testing can be considered if available, or generated as a result of testing according to Annexes VI to X.

SEWAGE TREATMENT PLANT SIMULATION TEST

At screening level, models such as SIMPLETREAT are used to predict the level of degradation in an STP based on simple biodegradation screening tests as described above. A STP simulation test should give a direct measure of substance removal under realistic operating conditions. The assessment of biodegradability and/or removal in sewage treatment plants should therefore be based on results from tests simulating the conditions in treatment plants such as the OECD 303A test or the newly proposed OECD TG on biodegradation in STPs. It should be noted that the former test does not give a direct measurement of degradation but rather removal of the test substance including both degradation and adsorption as characterised by a STP. Normally inflow and outflow DOC or specific analysis is used and the concentrations material may be used and a full mass balance obtained.

Data from non-standardised tests and/or tests not performed according to the principles of GLP may be used if expert judgement has confirmed them to be equivalent to results from the standardised degradation tests on which the calculation models, e.g. SimpleTreat, are based. The same applies to STP monitoring data, i.e. in-situ influent/effluent measurements.

ENVIRONMENTAL SIMULATION TESTS

The CSA will sometimes require the generation of a 'regional' or background steady state concentration that might arise from a particular emission or load to an environmental compartment. These are calculated using standard fugacity models that require inputs of the transport characteristics between environmental compartments and the degradation rates for each compartment. At screening level, these are estimated from simple screening data described above. Where refinement of these degradation rates is needed, data from environmental simulation testing can be used. The particular tests chosen should seek to simulate the compartment(s) of concern. These tests are requirements listed in Annexes IX to X. The decision on which specific test should be selected is considered in Sections R.7.9.4 and R.7.9.6.

In addition, the soil environment simulation test may also be used to further refine the local PEC soil where an initial concentration is calculated based on an assumption of a number of years of exposure, followed by an addition load from land spreading of sewage sludge. Both the initial concentration, and added concentration can be refined by a measure soil degradation rate constant from a simulation test.

FIELD DATA

A range of field investigation approaches such as mesocosms, lysimeters etc are described in Section R.7.9.4. These are not normally designed to measure just degradation processes and thus cannot be considered to yield a half-life that can be read directly against the criteria.

R.7.9.5.4 Information not adequate

The prerequisite for use of other information than those types specified by the information requirements of REACH is that such information alone or in combination with other information is:

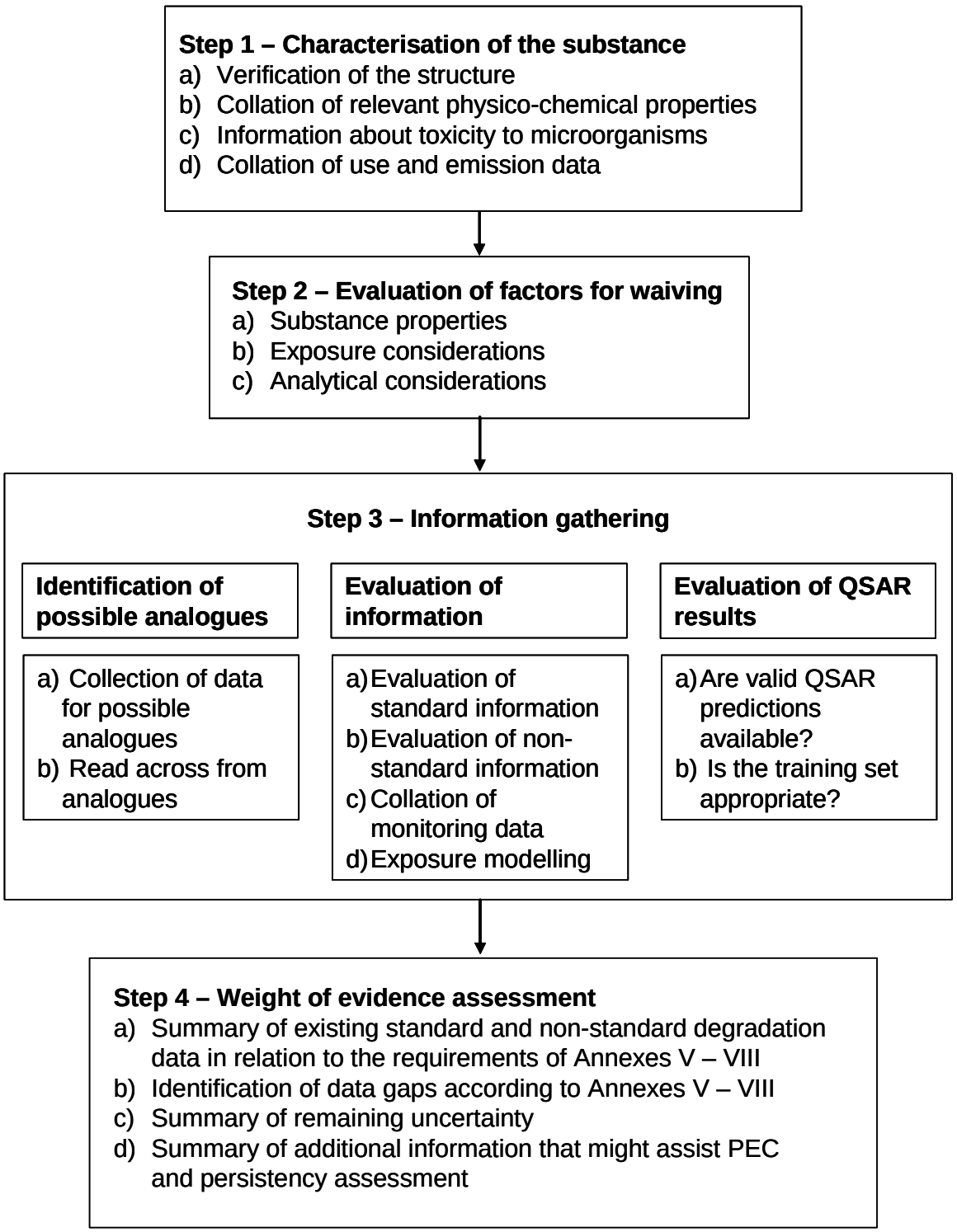
- equivalent to the results that would be obtained by standard testing, and
- adequate for the three regulatory endpoints: Classification and Labelling, PBT assessment and chemical safety assessment. The equivalence and adequacy will have to be substantiated by a Weight of Evidence approach using expert judgement and making best use of all existing information.

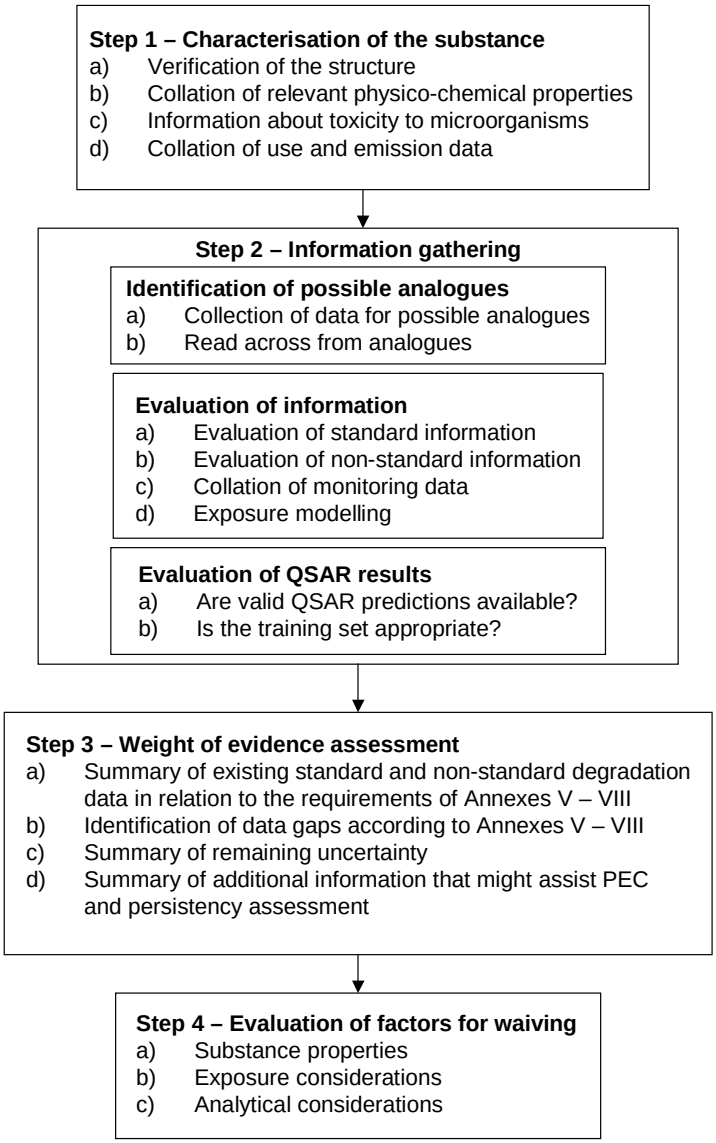
Weight of Evidence is closely linked to "integrated testing strategies (ITS)", in that the available evidence can help to determine the subsequent testing steps. Results from these subsequent tests affect the Weight of Evidence, which leads to a new decision on whether there is any need of further testing, and so on. The ITS's are designed to be flexible and applied on a case-by-case basis.

The following scheme outlines a systematic approach how to use all available degradation data on a Weight of Evidence decision (Figure R. 7.9-2). It provides a step-wise procedure for the assessment of different types of information, which might be helpful to come to an overall conclusion that may include the requirement for additional data. The scheme proposes a flexible sequence of steps, the order of which depends on the quality and quantity of data. Step 1, which is a collection of information on physico-chemical properties rather than an assessment of available information, is a prerequisite for the further assessment of other information. All steps are associated with three distinct activities: (i) the gathering of information, (ii) the evaluation of the quality of a distinct piece of information, and finally (iii) the overall assessment of all available information.

Weight of Evidence is a decision-making activity aiming at concluding on degradation of a substance based on integration of information from different sources and various aspects of uncertainty. It will often require expert judgement. To make this expert judgement transparent and comprehensible it is essential that all information used, all steps carried out in the evaluation process and all conclusions drawn are fully documented and justified.

Figure R. 7.9-1: A Weight of Evidence Approach for Assessing Degradation





STEP 1 – CHARACTERISATION OF THE SUBSTANCE

Initially it is important gather as much data about the chemical. This includes its CAS number, chemical formulae, chemical structure, purity and whether there are any known isomers.

Information on the following physico-chemical properties determined using the relevant OECD technical guidelines identified is also desirable: vapour pressure, water solubility, absorption - desorption using a batch equilibrium method, partition coefficient (n-octanol/water), dissociation constants in water, partition coefficient (n-octanol/water) - HPLC method, and Estimation of the Adsorption Coefficient (Koc) on Soil and on Sewage Sludge using High Performance Liquid Chromatography (HPLC).

Prior to assessing existing biodegradability data or requiring new biodegradation data it is important to assess information about the chemicals toxicity to microorganisms. Data from tests such as the activated sludge respiration inhibition test (OECD 209) are appropriate.

Finally, any information that can be gathered about the use and emission of the chemical will help determine the potential relevance of existing data, and it will also assist in prioritising additional degradation data requirements in Steps 2 and 3.

STEP 2 – EVALUATION OF FACTORS FOR WAIVING

There are a number of factors for waiving testing based on substance and exposure properties. These include:

Biodegradability studies are not required for inorganic chemicals as they cannot be tested for biodegradability.

Hydrolysis tests are not required for readily biodegradable chemicals, as the test will provide little additional information since rapid mineralisation of the chemical in the environment is assumed. In addition, if the chemical does hydrolyse this will occur in the ready biodegradation test and if it is accompanied with mineralisation >60% then it is unlikely that any terminal degradation products will exist. Hydrolysis tests are also difficult to conduct with chemicals that are highly insoluble in water and their relevance is likely to be low as such chemicals are unlikely to be associated water in the environment.

Simulation studies in surface water, soil and sediment are not required for readily biodegradable chemicals as it is assumed that they will undergo rapid degradation in the environment. Specific simulation studies are also not required if direct or indirect exposure is unlikely. When it is not necessary for PBT-assessment (e.g. the substance not either vB or not B or T) it may not be required for risk assessment purposes either if the exposure is so low that no refinement of the PECregional is indicated.

Identification of degradation products are not required for readily biodegradable substances as the 60% pass criteria assumes that the remaining 40% has been assimilated into new microbial biomass and any transient metabolites have been degraded.

STEP 3 – INFORMATION GATHERING

For chemicals where known analogues exist, relevant physico-chemical and degradation data need to be collated. In the case of biodegradation, where the biochemistry of biodegradation is known, analogues can include chemicals that are know to be degraded through identical mechanisms e.g. β-oxidation of certain hydrocarbons. It is also known that different pathways for biodegradation can exist for closely related analogues. Particular care will need to be taken with respect to differences in physico-chemical properties as simple structural changes to a chemical molecule can alter the behaviour of the chemical in the environment.

In the substance dossier mixed types of information is usually available. The information could be arranged according to information type each with its characteristics according to accuracy, interpretability and relevance for the particular regulatory type of decision:

- monitoring studies and field studies,
- simulation test data,
- inherent biodegradability data,
- ready and modified ready biodegradability studies

- enhanced screening studies indicating lack of persistency
- non-standard test data (including pure microbial culture data)
- poorly described test data
- marine biodegradability data
- abiotic degradation data
- sewage treatment plant removal data
- QSAR data

It should always be considered that a combination of information sources should give the most comprehensive assessment. When no reason can be found for lack of agreement between relevant and reliable testing and non-testing data then the non-testing data should normally not be decisive.

For substances where a range of degradation data is available, a Weight of Evidence approach should be employed. When more than one simulation test result is available, a suitable half-life in the higher end of the observed range should be selected taking into account the realism, relevance, quality and documentation of the studies in relation to environmental conditions (e.g. test substance concentration and temperature). When more than one screening test result is available, positive test results should be considered valid, irrespective of negative results, when the scientific quality is good and the test conditions are well documented, i.e. guideline criteria are fulfilled, including the use of non-adapted inoculum (cf. OECD, 2001c). It should also be noted that the results of screening tests may be negative due to toxic effects of the test substance, whereas simulation tests employing a low concentration of the test substance may give a more realistic estimate of the degradation in the environment.

When judging poorly reported or non-standard data (e.g. biochemical studies using mixed or pure culture) then the following information should be extracted in order to maximise the potential use of the data:

- The source and density of the inoculum should be defined; ideally this should not be taken from an industrial site and the density should be equivalent to that of a ready biodegradation test.
- Any pre-treatment of inoculum including pre-exposure to the test chemical.
- The test chemical, its purity and the concentration that is used in the test.
- The motivation for the study (e.g. isolation of competent microorganism or determination of the pathway for biodegradation)
- The analyte being measure (e.g. parent compound, DOC, BOD or CO_2 evolution)
- Either a removal percentage over a define time period or a degradation rate.

An example review of published literature has been provided for Toluene in the case studies provided with this guidance.

For chemicals that have been identified as readily biodegradable, any known metabolites of these compounds can also be considered as readily biodegradable. The public domain literature and the Minnesota Biodegradation Database might assist in identifying such metabolites (http://umbbd.msi.umn.edu/).

For chemicals where monitoring data exist it is important to gather these data together with appropriate metadata (e.g. sample points, dates, times, frequency, relevant hydrogeological and meteorological data etc.) associated with the monitoring programme.

Using the information gathered up to this point, it may be possible to model the exposure of the chemical at this stage to 1) identify environmental compartments of concern to determine the relevance of the available information and 2) to determine whether any available monitoring data supports the exposure model predictions.

The reliability of the prediction of a QSAR model should be taken into account based on an evaluation of the validation status for the models (sensitivity and specificity etc.) and based on an evaluation of whether the prediction falls within the applicability domain of the model. Similar considerations apply when judging the robustness of chemical categories relating to degradability. Often use of predictions from more QSAR models – if feasible supported by read-across or chemical categorisation - may enhance the overall possibility to make a robust overall prediction of ready biodegradability (see also Section R.7.9.4.1).

By using all available degradability test data, it may be possible to establish a comprehensive evaluation of the degradability of the substance. For example in particular ready biodegradation test data that demonstrated significant mineralisation (>40%) but fails to reach the pass criterion for ready biodegradability may exist. In certain cases where such data are available together with other evidence of biodegradation such as through the use of a valid QSAR and/or other test data that indicating rapid degradation without the presence of any significant metabolites, then this could together be used as evidence for non-persistence.

STEP 4 – WEIGHT OF EVIDENCE ASSESSMENT

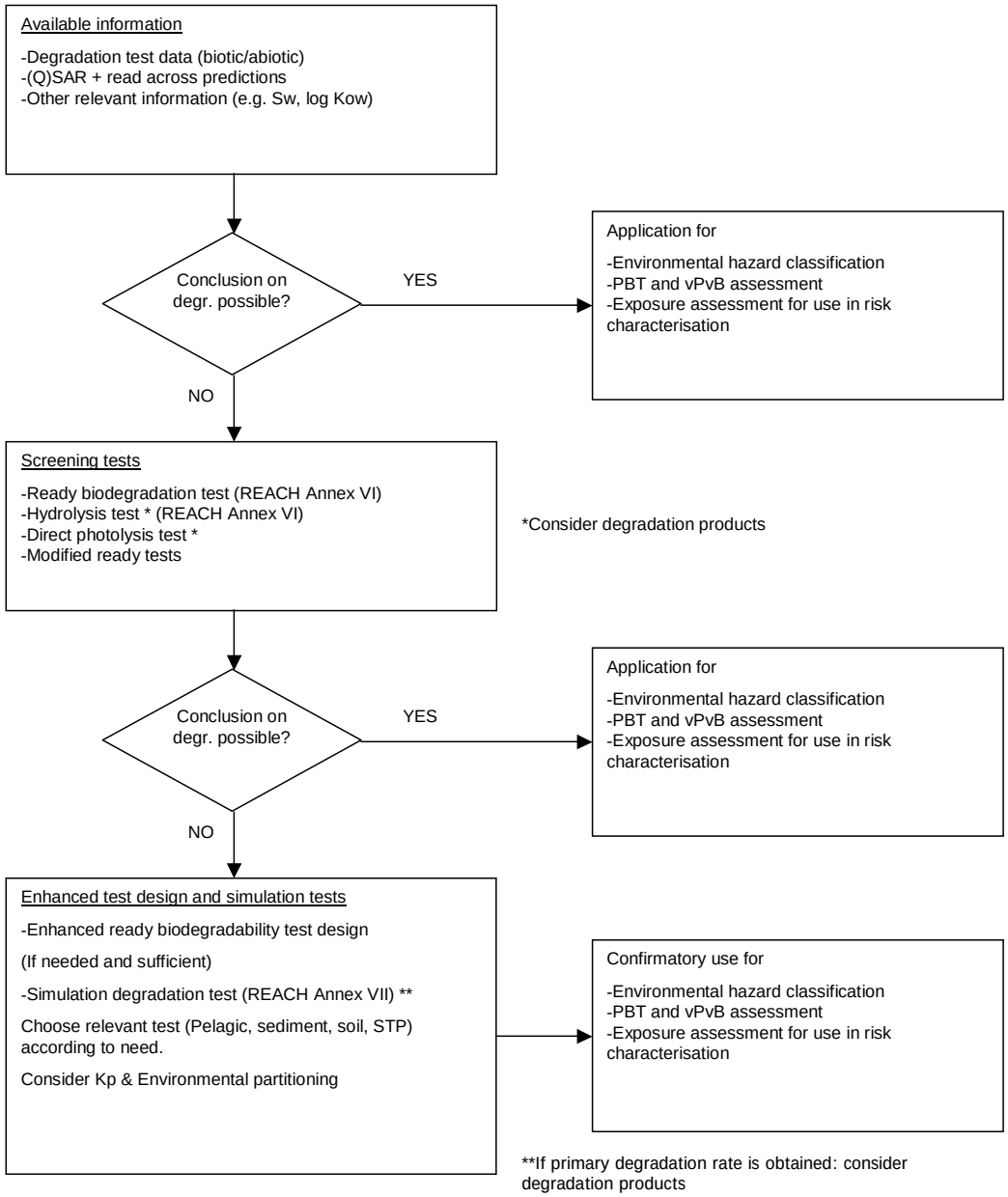
Once all the relevant information has been gathered in relation to the requirements of REACH, it needs to be determined whether sufficient information exists to draw conclusions for each of the three regulatory endpoints: hazard assessment (e.g. for classification and labelling), exposure assessment (for determination of the PEC) and persistency assessments (for PBT/vPvB assessment).

If insufficient information exists then the data gaps for each regulatory endpoint need to be identified together with a summary of any remaining uncertainty. For substances at tonnages that require simulation data, the most appropriate environmental compartments to support both P/vP assessment and exposure assessment should be identified.

R.7.9.6 Integrated Testing Strategy (ITS) for degradation/biodegradation

The ITS presented in Figure R. 7.9-2 attempts to summarise the approach required to maximise the use of degradation data against all three regulatory endpoints. The scheme starts with collating all available information before requiring tests at the screening and simulation test levels.

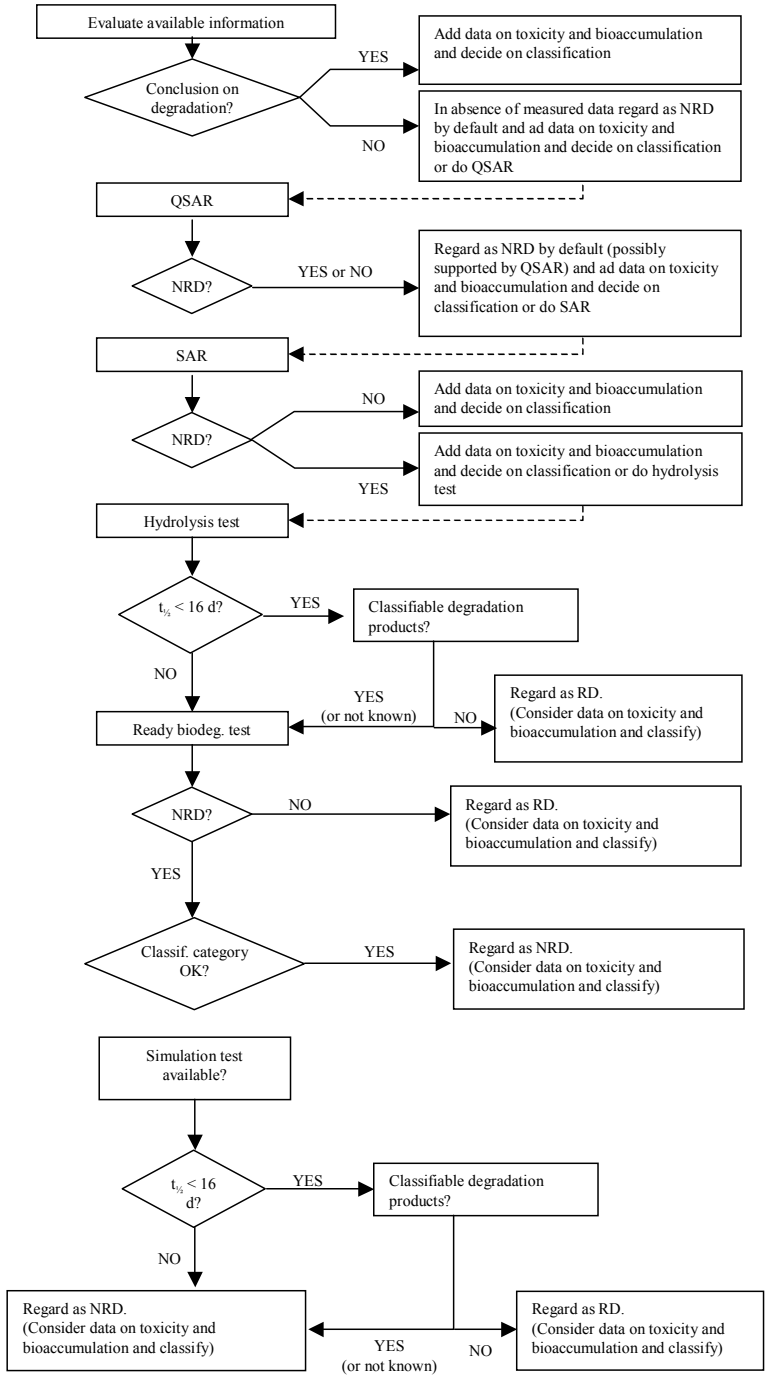
Figure R. 7.9-2: Overview decision scheme on degradation for the three regulatory needs Environmental hazard classification, PBT/vPvB assessment and Exposure assessment for use in risk characterisation



R.7.9.6.1 Classification and Labelling

An ITS to determine the suitability of degradation data on classification and labelling in provided in Figure R. 7.9-3.

Figure R. 7.9-3: An ITS for the use of degradation data in C&L.



Hazard classification should be considered regardless of the tonnage level and based on available information (GHS, Annex 9 [1]). Information on ready biodegradability is required already at a tonnage level of 1 t per year for the purpose of environmental hazard classification of a substance (OECD Test Guidelines 301 A-F, or OECD TG 310, or QSAR predictions). The choice between the six OECD 301 test guidelines, or the OECD TG 310 *head space variant* of OECD TG 301B, depends on the characteristics of the substance (see OECD introduction 'Degradation of Organic Chemicals' [2] and information in the individual test guidelines).

R.7.9.6.2 Chemical safety assessment

A chemical safety assessment (CSA) under REACH, including environmental hazard assessment and PBT/vPvB assessment, only has to be carried out for substances with an annual tonnage exceeding 10 tonnes per registrant. An exposure assessment (PEC characterisation) as well as a risk characterisation (PEC/PNEC ratios) has to be carried out if the substance is considered as dangerous.

Table R. 7.9-3 shows the relevant information on the ITS on degradation and which at a minimum should be available for each annual tonnage level above 10 tonnes per registrant.

Table R. 7.9-3: Required test data of interest for the ITS on biodegradation

Tonnage band (t/y/registrant)	**Required degradation data**	**Other relevant information**
10-100	Ready biodegradability Hydrolysis	Log K_{OW} Vapour pressure Water solubility Adsorption/desorption
100-1000	Ready biodegradability Hydrolysis Simulation of biodegradability in water[1] Simulation of biodegradability in sediment[2] Simulation of biodegradability in soil[3]	Log K_{OW} Vapour pressure Water solubility Adsorption/desorption Dissociation constant Degradation products BCF[4]
>1000	Ready biodegradability Hydrolysis Simulation of biodegradability in water[1] Simulation of biodegradability in sediment[2] Simulation of biodegradability in soil[3] Further testing shall be proposed if the CSA indicates a need for additional data on the degradation of the substance	Log K_{OW} Vapour pressure Water solubility Adsorption/desorption Dissociation constant Degradation products BCF[4]

1. Not needed if the substance is highly insoluble in water and/or is readily biodegradable (see Section R.7.9.2)
2. Not needed if the substance is readily biodegradable and/or direct and indirect exposure of sediment is unlikely (see Section R.7.9.2)
3. Not needed if the substance is readily biodegradable and/or direct and indirect exposure of soil is unlikely (see Section R.7.9.2)
4. Not needed if the substance has a low potential for bioaccumulation (for instance a log K_{ow} <3) and/or a low potential to cross biological membranes and/or direct and indirect exposure of the aquatic compartment is unlikely.

An exposure assessment can be carried out on the basis of information on ready biodegradability. If an environmental risk assessment of a substance leads to the conclusion *no risk*, using only information on ready biodegradability, then there is no need for further testing of the biodegradability.

However, further testing of the biodegradability (and/or ecotoxicity) of the substance may be required, if the risk assessment indicates a potential risk to one or more environmental compartments.

In the exposure assessment, rates for the biodegradation in the various compartments are used for the derivation of the associated PEC-values. These compartments include:

- Sewage treatment plant
- Freshwater
- Freshwater sediment
- Marine water
- Marine water sediment
- Soil

Additional consideration will be needed to whether or not inherent biodegradation test data (OECD 302) or sewage treatment simulation test data are required to refine the PEClocal and PECregional. These tests are not currently required under the REACH Annexes but can be used to refine the PEC and may help to determine whether either simulation tests are required or which simulation test may be the most relevant.

Table R. 7.9-5 shows an approach for selection of additional biodegradability tests, which may either simulate realistic conditions in the external environment (freshwater, marine or soil) or simulate the biodegradation and removal of the substance in the sewage treatment plant (estimates of effluent concentration, e.g. based on CAS test).

Table R. 7.9-4: Selection of appropriate biodegradation studies for PEC assessments

Relevant environmental compartment [1]	**Recommended biodegradation studies**
Freshwater	Freshwater simulation test (e.g. OECD 309) and/or CAS test (OECD 303)
Freshwater sediment	Freshwater water/sediment simulation test (e.g. OECD 308) and/or CAS test (OECD 303)
Marine water	Marine water simulation test (e.g. OECD TG 309) and/or CAS test (OECD 303)
Marine water sediment	Marine water sediment simulation test (e.g. OECD 308) and/or CAS test (OECD 303)
Soil	Soil simulation test (e.g. OECD 307)

[1]: The relevant environmental compartment(s) may be identified on the basis of an analysis of the intrinsic properties of the substance, modelling of transport and fate.

R.7.9.6.3 PBT/vPvB assessment

The information gathered through the steps outlined in the previous sections enables an assessment to be carried out for PBT/vPvB. Guidance for this is given in chapter R.11.

R.7.9.7 References on biodegradation

United Nations (2005) Global Harmonised System (GHS) for classification and labelling (GHS). ST/SG/AC. 10/30/Rev. 1. United Nations, New York and Geneva.

Howard, P., Meylan, W., Aronson, D., Stiteler, W., Tunkel, J., Comber, M., Parkerton, T. (2005). A new biodegradation prediction model specific to petroleum hydrocarbons, Environmental Toxicology &Chemistry, 24, 8, 1847-1860

Pavin, M. & Worth, A. P. (2006). Review of QSAR Models for Biodegradation. European Commission Joint Research Centre Report EUR 22355 EN. European Communities.

Howard et al (1992)"Predictive Model for Aerobic Biodegradability Developed from a file of evaluated Biodegradation Data", Environ.Toxicol. Chem., 11, 595-603

Boethling R.S., P.H. Howard, W. Meylan, W. Stitler, J. Beaumann & N. Tirado (1994): "Group Contribution Method for Predicting Probability and rate of Aerobic Biodegradation", Env. Sci. Technol. 28, 459-65.

TemaNord (1995):"Environmental Hazard Classification - data collection and interpretation guide (2nd edition)", Nordic Council of Ministers, Copenhagen, pp. 166

Müller M. (2005). Study for validation of Models for estimation of Atmospheric Degradation Based on data Set of 768 Substances. Final report. Frauenhofer Inst., Schmallenberg, Germany.

OECD. (1993). Environment Monograph No. 68 "Structure-Activity relationships for Biodegradation, OECD, Paris (available at http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html)

Grice, G. D., and Reeve, M. R. (1982). *Marine Mesocosms: Biological and chemical Research in Experimental Ecosystems*. Springer-Verlag, NY.

Lauth, J. R., Scott, G. I., Cherry, D. S., and Buikema, A. L. (1996). A modular estuarine mesocosm. Environmental Toxicology and Chemistry, 15 (5), pp 630-637.

Forney L.J., Liu W.-T., Guckert J.B., Kumagai Y., Namkung E., Nishihara T., Larson R.J. Structure of microbial communities in activated sludge: Potential implications for assessing the biodegradability of chemicals (2001) Ecotoxicology and Environmental Safety, 49 (1), pp. 40-53

OECD. (2000)., Series on Testing and Assessment no. 22, Guidance document for the performance of out-door monolith lysimeter studies. OECD Environmental Health and Safety Publications

Dubus I.G., Beulke S., Brown C.D., Gottesbüren B. & Dieses A. (2004). Inverse modelling for estimating sorption and degradation parameters for pesticides. Pest Management Science, 60:859-874.

Cornelese A, Boesten JJTI, Leistra M, Linden van der AMA, Linders JBHJ, Pol JW, Verschoor AJ (2003) Monitoring data in pesticide registration RIVM rapport 601450015, 27 p in English.

Knepper, T. P., Barcelo, D., and de Voogt P. (2003). Analysis and fate of surfactants in the aquatic environment. Comprehensive analytical chemistry, Volume LX, Elsevier, ISBN 0-444-50935-6.

Lyman et al. (1990), Lyman, W.J., Reehl, W.F. and Rosenblatt, D.H. Handbook of Chemical Property Estimation, Environmental Behavior of Organic Compounds, American Chemical Society, Washington, DC 1990.

Culp, J. M., Lowell, R. B., and Cash, K. J. (2000). Integrating mesocosm experiments with field and laboratory studies to generate weight of evidence risk assessments for large rivers. Environmental Toxicology and Chemistry, 19 (4), pp 1167-1173.

OECD. (2006a). Guidance document on simulated freshwater lentic field tests (outdoor microcosms and mesocosms). OECD Series on testing and assessment Number 53, ENV/JM/MONO (2006) 17.

Morrall S.W., J.C. Dunphy, M.L. Cano, A. Evans, D.C. McAvoy, B.P. Proce, W.S. Eckhoff (2006) Removal and environmental exposure of alcohol ethoxylates in US sewage treatment, Ecotoxicology and Environmental Safety 64, p. 3-13.

Eadsforth, C.V., A.J. Sherren, M.A. Selby, R. Toy, W.S. Eckhoff, D.C. McAvoy, E. Matthijs (2006) Monitoring the environmental fingerprints of alcoholethoxylates in Europe and Canada, Ecotoxicology and Environmental Safety 64, p.14-29.

Belanger, S.E., P.B. Dorn, R. Toy, G. Boeije, S.J. Marshall, T. Wind, R. van Compernolle, D. Zeller (2006) Aquatic risk assessment of alcohol ethoxylates in North America and Europe, Ecotoxicology and Environmental Safety 64, p.85-99.

FOCUS. (2000). "FOCUS groundwater scenarios in the EU review of active substances". The report of the work of the Groundwater Scenarios Workgroup of FOCUS (FOrum for the Co-ordination of pesticide fate models and their USe), Version 1 of November 2000. EC Document Reference Sanco/321/2000 rev.2, 202pp.

Feijtel, T., Struijs, J. and Matthijs. (1999). Exposure modelling of detergent surfactants – Prediction of 90th percentile concentrations in the Netherlands. Environmental Toxicology and Chemistry, Vol.18 no. 11, p 2645-2652.

Roije E., H. Loonen. M. Müller, G. Klopman & W.J. G.M. Peijnenburg (1999): “Evaluation and Application of Models for the Prediction of ready Biodegradability in the MITI-I test", Chemosphere 38 (6), 1409-17.

Roije, E. (2005): Presentation to EU TCNES QSAR WG, January 2006: “QSAR Experience Project 177 SIAM substances vs. BIOWIN models”.

Tunkel J. et al. (1999): “Predicting Ready Biodegradability in the Japanese Ministry of International Trade and Industry Test”, Environ. Toxicol. Chem, 19(10), 2478-85

Boethling R.S., D. G. Lynch & G. C. Thom (2003): "Predicting Biodegradability of Premanufacture Notice Chemicals", Environ. Toxicol. Chem, 22(4), 837-44.

Boethling R.S., D.G. Lynch, J.S. Jaworska, J. L. Tunkel, G. C. Thom & Simon Webb (2004): "Using BIOWIN, Bayes and batteries to Predict Ready Biodegradability, Environ. Toxicol. Chem 23(4), 911-20.

OECD (2006b). Introduction to the section of the OECD Guidelines for Testing of Chemicals on Degradation and Accumulation

OECD (2006c): Guidance document on validation of (Q)SAR models

ENV/JM/TG(2004)26Rev1: ”Comparison of SIDS Test Data with (Q)SAR Predictions for Acute Aquatic Toxicity, Biodegradability and Mutagenicity on Organic Chemicals Discussed at SIAM11-18", , Paris, Oct. 2004.

EU WG on Substances of very High Concern (2002): Working document: SHC/TS 2-3/029, Identification of potential PBTs and vPvBs by use of QSARs. Danish EPA, Copenhagen 2002-06-27

EU (2003) Technical Guidance Document on Risk Assessment.

OECD (2002). Report of the OECD/UNEP workshop on the use of multimedia models for estimating overall environmental persistence and long-range transport in the context of PBTs/POPs assessment. OECD series on testing and assessment n°36.

Appendices to Section 7.9

Appendix 7.9-1: International Guidelines for Assessing Biodegradability

Appendix 7.9-2: Reporting Requirements

Appendix 7.9-3: Testing Biodegradability of Poorly Water Soluble Substances

Appendix 7.9-4: Guidance for Testing of Mixtures for biodegradation

Appendix 7.9-1 International Guidelines for Assessing Biodegradability

Method	Test duration	Inoculum	Test conditions	Measurements	Limitations
Ready Biodegradability Tests					
OECD 301A DOC die away (ISO 7827)	Up to 28 days	Micro-organisms (~10^7 – 10^8 cells/ml) in surface waters, unchlorinated sewage treatment works effluents or activated sludge. Not pre-adapted inoculum	Agitation in the dark or diffuse light under aerobic conditions at 20-24^{o}C	DOC removal	Test substance has to be soluble, non-volatile, not sorbed to vessel or sludge and non-toxic at test conc.
OECD 301B CO_2 evolution test (ISO 9439, OPPTS 835.3120)	Up to 28 days	Micro-organisms (~10^7 – 10^8 cells/ml) in surface waters, unchlorinated sewage treatment works effluents or activated sludge. Not pre-adapted inoculum	Agitation in the dark or diffuse light under aerobic conditions at 20-24^{o}C	CO_2 production	Test substance must be non-toxic at test concentration.
OECD 301C Modified MITI Test	Up to 28 days	Micro-organisms (~10^7 – 10^8 cells/ml) in surface waters, unchlorinated sewage treatment works or industrial effluents or activated sludge. Not pre-adapted inoculum	Agitation in the dark under aerobic conditions at 24-26^{o}C	O_2 uptake	Test substance has to be non-toxic at test concentration, subject to interference from nitrification.
OECD 301D Closed bottle test (ISO 10707)	Up to 28 days	Micro-organisms (~10^5 cells/ml) in surface waters or unchlorinated sewage treatment works effluents Not pre-adapted inoculum	Agitation in the dark under aerobic conditions at 20-24^{o}C	O_2 uptake	Test substance has to be non-toxic at test concentration, subject to interference from nitrification.
OECD 301E Modified OECD screening test (ISO 7827)	Up to 28 days	Micro-organisms (~10^7 – 10^8 cells/ml) in unchlorinated sewage treatment works effluents Not pre-adapted inoculum	Agitation in the dark or diffuse light under aerobic conditions at 20-24^{o}C	DOC removal	Test substance has to be soluble, non-volatile, not sorbed to vessel or sludge and non-toxic at test conc.

Method	Test duration	Inoculum	Test conditions	Measurements	Limitations
OECD 301F Manometric respirometry test (ISO 9408)	Up to 28 days	Micro-organisms (~10^7 – 10^8 cells/ml) in surface waters, unchlorinated sewage treatment works effluents or activated sludge Not pre-adapted inoculum	Agitation in the dark or diffuse light under aerobic conditions at 20-24°C	O_2 uptake	Test substance has to be non-toxic at test concentration, subject to interference from nitrification.
OECD 310 (Headspace test) ISO 14593	Up to 28 days	Inoculum of aerobic mixed micro-organisms (approx 10^7-10^8 cells/l). Not pre-adapted inoculum	Batch culture, aerated aquatic test using the test chemical as the sole carbon source at 20-25°C. Assesses ultimate biodegradation.	CO_2 production in sealed vessels giving % degradation	Test substance must be non-toxic at test concentration.
Simulation Tests for Freshwater and Sediment Systems					
OECD 308 Aerobic and anaerobic transformation in aquatic sediment systems	Less than 100 days	Microorganisms in sediment (not pre-adapted)	Static test with natural water and sediment, with non-volatile ^{14}C labelled compounds at natural levels.	Chemical analysis of transformation products or $^{14}CO_2$ analysis where labelling used.	Simulates suspended sediment only. Test substance has to be non-toxic, non-volatile and soluble. Site specific with respect to sediment. Sorption to sediment may be misleading if ^{14}C not used.
OECD 309 Aerobic mineralisation in surface water	Up to 90 days for the batch test	Microorganisms in surface water (not preadapted) May include suspended sediment and/ or semi-continuous operation			

Method	Test duration	Inoculum	Test conditions	Measurements	Limitations
ISO 14592-1 (OPPTS 835.3170)	No fixed duration	Micro-organisms in surface water samples filtered through 100 um filter for a 'pelagic test' which may be amended with an aerobic sediment slurry from the study site for a 'suspended sediment test'.	Agitation in the dark or diffuse light under aerobic conditions at field temperature or 20-25^{o}C	Specific chemical or radio-chemical analysis (and DOC or TOC if possible) giving 1st order rate const.	Test substance has to be non-toxic, non-volatile and soluble. Site specific with respect to sediment. Sorption to sediment may be misleading if ^{14}C not used.
ISO 14592-2	No fixed duration but <60 days	Micro-organisms in surface water	Natural diffuse daylight or constant illumination of artificial white light (400-700 nm) with an energy of 50 uE/m^2/s at the water surface	Specific chemical or radio-chemical analysis giving 1st order rate const.	Test substance has to be non-toxic, non-volatile and soluble. Site specific with respect to sediment if used – glass beads may not be representative of sediment. Sorption to sediment may be misleading if ^{14}C not used.
OPPTS 835.3180 Sediment/ water microcosm	Less than 60 days	Natural microbial assemblage.	Sediment microcosms using intact cores with (semi) continuous water replacement. ^{14}C labelling at environmentally realistic levels recommended.	Chemical analysis of transformation products or $^{14}CO_2$ analysis where labelling used.	Test substance has to be non-toxic, non-volatile and soluble. Site specific with respect to sediment. Sorption to sediment may be misleading if ^{14}C not used.

Method	Test duration	Inoculum	Test conditions	Measurements	Limitations
Sewage Treatment Simulation Tests					
OECD 303A Aerobic sewage treatment: coupled unit test (ISO 11733)	Up to 12 weeks	Aerobic sewage	Elimination of test chemicals (20 mg l^{-1} DOC) from continuously fed laboratory scale coupled sewage treatment units.	DOC or COD giving % degradation	Test substance must be water soluble and non-volatile.
Primary Biodegradability Tests					
OPPTS 835.3220 Porous Pot Method,	At least 21 days	Activated sludge mixed liquor from a domestic plant.	Test and control pots filled with inoculum and 10-20 mgC/l test substance.	Primary biodegradation determined by test chemical removal, DOC analysis provides measure of ultimate biodegradation.	Test substance has to be soluble, non-volatile, not sorbed to vessel or sludge and non-toxic at test conc.
Simulation Tests for Marine Waters					
OECD 306 (ISO 7827 and 10707, OPPTS 835.3160)	Up to 60 days	Micro-organisms[2] in test seawater Not pre-adapted inoculum	Agitation in the dark or diffuse light under aerobic conditions at 15-20^{o}C. Concentrations 5-40 mg DOC l^{-1}	DOC	Test chemical must be non-toxic at test concentrations, soluble and not sorbed by vessel. Closed bottle test subject to interference from nitrification. High nutrient concentrations with respect to seawater
Simulation Tests for Soil					
OECD 307 Aerobic and anaerobic transformation on soil	Up to 120 days, longer under some circumstances				

Method	Test duration	Inoculum	Test conditions	Measurements	Limitations
Inherent Biodegradation Tests – Water					
OECD 302A Modified SCAS test (OPPTS 835.3210)	Months (often up to 120 days).	Settled domestic sewage and activated sludge. Inoculum to be sourced from a domestic treatment plant	Test chemical (20 mg DOC l^{-1}) aerated with settled domestic sewage and activated sludge (*ca.* 2500 mg l^{-1} TSS) for 23h at 20-25°C. Aeration stopped, sludge settled and supernatant removed. Fresh sewage and test chemical are added and the cycle repeated. ^{14}C-radiolabelled chemicals can be used for increased sensitivity.	DOC CO_2 production in sealed vessels giving % degradation. Potential to measure $^{14}CO_2$	Test substance must be non-volatile, not lost by foaming and non-toxic at test conc. Sorption potential needs to be determined.
OPPTS 835.5045 Modified SCAS for insoluble and volatile chemicals	Months (often up to 120 days).	Settled domestic sewage and activated sludge.		CO_2 production in sealed vessels giving % degradation Potential to measure $^{14}CO_2$	
OECD 302B Zahn Wellens (ISO CD9888) (OPPTS 835.3200)	28 days	Inoculum of 200 - 1000 mg l^{-1} (TSS) of activated sludge. Unadapted or pre-adapted inoculum	Aerated batch culture, using the test chemical as the sole carbon source (50 – 100 mg l^{-1} DOC) and with the inoculum at 20-25°C. Assesses ultimate biodegradation.	DOC or COD or Specific analysis for primary transformations	Test substance must be non-volatile, not lost by foaming and non-toxic at test conc. Sorption potential needs to be determined.

Method	Test duration	Inoculum	Test conditions	Measurements	Limitations
OECD 302C MITI (II)	14-28 days	Aerobic mixed, specially grown, unadapted micro-organisms at 100 mg l^{-1} (TSS, or approx. 3 $\times 10^7$ - 3 $\times 10^8$).	Agitated batch culture, using the test chemical as the sole carbon source (30 mg ThOD/l) with inoculum. Assesses ultimate biodegradation.	O_2 demand and possibly specific chemical analysis	Test substance must be non-volatile, not lost by foaming and non-toxic at test concentration.
OPPTS 835.3100 Aerobic aquatic biodeg	28 days after pre-adaptation	Pre-adapted inoculum	Agitated aerated aquatic test using test chemical (10 mg l^{-1} DOC) pre-adapted inoculum from a medium concentration of aerobic mixed micro-organisms at 20-25°C. ^{14}C labelled compounds may be used	DOC removal and CO_2 evolution ^{14}C provides mass balance phase distribution data	Test substances must be soluble and non-volatile.
OPPTS 835.5045 Modified SCAS test for insoluble and volatile chemicals	40 to 120 days	Settled domestic sewage and activated sludge Unadapted or pre-adapted inoculum	Same principle as for OECD 302A but with a volatiles trap on the aeration unit and additional analytical requirements for trapped volatiles and sludge solids. 20 mg l^{-1} DOC test concentration at 20-25°C. ^{14}C labelled compounds may be used.	DOC. Specific analysis can provide primary transformation data. Kinetic data and half-life determination available. >20% removal of DOC =inherent biodegradation, >70% =ultimate biodegradation.	Additional analytical requirements.
Inherent Biodegradation – Soil					
OECD 304A (ISO 14239 – biometer system) OPPTS 835.3300	Up to 64 days	Disturbed soil – alfisol, spodosol, entisol. In special cases can use soil with high silt fraction content or soil with high clay content (30%).		CO_2 evolution giving % degradation	

Method	Test duration	Inoculum	Test conditions	Measurements	Limitations
Anaerobic Degradation Test Methods					
OECD 311 ISO 11734	Up to 60 days	Washed digester sludge at 1-3 /l in nutrient amended anaerobic medium, containing a redox indicator in sealed vessels.	Batch culture with test concentration of 20-100 mg l^{-1} as OC, at 35°C. Assesses ultimate biodegradation	Total gas production (CH_4+CO_2) using a pressure transducer and DIC	Test substance must be non-toxic at test concentration.
OPPTS 835.3400 Anaerobic biodegradability of organic chemicals	Up to 56 days.	Sludge from an anaerobic sludge digestor. Recommendations are for a well-mixed primary sludge from a digester with a retention time of 15 to 25 days.	Test sample concentrations at around 50 mg l^{-1} with tests carried out at 35°C.	CO_2 and CH_4 production.	Not applicable to toxic chemicals, reproducibility not yet fully defined. Uses high concentrations of test substances.

Appendix 7.9-2 Reporting Requirements

Hydrolysis Test Requirements (OECD 111)

The test report should include the following information:

- Test substance:
    - common name, chemical name, CAS number, structural formula (indicating position of label when radiolabelled material is used) and relevant physico-chemical properties;
    - purity (impurities) of test substance;
    - label purity of labelled chemical and molar activity (where appropriate).
- Buffer solutions:- buffers and waters used;- molarity and pH of buffer solutions.
  Test conditions:
    - amount of test substance applied;
    - solvents (type and amount) used for application of the test substance;
    - volume of buffered test substance solutions incubated;
    - description of the incubation system used;
    - pH and temperature during the study;
    - sampling times;
    - method(s) of extraction;
    - methods for quantification and identification of the test substance and its hydrolysis products in the buffer solutions;
    - number of replicates.
- Results:
    - repeatability and sensitivity of the analytical methods used;
    - recoveries;
    - replicate data and means in a tabular forms;
    - mass balance during and at the end of the studies (when labelled test substance is used);
    - results of preliminary test;
    - discussion and interpretation of results;
    - all original data and figures.

The following information is only required when the hydrolysis rate is determined:

- plots of concentrations versus time for the test substances and, where appropriate, for the hydrolysis products at each pH value and temperature;
- tables of results of Arrhenius equation for the temperature 20 °C/25 °C, with pH, rate constant [h-1 or day-1], half-life or DT50, temperatures [°C] including confidence limits and the coefficients of correlation (r2) or comparable information;
- proposed pathway of hydrolysis.

Ready biodegradability test requirements (OECD 301 series and OECD 310)

- Test substance:
    - physical nature and, where relevant, physico-chemical properties;
- Test conditions:
    - inoculum: nature and sampling site(s), concentration and any pre-conditioning treatment;
    - proportion and nature of industrial waste water in sewage, if known;
    - test duration and incubation temperature;
    - in the case of poorly soluble test substances, methods of preparation of test solutions/suspensions;
    - test method applied; scientific reasons and explanation for any change of procedure;
    - details of controls.
- Results:
    - data in tabular form;
    - any observed inhibition or toxicity;
    - any observed abiotic degradation;
    - specific chemical analytical data, if available;
    - analytical data on intermediates, if available;
    - the graph of percentage degradation against time for the test and reference substances to include the lag phase, degradation phase, the 10-d window and slope (see Annex I for definitions);
    - percentage removal at plateau, at end of test, and/or after 10-d window.
- Discussion of results

Marine Biodegradability Test Requirements (OECD 306)

- Test substance:
    - physical nature and, where relevant, physico-chemical properties;
- Test conditions:
    - location and description of the sampling site; pollution and nutrient status (colony count, nitrate, ammonium, phosphate if appropriate);
    - characteristics of the sample (date of sampling, depth, appearance, temperature, salinity, DOC (optional), delay between collection and use in the test;
    - method used (if any) for ageing of the seawater;
    - method used for pre-treatment (filtration/sedimentation) of the seawater;
    - method used for DOC determination;
    - method used for specific analysis (optional);
    - method used for determining the number of heterotrophs in the seawater (plate count method or alternative procedure) (optional);
    - other methods (optional) used to characterise the seawater.
- Results:
    - the course of the degradation test is represented graphically in a diagram showing the lag phase (tL), slope, and time (starting from the end of the lag phase) to reach 50 per cent removal (t50). The lag phase may be estimated graphically as shown in the figure in the "Validity and interpretation of results" section or conveniently taken as the time needed for 10 per cent degradation;
    - percentage degradation measured after 60 days, or at end of test.
- Discussion of results.

Inherent Biodegradability Test Requirements (OECD 302 Series)

The test report should include the following information:

- Test substance:
    - physical nature and, where relevant, physico-chemical properties;
- Inoculum:
    - source, concentration, pre-treatment and status of adaptation.
- Test conditions:
    - analytical methods used;
    - procedure control and compound used in the control.

- Results:
    - biodegradation curve;
    - toxicity evaluations;
    - the degree of biodegradation attained at the end of the test after 28d, or earlier if complete degradation is attained in less than 28d, as "inherent biodegradability in the static test after x days";
    - any significant difference between the DOC (or COD) in the first sample at 3h after starting the test and the value calculated from the amount of test compound added as "adsorbed by the activated sludge" (OECD 302B);
    - the adaptation phase (days), the biodegradation phase (days) and the endpoint of biodegradation reached after x days as identified from the biodegradation curve.
- Discussion of the results.

Appendix 7.9-3 Testing the Biodegradability of Poorly Water Soluble Substances

This appendix discusses the technical issues associated with conducting biodegradability assays with poorly water-soluble substances and the data-reporting requirements that would improve confidence in the data generated for such substances. The OECD and ISO Guidance 10634 (1995) for testing poorly water-soluble substances will form the basis of discussion. Whilst the focus of this document will be towards methods for assessing the ready biodegradability of poorly water-soluble substances (OECD 301 series and the OECD 310 test) the issues equally apply to other biodegradability assays.

OECD Evaluation of the Biodegradability of Poorly Soluble Substances

OECD requires that when assessing biodegradability of poorly soluble compounds OECD the following aspects should receive special attention (OECD, 1992: Annex III):

- While homogeneous liquids will seldom present sampling problems, it is recommended that solid materials be homogenised by appropriate means to avoid errors due to non-homogeneity. Special care must be taken when representative samples of a few milligrams are required from mixtures of chemicals or substances with large amounts of impurities.
- Various forms of agitation during the test may be used. Care should be taken to use only sufficient agitation to keep the chemical dispersed, and to avoid overheating, excessive foaming and excessive shear forces.
- An emulsifier which gives a stable dispersion of the chemical may be used. It should not be toxic to bacteria and must not be biodegradable or cause foaming under the test conditions.
- The same criteria apply to solvents as to the emulsifiers.
- It is not recommended that solid carriers be used for solid test substances but they may be suitable for oily substances.
- When auxiliary substances such as emulsifiers, solvents and carriers are used, a blank run containing the auxiliary substance should be performed.
- Any of the four respirometric tests (301 B, 301 C, 301 D, 301 F) can be used to study the biodegradability of poorly soluble compounds.

Whilst OECD raise a series of valid issues that require careful considerations in testing the biodegradability of poorly soluble substances they do not constitute explicit guidance. The only critical guidance provided is the applicability of a restricted range of the 301 test series (point 7) and the requirement of additional control vessels where emulsifiers, solvents and carriers are used (point 6). Tests conducted with draft OECD 310 test "Ready Biodegradability – CO_2 in sealed vessels (Headpsace Test)" are also suitable for assessing the biodegradability of poorly soluble substances.

Whilst advocating the use of emulsifiers, solvents and carriers, none are specifically identified and no guidance is provided regarding the acceptable level of each that can be introduced into the test system. Consequently, numerous approaches of introducing the test substance can be applied and this will make it difficult to identify a set of core acceptable or workable solutions.

ISO Guidance for the preparation and treatment of poorly water-soluble organic compounds for the subsequent evaluation of their biodegradability in aqueous medium

In 1995 the International Standards Organization (ISO) concluded that the development of a single method for evaluating the biodegradability of poorly water-soluble organic substances might not be realized in the immediate future. Consequently, ISO proposed a series of methods where the final selection was based on a judgment of the physico-chemical properties of the test substance (ISO, 1995).

The ISO standard (1995) addressed four techniques for preparing poorly water-soluble substances and introducing them into the test apparatus. It must be noted than for water-soluble test substances compounds are usually introduced into the test medium via a concentrated stock solution. The methods proposed by ISO for poorly soluble substances were 1) direct addition, 2) ultrasonic dispersion, 3) adsorption on an inert support, and 4) creating a dispersion or emulsion. All of these techniques proposed by ISO are suitable for including within the OECD 301 and 310 test guidelines. ISO does not provide any advice about the use of suitable poorly soluble reference standards. Each of the ISO methods will be described below with a brief commentary or assessment.

DIRECT ADDITION

ISO proposed introducing the test compound by either 1) weighing the substance directly into the test vessel, 2) weighing the test compound on to an inert support (typically a glass cover slip or piece of foil) and introducing this into the test vessel, or 3) preparing a solution of the test substance in a volatile solvent are removing the solvent prior to testing.

Direct addition is applicable for a variety of substances e.g. crystalline solids and non-viscous liquids. These are introduced using either high precision micro-pipettes or direct weighing. In the case of direct weighing some replicate-to-replicate variability can be expected for crystalline compounds as they are usually being introduced at the very low mg weight range. Whilst direct pipetting using viscous liquids can be problematic, the use of a cover slip or foil can over come this. However care should be taken to ensure that the cover slip remains face up, if this becomes inverted then the microbiota will not be able to access the test substance.

It must be noted that control flasks will be needed where carrier solvents have been used to ensure that all the solvent has been eliminated. In this case the same volume of the solvent needs to be introduced into the test system as in the test flask, but without the test substance. Even low levels of respiration associated with the solvent will need to be accounted for when interpreting data from the test flasks. Whilst controls should be used for cover slips etc. it is unlikely that any background respiration will be observed.

Direct addition, particularly via direct weighing (or pipetting) or using a support, should act as a ‘bench mark’ and be applied in the assessment of all poorly water-soluble substances i.e. they should be used in parallel to any of the other guidance methods recommended by ISO. Direct addition is likely to give the most conservative estimate of biodegradation.

ULTRASONIC DISPERSION

ISO (1995) recommend that a dispersion of the compound can be prepared using an ultrasonic probe prior to introducing it into the test vessel. Specific guidance are provided with respect to the frequency of the ultrasonication required to make a 20 times concentrated stock solution, however total carbon analysis is required to confirm the concentration achieved.

It must be noted that this approach is not suitable for substances that undergo thermal decomposition and that a stable emulsion is rarely formed. Consequently, this may not be the most appropriate approach recommended within the ISO guidance. This is particular true when stable emulsions cannot be formed and large numbers of sacrificial test flasks are being prepared as the possibility exists for introducing reduced concentrations to each flask with time i.e. a concentration gradient. If this technique is to be applied to tests using sacrificial analysis (e.g. OECD 310) the test flaks need to be sacrificed randomly for analysis at each time point.

ADSORPTION ON TO AN INERT SUPPORT

ISO (1995) recommend the use of silica gel, glass filter or any other non-biodegradable inert supports that do not release organic carbon into the test media. Supporting evidence is required to demonstrate that the support is inert and carbon free and the amount of support used should be minimal. Silica-based gels that are used for chromatography represent an inert support that has been used successfully.

The test compound is usually introduced into the inert support at the required concentration via a carrier solvent (e.g. acetone or dichloromethane). Rotary evaporation and oven drying are then used to remove the solvent. A parallel procedure is required using the inert support and carrier solvent without the test substance for use in the control test flasks. Inert supports can also be used with insoluble solids.

Prior to testing the carbon level of the inert support containing the test chemical or the specific chemical contained in the inert support needs to be quantitatively determined and compared to nominal. The required amount of the inert support can then be directly weighed into the test vessel. Any biodegradation of the solvent should be taken into account through the use of parallel control vessels.

This procedure is applicable for compounds that will not be lost during the rotary evaporation and oven drying procedures. It does enable the amount of material to be directly weighed into the test flask to be increased thus increasing accuracy between replicate test flasks.

DISPERSION WITH AN EMULSIFYING AGENT.

ISO (1995) recommend using emulsifying agents to enhance the available of the poorly soluble test substance that are non-biodegradable and non-toxic under the conditions of the biodegradation test. Synperonic PE/P94, Synperonic PE/P103 or Tween 85 have been identified as commercial substances that could be used as emulsifying agents. Carrier solvents that are also non-toxic and non-biodegradable are also required to form these emulsions.

ISO recommends that three emulsions be prepared prior to selecting the most homogeneous emulsion for use in the biodegradation test. Very clear guidance is also provided that states that the degradation observed in the control vessel (solvent and emulsifier with no test compound) must not exceed 10% of the degradation observed in the test flasks for the test to be consider valid.

Supporting evidence should be provided to demonstrate that neither the solvent or the emulsifying agents are toxic to microbes or are biodegradable.

Minimum Test and Data Requirements for Poorly Water Soluble Substances

The following information should be reported:

- Information on the chemical's water solubility, vapour pressure and adsorption characteristics are essential.
- The solubility of the chemical in other solvents should be stated (especially those being used to disperse the chemical in emulsifications and on to inert supports).
- The chemical structure or formula should be identified in order to calculate theoretical values and/or check measured values of parameters, e.g. ThOD, $ThCO_2$, DOC, TOC, and COD. Information on the purity or the relative proportions of major components of the test material is required in order to interpret the results obtained, especially when the result lies close to the pass level.
- Information on the toxicity of the test substance, or any emulsifiers or carrier solvents, to bacteria may be very useful for selecting appropriate test concentrations and preparation strategies.
- Any pre-treatment of the compound before the test.
- The method of test substance introduction should be described in detail with supporting evidence especially regarding the use of solvents, emulsifiers and inert supports.
- Nominal versus measured carbon concentrations where inert supports and emulsions are used to generate concentrated stock preparations of the test substance prior to use. This should include the degree of recovery.
- Duration of any pre-treatment.
- Rate of degradation observed in the control flasks (treatment minus test substance).
- Suitable positive reference poorly soluble data (see below).

Conclusions & Recommendations on biodegradability testing of poorly water-soluble chemicals

There is no single method for assessing the biodegradability of poorly water-soluble substances. The state of the science has not changed since ISO published its guidance in 1995. A combination of approaches should be used and these should at the very minimum be compared to biodegradation observed by direct addition. Direct addition will usually provide the most conservative estimate of biodegradation.

Normal positive reference substances such as sodium acetate, sodium benzoate, aniline or glucose offer little support in the assessment of poorly soluble substances other than demonstrate that the inoculum is active. In order to 'bench mark' methods to assess poorly soluble substances common poorly soluble reference substances should be used. Two examples are provided in the Annexes of the ISO guidance. These are biodegradation curves for diisooctylphthalate (where adsorption on inert support and dispersion with an emulsifying agent enhances degradation compared to direct addition) and anthraquinone (where adsorption on inert support and dispersion with an emulsifying agent enhances degradation compared to direct addition). In both cases the use of ultrasonication did not provide any significant benefit.

Greater confidence in the methods for increasing the availability of poorly soluble substances will be gained by using either diisooctylphthalate or anthraquinone as a positive control. The reference control should be introduced to the test system by direct addition and the choice of preparation. Therefore for any given biodegradation assessment there will need to be the following series of flasks:

- Blank Control (inoculum & media with no test compound);
- Positive reference for biodegradation (sodium acetate, sodium benzoate, aniline or glucose);
- Poorly soluble positive control (either diisooctylphthalate or anthraquinone introduced by direct addition);
- Test substance (introduced by direct addition for conservative assessment);
- Direct addition control;
- Test substance with choice of introduction (e.g. adsorption on an inert support);
- Poorly soluble positive control using the same choice of introduction as the test substance; and
- Choice of introduction control (e.g. inert support and solvent without the test substance).

The above set of flasks appears onerous but they do not constitute a great deal of extra effort or expense. The long-term value of providing the additional information will be one of greater confidence in assessing poorly-soluble material against agreed bench mark standards.

Appendix 7.9-4 Guidance for Testing of Mixtures (e.g. UVCB Petroleum Substances) for biodegradation

(Due to derivation from natural crude oils and subsequent production from use of various refining processes, petroleum substances are complex mixtures of hydrocarbons, often of variable composition. Many petroleum substances are produced in very high tonnages to a range of technical specifications, with the precise chemical composition of unique structures, rarely if ever characterised. Since these materials are typically separated on the basis of distillation, the technical specifications usually include a boiling point range. These ranges correlate with approximate carbon number ranges, while the nature of the original crude oil and subsequence refinery processing influence the types of hydrocarbon structures present. The CAS definitions established for the various petroleum substance streams generally reflect this detail, including final refinery process; boiling range; carbon number range and predominant hydrocarbon types present.

For most petroleum substances, the complexity of the chemical composition is such that that it is beyond the capability of routine analytical methodology to obtain complete characterisation. Typical substances may consist of predominantly mixtures of straight and branched chain alcanes, single and multiple naphthenic ring structures (often with alkyl side chains), single and multiple aromatic ring structures (often with alkyl side chains). As the molecular weights of the constituent hydrocarbons increase, the number and complexity of possible structures (isomeric forms) increases exponentially.

Environmental testing strategies for petroleum substances must necessarily reflect the complexity of their composition. Reflecting the properties of the constituent hydrocarbons, petroleum substances are typically hydrophobic and exhibit low solubility in water. However, individual constituent hydrocarbons will exhibit a wide range of water solubilities. 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 in water will be different from the composition of the parent substance. The complex composition and generally low water solubility impacts the choice and conduct of biodegradation studies. A further complication is the volatility of constituent hydrocarbons, which shows a wide variation across the range of carbon numbers and hydrocarbon structures present in petroleum substances. It has been the practise to assess the inherent hazards of petroleum substances by conducting testing in closed systems (going to great lengths to ensure that volatile losses are minimised), even though under almost all circumstances of release into the environment, there would be extensive volatilisation of many of the constituent hydrocarbons.

BIODEGRADATION TESTING METHODS

Lower molecular weight hydrocarbons tend to be readily biodegradable in standard OECD tests, and although biodegradability decreases as molecular weight increases (corresponding to decreasing water-solubility and thus reduced bioavailability) hydrocarbons are generally regarded as being inherently biodegradable. The initial metabolites of hydrocarbons will be carboxylic acids and hence of less concern than the parent structures.

Typically, laboratory studies of the aquatic biodegradability of petroleum substances have evaluated the biodegradation potential of the whole substance, not just the portion which is soluble in water. To achieve adequate sensitivity, most biodegradation tests utilise higher concentrations of substances than would commonly be found in the environment. For a petroleum substance, this

means that there will be a large proportion of the substance in the undissolved phase and hence, not fully available to the degrading organisms. This will result in an underestimate of its true potential to biodegrade in the environment. It is also likely that the rate of biodegradation will be affected; firstly, the rate of biodegradation is likely to be limited by the rate of dissolution and solubility of individual hydrocarbon components. Secondly, the fact that petroleum substances contain a complex mixture of components results in a stepwise, sequential adaptation of the microorganisms to utilise individual hydrocarbons, again resulting in deviation from 'typical' kinetics. For these reasons, typical logarithmic growth phase (Monod) biodegradation kinetics (which are assumed to occur in RB tests) may not be observed with petroleum substances, so that even if individual components are readily biodegraded, the petroleum substance may not achieve the '10-day window' defined by OECD [Deneer et al, 1988].

Some modifications of test methods to enhance dissolution rates may improve this situation. Guidance on approaches to the testing of poorly soluble substances has been published [Whitehouse and Mallet, 1994]. Experimental methods include ultrasonic dispersion, addition of an inert dispersant or emulsifier to assist in dispersion, or addition of the test substance on an inert support (to increase the surface area and hence aid access of the microorganisms). See Section R.7.9.4.1.

Several accepted methods for determining biodegradation potential are unsuitable for poorly soluble substances (because they are based on measurement of total dissolved organic carbon) or are unsuitable for volatile substances (because volatile components are lost by evaporation, rather than biodegradation).

Three basic types of biodegradation test are used to estimate the relative biodegradability of substances, viz. ready, inherent and primary biodegradation methods. The use of these procedures in testing petroleum substances is dealt with in the following paragraphs. Usually only ready biodegradation data are used for classification, although, for example under the GHS scheme, other types of information may be used e.g. simulation test data or primary degradation data and consideration of degradation products.

The rationale for using standard laboratory tests to assess biodegradation potential of mixtures has been discussed in an EU workshop [European Chemicals Bureau, 1996]; it was agreed that the available methods were suitable for evaluating the biodegradation potential of mixtures comprising homologous series of hydrocarbons (like the petroleum substances), although such methods were not judged generally applicable for mixtures (e.g. preparations).

READY BIODEGRADABILITY TESTS

These are the most stringent of the commonly used laboratory tests, measuring complete mineralisation or Ultimate Biodegradation of the test substance (oxidation to carbon dioxide and water) using an unadapted inoculum [15]over a 28-day period. Ready Biodegradability is defined in terms of the pass/fail criteria agreed for each of the six test methods published by OECD (and subsequently adopted by the EU) [EU, 1967; OECD, 2000]; in particular, the required level of biodegradation must be obtained within 10 days of 10% biodegradation being achieved. In all the 28-day biodegradation tests, the mineral salts concentration, temperature and pH are tightly controlled, and the microbial inoculum is not allowed to be pre-exposed to the test substance. In addition to the OECD methods, there is a surrogate procedure whereby if the BOD5:COD ratio is

[15] The ready biodegradation testing implies use of inoculum from municipal STPs – and thus the adaptation that occurs in domestic STPs is implicitly taken into account

0.5 or higher, the substance is regarded as being readily biodegradable. Because of the stringency of these test methods, it is presumed that any substance demonstrating Ready Biodegradability will be rapidly biodegraded if released into the aquatic environment.

The Modified Sturm test (OECD 301B) for non-volatile substances and the Respirometric Method (OECD 301F) are the most commonly used methods for petroleum substances. More recently a test guideline that addresses the biodegradation of volatile substances has also been published, OECD 310.

INHERENT BIODEGRADABILITY TESTS

These laboratory methods are less stringent than the Ready Biodegradability tests, and hence, increase the likelihood of observing biodegradation within a specific test system. The extent of complete oxidation of the test substance to carbon dioxide and water is still measured.

Inherent Biodegradability is again defined in terms of the percentage biodegradation recorded in the test; it can be presumed that substances demonstrating Inherent Biodegradability will not persist if released into the aquatic environment.

Unfortunately, the currently available Inherent Biodegradation test methods defined by OECD [OECD, 2000] are not suitable for petroleum substances [CONCAWE, 1992]. However, following development and validation of a new Inherent Biodegradation test within ISO [Battersby, 1997;ISO, 1996], CONCAWE has recently validated a version of this Headspace Method, adapted to make it more suitable for petroleum substances; the results of this trial have recently been published [Battersby, et al, 1999]. This method is still under discussion as regards its suitability.

PRIMARY BIODEGRADATION TESTS

Originally developed for evaluating the biodegradability of two-stroke outboard engine lubricants, the CEC L-33-A-93 biodegradation method [CEC, 1995] has been extensively used in the oil industry for assessing the biodegradation potential of a wide range of oil products. The test estimates biodegradation on the basis of a specific change in chemical composition, viz. loss of the parent substance rather than mineralisation. Similar tests can also be conducted using specific GC and CG-MS analytical methods, although as the substance becomes more complex. Results obtained using these procedures are generally of limited value for classification purposes, but may in specific cases provide useful information on comparing the relative biodegradability between substances as well as providing data to support persistence and risk assessment. In such cases the degradation products should also be assessed to the extent necessary for the purposes of the assessment.

ABIOTIC DEGRADATION

Hydrolysis is not an important fate process for petroleum substances since hydrocarbons do not undergo reaction with water. However, degradation of unsaturated hydrocarbons, notably aromatic hydrocarbons by reaction with sunlight in the presence of oxygen can be a significant removal process where such substances are present in, or near the surface of water. Whilst current criteria for environmental hazard classification do not address photodegradation, this is a significant fate process for a number of aromatic hydrocarbons present in certain petroleum streams. The significance of the issue for risk assessment has been reviewed (CONCAWE Ecology Group paper,

2006). The rate of direct photolysis of chemicals in water is highly dependent on the latitude, season and the shadowing effect of the water column plus suspended material in the water column.

R.7.9.8 References to Appendices

Deneer, J.W. et al (1988) The joint toxicity to Daphnia Magna of industrial organic chemicals at low concentrations. Aquatic Toxicology, 12, 33-38

Whitehouse, P. and Mallet, M, (1994) Aquatic toxicity testing for notification of new substances: an advisory document on dealing with 'difficult' substances. Report No. CP 722(P). London: HMSO

European Chemicals Bureau (1996) Draft summary report of a workshop on biodegradation. Doc. No. ECBI/17/96-Add.2 Revised. Workshop June 12-14, 1996, Arona.

EU (1967) Council Directive of 27th June 1967 on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances (67/548/EEC). Official Journal of the European Communities No. L196, 16.08.1967

OECD (2000) OECD guideline for testing of chemicals. Volume 1 and 2. Paris: Organization or Economic Cooperation and Development

CONCAWE (1992) Ecological testing of petroleum products: test methodology. Report No. 92/56. Brussels: CONCAWE.

Battersby, N.S. (1997) The ISO headspace method CO2 biodegradation test. Chemosphere, 34, 8, 1813-1822.

ISO (1996) Water quality – evaluation of ultimate aerobic biodegradability of organic compounds in aqueous medium – method by analysis of released inorganic carbon in sealed vessels. Draft International Standard ISO/DIS 14593. Geneva: International Organization for Standardization.

Battersby N.S, et al, (1999) An inherent biodegradability test for oil [products: description and results of an international ring test. Chemosphere, 38, 14, 3219-3235.

CEC (1995) Biodegradability of two-stroke cycle outboard engine oils in water. Test Method CEC L-33A-93. Brussels: Co-ordinating European Council.)