RIVM-report 601050001/2007

Selected Integrated Testing Strategies (ITS) for the risk assessment of chemicals

T.G. Vermeire, T. Aldenberg, Z. Dang, G. Janer,
J.A. de Knecht, H. van Loveren,
W.J.G.M. Peijnenburg, A.H. Piersma, T.P. Traas,
A.J.Verschoor, M. van Zijverden, B. Hakkert

Contact
T.G. Vermeire
Expertise Centre for Substances (SEC)
theo.vermeire@rivm.nl

This investigation has been performed by order and for the account of the Director General of RIVM and the Directorate-General for Environmental Protection, Directorate for Chemicals, Waste and Radiation Protection (VROM/DGM/SAS), within the framework of projects S/601050 (ITS) and M/601200 (Risk assessment tools), respectively.

RIVM, P.O. Box 1, 3720 BA Bilthoven, telephone: 31-(0)30 274 9111; telefax: 31-(0)30 274 29 71

Abstract

Selected Integrated Testing Strategies (ITS) for the risk assessment of chemicals

RIVM's investigation of integrated testing strategies (ITS) to reduce the use of experimental animals in the risk assessment of chemical substance focuses on the application of alternative and improved test methods. This research is highly desirable for the contribution it can make to the successful implementation of the new EU legislation for industrial chemicals, REACH, as of June 1, 2007. In this way RIVM will contribute to EU and OECD research into testing strategies.

Integrated, or Intelligent, Testing Strategies (ITS) are strategies for the effective testing of the hazards of chemical substances. Such strategies show what tests or mathematical methods should be used for a particular substance, and in what order. ITS are intended as an answer to the ever-increasing demand for testing in regulations for a great number of substances with limited databases. The focus of ITS is especially on the development of strategies on the basis of test methods at cellular level (*in vitro*) and mathematical methods (*in silico*). The mathematical methods are needed for the assessment of exposure and of the relation between effects and chemical structure. Some tests with experimental animals (*in vivo*) will also remain necessary. Knowledge on the effects of chemical substances can be derived with sufficient certainty by smartly coupling these methods with each other. In this way, the expectation is that chemical substances will be assessed cheaper and faster, with less use of experimental animals.

The report first describes how to deal with uncertainties in the results of tests and methods used in each step of a particular ITS. Next it focuses on testing strategies for the assessment of: 1) environmental degradation 2) sensitization and 3) adverse effects on fertility and progeny.

Key words: ITS, integrated testing strategy, risk, chemicals, weight-of-evidence

Rapport in het kort

Teststrategieën voor de risicobeoordeling van chemicaliën

(**Selected Integrated Testing Strategies (ITS) for the risk assessment of chemicals)**

Het RIVM heeft teststrategieën onderzocht (ITS) om de risicobeoordeling van chemicaliën proefdiervriendelijker te maken. Het wil hiermee bijdragen aan onderzoeksprogramma's hiernaar van de EU en de OESO. Het accent ligt daarbij op de toepassing van alternatieve methoden en de verbetering van bestaande testmethoden. Dit onderzoek is hard nodig om de nieuwe Europese wetgeving voor industriële stoffen, REACH, te laten slagen, welke op 1 juni 2007 in werking treedt.

ITS staat voor geïntegreerde of intelligente Teststrategieën. Het zijn strategieën op papier waarmee chemische stoffen zo effectief mogelijk kunnen worden getest op mogelijke gevaren. Die strategieën maken inzichtelijk welke testen of wiskundige methoden voor een bepaalde stof moeten worden gebruikt, en in welke volgorde. ITS wil een antwoord zijn op de toenemende testvereisten in de regelgeving voor grote aantallen stoffen waarover weinig fysisch-chemische en (eco)toxicologische gegevens bekend zijn. Het gaat dan vooral om de ontwikkeling van strategieën waarvoor testmethoden op celniveau (*in vitro*) en wiskundige testmethoden (*in silico*) gebruikt worden. De wiskundige methoden zijn nodig om de blootstelling aan stoffen en de relatie tussen effecten hiervan en chemische structuur in te kunnen schatten. Daarnaast blijft proefdiergebruik (*in vivo*) in sommige gevallen nodig. Door de resultaten van al deze methoden op slimme wijze aan elkaar te koppelen kan met voldoende zekerheid inzicht worden verkregen in de effecten van chemische stoffen. En daarmee kunnen, naar verwacht, stoffen met minder proefdieren, goedkoper en sneller op hun veiligheid worden beoordeeld.

Het rapport beschrijft hoe in elke stap van de teststrategie zo verantwoord mogelijk kan worden omgegaan met de onzekerheden in de gebruikte resultaten van testen en methoden. Vervolgens worden teststrategieën besproken voor de beoordeling van 1) de afbraak van chemicaliën in het milieu, 2) overgevoeligheid van huid en ademhalingswegen en 3) nadelige effecten op de vruchtbaarheid en op het nageslacht.

Trefwoorden: ITS, geïntegreerde teststrategie, risico, chemicaliën, weight-of-evidence

Contents

Summary

This report is a contribution towards EU and OECD research programs to reduce the use of experimental animals in testing strategies for the risk assessment of chemical substances. The focus is on the application of alternative and improved test methods. This research is highly wanted for the successful implementation of the new EU legislation for industrial chemicals REACH. REACH will be implemented in June, 2007. The ultimate aim of REACH is to improve the protection of human health and the environment through a better and earlier identification of the properties of chemical substances. Within the scope of REACH the use of alternative methods will become an important element in the Classification and Labelling, the assessment of PBT properties (Persistent, Bioaccumulative, Toxic) and risk assessment of chemicals both for human and environmental safety. This means that alternative methods (non-testing methods or *in vitro* tests) have to be developed that allow regulatory decisions to be made. These have until now been used but to a varying degree and in different ways for classification and labelling, risk assessment and PBT assessment of chemicals.

For this reason the European Commission initiated a REACH Implementation Project (RIP 3.3), which should provide a general decision making framework with Integrated (or Intelligent) Testing Strategies (ITS) for specific endpoints. Hence, information from alternative methods should be combined in the decision-making process that leads to the hazard assessment of the substance. At present, sufficient knowledge of alternative methods and their application in integrated testing strategies is lacking, both in- and outside RIVM. RIVM therefore, in parallel with RIP 3.3, initiated a 2-year project with the aim to build capacity in this area within RIVM. Secondly, it would be explored which alternative methods and ITS can be improved or developed, providing input to international work. This project is further supported by an ITS project for VROM (Netherlands Ministry of Housing, Spatial Planning and the Environment) and this report reflects results of both.

This report gives the background of ITS and a general outline of alternative methods and integrated testing strategies. ITS will benefit from a robust decision-analysis framework which deals transparently with uncertainties in each step of the Weight-of-Evidence (WoE) procedure in a particular ITS. This subject has been explicitly investigated. Subsequently, the results of the use of alternative methods and ITS for environmental degradation are presented as an element in the environmental risk assessment. Aspects of the use of alternative methods and ITS for sensitisation and reproductive toxicology as important elements in the human safety assessment which require a lot of experimental animals have also been explored.

WoE
Different methods for a qualitative and quantitative WoE are discussed. At present, there is no consensus method for applying a WoE approach to testing strategies in REACH. The WoE methods should allow the assessor to decide whether the information is good enough to give information for REACH on hazard, for classification and labellingLabelling and for PBT/vPvB assessment if applicable. Additional research efforts should clarify which WoE methods are most suitable. In this report, a Bayesian framework is explored. Prior knowledge about chemicals can be used in a WoE approach in a Bayesian framework and this approach has been explained in this report and demonstrated for biodegradation data. The advantage of the Bayesian analysis is that, once familiar with the terminology and notation, relatively simple calculations can demonstrate the influence of the prior knowledge on the outcome of the model predictions. The consequences of the additional information can then be evaluated

for increased certainty on the outcome of the test. This, however, emphasizes that the user of the information needs to decide which certainty is needed (or: uncertainty accepted), since the outcome of the analysis is expressed as probabilities.

Environmental degradation
In different regulatory frameworks, prediction and understanding of the fate of the chemicals are essential so that measures can be taken to avoid effects on humans and the environment. The current status of QSAR applications for environmental degradation with regard to Classification and Labelling, PBT assessment and risk assessment has been described. The report further concentrates on the usefulness of CATABOL as a tool for the assessment of biodegradation and biodegradation products within the scope of an ITS. CATABOL is able to predict identity and amounts of metabolites, which then can also be used in an ITS as currently used by Health Canada. In order to assess its extrapolation capability, CATABOL was subjected to a study with several groups of chemicals. The majority of the substances appeared to be out of the applicability domain. Nevertheless, this study confirms earlier findings that CATABOL is well able to predict the 'ready biodegradability' of chemicals. A comparison with experimental data for a number of pesticides and existing chemicals however reveals that the predictability of the metabolites formed is still limited. Nevertheless, CATABOL could still be useful in a weight of evidence approach or for targeting further testing.

Sensitisation
The section of sensitisation reflects the discussions in the RIP 3.3-2 Endpoint Working Group on ITS for this endpoint up to the end of 2006. The value of data which may be available and their contribution towards an ITS for sensitisation has been discussed. A tiered ITS approach consisting of non-animal methods, and possibly including *in vivo* testing as a last resort for a very small percentage of the total amount of chemicals, may be realistic for the near future. Given lack of available (Q)SARs and *in vitro* tests for respiratory sensitisation, it is not possible to provide any additional guidance on the evaluation of non-testing data for respiratory sensitisation. Therefore, the development of models and testing guidance for respiratory sensitisation should be stimulated.

Reproductive toxicology
A study is described focusing on the added value to risk assessment of the two-generation reproductive toxicity study when a subchronic study is available, the impact on risk assessment of the second generation within the two-generation reproductive toxicity study and the added value to risk assessment of the rabbit developmental study when a rat developmental study is available. The study will, among others, provide information on the proportion of reproductive toxic substances that would not be classified if a two-generation study would be missed, or information on the differences between the NOAELs obtained in a two-generation study versus a subchronic study. The results obtained from these analyses will be published in peer-reviewed journals.

1. Introduction

1.1 Regulatory background

Around 100,000 different substances are registered in the EU, of which around 30,000 are manufactured or imported in quantities above 1 tonne. The existing regulatory system inherent in current EU policy for dealing with the majority of these chemicals - known as 'existing' substances - has been in place since 1993 and has prioritised 140 chemicals of high concern. Although a programme of work has been drawn up, the current EU legislation on chemicals has several drawbacks. Firstly, a substantial number of existing chemicals which are marketed have not been adequately tested. Information related to their hazard potential is minimal (less than base-set), and they may be harmful to human health or the environment. This contrasts sharply with new chemicals which have to be notified and tested starting from volumes as low as 10 kg per year, discouraging research and invention of new substances. Secondly, there is a lack of knowledge on (mainly downstream) use and exposure. Thirdly, the present process of risk assessment and chemical management in general is relatively slow, and certainly too ineffective and inefficient to take care of the problem raised by the huge data gap in the field of the existing chemicals. And last but not least, the current allocation of responsibilities is not appropriate: public authorities are responsible for the risk assessment of substances, rather than the enterprises that produce or import them (JRC, 2005).
For this reason, the Commission proposed a new EU regulatory framework for the Registration, Evaluation, Authorisation and restriction of Chemicals (REACH) (EC, 2006) which covers both new and existing chemicals, and replaces approximately forty existing Community Directives and Regulations by one single regulation. The ultimate aim of REACH is to improve the protection of human health and the environment through a better and earlier identification of the properties of chemical substances. The basic elements of REACH are as follows:
Registration - In principle REACH covers all substances, but some classes of substances are exempted (e.g. radioactive substances, polymers and substances for research and development). The safety of substances is the responsibility of industry. Manufacturers and importers of chemicals are therefore required to obtain information on their substances in order to be able to manage them safely. The extent of the obligations depends upon the quantity of the substances manufactured or imported. For quantities of 1 tonne or more per year a complete registration has to be submitted. For substances of 10 tonnes or more per year, a chemical safety report (CSR) has to be included. Since one of the goals of REACH is to limit vertebrate testing and reduce costs, sharing of data derived from *in vivo* testing is mandatory.
The information on hazards and risks and how to manage them is passed up and down the supply chain. The main tool for downstream information is the safety data sheet (SDS), for dangerous substances only. A SDS contains information which is consistent with the chemical safety assessment. Relevant exposure scenarios are annexed to the SDS. The downstream user is required to apply appropriate measures to control risks as identified in the SDS.
Evaluation - Evaluation will be performed on registration dossiers, to check the testing proposals and the compliance with the requirements of registration. In addition, substances which are suspicious of being a threat to human health or the environment can be evaluated by a Member State.

Authorisation - Authorisation of use and placing on the market is required for all substances of very high concern (CMR, PBT, vPvB[1]), regardless of tonnage level.
Restrictions - Restrictions may apply to all substances, regardless of tonnage level.
Classification and Labelling inventory - Directives 67/548/EEC on Classification and Labelling of substances and 1999/45/EC on Classification and Labelling of preparations will be amended to align them with REACH.

1.2 The use of alternative methods within REACH

The consequence of REACH is that in a relative short time period the risk of a large group of chemicals has to be assessed, which implies that also a large amount of information on the fate and effects of chemicals has to become available. In principle, this can be achieved by conducting a large number of human toxicity and ecotoxicity studies as well environmental fate and behaviour studies. However, not only in REACH but in OECD as well, there is understanding that for reasons of animal welfare, costs and logistics, it is important to limit the number of tests to be conducted. In line with ANNEX XI of the REACH proposal, the generation of a comprehensive test dataset for every chemical will not be needed if these test data can be replaced by the following methods (for definitions: see section 2.1):

- Non testing methods:
  - The application of grouping (categories) and read-across
  -Computational methods (SARs, QSARs and biokinetic models)
- *In vitro* tests
- Existing experimental and historical data
- Substance-tailored exposure driven testing
- Optimised *in vivo* tests
- Weight-of-evidence (WoE) based on several independent sources of information.

This means that alternative methods (non-testing methods and *in vitro* tests) have to be developed as well as weight-of-evidence schemes that allow regulatory decisions to be made (Pedersen, 2003; Van der Jagt, 2004). These have until now been used but to a varying degree and in different ways for risk assessment, classification and labelling, and PBT assessment of chemicals. The benefits of using such non-testing methods can be claimed to include:

- *Avoiding the need for (further) testing*, i.e. information from non-testing methods has been used to *replace test* results.
- *Filling information gaps*, also *where no test* would be *required* according to current legislation
- *Improving the evaluation of existing test data* as regards data quality and for choosing valid and representative test data for regulatory use. Furthermore, use of non-testing data in addition to test data employing weight-of-evidence could increase the confidence in the assessments.

Thus, the use of non-testing information may improve the basis for taking more appropriate regulatory decisions (as well as for voluntary non-regulatory decisions taken by industry). In fact, use of non-testing information may decrease uncertainty, or even make it possible to

[1] CMR = classification of a substance as carcinogenic, mutagenic, reprotoxic category 1 and 2 according to Directive 67/548/EEC. PBT = persistent, bioaccumulative, and toxic according to REACH criteria. vPvB = very persistent, very bioaccumulative according to REACH criteria.

conclude on a classification or the need for more information in relation to hazard, risk and PBT assessment. Figure 1.1, taken from a report of a REACH Implementation Project, shows how these alternative methods could fit in the overall framework of the information strategy of the REACH proposal.

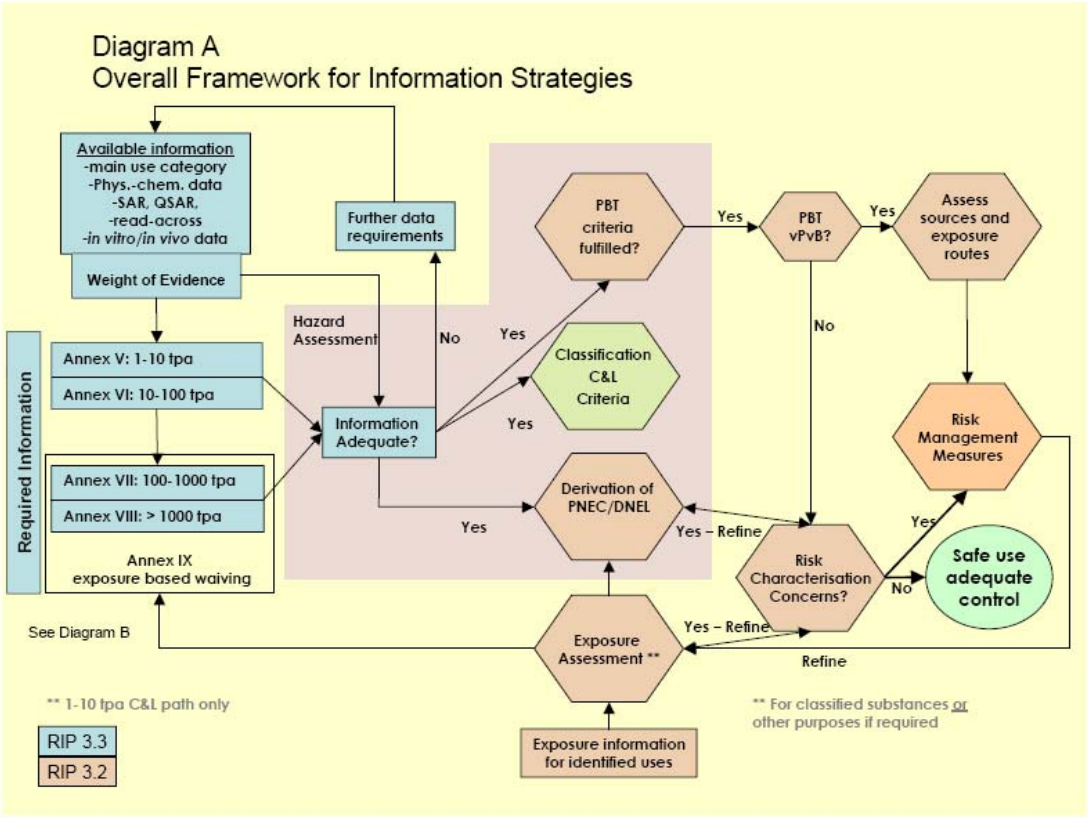


Figure 1.1: Overall framework for information strategies (TAPIR, 2005)[1].

Alternative methods are in several stages of development, verification and validation (Worth and Balls, 2002) and therefore often cannot be used as stand alone. Other information gaps will exist. It is therefore necessary to integrate all available information into so called integrated or intelligent testing strategies (ITS). In this way, all possible available information on a substance can be optimally used and further testing will only be required where essential information is lacking (Combes et al., 2003; Bradbury et al., 2004). ITS are nothing new: such strategies have been developed for both Classification and Llabelling and risk assessment in various regulatory frameworks (see for example EC, 2003 and 2004). However, so far the integration of alternative methods in these ITS is limited. A major challenge under REACH will be to develop a set of strategies that contain the necessary level of detail (e.g. objective decision rules for appropriate endpoints) for fulfilling the information requirements of REACH, while remaining consistent with the general framework laid down by the REACH proposal. For this reason, the European Commission initiated a REACH Implementation project (RIP 3.3) in 2005, which should provide guidance on the reasoned justification for asking derogations/waiving (or adaptations) from the standard testing regime by using a general decision making framework with ITS for specific endpoints. This project

[1] Annex IX in this diagram, is Annex XI in the current REACH proposal (EC, 2006).

will be finalized early 2007. Hence, information from alternative methods should be combined in the decision-making process that leads to the hazard assessment of the substance. It should be emphasized that other regulatory frameworks such as for food and feed additives, plant protection products, biocides and cosmetics will also benefit from this work.

1.3 Objectives

Within the scope of REACH the use of alternative methods will become an important element in the classification and labelling, PBT and risk assessment of chemicals both for human and environmental safety. At present, sufficient knowledge of alternative methods and their application in integrated testing strategies is lacking, both in- and outside the National Institute for Public Health and the Environment (RIVM). In 2005 therefore, parallel to RIP 3.3, a 2-year project was initiated at RIVM aiming at capacity building by investigating and analyzing existing alternative methods and integrated testing strategies. Because similar activities were being conducted in a project carried out for VROM, these projects were merged in order to handle more topics. The second aim was the improvement and development of ITS for selected topics. Following an exploratory phase in year one, a selection was made for further research on WoE methods and on ITS for the endpoints environmental degradation, sensitisation and reproduction toxicology. The project explicitly aimed to promote the cooperation with (inter)national bodies involved in the development and validation of alternative methods and ITS and to establish firm ground for further work of RIVM in this area.

This report addresses general aspects of ITS and in particular WoE methods. It also shows the results of the research into ITS for the endpoints environmental degradation, sensitisation and reproduction toxicology.

1.4 Readers guide

Chapter 2 gives a general outline of alternative methods and their integration into ITS. Next, chapter 3 discusses the general decision framework and WoE procedures needed for ITS. Chapter4 explores aspects of the use of alternative methods and ITS for environmental degradation, as a key element in the environmental risk assessment. The current status of QSAR applications for environmental degradation has been described. The report further concentrates on the usefulness of CATABOL as a tool for the assessment of biodegradation and biodegradation products within the scope of an ITS. In chapters 5 and 6, the use of alternative methods and ITS in human safety assessment for sensitisation and reproduction toxicology, respectively, is explored. The section of sensitisation reflects the discussions in the RIP 3.3-2 Endpoint Working Group on ITS for this endpoint up to the end of 2006. In chapter 6 a study is described focusing on the added value to risk assessment of reproductive toxicity tests. Whereas each of the chapters 2 to 6 will provide conclusions and recommendations, chapter 7 will present the overall conclusions and recommendations of this project.

2. Alternative methods and ITS

2.1 Non-testing methods (QSARs, SARs, read across, chemical category)

The principles of development and use of non-testing methods are based on the expectation that structurally similar chemicals will have similar physical attributes and biological effects. This underlying premise of similarity could be used in hazard and risk assessment when there are inadequate test data to estimate missing values. These non-testing methods include SARs, and QSARs and grouping approaches, including read-across and chemical category approaches.

2.1.1 (Q)SARs

Structure-activity relationships and quantitative structure-activity relationships, collectively referred to as (Q)SARs, are theoretical models that can be used to predict the physicochemical and biological properties of molecules. They are sometimes called '*in silico* models' because they can be applied by using a computer (Joint Research Centre, 2005).
A structure-activity relationship (SAR) is a (qualitative) association between a chemical substructure and the potential of a chemical containing the substructure to exhibit a certain biological effect. A quantitative structure-activity relationship (QSAR) is a mathematical model that relates a quantitative measure of chemical structure (e.g. a physicochemical property) to a physical property or to a biological effect (e.g. a toxicological endpoint) (Joint Research Centre, 2005).
Further explanation of the concepts of (Q)SAR can be found in (Nendza and Hermens, 1995).
In principle, (Q)SARs can be used to provide the following types of information which may be useful for regulatory purposes:
1. physicochemical properties
2. toxic potential and potency
3. environmental distribution and fate
4. biokinetic processes

2.1.1.1 QSAR models and application criteria

Under the current EU legislation for new and existing chemicals, the regulatory use of (Q)SARs is limited and varies considerably among the Member States. This is due to the fact that there is no agreement in the scientific and regulatory communities over the applications of (Q)SARs and the extent to which (Q)SAR estimates can be relied on. According to the ECETOC Task Force on (Q)SARs (ECETOC, 2003), the quality of models used for human endpoints 'is often poor because the endpoints are expressed through many different mechanisms, are receptor-mediated, involve multi-stage processes comprising ADME and are site-specific. At the present time, this complexity imposes severe limitations on the successful development of (Q)SARs suitable for non-congeneric sets of endpoints.'

Nevertheless, there are some QSARs which are already part of the EU legislation on chemicals, like the QSAR for bioconcentration of chemicals in fish (BCF). In addition, QSARs have been used by EU regulators to identify PBT and vPvB substances. Furthermore, QSAR estimates have sometimes been accepted by the EU when participating in the OECD Existing Chemicals Programme, in cases where experimental data was lacking (via the contribution of the USA to the OECD Existing Chemicals Programme, where the USA

applied read across or QSARs). So there is already regulatory implementation and acceptance of QSARs in the EU Member States, but further work needs to be done in the area of human health related toxicological endpoints. In November 2004, the OECD member countries agreed on the principles for validating (Q)SAR models for their use in regulatory assessment of chemical safety. (Q)SARs have been used in regulatory assessment of chemical safety in some OECD member countries for many years, but universal principles for their regulatory applicability are lacking. The agreed principles provide member countries with basis for evaluating regulatory applicability of (Q)SAR models and will contribute to their enhanced use for more efficient assessment of chemical safety (OECD, 2004a). OECD also summarized the experiences of OECD member countries in the regulatory use of (Q)SAR models in chemicals assessment (OECD, 2006). A report of the European Chemicals Bureau provides preliminary guidance on how to characterize (Q)SARs according to the OECD validation principles (Worth et al., 2005). Further OECD/EU guidance is currently being developed further within the scope of Reach Implementation Project 3.3.

2.1.1.2 Limitations of QSARs

In the past, non-testing methods for the regulatory purposes have not been used to provide definitive stand-alone information. Instead, they have been largely confined to screening or prioritizing chemicals for testing, and to supplementing existing animal test data. Validation, limited applicability domain, and poor availability of guidance are the major limitations of these approaches for regulatory testing.

2.1.1.3 Validation of QSARs

Validation of (Q)SAR models is still essential for their regulatory use. The OECD Principles for validation of QSAR models (OECD, 2004a, see 2.1.1.1) are indispensable for the assessment of the validation status and its regulatory applicability. These principles are listed in Table 2.1.

Table 2.1: OECD Principles for validation of (Q)SAR models (OECD, 2004a)

Principle	Explanation
1. A defined endpoint	Endpoint refers to any physico-chemical property, biological effect, environmental fate parameter
2. An unambiguous algorithm	Ensures transparency in the description of the model algorithm
3. A defined domain of applicability	Defines limitations in terms of types of chemical structures, physico-chemical properties and mechanisms of action for which models can generate reliable predictions
4. Appropriate measures of goodness-of-fit, robustness and predictivity	Information needed on 1) the internal performance of the model determined by using a training set, and 2) the predictivity of the model, using an appropriate test set
5. A mechanistic interpretation, if possible	Assessment of mechanistic associations between the descriptors used in the model and the endpoint being predicted

2.1.1.4 Applicability domain of QSARs

Many (Q)SAR models have been developed and used principally with respect to modelling congeneric series of chemicals. Thus, these models have a limited applicability domain. This limitation could be solved by the development and use of batteries of models that cover complementary parameter spaces; new models employing available and new statistical approaches like iterative model development to increase the size of the applicability domain; and expert computational prediction systems based on a series of rules that cover a wide applicability domain. For reviews of methods for estimating the applicability domains of QSARs: see Jaworska et al.(2005) and Netzeva et al. (2005).

2.1.2 Read-across and chemical category

Grouping approaches are strongly linked to SAR concepts. Annex XI of the draft REACH regulation (EC, 2006) defines grouping approaches as follows:

'Substances whose physicochemical, toxicological and ecotoxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity may be considered as a group, or 'category' of substances. Application of the group concept requires that physico-chemical properties, human health effects and environmental effects or environmental fate may be predicted from data for a reference substance(s) within the group by interpolation to other substances in the group (read-across approach). This avoids the need to test every substance for every endpoint. The similarities may be based on:

- a common functional group,
- the common precursors and/or the likelihood of common breakdown products via physical and
- biological processes, which result in structurally similar chemicals, or
- a constant pattern in the changing of the potency of the properties across the category.

If the group concept is applies, substances shall be classified and labeled on this basis.

In all cases results should:

- be adequate for the purpose of classification and labelling and/or risk assessment.
- have adequate and reliable coverage of the key parameters addressed in the corresponding test method …..
- cover an exposure duration comparable to or longer than the corresponding test method….if exposure is a relevant parameter, and
- adequate and reliable documentation of the applied method shall be provided.'

Qualitative read-across involves the identification of a chemical substructure that is common to the two substances and the assumption that the presence or absence of a property for a substance can be inferred from the presence or absence of the same property for an analogous substance. Quantitative read-across involves the identification of a chemical substructure that is common to the two substances and the assumption that the known value of a property for one substance can be used to estimate the unknown value of the same property for another substance (TAPIR, 2005).

The main distinction between read-across and chemical category is that the former approach will normally be performed between one data-rich substance and a substance for which limited data are available. In the category approach similarity of a pattern for several chemicals will be evaluated. Read-across can be one tool to do this, but interpolation and extrapolation and (Q)SARs will also be considered to do this trend-analysis (Rila et al., 2006). Both approaches can be used to assess physicochemical properties, (eco)toxicity and

environmental fate. Based on experience in the use of read-across data under NONS the UK-HSE (Hanway, 2002a,b) has developed a strategy that, in their opinion, can be used to find scientific justification for using toxicological read-across data. In a recent RIVM project (Rila et al., 2006), both guidance documents have been applied on groups of chemicals (phthalates, butanes, aliphatic hydrocarbons) to assess a number of human and environmental endpoints. One of the limitations of current guidance is that it uses only qualitative and not quantitative guidance for deciding whether a category is robust. This means that the decision remains expert judgment. Quantitative data are normally needed for risk assessment purposes. With respect to the UK-HSE guidance the following points need attention: (1) the group to which the chemicals belong should be indicated, (2) the similarity of the 2D and 3D structures should be indicated, (3) whether the group should be assessed in an increasing or decreasing order or whether the chemicals should be considered as isomers should be indicated, (4) the expected metabolism/environmental transformation of the different structures should be described.

It s noted that validation of grouping approaches is not explicitly mentioned as requirement in REACH. As demonstrated above, these approaches heavily rely on expert judgment, which should be documented carefully. Guidance on the formation and use of chemical categories for fulfilling data requirements has been published by the OECD as part of the OECD Manual for Investigation for HPV chemicals (OECD, 2004b) and are currently further developed within the scope of REACH Implementation Project 3.3.

2.2 In vitro studies (including ex-vivo)

In vitro tests are based on the use of subcellular fractions, as well as cell and tissue cultures maintained for varying periods of time. The development of *in vitro* tests for different endpoints is being carried out by a wide range of research activities, and different methods are at different stages of standardization, documentation and validation (Joint Research Centre, 2005).

Recently, Groen et al. (2005a) made an inventory of alternatives in the field of *in vitro* testing methods aimed at effects observed after acute and short term exposure (acute systemic toxicity, skin and eye irritation/corrosion and skin sensitisation) and some reproductive endpoints. With respect to **acute toxicity** it was concluded that with the current *in vitro* methods a large reduction of animal use cannot be foreseen in the near future. For this purpose an ITS has to be developed comprising the use of physicochemical data, *in vitro- in vivo* data, computational methods, basal toxicity assays and complementary assays (for metabolism, transport, kinetics and target organ toxicity).
Alternative methods for **skin corrosion**, like the TER assay, human skin model assays (EPISKIN and Epiderm), have been validated and are already broadly accepted for regulatory use.

With the current **skin irritation** methods available a reduction in the number of test animals is possible to a limited extent, as corrosive substances should not be tested for irritation. Considering the stringent requirements for classification and labelling the use of non-validated methods is difficult if not impossible for that purpose, except for those compounds which are clearly irritating and may need no further testing. The forthcoming ECVAM validation study may determine whether the human skin model assay (e.g. Epiderm and EPISKIN) can adequately distinguish acute skin irritants form non-irritants for classification and labelling purposes.

The most commonly used and most developed test for **eye irritation** based on isolated organs are the bovine corneal opacity and permeability (BCOP) test, the isolated rabbit eye (IRE) test and the chicken enucleated eye test (CEET). Organotypic methods are the Hen's egg test chorio-allantoic membrane (HET-CAM assay) and the chorio-allantoic membrane vascular assay (CAMVA). Furthermore, there are numerous human corneal epithelium models (EpiOcular assay, HCE model) and cell based cytotoxicity methods (neutral red uptake, RBC haemolysis test). With these methods available, although not completely validated, the *in vivo* eye study can be waived for these substances that cause severe effects. For classification and labelling the HET-CAM test is used provided that the chemical under investigation belongs to classes of chemicals for which the model has been shown relevant (Rila et al., 2005).
The *in vitro* tests for **skin sensitisation** are described in more detail in chapter 5, and are all non-validated. Rila et al. (2005) conclude that replacement of current *in vivo* methods by a single *in vitro* test is currently not possible due to the complexity of the mechanism of skin sensitisation. However, combination of several *in vitro* tests, covering all relevant mechanistic steps of skin sensitisation, into a battery can likely lead to replacement of the vivo tests. For this reason this endpoint was selected for further development within the current project (see chapter 4).

In the area of **reproductive toxicology**, it is not possible to model the whole of the reproductive cycle *in vitro* with one approach, the components need to be studied individually and then integrated into testing strategies. Progress in this area will depend amongst others on the outcome of the ReProTect project with is an integrated project in the 6th Framework Program that has recently started.

2.3 Optimisation of in vivo testing

As concluded by OECD and ECVAM the opportunities for streamlining individual assays are very limited but *in vivo* testing can be made more efficient by: a) only performing tests that provide relevant data; b) eliminating redundant tests; c) using one sex; d) applying some tests simultaneously to the same animals; and e) making greater use of screens and preliminary testing. This view has also been put forward by ECVAM (JRC, 2005). One has to realise that loss of sensitivity may be a consequence of these efficiency measures. On the other hand, if one pursues the same sensitivity, more tests, e.g. screening tests, will be required which may not result in the reduction of use of experimental animals desired. With regard to testing only one sex: this is allowed if one either knows there is no sex difference or one is convinced that the most sensitive sex is included in the test. Note that for endpoints such as sensitisation and irritation no discrimination between sexes is made.

Initiatives in the areas of systemic toxicity, sensitisation and ecotoxicity, have been taken to explore and implement **reduction** strategies. For example, retrospective analysis of the ecotoxicological data for chemical substances in the New Chemicals Database revealed that reduction between 55-70% in the number of fish used for acute fish toxicity testing would be feasible (Jeram et al., 2005). A detailed discussion on the (regulatory) acceptability of these proposals still needs to take place.

Optimisation of *in vivo* testing also refers to the refinement of animal tests. An example is the replacement of the acute oral toxicity OECD study 401 by three alternative animal tests, the Fixed Dose Procedure (OECD guideline 420,), the Acute Toxic Class method (OECD 423,) and the Up and Down procedure (OECD 425,). All these alternative methods result in a reduction of animal use by two third (Botham, 2004). It is noted that such tests provide

sufficient information for the purpose of classification and labelling, but add little to information needed for the assessment of risks from acute exposure and for estimation of the doses to be applied in repeated dose tests. Dose-range finding studies for expensive repeated dose tests are essential and do require a significant number of experimental animals. Another example is the Local Lymph Node Assay (LLNA) for skin sensitisation (OECD 429), which reduces animal numbers and suffering and has already been implemented into regulation, and to the substitution of lethality as an endpoint in an animal test. The reduced LLNA (rLLNA) uses even less animals, but in contrast to the LLNA it does not provide information on the potentcy.

2.4 Toxicogenomics

Toxicogenomics is defined as a study of the response of a genome to hazardous substances, using 'omics' technologies such as genomic-scale mRNA expression (transcriptomics), cell and tissue-wide protein expression (proteomics), and metabolite profiling (metabolomics), whole organisms responses (metabonomics) in combination with bio-informatic methods and conventional toxicology.

In relation to chemical hazard/risk assessment, this emerging science could provide tools for e.g. improving the understanding of mechanisms of toxicity, identification of biomarkers of toxicity and exposure, reducing uncertainty in grouping of chemicals for assessments, (Q)SARs, inter-species extrapolation, effects on susceptible populations, and possibly provide alternative methods for chemical screening, hazard identification and characterisation (OECD/IPCS, 2005).

The main conclusion from a OECD/IPCS workshop in 2004, focusing on ecotoxicogenomics (including considerations from an earlier IPCS workshop on toxicogenomics in 2003), was that molecular-based approaches for studying the impact of chemicals on human and wildlife populations will have an important role in hazard and risk assessment. However, regarding regulatory purposes toxicogenomic approaches are recognised as not yet developed enough for direct replacement of existing approaches, but it could give supportive evidence on a case by case basis. As such, these advanced technologies hold great promise for future application providing solid experience, proven demonstrations, and sufficient harmonization. Like for biomarkers, these techniques can provide additional evidence of both exposure to and effects of pollutants in individuals, but they need to be linked to changes in growth rates, reproductive output and viability of offspring of organisms if robust predictions of the likely impacts of pollutants are to be made. Environmental risk assessment might also benefit from additional relationships between community responses (those extending responses on individuals and populations) and measures for exposure and effects (OECD/IPCS, 2005).
It has been argued that smart combination of different activities, including a suit of omics-related techniques, classical studies in the field of toxicology and ecotoxicology, high throughput systems, pattern analysis, data management systems, and improved statistics and mathematics will enable the evolution of systems toxicology (Waters and Fostel, 2004). This kind of analysis will eventually allow us to analyze many toxicological interactions in living organisms and ecosystems under stress in an integrative approach (Gant and Zhang, 2005). Although uncertainty reduction can in theory be obtained by such an approach, effective use of these complex data in risk assessment and environmental management is another issue.

2.5 Exposure based waiving (including Threshold of Toxicological Concern)

The principle behind any potential waiving is that there are situations when human or environmental exposures are so low that the acquisition of additional effects information does not necessarily lead to an improvement in the ability to manage risk. In the ANNEXES VI-X of the REACH proposal (EC, 2006) specific rules are presented when standard information may be omitted, triggered, replaced or adapted. Annex IX and X include examples of waiving of certain tests based on exposure criteria. In addition, Annex XI presents the possibility of the waiving of certain effects information in Annex VIII, IX and X based on exposure considerations. The approach is promising, especially if combined with (Q)SAR or read-across, but it requires further investment in the development of exposure models and it also precise information on the use pattern of the chemicals (e.g. downstream use information), which is one of the current bottlenecks.

An example where effects information based on exposure considerations has been incorporated in the legislation include a Community procedure for flavouring substances used or intended for use in or on foodstuffs where the European Food Safety Authority (EFSA) has implemented the concept of exposure based waiving (EBW) and exposure based triggering (EBT). In this approach the concept of the Threshold of Toxicological Concerns (TTC) is being applied in a risk assessment process to justify the waiving of testing for flavouring substances. The TTC concept relies on the assumption that one can identify a concentration threshold below which the risk of any chemical for any harm is acceptably low. The concept that there are levels of exposure that do not cause adverse effects is inherent in setting acceptable daily intakes (ADIs) for chemicals with known toxicological profiles. The TTC principle extends this concept by proposing that a *de minimis* value can be identified for many chemicals, in the absence of a full toxicity database, based on their chemical structures and the known toxicity of chemicals which share similar structural characteristics. This means that if exposure information shows that TTCs will not be reached in the human body, in food or in the environment, this could be used as screening tool to set aside a chemical as being of ‘low concern’ or low priority for testing. If the measured or predicted exposure concentration comes close to the TTC, this could trigger the need to obtain further information on the toxicity of the chemical. The use of the TTC concept could however only be used to limit testing, when adequate information on the use of and exposure to chemicals is available (JRC, 2005; Barlow, 2005).

Overall, the decision to waive the generation of human effect information could be based on:

- Where a substance is used; e.g. when it is restricted to a well-characterised place, or situation with a specific group of workers.
- How a substance is used; e.g. when it is used in closed systems or when a limited amount is used per day, due to the type of use or when it is used in strictly controlled ‘permit to work’ systems with extensive personal protection equipment.
- The intensity in which a substance is used; e.g. infrequent use due to the function of the substance, or the use by only small numbers of workers are exposed which can be adequately protected or the wide spread use of some additives in small concentrations in products.
- The expected exposure route; e.g. an inhalation test could be waived if exposure is only dermal.

- The substance characteristics; e.g. liquid with very low vapour pressure, Solid produced/used in solution or dispersion only or Solid produced as non-abrasive large granules or flakes (e.g.marbles).

With respect to the environment, tests can be waived when information is available that one or more environmental compartments or a specific group of animals are not exposed. Waiving can also be based on the substance property, e.g. showing that the substance is unlikely to cross biological membranes (MW >800 or molecular diameter >15 Å); is highly insoluble (<10 μg/l) or that a substance degrades too fast to cause long-term effects. In case ingestion is not considered to play an important role (e.g. log Kow <5), the equilibrium partitioning approach could be used to derive the PNEC for sediment and soil organisms, without further testing (TAPIR, 2005).

The Working Group of RIP 3.3-1 considered the following general principles appropriate if the acquisition of endpoint-related information is to be focussed, relevant and necessary if exposure based waiving is to be applied it should meet the following principles (TAPIR, 2005):

1. Information specified in Annex VI requirements are the minimum requirements as a starting point for consideration of exposure-informed waiving and triggering.
2. A full understanding of foreseeable conditions of exposure across the substances' use cycle such that these can be described, with confidence, as being 'negligible'. For consumer uses, this understanding would extend to reasonably conceivable unintended uses. Workplace exposures would include those occurring during maintenance activities (but not those activities controlled by permit-to-work type systems). The understanding would be sufficient to describe exposure at a detailed level.
3. An understanding of the extent to which different physico-chemical forms and presentations of the substance (if relevant) would affect exposure.
4. A process that is transparent and requires justification for key assertions, particularly measures established to reduce and/or manage exposures or those, which seek to quantitatively describe exposure (other than via prediction using established models).
5. A process that does not recognise the utility of the role that PPE has in controlling exposure, apart from in those situations involving either short-term and/or infrequent exposures and when other control options have been exhausted.
6. A process that utilises a DNEL/PNEC as the basis for evaluating whether risks could be considered as tolerable (and hence which would, in turn, inform the extent to which the acquisition of further effects information was necessary).
7. A process that distinguishes between consumer and worker risks, particularly in those situations where substances are only manufactured/marketed for industrial use.
8. An assumption that considerations of exposure, taken within the context of a risk based decision-making framework, will both inform whether information on an endpoint is appropriate and/or necessary.

If exposure-based arguments are used as a basis for a reduced data set, it is of course essential for registrants to remain aware of this in the years following registration. In particular, any changes in circumstances must be reviewed. This might include changes to the plant and to the process, new users, a new site for production and further processing, the batch size, the number of batches per day, the level of the emissions and the number of days of emission per year.

2.6 Integrated testing strategies

2.6.1 Definition of ITS

An ITS can be defined as follows: 'An Integrated Testing Strategy is any approach to the evaluation of the hazard which serves to reduce, refine or replace an existing animal procedure, and which is based on the use of two or more of the following: physicochemical data, *in vitro* data, human data (for example, epidemiological, clinical case reports), animal data (where unavoidable), computational methods (such as quantitative structure activity relationships (QSARs) and biokinetic models' (Blaauboer et al., 1999). It should be added that the ITS approach for any particular endpoint depends on its goal and on the data requirements of a particular regulatory framework. Important goals are classification and labelling, PBT of vPvB assessment and risk assessment. ITS are hierarchical in nature starting by making maximum use of existing effects and exposure data. Key to the resulting decision schemes is the Weight-of-Evidence process to be followed which should be as explicit as possible in order to determine the uncertainty in their outcome.

2.6.2 Weight-of-evidence (WoE)

In determining whether data are adequate to draw conclusions on endpoints, a weight of evidence (WoE) approach is recommended. The WoE approach is closely linked with the issue of uncertainty. WoE can be defined as 'a decision making activity, often by an expert able to integrate all aspects of uncertainty (about data quality, model uncertainty, etcetera). The weight associated to each evidence (fact) discussed is simply the subjective probability that facts is T*rue*' (Appendix 3 of TAPIR, 2005). WoE is closely linked to 'testing strategies' in that the available evidence can help to determine the subsequent testing steps. WoE is also recommended by OECD in the SIDS programme and is given prominence in the GHS for classification purposes (TAPIR, 2005). A recent example of a WoE approach has been provided by Health Canada in a framework for carcinogenicity and mutagenicity (TERA, 2005). This framework considers three lines of evidence, empirical data, QSAR model predictions and SAR model predictions, and these are weighted based principally on predictive power of the relevant or underlying bioassays. The WoE concept will be further explored in Chapter3.

2.6.3 An example of an ITS

(Blaauboer et al., 1999) presented an integrated decision strategy for evaluating the human toxicity of chemicals on the basis of their structure, *in vitro* toxicity data, and biokinetic modelling. Their scheme is subdivided in four stages (see Figure 2.1). Stage I focuses on the use of *in silico* methods, (QSARs, including physicochemical properties and existing knowledge). Stage II concentrates on the use of *in silico* pharmacokinetic models optimized and added with information from *in vitro* tests for e.g. acute toxicity, metabolism. In stage III the focus is on tissue-specific toxicity. Only at stage IV, if required, limited *in vivo* tests would be applied. According to Groen et al. (2005b), this tiered approach may have its advantage for hazard identification and characterization of chemicals. With the present status of potential alternatives, especially regarding validation and predictability for the *in vivo* situation, the alternatives are not applicable, yet, without compromising the protection of human health. It is also essential to consider whether the total strategy covers all aspects of the endpoint or only a part. In the latter case an *in vivo* test must be the logical last step unless a lower safety level is accepted.

Based on the evaluation scheme presented by Blaauboer et al. (1999), Groen et al. (2005b) explored the practical feasibility of actually predicting *in vivo* dose levels (associated with no or small adverse effects) from *in vitro* test results. Basically, two approaches were tested, i.e:

- An empirical approach describing *in vitro – in vivo* (IVIV) relationships;
- A mechanism-based approach linking *in vitro* data to *in vivo* data by describing the kinetics of a compound both in the *in vitro* test system as well as *in vivo*.

As both *in vitro* and *in vivo* data were available from an ECVAM validation study for the whole embryo culture embryo toxicity test (WEC test), this *in vitro* system was used for the pilot study on *in vitro* - *in vivo* extrapolation.For the empirical approach a so-called Bench Mark Concentration (BMC) was deduced for 7 compounds in the WEC model. Parallel a so-called Bench Mark Dose (BMD) was deduced for the same compounds on the basis of available data from *in vivo* teratogenicity studies. Subsequently, it was tested whether a relationship between BMC and BMD could be observed.
For the mechanism based approach, a PBPK model is necessary for translating *in vitro* toxicity concentrations/doses (Bench Mark Concentration or BMC) into corresponding *in vivo* toxicity doses (Bench Mark Dose or BMD). In essence, the PBPK model helps to translate which dose in the mother will lead to a certain concentration/dose in the embryo, either *in vitro* as well as *in vivo*. In Figure 2.1, the specific areas of interest of the project are indicated in the scheme as used by (Blaauboer et al., 1999).

The preliminary analysis indicated that both approaches have potential for application into risk assessment. The BMC from the WEC test seemed to predict the *in vivo* BMD quite well. Nevertheless, it needs further discussion whether it is feasible to implement this approach in risk assessment and regulatory settings. The mechanism based approach was more complicated and for this situation did not lead to added value for the risk assessment. However, PBPK modelling might have benefits over the empirical approach when regarded from a regulatory point of view. A PBPK model has the advantage that it is more dynamic than the BMC-BMD as it may serve various toxicological endpoints and various exposure scenarios. The BMC-BMD relationship has to be determined per endpoint and per test system. Moreover it is only valid for a restricted set of exposure scenarios.
The outcome of this analysis makes clear that the best achievable goal of this approach is reduction of the number of animals to be tested for risk assessment. In this regard such approaches are only useful to apply as a part of ITS.

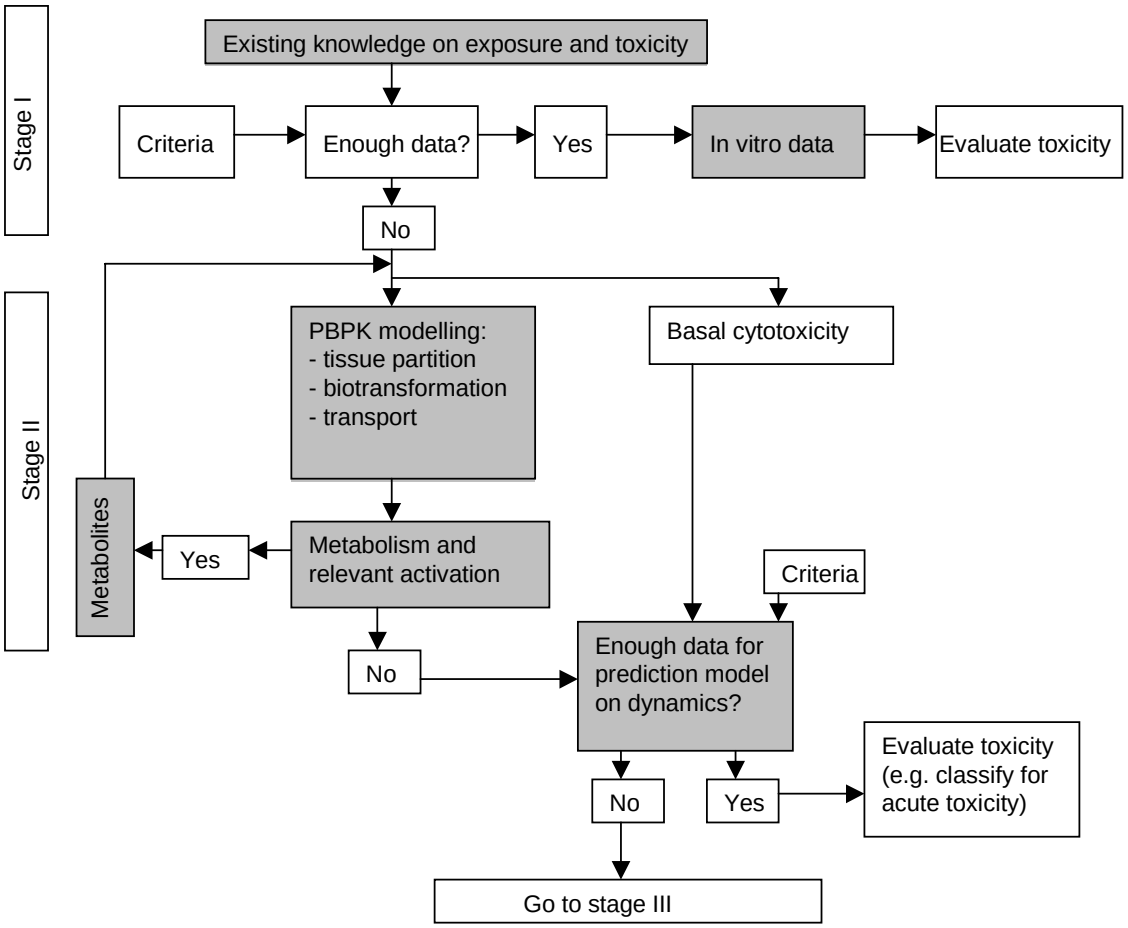


Figure 2.1: General ITS scheme for human toxicology by (Blaauboer et al., 1999). The subjects of the project of Groen et al. (2005b) are indicated by shaded boxes.

3. Decision analysis and Weight-of-Evidence procedures

A Weight-of-Evidence procedure (WoE) is needed to determine whether data are adequate to draw conclusions on endpoints. The combination and interpretation of data, often by an expert, leads to a judgment on the quality of the overall body of information that is present. WoE procedures are not standardised and terminology is not clear. In addition, an explicit discussion of the uncertainty in the data is often lacking. In our view, a WoE approach should be built on a flexible framework that can be used over all endpoints and integrates elements of statistical and decision-making theory.

3.1 Introduction

The 'Weight-of-Evidence' approach (WoE) is often mentioned in the risk assessment literature, without adequate documentation (Weed, 2005). In many cases, it is not clear which methods were used, how they were applied to the scientific evidence, what the results were and how these were used to make decisions in a specific risk assessment. The scientific evidence gathered in the course of a risk assessment is often of a variable nature: strong studies versus weak studies, *in vitro* studies versus *in vivo* studies, animal studies versus human studies etcetera. The WoE approach taken should make it transparent which interpretative methods were used and how these were applied to scientific evidence and expert judgment. First of all, the definition of WoE is used in different ways and both qualitative and quantitative weighting methods are in use. An important issue in WoE is the influence of values on expert judgment that needs to be recognized and made explicit as far as possible.

3.1.1 Use and definitions of WoE

Based on a review of selected PubMed papers from 1994-2004, Weed (2005) deduced the following three uses of the phrase 'weight of evidence'.

3.1.1.1 Metaphorical use of WoE

In many cases, WoE in a risk assessment is used as a metaphor for a synthesis or interpretation of the available evidence, without going into detail on the interpretative methods or weighting that were used. In such cases, the lack of transparency on the WoE prevents insight in the weighting of the evidence; it is not clear what role is assigned to the various types of available information and expert judgment.

3.1.1.2 WoE means reviewing all available evidence

In many guidelines for risk assessment (EPA, EU), the hazard assessment contains a summary of the results of the individual lines of evidence to provide a conclusion on a specific endpoint. In some papers, WoE refers to a situation where all available evidence is used, both the positive or significant results and the negative or non-significant results, whether from standardized tests or not. It is often not specified what is meant by 'all evidence', even if this is explicitly mentioned as opposed to using some evidence. There may have been a selection process nonetheless, e.g. on quality, peer review or other standards.

3.1.1.3 WoE refers to accepted methods of summarizing and interpreting science

Existing methods for summarizing scientific data are often used in WoE approaches, where qualitative and quantitative methods can be used and weighted in various ways. Some of these methods are systematic narrative reviews, causal inference (based on statistically significant results) or statistical methods for weighting different lines of evidence. However, expert opinion is often used as well without clear criteria for weighting this against the available information.

3.1.1.4 Conclusions on a definition of WoE

The following conclusions can be drawn in order to better define WoE:

- It should be documented what type of information is assessed and why (all, a subset or selected set) and which quality criteria are used.
- The interpretative methods should be transparent.
- The weighting procedure (e.g. qualitative vs. quantitative evidence) should be transparent.
- A separation between scientific evidence and (value driven) expert judgment is needed.

3.1.2 Methods for qualitative or quantitative WoE

WoE methods are linked to current practice in a certain scientific field. The use of epidemiological evidence is clearly more prevalent in medical science than in environmental science. Some general methods are available that are applicable in any field of science.

3.1.2.1 Establishing causal relationships

One of the challenges of WoE is to combine evidence from *in vitro, in vivo* and field or epidemiological observations. Establishing causality should be preceding any following statistical weighting of various lines. However, it may need to be combined with statistical techniques to quantify increased plausibility of observed effects if multiple lines of evidence are present. The considerations here strongly depend on the testing strategy currently being outlined in each Endpoint Working Group under REACH, but should consider what essential types of information can be derive from each type of information. An example of such considerations is given below (modified after Proctor et al., 2002).

- *In silico* estimation methods (e.g. (Q) SARs) or other statistical methods
    - Structural similarity with other chemicals
    - Mode of action similarity
- *In vitro* estimation methods or biomarkers
    - Structural similarity with other chemicals
    - Mode of action similarity
    - Comparable toxicology or metabolism in other species
    - Severity of effects, dose-response relationship
- *In vivo* estimates
    - Consistent results when multiple studies are available
    - Consider routes of exposure and exposure levels
    - Comparable toxicology or metabolism in other species
    - Causal criteria satisfied (not exhaustive)
        - Temporal relation consistent with cause and effect
        - Statistically significant effect
        - Dose response relation evident
        - Free from bias or confounding factors

    - Biologically plausible
- Field or epidemiological supporting evidence
    - Consistent results when multiple studies are available
    - Consider routes of exposure and exposure levels.

For each of the studies mentioned above, separate quality criteria can be used to assign a reliability score to a study, e.g. scored according to the Klimisch et al. (1997) criteria.

3.1.2.2 Quantitative weighting schemes

Scoring weighting schemes
Once causal criteria and quality criteria have been established, and the steps in a testing strategy are defined, a score can be assigned to each study and step in the procedure, thus making a relative ranking possible between chemicals and between the steps in an ITS. Such a relative ranking approach was used by Calabrese et al. (1997) and Menzie et al. (1996). Scoring methods are easy to apply and can take into account different types of information or criteria. However, there is also a number of disadvantages associated with it:

- The ranking is relative. If there is little comparison with other cases, what is a 'good' score, in other words: when does the WoE leads to a score that will allow decision making?
- The list of test properties and criteria should always be scored in the same way if a fair comparison is expected. However, scoring always entails some qualitative judgements on quality or compliance with criteria.

Probabilistic weighting schemes
Probabilistic weighting of different types of information (e.g., expert judgement and test results) is possible using classical or Bayesian statistical frameworks. To use these methods properly, statistical information about the predictability of routine tests should be available, as generated in small or large scale test verification schemes (Eriksson et al., 2003). Of course, such statistical information would have to be generated for new tests or alternatives.

Bayesian statistics allows the weighting of prior information (including expert information) and information from testing. The interesting notion here is that this method allows successive updating of the prediction probability if new, additional information is found or generated. This means that in an ITS, the result of sequentially generating new information will show if the confidence has increased or decreased (Eriksson et al., 2003). More details are given in the TAPIR report of RIP3.3 phase 1 (TAPIR, 2005), and are elaborated in section 3.3.

3.1.3 Conclusions on WoE approaches

At present, there is no consensus method for applying a WoE approach to testing strategies in REACH. An ITS can follow quite different approaches, depending on the available information. In many cases, information from several sources will be available that should be weighted with one of the above mentioned methods. In other situations, a step-wise procedure may be followed where each additional step in the ITS should lead to increased knowledge on the hazard associated with a chemical. The WoE methods should allow the assessor to decide whether the information is good enough to give information for REACH on hazard, for classification and labelling and for PBT/vPvB assessment if applicable. Additional research effort should clarify which WoE methods are most suitable for this. In the remainder of this chapter, the outline and methodology of a quantitative framework that is

built on Bayesian statistics will be developed (Press, 2003, Campbell and Machin, 1993) and the use of supporting decision theory (Winkler, 2003; TAPIR, 2005).

3.2 General background for a quantitative WoE

3.2.1 A probabilistic framework

In a WoE, one has to deal with uncertainty about the effects of chemicals on a specific endpoint. This can be expressed as 'it is likely that this chemical is biodegradable', for instance if a biodegradation model is used instead of an experimental test. When this is quantified, one deals with probabilities. This can make it easier to communicate the nature of the uncertainty concerned. Instead of a long verbal argument, the judgement can be summarised in the form of a probability. A statistical framework can express our expectation on the outcome of a model, a test of a combination of both, as a probability.
In a quantitative WoE approach, probabilistic weighting schemes are used to illustrate general principles. For the sake of the argument, it is considered that a test for a certain endpoint results in either a positive result or a negative result. Examples of this are biodegradation tests ('ready biodegradable' or 'not ready biodegradable') or the Ames test for mutagenicity ('*in vitro* mutagenic' or 'not mutagenic *in vitro*').

3.2.2 Bayesian methods

Most Bayesian statistical models involve random variables at only two *levels*, or *stages*: Y and Θ ('theta'), cf. Carlin and Louis (2000, page 10), Gelman et al. (2004, page 7). These quantities correspond to *variables* and *parameters* in data analysis, or *observables* and *unobservables*.
In Bayesian statistics, the available data are taken to be given (fixed), while the parameters, and therefore the model, are uncertain. By contrast, in classical sampling statistics, parameters are presumed to be fixed, but unknown. The data are thought to be a particular instance of a conceptually infinite set of repeated samples from this fixed unknown model. This leads to different methods, in general, but often parallel results exist.

Bayes' Theorem, also called Bayes' Rule, can be concisely written as:

$$p(\theta \mid y^*) = const \cdot p(y^* \mid \theta) \cdot p(\theta) ,$$

where p stand for probability, θ for parameter value(s), and y^* for data value(s).

The standard Bayesian interpretation in practical applications is to apply the rule for sequential learning of probable parameter values, given the data ('updating'):

$p(\theta)$ is the *prior* parameter distribution,
$p(y^* \mid \theta)$ is the *likelihood* of parameter values, and
$p(\theta \mid y^*)$ is the *posterior* parameter distribution.

So, in order to get the analysis started, you have to supply a prior parameter distribution[1], the choice of which has been the subject of much debate. In some cases, a genuine prior distribution is available from previous studies, literature surveys, medical records, logical considerations, expert judgment, etc. However, in the majority of cases, not much is known about model parameters initially. This causes the analysis to depend on the choice of priors, at first sight destroying the intuitively required objectivity of statistical procedures. Bayesian statisticians have combated this at first sight fatal objection, by showing that classical statistical procedures involve such subjectivity in hidden assumptions, making their influence even more difficult to analyze. In general, the more data are available, the less influential the choice of the prior distribution will be.

One would use a Bayesian analysis:

1. if one wants to incorporate prior parameter knowledge (ranges, probable values, correlations) into the analysis in a systematic way.
2. if a systematic approach is needed to compare model predictions of observable variables with data and improve the match;
3. if the predictive power of the model is to be assessed, or if alternative competing models are to be judged.

3.2.3 Probability and decision making

Specific features are needed to help to decide if additional information is needed. For instance, accepting a certain test result can be based on cut-off probabilities. The difficulty remains that the risk analyst still has to decide on appropriate decision cut-off probabilities. This choice in itself is subjective but at least the effect of using different cut-off values can again be analysed in a quantitative way.

If adding information leads to an increase in the probability that a certain test result is positive, more certainty is attained on the outcome of the test and one could decide to stop data gathering. If adding information does not increase the probability that a certain test result is positive, it could be decided to gather more information as part of an intelligent testing strategy. A probabilistic framework can make such decisions more transparent, by showing the probabilities on which the decisions are based.

As part of integrated test strategies, it can be very helpful to consider the additional consequences of the decision to go for further testing or not (Winkler, 2003). A fictitious example on soil toxicity will be considered where the cost of potential environmental effects is ignored for the sake of clarity.
If a substance turns out to be toxic, while the preliminary results indicated ‘not classified for soil toxicity’, it could be decided not to pursue further information collection. If the substance would be toxic in that case, high costs could ensue, e.g. due to the obligation for local soil cleanup and the additional cost of further testing and registration. The consequences of the decision whether or not to pursue information gathering (or testing) on soil toxicity can be expressed in a pay-off table (Table 3.1). The pay-off of no further testing, if you consider the current (non-testing) information reliable, would be €20,000 due to no further data gathering and testing. If you decide to collect more data, it would cost you €20,000 if the substance was indeed not classified as toxic in the end. However, if the substance *is* toxic and classified

[1] In the case of a biodegradation example: the prior probability that a substance is ready biodegradable or not ready biodegradable, based on e.g. expert judgment, training set of a specific model etc.

dangerous while you stopped when you thought it was harmless, it could cost the registrant maybe €500,000 for soil cleanup, if this was found out later.

Table 3.1: Pay-off table for fictitious soil toxicity example

		Test Result	
Current state: 'not toxic'		Not toxic to soil organisms	Toxic to soil organisms
Decision	No further Information collection	20.000	- 500.000
	Further Information collection	-20.000	0

If further testing was pursued and it was found toxic, it would infer costs but the registrant could enhance his 'green image' so the net result could be zero.

Of course, the uncertainty about the test result needs to be considered in order to make the right decisions. Information can be gathered on the probability that a certain chemical is toxic from a database analysis of that specific chemical class. Suppose that for that class of chemicals, the probability (*P*) to be classified as toxic is 0.2, and the probability *(not classified as toxic)* is 0.8. Then it is possible to evaluate the pay-off in a quantitative way. The 'expected value' of not testing in this fictitious case is €20,000*0.8 + (-€500,000 * 0.2) = -€84,000. The expected value of testing in this case is -€20,000*0.8 + 0 = -€16,000. So in both cases, the pay-off is negative but the cost of the relatively unlikely event that the chemical is toxic still leads to a strong negative pay-off. In this case, the decision should be to pursue further information collection.

Of course, more factors may need to be taken into account, for instance the uncertainty about the current state of knowledge on soil toxicity ('not classified as toxic') and the addition of the 'value' of the animal lives that were saved by no further testing. If that would be added, a non-monetary factor (animal welfare) would need to be expressed in an equivalent amount of money.

The attitude towards risk should also be involved as well. If the consequences of a bad decision (i.e. no further testing while the substance is harmful) are large, either in monetary terms or in terms of affected organisms, a low probability of that event occurring will for many be a reason to seriously consider the risk of such a relatively unlikely event.
These additional factors make clear that, although statistics play the major role in calculating probabilities that can assist the WoE approach, the combination with the different considerations (utility arguments) yields the final result.

The requirements for a decision making problem in a Bayesian framework can now be listed (Winkler, 2003):
1. *A,* the set of decisions or acts

2. *S,* the different outcomes of the test
3. The probability of the test outcome *P(θ)*
4. Some form of linking each pair of *(a,θ)* to considerations about pay-off, costs etcetera (utility arguments as in the pay-off table). This is formally the utility function (more complex than in the pay-off table) that associates a utility with each pair of *(a,θ)* where *a* is an action and *θ* is a different outcome of a test.

The use of utility functions in decision making allows for explicitly valuing the additional arguments that are important in the decision making: direct costs of testing, costs of making the wrong decision, costs of animal lives etcetera. These issues will be developed in more detail in a planned EU FP6 project OSIRIS.

3.3 A quantitative WoE approach

3.3.1 Introduction on biodegradation examples

For many chemicals, biodegradation data are lacking or require confirmation. Biodegradation models have been developed that predict the probability of rapid biodegradation. The so called BIOWIN models include decision rules that translate probabilities in ready biodegradable or non-ready biodegradable (US EPA, 2000; see also section 4.3). Six different BIOWIN models are available. In the EU, a combination of BIOWIN models is used to classify substances as potentially persistent or not. Their predictive power for a set of 110 notified substances was evaluated by Posthumus et al. (2005). A similar exercise was done by Boethling et al. (2004) on a different data set of premanufacture notice substances. Posthumus et al. were interested in trying out different combinations of BIOWIN models to empirically find an optimum predictive combination for regulatory purposes. This meant that a combination of models was tried out, where a high percentage of correct predictions for ready biodegradation was combined with a low percentage of false positives (predicted to be biodegradable, while in fact, they are not).

Combining the results of several tests or models to increase the predictivity is called a '*battery approach'* to testing. Decision rules are needed to interpret the combined results, e.g. both model A and model B need to predict 'ready biodegradable' in order to score a substance as such. Other combinations ('and, or') with different numbers of models are of course possible. Interestingly, this paper did not consider the implications of the outcome of model A for the degree of belief in the combined outcome of model A and model B. If the outcome of model A would be 'ready degradable', this prediction (with a certain probability) would then be the input (prior probability) for model B.
Boethling et al. (2004) systematically analyzed the incorporation of prior knowledge about the prediction of ready degradability. Prior knowledge on degradability comes from analyzing the model outcome for the training set or specific validation data sets. Bayesian analysis was chosen to show that prior probabilities on biodegradation are updated by a first model prediction which posterior prediction then functions as prior information for the next step in the battery. The sequential updating of the predictions can decrease the uncertainty one has on the predicted biodegradation and improve the predictive power of the battery. This approach is explained in the next section.

3.3.2 Applying Bayesian statistics to biodegradation data

3.3.2.1 Bayesian inference explained from a 2x2 table

The outcomes of the application of one particular model for a given set of substances with known experimental biodegradability can be analyzed on the basis of Bayesian inference applied to a 2x2 table. Posthumus et al. (2005), Table IV, have scored the biodegradation predictions of different models for 110 substances with known biodegradability determined in the laboratory.

Model BIOWIN-3 is considered with cut-off value 2.75, which is Line 6 in Table IV (Posthumus et al., 2005). The raw data can be re-arranged in a 2x2 table as in Table 3.2.

Table 3.2: Ready Biodegradability outcomes of Model BIOWIN-3 (cut-off 2.75) for 110 Ready and Non-ready biodegradable substances in the laboratory (Experimental).

BIOWIN-3 (2.75)	Experimental results		
	NRB-Exp (θ_0)	RB-Exp (θ_1)	Total
	77	33	110
Model results:			
NRB-Mod (y_0)	**62**	**7**	69
RB-Mod (y_1)	**15**	**26**	41

Here NRB and RB stand for Non-Ready Biodegradable and Ready Biodegradable, respectively, also coded as 0 and 1. 'Mod' is the model under consideration, and 'Exp' is the experimental outcome in the laboratory. The standard Bayesian variable notation is employed: *y* for data, in this case model data, and θ (theta) for the state of nature, here the unknown ready biodegradability of a new substance.

The Bayesian calculation and terminology is derived from 2x2 table inference used in medical statistics (Campbell and Machin, 1993, Chapter3, Diagnostic tests). In the medical situation, one is interested in the question whether, if a person test positive for a disease, the disease is actually present. The disease 'parameter' is the state of nature indicated by θ. The 'model' test outcome, e.g. a high blood pressure, is the data *y*. One is interested in the probability of having the disease given the model test outcome:

$$p(\theta \mid y).$$

This probability is called the 'posterior' probability, which is calculated through Bayes' Theorem, as explained below. One has four possible combinations: $p(\theta_0 \mid y_0)$, $p(\theta_1 \mid y_0)$, $p(\theta_0 \mid y_1)$ and $p(\theta_1 \mid y_1)$. The vertical bar is shorthand for 'given'. These are the chance of the substance being NRB (θ_0) under the condition that the model predicted NRB (y_0), etcetera.

In medical statistics, a ‘positive’ result is when the model test indicates disease. In the environmental situation, this is less obvious. The current literature (Boethling et al., 2004, Hulzebos et al., 2005; Posthumus et al., 2005) calls RB to be positive, since the substance is positively biodegradable. Obviously, NRB is the worrisome feature analogous to disease.

The *sensitivity* of a model prediction is the probability of a ‘positive’ RB model outcome given that the substance is RB in the laboratory. This is denoted as:

$$p(y_1 \mid \theta_1).$$

The *specificity* of a model prediction is the probability of a negative RB model outcome, i.e. NRB, given that the substance is indeed NRB in the lab, which is:

$$p(y_0 \mid \theta_0).$$

These names are typical for medical applications of the statistics. However, due to the transfer of the medical concepts to environmental assessment, one must conclude that the arbitrariness of what to call positive or negative would generate confusion about what model-laboratory combination to call the *sensitivity* of the model prediction, and which one the *specificity*. So, the latter one, $p(y_0 \mid \theta_0)$, is proposed to be called the *NRB-specificity*, and the former, $p(y_1 \mid \theta_1)$, the *RB-specificity*. This frees the term *sensitivity* for a sensitivity analysis of the calculations to come.

The importance of these *specificity statistics* is that they yield consistent results for a particular model applied to a variety of different classes of substances. The specificity is a true model property. It indicates how well the model performs for a given kind of substance.

The other two combinations for the probability of a model outcome given an experimental property in the laboratory also have names in medical statistics: *False Positives* and *False Negatives*. Similarly they have to be renamedto avoid what to call positive: NRB or RB. Thus, $p(y_0 \mid \theta_1)$, i.e. the probability of a false NRB-model outcome, when the substance is actually RB in the laboratory, is called the *False NRB-model* predictions, instead of False Negatives. Accordingly, $p(y_1 \mid \theta_0)$, i.e. the probability of a false RB-model outcome, when the substance is actually NRB in the laboratory, is the *False RB-model* predictions, called *False Positives* in medical applications.

Table 3.3: Same as Table 3.2, with indication of Model-specificity entries and False Model outcomes.

BIOWIN-3 (2.75)	Experimental results		
	NRB-Exp (θ_0)	RB-Exp (θ_1)	Total
	77	33	110
Model results:			
NRB-Mod (y_0)	***NRB-specificity*: 62**	***False NRB-model*: 7**	69
RB-Mod (y_1)	***False RB-model*: 15**	***RB-specificity*: 26**	41

The first matrix to compose then, is the matrix of *conditional probabilities*, $p(y \mid \theta)$, as shown in Table 3.4.

Table 3.4: Conditional probabilities of Model outcomes given the experimental knowledge; columns sum to 1.0.

BIOWIN-3 (2.75)	Experimental results		
	NRB-Exp (θ_0)	RB-Exp (θ_1)	Total
	77/110 = 0.700	33/110 = 0.300	1.000
Model results:			
NRB-Mod (y_0)	**62/77 = 0.805**	**7/33 = 0.212**	1.017
RB-Mod (y_1)	**15/77 = 0.195**	**26/33 = 0.788**	0.983
Sum:	1.000	1.000	2.000

Since the model either says NRB or RB, these fractions sum to 1.0 in the vertical direction. The horizontal rows are called *likelihoods* in statistics. They *do not* sum to 1.0. For example, the first row: 0.805 and 0.212 is the *likelihood* of a substance being NRB or RB respectively in the laboratory, given that the model predicted NRB (y_0). Thus the likelihood of states θ_0 versus θ_1, given that the model outcome is $y = y_0$, is given by the numbers:

$$p(y_0 \mid \theta_0) \text{ and } p(y_0 \mid \theta_1)$$

The well-known method of *Maximum Likelihood* says that when the model predicts NRB, then NRB-Exp is the most likely property of the substance. Similarly, if the model yields RB, it is more likely the substance will be RB in the laboratory as well.

In Table 3.4, the so-called *prior probabilities*, the fraction of NRB-substances versus RB-substances present in the study are also calculated:

$$p(\theta_0) \text{ and } p(\theta_1),$$

which are 77/110 = 0.700 and 33/110 = 0.300 respectively. In medical statistics this is called the *prevalence* of a disease.

The crux of a Bayesian analysis is that the model performance should be constant and relatively independent of the prior probabilities, as will be demonstrated.

The next step is to form a matrix of *joint probabilities* by multiplying the prior probabilities with the conditional probabilities:

$$p(y,\theta) = p(y \mid \theta) \cdot p(\theta),$$

for all combinations. This yields Table 3.5.

Table 3.5: Joint probabilities of model outcomes and experimental results; all 4 entries sum to 1.0.

BIOWIN-3 (2.75)	Experimental results		
	NRB-Exp (θ_0)	RB-Exp (θ_1)	Total
	77/110 = 0.700	33/110 = 0.300	1.000
Model results:			Horizontal sums:
NRB-Mod (y_0)	**62/110 = 0.564**	**7/110 = 0.064**	0.627 = 69/110
RB-Mod (y_1)	**15/110 = 0.136**	**26/110 = 0.236**	0.373 = 41/110
Sum:			1.000 (Total joint)

So, the upper-left joint entry is calculated as 62/77 x 77/110 = 62/110 = 0.564. This is the joint probability of a NRB-model result and NRB-experimental substance. Note that all four entries sum to 1.000. Obviously, this is the initial matrix at the start divided by the total number of substances.

The horizontal sums are the *marginal probabilities* of a NRB-model result and a RB-model result. They are denoted as: $p(y_0)$ and $p(y_1)$. We have:

$$p(y_0) = p(y_0 \mid \theta_0) \cdot p(\theta_0) + p(y_0 \mid \theta_1) \cdot p(\theta_1)$$

The final Bayesian step is calculated as the joint probabilities divided by the *y*-marginals:

$$p(\theta \mid y) = p(y, \theta) / p(y) .$$

These relationships combine to the celebrated Bayes' Theorem, as it is often presented:

$$p(\theta_0 \mid y_0) = \frac{p(y_0 \mid \theta_0) \cdot p(\theta_0)}{p(y_0 \mid \theta_0) \cdot p(\theta_0) + p(y_0 \mid \theta_1) \cdot p(\theta_1)},$$

$$p(\theta_1 \mid y_0) = \frac{p(y_0 \mid \theta_1) \cdot p(\theta_1)}{p(y_0 \mid \theta_0) \cdot p(\theta_0) + p(y_0 \mid \theta_1) \cdot p(\theta_1)}$$

and similarly for the other two combinations. These are the posterior probabilities of a state of nature (NRB of RB in the lab), given a model outcome (NRB or RB).

In the medical and environmental statistics references cited, the posterior is also called the *predictive value* of the model. The results are presented in Table 3.6.

Table 3.6: Posterior probabilities of the experimental biodegradability given one of two model outcomes; entries sum row-wise to 1.0.

BIOWIN-3 (2.75)	Experimental results		
	NRB-Exp (θ_0)	RB-Exp (θ_1)	Total
	77/110 = 0.700	33/110 = 0.300	1.000
Model results:			Horizontal sums:
NRB-Mod (y_0)	**62/69 = 0.899**	**7/69 = 0.101**	1.000
RB-Mod (y_1)	**15/41 = 0.366**	**26/41 = 0.634**	1.000

The upper-left entry $p(\theta_0 \mid y_0)$ is calculated as (62/110) / (69/110) = 62/69, or in the Bayesian formula:

$$\frac{(62/77) \cdot (77/110)}{(62/77) \cdot (77/110) + (7/33) \cdot (33/110)} = \frac{62/110}{62/110 + 7/110} = 62/69 = 0.899 .$$

The usual interpretation of this so-called Bayesian *updating* is that the *prior* probability:

$$\{ p(\theta_0),\ p(\theta_1) \}$$

of experimental (laboratory) biodegradability is converted to a *posterior* probability:

$$\{\, p(\theta_0 \mid y_0),\ p(\theta_1 \mid y_0) \},$$

when y_0, i.e. NRB-Model, is observed. Similarly, the prior changes to the posterior:

$$\{\, p(\theta_0 \mid y_1),\ p(\theta_1 \mid y_1) \},$$

When the model yields y_1, that is: RB-Mod.

Numerically, when the model predicts NRB (y_0), the prior probability of {0.700, 0.300} for NRB-Exp versus RB-Exp, is updated to the posterior probability: {0.899, 0.101}. Similarly, if the model says RB-Mod (y_1), then the posterior probability is {0.366, 0.634} for NRB-Exp versus RB-Exp.

3.3.2.2 Bayesian calculations on biodegradation data

The calculations on the biodegradation example are summarized in Table 3.7.

Table 3.7: Excel sheet copy of Bayesian calculation for model BIOWIN-3 (cut-off 2.75), with prior probability given by the experimental outcomes for 110 substances.

BIOWIN-3 (2.75)

	NRB-Exp	RB-Exp		
Prior:	**0.7**	**0.3**		
NRB-Mod	0.8052	0.2121		
RB-Mod	0.1948	0.7879		
			Sum:	
Joint:	0.5636	0.0636	0.627273	
	0.1364	0.2364	0.372727	80.0%
NRB-Mod	**89.9%**	**10.1%**	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**36.6%**	**63.4%**	*Posterior (N)RB-Exp given RB-Mod*	

In practical Bayesian applications, the usual stance is that the prior distribution (lila) and the model performance conditionals (light blue), (N)RB-specificity and False (N)RB-model outcomes, are independent from each other. This means that for another class of substances, the prior NRB-ness may be quite different, while the model characteristics are assumed to stay the same. This is exemplified in the next spreadsheet run, where the prior of NRB-ness is changed to {0.9, 0.1}, as in Table 3.8.

Table 3.8: Excel sheet copy of the same model as in Table 3.7, but with a different prior distribution.

BIOWIN-3 (2.75)

	NRB-Exp	RB-Exp		
Prior:	**0.9**	**0.1**		
NRB-Mod	0.8052	0.2121		
RB-Mod	0.1948	0.7879		
			Sum:	
Joint:	0.7247	0.0212	0.745887	
	0.1753	0.0788	0.254113	80.3%
NRB-Mod	**97.2%**	**2.8%**	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**69.0%**	**31.0%**	*Posterior (N)RB-Exp given RB-Mod*	

Clearly, when the model says NRB, the evidence for NRB-Exp (Exp) is overwhelming, but now also, when the model predicts RB, there is a 69% chance that the substance is NRB, despite the fact that the RB-specificity is quite high: 79% (lower-right entry in the light-blue matrix).

The result of a class of substances that is very likely to be RB is presented in Table 3.9.

Table 3.9: Same as Table 3.8, but with the prior probability with respect to biodegradability the other way around.

BIOWIN-3 (2.75)

	NRB-Exp	RB-Exp		
Prior:	**0.1**	**0.9**		
NRB-Mod	0.8052	0.2121		
RB-Mod	0.1948	0.7879		
			Sum:	
Joint:	0.0805	0.1909	0.271429	
	0.0195	0.7091	0.728571	79.0%
NRB-Mod	**29.7%**	**70.3%**	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**2.7%**	**97.3%**	*Posterior (N)RB-Exp given RB-Mod*	

If the model predicts NRB, the posterior (predictive value) for RB is higher than for NRB.

It is concluded that the model performance is summarized by the conditional matrix (light blue), while the predictive, i.e. posterior, value is influenced by both the model performance as well as the prior probability of the class of chemicals. A good (N)RB-specificity can be counteracted by a low prevalence of the corresponding laboratory occurrence.

3.3.2.3 Applying Bayesian methodology to the 'Battery' approach of model couplings

The idea behind the battery approach of model applications is that, knowing that models are not perfect, one may gain precision by sequentially applying different models. So, model #1 is applied first based on some prior information, then the posterior (N)RB-Exp outcome is used *as a prior input* for model #2. The result is found to display improved information with respect to the models applied separately. The next calculations illustrate this for the models BIOWIN-3 (2,75) and BIOWIN-6 (0.5).

First, the second model is run separately (Posthumus et al., 2005, Table IV, line 8). Since the dataset is the same, one has the same prior. This is presented in Table 3.10.

Table 3.10: Bayesian analysis of the 2x2 results of applying model BIOWIN-6 (cut-off value 0.5) to the same data as in Table 3.7.

BIOWIN-6 (0.5)

	NRB-Exp	RB-Exp		
Prior:	**0.7**	**0.3**		
NRB-Mod	0.8701	0.4545		
RB-Mod	0.1299	0.5455		
			Sum:	
Joint:	0.6091	0.1364	0.745455	
	0.0909	0.1636	0.254545	77.3%
NRB-Mod	**81.7%**	**18.3%**	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**35.7%**	**64.3%**	*Posterior (N)RB-Exp given RB-Mod*	

Note that BIOWIN-6 (0.5) does better on a NRB substance than BIOWIN-3 (2.75), because the upper left entry in the conditional (light blue) matrix, which is the ***NRB-specificity* = 0.8701** is higher than for BIOWIN-3 (2.75): 0.8052. So model #2, BIOWIN-6 (0.5), can be trusted more for a NRB substance than model #1, BIOWIN-3 (2.75).

Since the entries in the conditional matrix sum to 1.0 column-wise, this implies that the False-RB-Mod prediction rate (False 'positives') is *lower*: 0.1948 for BIOWIN-3 (2.75) versus 0.1299 for BIOWIN-6 (0.5).

Remarkably, the same is not true for the RB-specificity. The ***RB-specificity* = 0.5455** for the new model, BIOWIN-6 (0.5), which is less than 0.7879 for the first model, BIOWIN-3 (2.75). Hence, BIOWIN-6 (0.5) is a *worse* RB-predictor, which implies that the False-NRB-Mod prediction rate (False 'negatives') is *higher*: 0.4545 for BIOWIN-6 (0.5) versus 0.2121 for BIOWIN-3 (2.75).

Yet, the example shows that, if BIOWIN-6 (0.5) comes out RB-Mod, it predicts RB-Exp versus NRB-Exp with odds 64.3 to 35.7, because the False-positive rate (False RB-Mod) 0.1299 is much lower compared to the moderately bad RB-specificity, 0.5455. It also helps that the prior to have a RB substance (RB-Exp) is relative *high*: 0.3 versus 0.7 (NRB-Exp), because 0.3 x 0.5455 = 0.1636 is *higher* than 0.7 x 0.1299 = 0.0909. So, the model still performs not as bad as expected for a RB-Mod outcome.

Obviously, this changes when the prior is different. If the prior for NRB-Exp has odds 9:1, than the ratio of the moderately bad RB-specificity (0.5455) to the False positive RB-Mod rate (0.1299), is destroyed by the prior odds 9:1, resulting in a RB-Mod outcome to yield a 68.2 to 31.8 posterior evidence that the substance is NRB (NRB-Exp), c.f. Table 3.11. Apparently, a bad (N)RB-specificity hits harder the more unlikely the occurrence of a (N)RB substance is.

Table 3.11: Application BIOWIN-6 (cut-off 0.5) with a prior distribution denoting a large probability of Non-biodegradability, yields a high posterior experimental NRB even if the model indicates that it is RB.

BIOWIN-6 (0.5)

	NRB-Exp	RB-Exp		
Prior:	**0.9**	**0.1**		
NRB-Mod	0.8701	0.4545		
RB-Mod	0.1299	0.5455		
			Sum:	
Joint:	0.7831	0.0455	0.828571	
	0.1169	0.0545	0.171429	83.8%
NRB-Mod	**94.5%**	**5.5%**	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**68.2%**	**31.8%**	*Posterior (N)RB-Exp given RB-Mod*	

The next display (Table 3.12) shows, what happens when one makes a prior-posterior battery coupling of BIOWIN-3 (2.75) to BIOWIN-6 (0.5). It is assumed that the first model, BIOWIN-3 (2.75), came out RB-Mod. The posterior evidence over (N)RB-Exp is 36.6 to 63.4. This posterior is entered as the prior input for the second model, BIOWIN-6 (0.5).

Table 3.12: Bayesian Battery approach of model coupling: first BIOWIN-3 (2.75) is applied, then, when the model indicates RB, the posterior experimental property is taken as the prior probability for the second model BIOWIN-6 (0.5); if the outcome of the second model is consistent, the uncertainty is reduced, when not, the information is weakened.

BIOWIN-3 (2.75)

	NRB-Exp	RB-Exp		
Prior:	**0.7**	**0.3**		
NRB-Mod	0.8052	0.2121		
RB-Mod	0.1948	0.7879		
			Sum:	
Joint:	0.5636	0.0636	0.627273	
	0.1364	0.2364	0.372727	80.0%
NRB-Mod	89.9%	10.1%	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	***36.6%***	***63.4%***	*Posterior (N)RB-Exp given RB-Mod*	

BIOWIN-6 (0.5)

	NRB-Exp	RB-Exp		
Prior:	***36.6%***	***63.4%***	***RB-Mod posterior***	
NRB-Mod	0.8701	0.4545		
RB-Mod	0.1299	0.5455		
			Sum:	
Joint:	0.3183	0.2882	0.606589	
	0.0475	0.3459	0.393411	66.4%
NRB-Mod	**52.5%**	**47.5%**	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**12.1%**	**87.9%**	*Posterior (N)RB-Exp given RB-Mod*	

Now if the second model confirms the first result (RB-Mod), the conclusion that it *is* a RB substance (RB-Ex) is amplified with an overwhelming 87.9 to 12.1%. But, if the second model is in conflict by yielding NRB-Mod, one no longer has a clue what to think: it may be NRB or RB with a nearly 50-50 chance (52.5 to 47.5).

One might immediately ask, whether it makes a difference to first apply model #1, then feed it to model #2, versus first applying model #2, then feeding it into model #1. This does not make a difference, as long as one keeps track of the appropriate model outcome combinations.

The next run first applies BIOWIN-6 (0.5), then, assuming that the model outcome is RB (RB-Mod), it is fed to BIOWIN-3 (2.75) (Table 3.13). If the result is confirmed (RB-Mod), the same result is obtained as the above run for getting RB-Mod twice.

Table 3.13: Bayesian Battery approach of model coupling similar to Table 3.12, but now with the model sequence reversed: first BIOWIN-6 (0.5) is applied, then BIOWIN-3 (2.75). If the model outcome is RB-Mod twice, the same posterior probability is obtained as in Table 3.12.

BIOWIN-6 (0.5)

	NRB-Exp	RB-Exp		
Prior:	**0.7**	**0.3**		
NRB-Mod	0.8701	0.4545		
RB-Mod	0.1299	0.5455		
			Sum:	
Joint:	0.6091	0.1364	0.745455	
	0.0909	0.1636	0.254545	77.3%
NRB-Mod	81.7%	18.3%	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**35.7%**	**64.3%**	*Posterior (N)RB-Exp given RB-Mod*	

BIOWIN-3 (2.75)

	NRB-Exp	RB-Exp		
Prior:	**35.7%**	**64.3%**	***RB-Mod posterior***	
NRB-Mod	0.8052	0.2121		
RB-Mod	0.1948	0.7879		
			Sum:	
Joint:	0.2876	0.1364	0.423933	
	0.0696	0.5065	0.576067	79.4%
NRB-Mod	**67.8%**	**32.2%**	*Posterior (N)RB-Exp given NRB-Mod*	
RB-Mod	**12.1%**	**87.9%**	*Posterior (N)RB-Exp given RB-Mod*	

If the new result, though, is NRB-Mod, it gives emphasis to NRB-Exp. This result would also be obtained, if one would run BIOWIN-3 (2.75) first, getting NRB as outcome, then feed that posterior to BIOWIN-6 (0.5) as a prior, obtaining conflicting RB from the latter.

3.3.3 Conclusions on WoE in biodegradation studies

Prior knowledge about chemicals can be used in a WoE approach in a Bayesian framework. Such prior knowledge could be derived e.g. from experimental results for a class of similar chemicals, or information from several joint model predictions. The advantage of the Bayesian analysis is that, once familiar with the terminology and notation, relatively simple calculations can demonstrate the influence of the prior knowledge on the outcome of the model predictions. The consequences of the additional information can then be evaluated for increased certainty on the outcome of the test.

There are two ways of combining predictions of several models. One is the prior-posterior coupling as illustrated above. If the models agree, this has a strengthening effect on the conclusions. If they conflict, the information is not amplified but weakened. The second way

of model coupling is the logical way of combining two or more models into a new model, by AND and OR combinations of the results, as published in Posthumus et al. (2005). The difficulty is to implement this coupling in such a way that the resulting model has more predictive power than the component models, irrespective of the prior knowledge with regard to the substances it will be used on. With Bayesian statistics this can be investigated.

3.3.4 Further developments needed

The current analysis has shown that the use of Bayesian statistics allows quantifying the information value of additional information in sequential steps of an ITS. Although one can improve the predictive value by applying a model battery, as demonstrated in the biodegradation example, it is still necessary to decide if one is satisfied with a model outcome, or need to conduct further testing. This decision is not only a matter of agreeing on cut-off probabilities for the uncertainty one is willing to accept. The pay-off example on soil toxicity demonstrates that the additional considerations on costs and magnitude of the potential risk are important as well. It is therefore needed to expand the statistical framework towards decision making. The use of utility functions in decision making allows for explicitly valuing the additional arguments: direct costs of testing, costs of making the wrong decision, costs of animal lives etcetera.

The current application as demonstrated for biodegradation data needs to be expanded towards different assessment endpoints. Whenever the model outcomes are discredited through cut-off points, the same Bayesian ‘table-oriented’, i.e. categorical, analysis is applicable. Although the updating process will in theory always be the same, variations are needed to deal with different types of information such as discrete or continuous data, expert judgement etcetera. For example, it may be that a substance property has more than two classes, and/or that the model predictions employ multiple cut-off values, or are essentially continuous. The advantage of the Bayesian method is that the handling of prior, model-specific, and posterior information is the same in all cases.

Further insight is needed into the value and independence of information. It follows from the above analysis, that applying the same model twice in a battery approach does cause the information content to improve, which is not justified, of course. The analysis implicitly assumes, that the second model, although having similar, or even identical properties, is an independent model coming from a different source. Now, some of these models, e.g. in the BIOWIN cases, are mathematical modifications developed on the same data. It is essential to somehow propagate information with regard to the composition of the training data, and its associated domain of application. The relationships between logical model couplings and battery couplings, or combinations, will be essential to understand hidden multiple use of information.

4. Intelligent testing strategy for assessing environmental degradation

4.1 Introduction

During production and use organic chemicals can be released into sewers, soil, surface water, sea, air, dumped or incinerated after use. Their fate and potential environmental hazard is strongly determined by the potential of degradability. Substances that do not rapidly degrade have a higher potential for longer term exposures and may consequently have a higher potential for causing long term adverse effect on biota and human than degradable substances. Prediction and understanding of the fate of the chemicals are therefore essential so that measures can be taken to avoid effects on humans and the environment. For this reason information on the biodegradability is used for different regulatory purposes: (1) environmental hazard classification, (2) PBT and vPvB assessment and (3) exposure assessment for use in the risk characterization.

Transformation of chemicals in the environment involves abiotic degradation and biodegradation. Abiotic degradation includes hydrolysis, oxidation, reduction, and photolysis. Biodegradation is defined as the transformation of substances caused by micro-organisms. Primary biodegradation of a molecule refers to any microbial process which leads to the formation of metabolites and thereby contributes to the degradation of the original substance. Ultimate biodegradation is known as the complete mineralization of a substance into carbon dioxide, water, and mineral salts.

In this chapter an overview is given of methods that are prescribed or can be used for the assessment of biodegradability. Experiments (paragraph 4.2) and estimation models (paragraph 4.3) are described in the context of regulation (paragraph 4.4) and their position in an intelligent testing strategy (paragraph 4.6). Metabolites are addressed as a topic for which more guidance is needed (paragraph 4.5). One promising estimation model able to predict identity and amounts of metabolites, CATABOL, was subjected to a verification study with several groups of chemicals.

The results are presented in paragraph 4.7.

4.2 Biodegradation tests

In order to investigate biodegradation, standardized biodegradation tests have been developed by different organizations (OECD, ISO, EU, US-EPA and STM), which can roughly be divided into three groups:

- Screening (ready or ultimate biodegradation) tests
- Intermediate (inherent or primary biodegradation) tests
- Definitive (simulation) tests

Screening studies
A positive result in the screening studies can be considered as indicative of rapid ultimate degradation in most aerobic environments including biological sewage treatment plants (Struijs and Stoltenkamp, 1994) and may take away the necessity for further testing.

A negative result in a test for ready biodegradability does not necessarily mean that the chemical will not be degraded under relevant environmental conditions, but it means that it should be considered to progress to the next level of testing, i.e. either an inherent biodegradability test or a simulation test.

Inherent or primary biodegradation tests
Using favourable conditions, the tests of inherent biodegradability have been designed to assess whether the chemical has any potential for biodegradation under aerobic conditions. Compared with the ready biodegradability tests, the inherent biodegradability tests are usually characterized by a high inoculum concentration and a high test substance concentration. A negative result will normally be taken as an indicator of that non-biodegradability (persistent) should be assumed for precautionary reason.

Simulation studies
Compared to ready and inherent biodegradability tests, simulation tests are higher tier tests that are more relevant to the real environment. These tests aim at assessing the rate and extent of biodegradation in a laboratory system designed to represent either the aerobic treatment stage of STPs or environmental compartments like surface water, sediment, and soil. They usually employ specific or semi-specific analytical techniques to assess the rate at which a substance undergoes degradation and to provide insight into subsequent metabolite formation and their decay. The fate of chemicals in STPs can be studied in the laboratory by using the simulation tests: activated sludge units (OECD 303) and biofilms (TG 303B). Simulation tests in soil (OECD 307), in aquatic sediment systems (OECD 308), and in surface water (OECD 309) have been also included in the guidelines of OECD (OECD, 1981-2006). No specific pass-levels have been defined for the simulation tests. Simulation tests are especially useful, if it is known from other tests that the test substance can be mineralized and that the degradation covers the rate determining process.

The tests complexity and the economic consequences bond to the tests increases from the simple screening test for ready biodegradability to the more complex simulation tests. For this reason, the standard information requirements within REACH are based on the tonnage of chemicals. The requirements for these tonnage-driven degradation tests are listed in Table 4.1.

Table 4.1. REACH tonnage-driven degradation tests requirements

Tonnage (tpa)	Degradation tests
1-10	Ready biodegradation test
10-100	Ready biodegradation test, hydrolysis test
100-1000	Ready biodegradation test, hydrolysis test, and simulation test, identification of the most relevant degradation products
>1000	Ready biodegradation test, hydrolysis test, and simulation test. Further confirmatory testing on rates of biodegradation with specific emphasis on the identification of the most relevant degradation products

4.3 Estimation models

Under the current EU legislation for new and existing chemicals, the regulatory use of estimation models or (Q)SARs is limited and varies considerably among the Member States, which is probably due to the fact that there is no agreement in the scientific and regulatory communities over the applications of (Q)SARs and the extent to which (Q)SARs estimates

can be relied on. In contrast, it is anticipated that these non-testing methods like (Q)SARs and read-across, in the interests of time- and cost-effectiveness and animal welfare, will be used more extensively under the future REACH system. Below, the current status of QSAR application for abiotic degradation and biodegradation is briefly reviewed.

(Q)SARs for abiotic degradation
QSAR models for abiotic degradation have been developed in the past decade. These models are mainly used for predication of abiotic degradation in the atmosphere and water. Two types of abiotic degradation, hydrolysis and oxidation, are the main focus of the QSAR model development. The application of (Q)SAR models for hydrolysis focus only on potentially hydrolysable substances. These substances include brominated alkanes, esters, carbamates, and para-substituted benzonitriles. The HYDROWIN program incorporates models for the estimation of the rate of hydrolysis of the above classes of substances in the environment. However, this program has not been published and is not evaluated in the EU-QSAR project.

(Q)SARs for oxidation applied to two categories of chemical species:
1. chemically well-defined species, such as the hydroxyl radical, the carbonate radical, singlet oxygen, chlorine dioxide, and ozone, and
2. groups of oxidants with no well-defined chemical structure, such as the different families of radicals and excited triplet states derived from the dissolved natural organic matter.
For the first category, (Q)SARs can be used directly to estimate second-order rate constants of a specific compound belonging to a congener series and, by subsequently applying equations, to obtain pseudo-first-order rate constants and half-lives for the transformation of such a compound under defined environmental conditions. For the second category of oxidants, (Q)SARs have been derived for model oxidants such as 2-cyanopropyl-2-peroxyl radical, substituted phenoxyl radicals, and the excited triplet states of three aromatic ketones. These (Q)SARs cannot be used directly to estimate environmental transformation rates of organic contaminants, but they are useful to get insight into the transformation mechanisms and energetics. They can also help in constructing (Q)SARs for the groups of environmental oxidants they represent. For the detailed information on (Q)SAR application for oxidation, see the review paper by (Boethling et al., 2003).

(Q)SARs for biodegradation
(Q)SARs for biodegradation could potentially be used either to supplement experimental data or to replace testing. Current generally applicable biodegradation models focus on the estimation of readily and non-readily biodegradability in screening tests. This is because most experimental data are from such tests. There are far fewer data that are both quantitative and environmentally relevant, such as measured half-lives and rate constants (Nendza, 2004).
In the past decade, the development of QSAR modelling is mainly via three approaches: group contribution approaches, statistical/chemometric approaches, and expert system/AI approaches. Table 4.2 summarizes the most often used QSAR models for biodegradation.

There are six models in BIOWIN. A description of these six BIOWIN models and their application for biodegradation can be found in Posthumus et al. (2005), Hulzebos et al. (2005), and Hulzebos and Posthumus (2003). Briefly, BIOWIN probability models includes the linear and non-linear BIODEG and MITI models for estimating the probability of rapid aerobic biodegradation and an expert survey model for primary and ultimate biodegradation estimation. Another model is MultiCASE, which combines a group-contribution model and an expert system to simulate aerobic biodegradation pathways (Klopman and Tu, 1997). This model has also been used by Rorije et al. (1998) to predict anaerobic biodegradation.

A promising model which can be used for quantitative assessment of biodegradability in biodegradation pathways of chemicals is CATABOL. The model allows for identifying potentially persistent catabolic intermediates, their molar amounts, solubility (water solubility, logKow, BCF) and toxic properties (acute toxicity, phototoxicity, mutagenicity, ER/AR binding affinity). Presently, the system simulates the biodegradability in MITI-I OECD 301 C and Ready Sturm OECD 301 B tests. Other simulators will be available in the program upgrades. The latest version of CATABOL (version 5) enables to establish to which degree chemicals belong to the domain of the biodegradation simulator. A more detailed description of CATABOL and its performance for several groups of chemicals is given in paragraph 4.7.

An evaluation of the predictions of the models for the high production volume chemicals showed that all models are highly consistent in their prediction of not-ready biodegradability, but much less consistency is seen in the prediction of ready biodegradability. This complies with the observation that the models show better performance in their predictions of not-ready biodegradability (Rorije et al., 1999).

Table 4.2. QSAR models for biodegradability

<table>
<tr><td rowspan="7">Group contribution approaches</td><td rowspan="6">Biodegradability probability program BIOWIN</td><td rowspan="2">BIODEG models</td><td>BIOWIN 1</td></tr>
<tr><td>BIOWIN 2</td></tr>
<tr><td rowspan="2">Expert survey models</td><td>BIOWIN 3</td></tr>
<tr><td>BIOWIN 4</td></tr>
<tr><td rowspan="2">MITI models</td><td>BIOWIN 5</td></tr>
<tr><td>BIOWIN 6</td></tr>
<tr><td>MULTICASE anaerobic biodegradation program</td><td colspan="2">To model anaerobic aquatic biodegradation rates</td></tr>
<tr><td>Statistical or chemometric approaches</td><td colspan="3">Ready biodegradability is modelled more adequately than not-ready biodegradability.</td></tr>
<tr><td rowspan="3">Expert system/AI approaches</td><td>Biodegradability evaluation and simulation system</td><td colspan="2">Qualitative, aerobic biodegradation pathways
The model needs to be validated.</td></tr>
<tr><td>MultiCASE/META</td><td colspan="2">Aerobic biodegradation pathways</td></tr>
<tr><td>CATABOL</td><td colspan="2">Quantitative assessment of biodegradability</td></tr>
</table>

4.4 Regulation

Classification and labelling
In the framework of classification and labelling, a substance is considered to be not rapidly degradable unless at least one of the following is fulfilled (UN, 2003):
(a) the substance is demonstrated to be readily biodegradable in a 28-day test for ready biodegradability. The pass level of the test (70% DOC removal or 60% theoretical oxygen demand) must be achieved within 10 days from the onset of biodegradation, if it is possible to evaluate this according to the available test data (see Appendix I). If this is not possible, then the pass level should be evaluated within a 14 days time window if possible, or after the end of the test; or
(b) the substance is demonstrated to be ultimately degraded in a surface water simulation test with a half-life of <16 days (corresponding to a degradation of >70% within 28 days); or
(c) the substance is demonstrated to be primarily degraded (biotically or abiotically) in the aquatic environment with a half-life <16 days (corresponding to a degradation of >70% within 28 days) and it can be demonstrated that the degradation products do not fulfil the criteria for classification as hazardous to the aquatic environment. When these data are not available rapid degradation may be demonstrated if either of the following criteria is justified:
(d) the substance is demonstrated to be ultimately degraded in an aquatic sediment or soil simulation test with a half-life of <16 days (corresponding to a degradation of >70% within 28 days); or
(e) in those cases where only BOD5 and COD data are available, the ratio of BOD5/COD is greater than or equal to 0.5. The same criterion applies to ready biodegradability tests of a shorter duration than 28 days, if the half-life furthermore is <7 days.
If none of the above types of data are available then the substance is considered as not rapidly degradable. This decision may be supported by fulfilment of at least one of the following criteria:
(i) the substance is not inherently degradable in an inherent biodegradability test; or
(ii) the substances is predicted to be slowly biodegradable by scientifically valid QSARs. The probability of <0.5 for BIOWIN 2 and 6 or <2.7 for BIOWIN 3 are used to consider a substance as potentially persistent, the reverse (i.e. >0.5 or 2.7) can be used as a criterion to classify a substance as readily biodegradable, or
(iii) the substance is considered to be not rapidly degradable based on indirect evidence, as e.g. knowledge from structurally similar substances; or
(iv) no other data regarding degradability are available.

PBT assessment
A testing strategy for the persistency criterion is described in the Technical Guidance Documents for new and existing substances (EC, 2003) that include final criteria based on experimental data derived from the marine or freshwater pelagic and benthic environments, screening criteria for identifying potentially persistent substances on the basis of the experimental data, and a preliminary assessment based on estimated data. The environmental simulation data would normally be required unless there is compelling evidence from other degradation testing or (Q)SAR that a substance will degrade sufficiently rapidly so as not to meet the criteria. Readily biodegradable substances (fulfilling or not fulfilling the 10-day window criterion) are considered as not persistent in the PBT assessment. When results are available showing that a substance does not fulfil the criteria for inherent biodegradability this is a clear indication that the substance will not biodegrade in the marine environment either. The substance will then be regarded as potentially persistent. For substances that pass the criteria for inherent biodegradability tests this does not necessarily indicate that they will

not be persistent under environmental conditions. However, in order to make the best use of available information it is accepted to use the results of two specific tests when they fulfil certain criteria (see EC, 2003).

A preliminary use of (Q)SAR estimates for identifying substances with a potential for persistency is described in the TGD (EC, 2003). The combined use of results of three estimation models in the EPI suite (US-EPA, 2000) for identifying potentially persistent substances is proposed as follows:

- Non-linear model prediction (BIOWIN 2): does not biodegrade fast (probability <0.5), or
- MITI non-linear model prediction (BIOWIN 6): does not biodegrade fast (probability <0.5), and
- Ultimate biodegradation time (BIOWIN 3): ≥ months (value <2.2).

For substances where estimation results are clearly below the limits, further information will normally not be required for the PBT and vPvB assessment, and they may be considered as not persistent. However, for borderline cases (e.g. when the estimate of the ultimate degradation time gives a result in the range 2.2 - 2.7) development of decision rules is currently discussed in the European Commission Working Group on PBT assessment. A substance fulfilling the criteria above is considered potentially persistent. Compelling evidence that a substance was not persistent would also include the knowledge that the substance was rapidly degraded, either through primary or ultimate degradation, in a test conducted under the conditions of the ready biodegradability test. For the PBT and the vPvB assessment the substance is considered to be not persistent if the pass-levels of 60 or 70% has been reached within the 28-day test duration irrespective whether the 10-day or 14-day time window has been fulfilled or not.

Where data are available from inherent biodegradation testing showing no or limited degradation (<20%) this may be taken, on a case-by-case basis, as an indication that biodegradation will not occur under environmental conditions, and the substance may be considered as persistent or very persistent. In some circumstances this may indicate that an environmental simulation test is not necessary since it would be unlikely that there would be sufficiently rapid degradation in such a test to meet the criteria for lack of persistence. Where >20% biodegradation is observed, an environmental simulation test should be considered. The choice of environmental media to use in a simulation test would be indicated by the compartment(s) of primary importance. To judge this, some knowledge of the physico-chemical properties, partitioning behaviour and emission pattern are important factors to consider. The TGD suggests for example by reference to the testing strategy for biocides for substances release to surface water that a Kp in sediment above 2000 should trigger consideration of performing a sediment simulation degradation test (e.g. OECD TG 308) in addition to the normal consideration of a pelagic simulation test (e.g. OECD TG 309). An attempt to develop a stepwise approach for identification of compartment(s) of primary importance was proposed by ECETOC (2004) and at the Simulation Testing of Environmental Persistence (STEP) workshop held in Rotterdam in October 2004.

The following decision criteria are broadly in line within that used by the Commission sub-group on PBT/vPvB substances:

- Evaluation of (Q)SAR data. Where this clearly indicates rapid degradation, consider (as a screening approach for prioritisation purposes only) as not persistent unless experimental data indicate otherwise. In other circumstances, seek further experimental data.

- First tier degradation screening data that could include ready biodegradation testing with modifications if appropriate and/or hydrolysis testing, although for PBT testing, both may be difficult for substances of low water solubility. Where data indicate that the persistence criteria are not met, no further testing is necessary. Otherwise, seek confirmatory testing.
- Confirmatory degradation testing, which would normally be an environmental simulation test in an appropriate environmental media. (An inherent biodegradation test may be considered prior to a simulation test to determine whether degradation may be expected.)

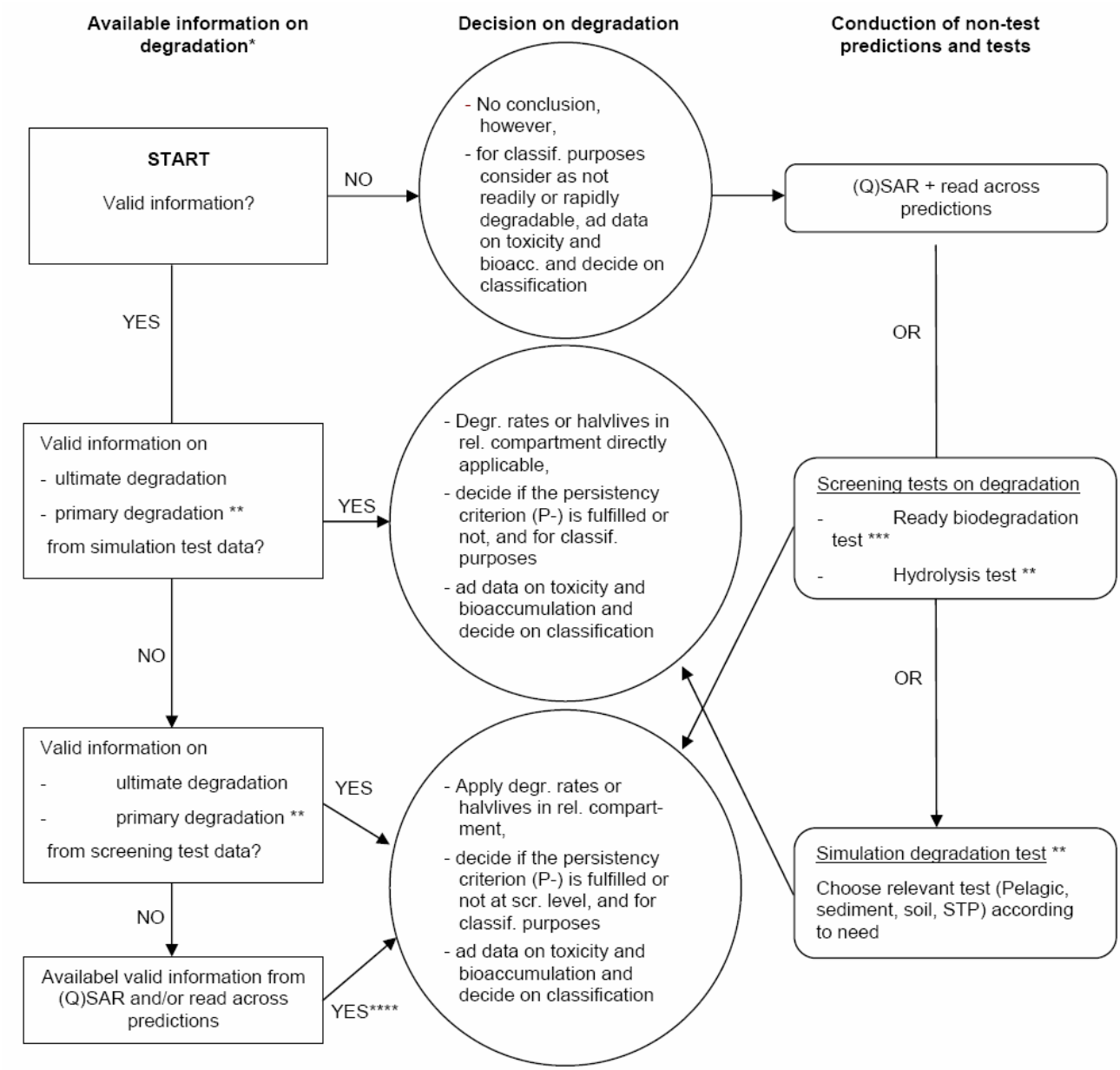
This approach is in line with the decision scheme on degradation proposed in the final TAPIR Report RIP 3.3-1, shown in Figure 4.1.

Exposure assessment
For the assessment of the degradation of a substance in surface water, sediments and soils, the default rate constants for substances given in the TGD can be used. These default rate constants are designed to cover the standard degradation data available i.e. data on ready and inherent biodegradability. Data on abiotic degradation, both hydrolysis and photolysis, is also considered. If the primary degradation has been measured in a ready biodegradability test, and/or in the hydrolysis or photolysis tests, the default degradation rate constant can be applied to the parent substance but the degradation products should also be assessed. The environmental degradation rates may be further investigated through the conduct of an appropriate environmental simulation test. The degradation half-lives and rate constants can be used directly in the multi-media modelling calculations after relevant corrections for temperature etcetera.

Based on the REACH proposal (REACH, Annex I), an exposure assessment is composed of exposure scenarios and estimation. An exposure scenario includes the description of emissions to the environmental compartments, sewage treatment systems, and the dilution in the receiving environmental compartment. For a great deal of industrial chemicals, emission to the sewage system takes place during at least some of the life-cycle step(s) of the chemical. Thus, the fate in the sewage treatment plants will often need to be determined. The exposure estimation entails an estimation of the emission, an assessment of chemical fate and pathways, and an estimation of exposure levels in relevant environmental compartments. Three aspects for exposure assessment need to be determined:

- Environmental release pattern including whether and to which degree and in which form the substance is emitted to the biological sewage treatment plants (STPs)
- Whether and to what degree the substance is degraded in a sewage treatment plant (if emission to a STP takes place)
- Whether and to what degree the substance is degraded and/or transformed in relevant environmental compartments.

These three considerations affect the calculation of both the local Predicted Environmental Concentration PEC_{local} (STP degradation) and the $PEC_{regional}$ (both STP and environmental degradation). Both calculations require, as a minimum, data on whether the substance is readily biodegradable, either experimental or estimated by (Q)SARs (Appendix I and II). there is no specific guidance on use of (Q)SAR estimates for exposure assessment in the TGD (EC, 2003). The application of (Q)SARs in exposure assessment is only possible under the following conditions: if a substance can be judged as being readily degradable or not readily degradable and if a degradation rate constant can be predicted with sufficient certainty.



* Note that the retrieval of existing data and generation of predictions using existing non-test methods may be done sequentially according to how easy it is to obtain the information. Nevertheless information should generally be gained from sources providing information from simulation studies, screening tests and (Q)SAR predictions / read across. In general it is not recommended to stop data retrieval/generation before all easily available information sources have been used, because often decisions for a regualtory endpoint may be taken by employing a weight of evidence approach (see further in the test regarding use of multiple data).

** When only primary degradation is demonstrated, consider the properties of the degradation products.

*** Incl. an enhanced test design if appropriate.

**** No specific guidance on use of QSAR estimated biodegradation results for derivation of degradation rates for exposure assessment is available in the TGD (EC 2003). However, when a substance can be judged as being readily degradable or not readily degradable or when a degradation rate constant can be predicted with sufficient certainty, such results may be used for establishing degradation rates for use in exposure assessment as described in chapter 3.3. Note, however that the currently available QSAR models are generally not sufficiently accurate for predicting that a substance is degraded fast.

Figure 4.1: Detailed decision scheme on the evaluation of degradation data in the frame work of the PBT assessment (TAPIR, 2005)

For the assessment of the degradation of a substance in surface water, sediments and soils, the default rate constants for substances given in the TGD can be used. These default rate constants are designed to cover the standard degradation data available i.e. data on ready and inherent biodegradability. Data on abiotic degradation, both hydrolysis and photolysis, is also considered. If the primary degradation has been measured in a ready biodegradability test, and/or in the hydrolysis or photolysis tests, the default degradation rate constant can be applied to the parent substance but the degradation products should also be assessed. The environmental degradation rates may be further investigated through the conduct of an appropriate environmental simulation test, as shown in Table 4.3 for surface water. The degradation half-lives and rate constants can be used directly in the multi-media modelling calculations after relevant corrections for temperature etcetera.

Table 4.3: First order rate constants and half-lives for biodegradation in surface water based on results of screening tests on biodegradability

Test result	**Rate constant k (d^{-1})**	**Half-life (d)**
Readily biodegradable	$4.7 \cdot 10^{-2}$	15
Readily, but failing 10-d window [b)]	$1.4 \cdot 10^{-2}$	50
Inherently biodegradable [c)]	$4.7 \cdot 10^{-3}$	150
Not biodegradable	0	∞

4.5 Degradation products

When assessing the biodegradation of organic chemicals, it may also be needed to consider the fate and toxicity of the resulting biodegradation products, especially when they have the potency to persist in the environment. The concentrations of these products in the different environmental compartments depend on numerous factors and processes, including how the parent compound is released to the environment; how fast it degrades; the half-lives of the degradation products; partitioning to sludge, soil, and sediment; and subsequent movement to air and water. In general, microbial degradation processes lead to the formation of more polar and more water soluble compounds. Hence, the resulting transport behaviour of degradation products may be different. The available data on pesticides demonstrate that in most cases degradation products are as toxic as or less toxic than the parent compounds. However, in some instances, degradation products can be more toxic. In general, the biggest increases in toxicity from parent to degradation products were observed for parent compounds that had a low toxicity. Possible explanations for an increase in toxicity are: (1) the active moiety of the parent compound is still present in the degradation product; (2) the degradation product is the active component of a pro-compound; (3) the bioaccumulation potential of the degradation product is greater than that of the parent; (4) the transformation pathway results in a compound with a different and more potent mode of action than that of the parent (Boxall et al., 2004).

Although, the EU TGD (EC, 2003) highlights that, where degradation occurs, consideration should be given to the properties (including toxic effects) of the products that might arise, that information does not exist for many compounds. REACH will introduce a range of required tests which could lead to metabolite investigations, e.g. hydrolysis is required for substances produced in quantities above 10 ten tonnes per year and biodegradation simulation tests in surface water, sediment and soil are required at production volumes above 100 tonnes

per year. There might be some concern that such a requirement will lead to an over emphasis on the behaviour of metabolites and that e.g. for such low production volumes, or in the case of inherently degradable substances such investigations will not be cost effective. As >100,000 chemicals are commonly used worldwide every day, pragmatic approaches are needed to identify the primary degradation products and those that are toxic, persistent, or bioaccumulative and/or which pose a risk to the environment. For this purpose guidance is needed to establish the criteria upon which metabolites of concern may be identified and to determine when a metabolite would not be of concern.

4.6 ITS for assessing biodegradation

As mentioned above, degradation route studies are complex and costly, and it is often very difficult to identify the minor degradation products in a system. An alternative or an additional tool to experimental testing might be to use of a QSAR model, like CATABOL to predict degradation pathways from the chemical structure of the parent compound. A decision framework, currently used by the Environment Canada for the identification of PBT properties of industrial substances and their probable stable biodegradation products, is using CATABOL. The framework seems to be a promising ITS tool to promote an evaluation of the level of concern of the metabolites (see Figure 4.2). The predicted metabolites by CATABOL can be selected based on:

- The predicted or measured molar amount of metabolite compared to parent is more than 10%.
- The probability of the metabolite occurring is high.
- The probability of the metabolite being stable is high.
- The degree of coverage by the model's training set.

The numbers in the flowchart illustrate the steps of this proposed evaluation process. The eight numbered paragraphs below explain the steps in Figure 4.2. The text is based on the description given in the draft report of the STEP workshop held in 2004 at Rotterdam, also to take into account the discussion and interpretation on the use of the flowchart (Bowmer and Leopold, 2006). Participants of the workshop recognized that the flowchart could be used in PBT context and a case can be made that even if a metabolite is stable, if it is not toxic, it could be taken off the PBT list. The group felt that the flowchart might also serve as a source of scientific arguments that could lead to the delisting of the parent material as a PBT, if it is shown that metabolites are not of concern.

1. The evaluation starts with ready biodegradability tests which are used to assess whether a chemical will undergo rapid ultimate degradation under stringent conditions in an aerobic aqueous medium.
   If substances are readily biodegradable the parent compound is likely to mineralize completely and in concomitant most likely no persistent degradation products are formed. However in exceptional cases biodegradation of 60-80% could still end up with a persistent degradation product of concern. Therefore, it might still be useful to run an estimation model like CATABOL to provide further evidence that no further investigation is warranted.
   If <20% degradation in a Ready Test is observed the parent may be considered stable and metabolites are not likely to be produced within relevant exposure assessment timeframes. This should however be backed up by data indicating lack of primary degradation, found e.g. in hydrolysis studies. It could still be possible that under more

favourable conditions in a simulation study the parent compound will degrade and that testing may be needed.
In cases where the Ready Tests show 20-60% degradation further investigation to determine stability and level of concern of the metabolites is warranted.

2. At this stage is should be considered to perform a simulation study or alternatively to run to generate information on the metabolites using predictive models that simulate biodegradation, like CATABOL. The probability of occurrence can be used to initially select those metabolites that are likely to be produced during biodegradation. At this stage is also useful to run CATABOL for less degradable compounds (<20% degradation) to investigate the occurrence of primary degradation.

3. In this step, the stability of a metabolite (i.e., from further transformation) should be determined either generated from the simulation study or obtained from the predictive models. In the case of CATABOL, the level of probability (e.g. >0.1) can be used to determine if the metabolite is stable and needs to be considered further. Other tools can be used to qualitatively determine metabolite stability, e.g. the BIOWIN model. Read across techniques should be used to determine if there are no stable metabolites (e.g., compare structural features of metabolite with known degradable functional groups). A rationale should be provided.

4. Once stable metabolites have been identified, only those expected to exist in sufficient quantity versus the parent compound need be considered. For example, Environment Canada uses CATABOL to determine if the molar quantity of stable metabolites is greater than 10%. If so, the metabolite is considered further.

5. It is also important to consider the actual amount of the substance that can enter the environment. In case a low amount of a substance is produced, the release of a metabolite may be quite low and therefore no further action for these metabolites may be considered. Conversely, high production volumes (e.g., >1000 tonnes/yr) may suggest that the release of metabolites could be significant, especially if the molar quantity of the metabolite is high vs. the parent. Further consideration of the metabolite may therefore be warranted.

6. The available information on metabolites that are likely to occur (due to to their stability and production volume) is examined, in order to assess whether these metabolites are of concern. It was also suggested to look at potential (eco)toxicity at this stage and that metabolic steps can lead to more polar or to less polar compounds as evidenced from HPLC-RAD chromatographs routinely produced during a simulation test. Preliminary information on toxicity can also be obtained with the help of QSARs and bioaccumulation.
Within the PBT assessment, logKow can be used to assess the bioaccumulation potential of a substance and give some indication of baseline toxicity provided the substance has a narcotic mode of action. If a metabolite has been determined not ‘of concern’ (e.g., low logKow/BCF, low toxicity), then no further action is required.
Simulation testing may not, at this point, have revealed the chemical identity of the important metabolites. This must now be performed for metabolites of concern in order that a chemical structure of the metabolite(s) is available for input into QSARs or for read-across purposes.

7. In the final step, metabolites of concern are included in the chemical assessment. A decision can be made to focus the assessment on the metabolites instead of the parent compound in the assessment or the assessment may include both the parent and important metabolites (e.g., when both the parent and metabolite(s) are of concern).

8. The final step involves identification by analytical chemical means of significant metabolites considered from the preceding steps to be ‘of concern’.

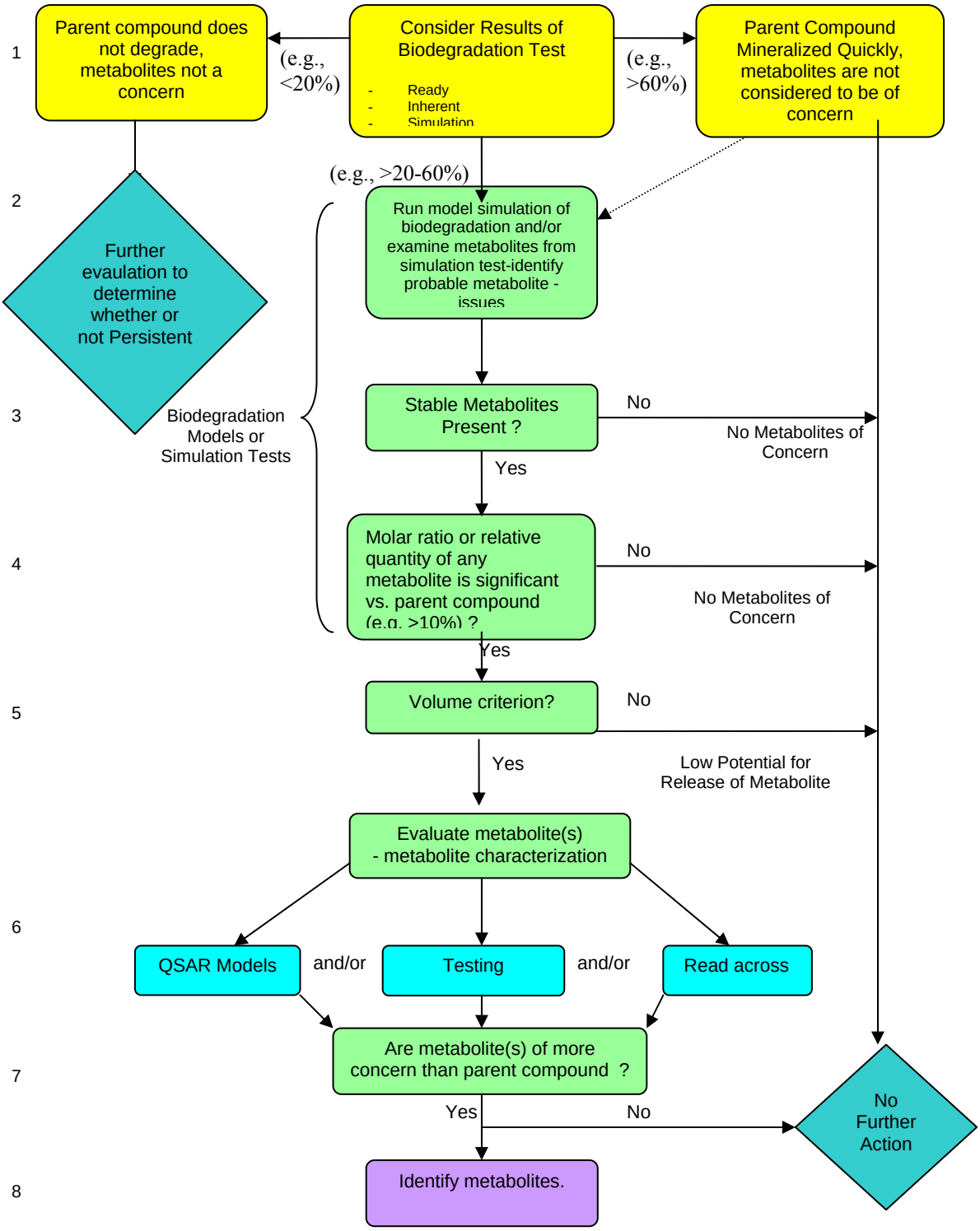


Figure 4.2: A framework used by Health Canada for evaluating whether metabolites are 'a cause for concern' for new substances assessment.

4.7 Study on the extrapolation capabilities of CATABOL

A promising model which can be used for quantitative assessment of biodegradability in biodegradation pathways of chemicals is CATABOL. This system generates most plausible biodegradation products and provides quantitative assessment for their physicochemical properties and toxic endpoints. The possibilities of QSARs in a framework of intelligent testing strategy have been described in the previous paragraphs.

Several papers have been published on how CATABOL has been developed (amongst others Jaworska et al., 2002). Its performance in BOD predictions has been described (Sakuratani et al., 2005). The model has been refined in order to predict a better environmental fate of perfluorinated chemicals (Dimitrov et al., 2004). An important issue in the application of QSARs for regulatory purposes is their reliability and the determination of their applicability domain. The OECD QSAR Validation Principles recommend that the models are only used for substances that fall within the applicability domain. Since CATABOL version 5 and later, it is possible to determine the applicability domain. The background of this new routine is described by Dimitrov et al. (2005).

Several groups of chemicals have been selected in order to gain insight in the extrapolation capability of CATABOL (Version 5.097) with respect to BOD prediction and prediction of metabolites:

- New substances (83)
- PBT substances (76)
- Pesticides (192)
- Existing chemicals listed on Annex I with R50 and R53 (59)

As for the majority of the substances the data were generated under different test conditions than in the MITI tests used to develop CATABOL, this study should not be considered as a validation or verification study but rather as a study into the possibility to extrapolate CATABOL results to biodegrability in other media and other types of experiments.

The major research questions were:

1. Is the applicability domain of CATABOL wide enough to warrant application to several groups of chemicals?
2. What is the reason that particular substances are ‘not in the domain’?
3. Is it possible to identify data gaps, or missing rules that are responsible for restrictions in the domain?
4. How good does CATABOL predict BOD for substances in the applicability domain and for substances not in the applicability domain?
5. What is the chance of false positives (unjustified ‘ready’ characterisation) and false negatives (unjustified ‘not ready’ characterisation)?
6. How good does CATABOL predict persistent metabolites for substances in the applicability domain and for substances not in the applicability domain?
7. Is it possible to identify data (gaps), or (missing) rules that could improve the prediction of the metabolic pathway?
8. What criteria could be set to define a reliable prediction of CATABOL (domain, map reliability, % error, other criteria) for BOD as well as the metabolic pathway?
9. Is there a relation between BOD predictions and half-lives obtained from soil simulation studies?

4.7.1 Description of CATABOL

CATABOL was created to predict the most probable biodegradation pathway, the distribution of stable metabolites and the extent of biological oxygen demand or CO_2 production compared to theoretical limits. It can be considered as a hybrid system, containing a knowledge-based expert system for predicting biotransformation pathway combined with a probabilistic model that calculates the overall BOD and/or extent CO_2 production. The CATABOL system is trained to predict biodegradation within 28 days on the basis of 743 chemicals from the MITI database and another training set of 109 proprietary chemicals from Procter & Gamble (P&G) obtained with the OECD 301C and OECD 301B tests, respectively. In the first database biodegradation is expressed as the oxygen uptake relative to theoretical uptake, while in the P&G database biodegradation is measured by carbon dioxide production. Version 5.097 used in this study only contains information of the MITI dataset.

CATABOL is based on two sources of information:

1. A training set containing 743 substances with measured BOD values in a MITI test.
2. A library of observed catabolic pathways of organic compounds from different sources including monographs, scientific articles and public web sites.

The MITI database

The largest available biodegradation database contains the so-called MITI-I test data, which comprises results of a single uniform biodegradation test for nearly 900 commercial chemicals. In the Japanese MITI I test a 'de-acclimated' inoculum is used, existing of a mixture of effluent, fresh surface water, estuarine marine water and soil extract (exclusively collected in Japan) which is extensively fed with glucose/pepton. This inoculum is not very representative for the real aquatic environment. That is also the reason that the MITI-I test, despite its high inoculum concentration in the Ready Test (30 mg/L dry matter active sludge) belongs to the most stringent methods. The MITI-I test is a screening test for 'ready' biodegradability in an aerobic aqueous medium and is described in OECD and EU test guidelines. The MITI-I test was developed in Japan, and it now constitutes one of the six standardised 'ready' biodegradability tests described by EU and OECD regulations. Biological oxygen demand (BOD) is measured continuously during the 28-day test period. The pass level for 'ready' biodegradability is reached, if the BOD amounts to ≥ 60% of theoretical oxygen demand (ThOD).

Catabolic pathways database

This database contains information on microbial biocatalytic reactions and biodegradation pathways for more than 550 chemicals. The collection includes the catabolism of C1-compounds, aliphatic hydrocarbons, alicyclic rings, furans, halogenated hydrocarbons, aromatic hydrocarbons and haloaromatics, amines, sulfonates, nitrates, nitro-derivatives, nitriles, and compounds containing more than one functional group. The catabolic pathways database was used to extract the principal transformations and train the system to simulate aerobic catabolism of training chemicals with measured BOD in MITI tests. For substances in the training set a measured BOD is available. Their transformation steps are based on an observed transformation scheme (for approximately 90 out of 743 substances) or on a pathway suggested by experts. The least squares estimates of the probabilities of occurrence of transformations were obtained by minimizing the sum of squares of residuals:

$$\min_{\mathbf{P}} RSS = \sum_{n=1}^{N} \left(BOD_n^{Obs.} - BOD_n^{Calc.} \right)^2 ,$$

where $BOD_n^{Obs.}$ and $BOD_n^{Calc.}$ are observed and predicted values of biodegradability, and N is the number of fitted data. The predicted biodegradability is given by the ratio of the calculated biological oxygen demand and its theoretical value:

$$BOD = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i}\Delta_{ij}^{O_2}\prod_{k=1}^{j}P_{ik}}{\sum_{i=1}^{I}\sum_{j=1}^{J_i}\Delta_{ij}^{O_2}}100,\%\,,$$

where $\Delta_{ij}^{O_2}$ are the biological oxygen demand of the transformations and P_{ik} are the probabilities of the transformations across the i^{th} pathway.

The probability hierarchy is used to create the most probable pathways and to predict BOD values for the training set. However, some transformations can be grouped because they have the similar source fragment and the same probability. Within these groups the hierarchy is established by expert judgement in which the effect of neighbouring groups is taken into account. For some transformations, fragments called ‘masks’ are attached to a source fragment. These inactivating fragments prevent the performance of a specific transformation. With the fitted probabilities the correlation between these predicted BODs and the observed BODs was 0.9.

For some of the substances in the training set the predicted BOD did not agree with the observed BOD. These structures are ‘out of domain’. The criteria for a good prediction have been connected to the reliability for a correct prediction of readily or not readily biodegradable. The areas for false positives (wrongly predicted as ‘readily biodegradable’) and false negatives (wrongly predicted as ‘not readily biodegradable’) represent the limitations of the applicability domain. A comparison of predicted and observed values for the training set, as well of areas out of the domain is sketched in Figure 4.3.

The properties of substances in the training set are crucial in the determination of the applicability domain. The applicability domain is defined as the group of chemicals for which the model is valid.
The applicability domain of CATABOL includes 3 components:

1. The general parametric requirements domain
2. The structure domain
3. The metabolisation domain.

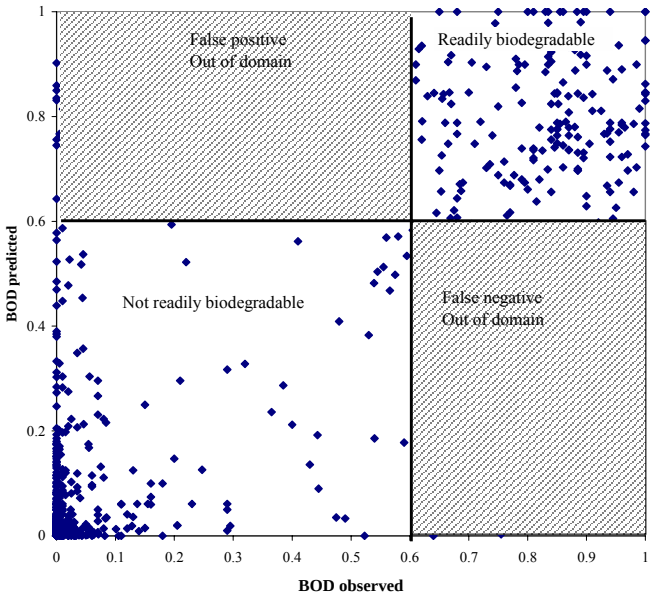


Figure 4.3: Comparison of observed and predicted BOD for substances in the training set.

The general parametric requirements restrict the applicability domain based upon variation of logKow and molecular weight of the training set. The range in these properties is given in Table 4.4. Although these substances do not directly affect the BOD, they may implicitly do so, for example through the bioavailability of the substance. The range of logKow (calculated by KOWWIN) and molecular weight for substances of the training set with a good prediction is given in Table 4.4.

Table 4.4: Range in general parameters for substances of the CATABOL training set with a good BOD prediction.

	min	max	median
Log Kow	-3.8	24	2.6
Molecular weight (g/mol)	44	959	168

The structure domain defines the structural similarity with chemicals that are correctly predicted by the model. It is based upon the principle that the properties of a substance depend on the nature of their atoms as well as on their arrangement. In order to check whether a new substance is in the structure domain its fragments are compared with those substances in the training set that had good BOD predictions. When the fragments of the substance of interest are not found in this group within the training set, the substance is considered ‘out of domain’. The limitations in the structure domain are very dependent upon the variety of structures in the training set; substances with unknown structural fragments are by definition ‘out of domain’. A technical description about how these molecular fragments are determined is described by Dimitrov et al. (2005). For substances that are ‘out of structure domain’ this does not mean that the structure is unknown to transformation library. A new substance, although ‘out of structure domain’ will be degraded according to the hierarchy and probabilities in the transformation library. However, the predicted BOD should be considered less reliable, because ‘out of structure domain’ only refers to the dissimilarity with substances in the training set that had a good BOD prediction

The third component of the domain is the ‘metabolisation domain’. A list of reactions included in the library is given in Tabel 4.5. The BOD is based on those pathways that can occur on familiar fragments of the molecule. Unknown fragments will remain as recalcitrant residues. Spontaneous reactions obtain a probability of 1, the probabilities of microbial reactions have been derived statistically from the training set. When a substance is ‘out of the metabolisation domain’, there is no pathway available for a particular (sub)structure. Structures that are unknown to the library do not contribute to the predicted BOD. Consequently, CATABOL is unable to mineralise the target substance and consequently the predicted BOD could be wrong.

The most severe violation of the components of the applicability domain is Metabolism Domain, followed by Structural Domain and finally the General Requirements.

Another measure of the quality of generated pathways is the reliability which is expressed in a value between 0 and 1. It is determined by making use of the reliability of transformations (their successive use versus their total use within the training chemicals). Reliability close to unity (1) means that all transformations used to generate a certain pathway were used correctly within the training set. The Reliability is close to 0 should be interpreted as a warning message that some of the used transformations may generate not realistic (not documented within the training set) pathways.

The interpretation of the combinations of ‘high reliability and out of domain’ or ‘low reliability and in domain’ needs some expert knowledge and should be solved case-by-case analysing causality for such a combination. Generally for BOD prediction: ‘high reliability and out of domain (General or Structural)’ is an indication that the prediction could be correct if the target chemical does not contain very ‘strange’ functionality. Substances with a ‘low reliability and in domain’ requires an analysis of the effect of the used transformations with low reliability on the predicted BOD.

Table 4.5: List of metabolic steps in the CATABOL library.

Spontaneous reactions	Microbially catalyzed reaction
Addition to ketenes and isocyanates	Alkyne hydrogenation
Alkaline salt hydrolysis	Aromatic ring cleavage
Aldehyde oxidation	Acetone degradation
Acyl halide hydrolysis	Aromatic ring oxidation
Alpha-pinene oxidation	Ammonium and iminium salt decomposition
Anhydride hydrolysis	Alkylammonium salt decomposition
Ammonium and iminium salt decomposition	Alkoxysilane hydrolysis
Alkoxide hydrolysis	Alkylphosphinite hydrolysis
Aromatic ring cleavage	Azo compounds reduction
Aziridine hydrolysis	Oxidative deamination and N-dealkylation
Benzotriazole tautomerism	Beta-oxidation
Carbamate hydrolysis	Baeyer-Villiger oxidation
Cyclopropane oxidative decyclization	Beckmann rearrangement
Cyanuric acid isomerization	Bisphenol A cleavage
Diketone and unsaturated ketone oxidation	Carboxylation
Geminal derivatives decomposition	Carbodiimide hydrolytic deamination
Hydrazine oxidation	Cycloalkadiene oxidative ring opening
Hydroxylation of substituted haloarenes	Diketone and unsaturated ketone oxidation
Hydroperoxide decomposition	Decarboxylation
Keto-enol tautomerism	Dehalogenation
Lactone hydrolysis and formation	Diarylketone oxidation
N-nitrosoamine hydrolysis and reduction	Dibenzofuran oxidative degradation
Nitrate ester denitration	Epoxidation
Oxidative denitrification of azides and hydrazones	Ester hydrolysis
Oxirane hydration	Furans oxidation
Primary hydroxyl group oxidation	Hexahydrotriazine hydrolytic ring opening
Phosphine oxidation	Imine reduction
Polyphosphate decomposition	Imidazole and triazole C-hydroxylation
Quinone reduction	Lactone hydrolysis and formation
Reductive deamination	Methyl group oxidation
Thiophosphate oxidative desulfuration	Nitrogroup reduction and nitrite release
Thiol-thion tautomerism	Nitrile and amide hydrolysis
Tetrahydrofuran oxidation	Omega oxidation
Thiol oxidation and reduction	Organotin compound oxidation
Thiolic acid and thioester hydrolysis	Oxidative desulfonation
	Oxidative thion desulfuration
	Oxidative S-dealkylation
	Organic sulfide S-oxidation
	Oxidative desulfuration
	Oxidative O-dealkylation
	Perfluoroketone degradation
	Pyridinium salt decomposition
	Phosphate hydrolysis
	Pyridine and azine ring oxidation
	Reductive deamination
	Sulfate hydrolysis
	Subterminal oxidation
	Sulfoxide reduction
	Sulfonyl derivative hydrolysis
	Thiol oxidation and reduction
	Tin and lead carboxylate hydrolysis

4.7.2 Selected data for analysing the extrapolation capabilities of CATABOL

When a new substance is entered, CATABOL predicts the degradation steps based on the hierarchy in the transformation library and it computes the BOD with the probabilities assigned to each transformation step. The applicability of the model for the substance of interest is also generated as output.

Selection of data

Several groups of chemicals have been selected in order to gain insight in the extrapolation capabilities of CATABOL (Version 5.097) with respect to BOD prediction and prediction of metabolites.

- Existing chemicals listed on Annex I with R50 and R53 (152)
- New substances (89)
- PBT substances (76)
- Pesticides (192)

A complete list of substance names and CAS number is given in Appendix III.

4.7.2.1 Chemicals listed on the Annex I entries

The research time was too limited to determine of each chemicals listed on Annex I whether the experimental data indicate that they are degradable or non-degradable. For this reason it was decided to select these chemicals which are classified with either R50 or R53. Chemicals that are classified with R50 only are very toxic but should be rapid degradable. This in contrast to chemicals classified with R50/53, which should be very toxic and not rapidly degradable **or** bioaccumulative. Chemicals classified with R53 only should always be not rapidly degradable **and** bioaccumulative.

Annex I has 3366 entries. For 1941 of these 3366 single chemical structures could be assigned. 1488 of these 1941 have at least one carbon atom, the others are inorganic, for which CATABOL can not predict anything. From the 523 chemicals classified with R50, 464 are also labelled as R53 and 59 chemicals with R50 only. For all of these 59 structures the SMILES notations were available. 148 substances of the 1488 chemicals are classified as R53 only. For 93 SMILES notation were available at RIVM or found on the SPARC website (http://sparc.chem.uga.edu/sparc/search/searchcas.cfm).

Thus a dataset of 59 chemicals that are rapidly degradable and 93 chemicals that are not rapidly biodegradable were available for evaluation of the CATABOL MITI-I %BOD prediction.

4.7.2.2 New Chemicals (89)

For a selection of 83 new substances notified in the Netherlands, a comparison is made with the predicted and the observed BOD Most of the observed BOD were determined in a modified Sturm test (OECD 301B). Only for 4 substances the BOD was determined in a MITI test (OECD 301C).

4.7.2.3 Potential PBT substances (76)

From the list of potential PBT and vPvB substances discussed by the WG group on PBT assessment, 76 substances were selected to compare the CATABOL predicted BOD and major metabolites formed with the experimental observations.

4.7.2.4 Pesticides

For the verification of CATABOL two sets of pesticides have been used.

1. A list of all registered plant protection products in the Netherlands is used, in order to cover a wide variety of chemical structures and functionalities. With this set of pesticides the relation between predicted BOD and soil parent half-lifetimes (the so called primary half-live) is assessed, as well as the occurrence of major metabolites. Biological pesticides (pyrethrins) are eliminated as well as polymeric structures (mancozeb, maneb, metiram), which can not be entered in CATABOL. Thus, a list of 192 substances is obtained. A diagram is sketched were the selection and verification procedure is drawn (Figure 4.4)
2. A list with substances that share major structural features. To gain insight in the extrapolative capabilities when applied to a well-defined domain of chemical structures, CATABOL was used to predict metabolite formation of 16 carbamates and 18 organo-phosphate esters. These two classes of chemicals were selected as they constitute two classes of chemicals that, given their emission patterns, and given their fate and effect profiles, may pose a risk for the environment. In Table 4.6 chemical names of the substances are given and it is indicated which substances were not in the list of registered pesticides.

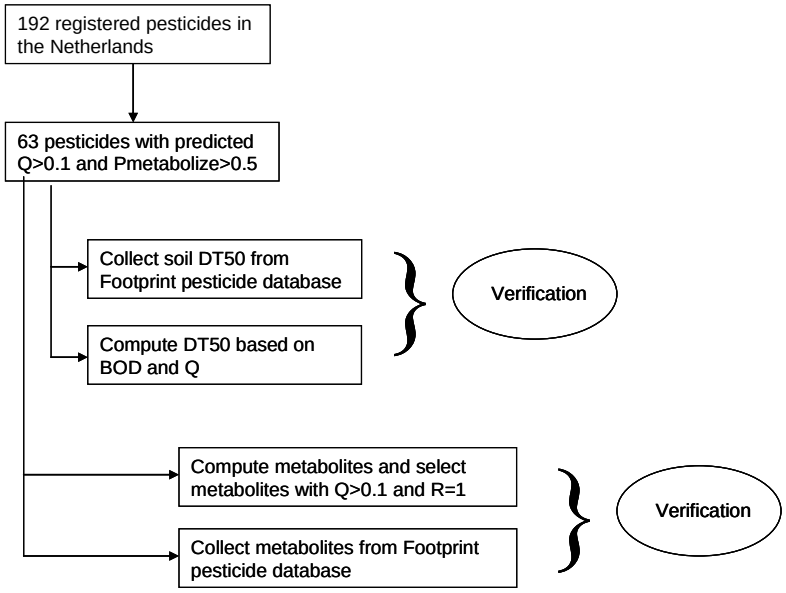


Figure 4.4: Overview of selected pesticides for CATABOL verification (Q = predicted quantity of parent compound after 28d in a MITI test).

The predicted DT50 is derived from the predicted BOD by the following formula:

$DT50_{BOD} = 14/BOD$

The primary half-life derived from this linear relationship with BOD is an approach for the estimation of the worst-case DT50. If the relationship is not linear but S-shaped or 'hockey-stick'-like with significant lag phase, the linear approximation tends to underestimate the biodegradation rates (i.e. overestimate the DT50). Moreover, the DT50 derived from BOD is determined by the slowest step in the degradation pathway, and thus can be considered as an upper limit value for the DT50 of the parent.

The formula above is based on the fact that a MITI test lasts 28 days. When a BOD of 1 is reached in 28 days, a complete degradation is obtained. As the shape of the BOD-curve in the period between 0 and 28 days is unknown a linear progression is assumed. The half-lifetime in that case would be 14 days. So for test substances with a lower BOD, the half-lifetime would increase with a factor equal to 1/BOD. Similarly, DT50 can also be estimated from predicted quantities of the substance after 28 days, by the following formula:

$DT50_Q = 14/(1-Q)$

Linear approximation of time dependent BOD or Q could be useful for not degradable chemicals within the studied period. Usually the relationship is not linear but S-shaped or like 'hockey-stick' with significant lag phase

Table 4.6: Selected carbamates and O-P and S-P esters. Pesticides not registered in the Netherlands are indicated in bold.

O-P and S-P esters	Carbamates
Chlorfenvinphos	Aldicarb
Diazinon	**Propoxur**
Parathion, methyl	1-Naphthalenol, methylcarbamate
Fenitrothion	**Nabam**
Malathion	Carbofuran
Methylazinphos	Methiocarb
Phosphoric acid, 2,2-dichloroethenyl, dimethyl ester	**Phenol,2-(1-methylethyl)-,methylcarbamate**
Dimethoate	**Trimethacarb**
Parathion	**Phenol, 2-(1-methylpropyl)-, methylcarbamate**
Fenthion	Methomyl
Dipterex	**1,3-Benzodioxol-4-ol,2,2-dimethyl-**
Ethoprophos	Pirimicarb
Profenofos	**Oxamyl**
Phosphamidon	**Butoxycarboxim**
Methamidphos	Thiodicarb - symmetrical carbamate
Demeton	**Benfuracarb**
Fonophos	
Phosmet	

Smiles are obtained from http://sparc.chem.uga.edu/sparc/search/searchcas.cfm (assessed: nov. 2006) or created manually. Experimental data, half-lives in soil and major metabolites are obtained from the Footprint database. The database used as the source for experimental data on metabolite formation was: http://www.herts.ac.uk/aeru/footprint/. The website provides the following indication of its content: 'The best sources of information currently available for pesticide properties are the monographs produced as part of the EU review process and these documents have been used in priority for putting together the FOOTPRINT PPDB. Where EU documents were not available, alternative sources were used:

- Databases and documents from various national government departments including the UK's PSD, Germany's Federal Environment Bureau, the EPA in the USA and the French Agritox database.
- On-line databases including ARIS, EXTOXNET, ARS/OSU, PAN, GLEAMS, etc.
- Manufacturers safety datasheets and environmental fact sheets, on- and off-line.
- Publications such as the Pesticide Manual.
- Data derived from research projects such as the Pandora data set.
- Peer reviewed scientific publications.

In a very limited of instances, data had to be retrieved from miscellaneous on-line sources. All data held in the FOOTPRINT PPDB are 'tagged' with a code so that their source and quality can be identified (see below)'.

4.7.3 Results –BOD predictions

4.7.3.1 Overview of BOD predictions

In Table 4.7 is shown how many substances in several groups of chemicals are considered 'readily' or 'not readily biodegradable' by CATABOL as well as the fraction out of domain and false positive and false negative predications. In the next paragraphs a further discrimination of predictions 'in the domain' and 'not in the domain' is presented.

Table 4.7: Overview of CATABOL performance and BOD predictions for several groups of chemicals

	Existing substances		New substances	PBT	Pesticides	Carbamates	O-P and S-P esters
	R50	R53					
n total	59	93	83	76	192	16	18
in domain (%)	43	39	9	43	7	25	22
not in general parameter domain (%)	0	0	4		3	0	0
not in structure domain (%)	57	61	73	43	93	75	78
not in metabolization domain (%)	0	4	2	8	7	0	0
in training set (MITI) (%)	0	18	0	16	2		
Predicted Ready Biodegradable	51	8	14	11	2		
Predicted Not Readily Biodegradable	49	92	69	89	98		
False positives (%)	-	8	5*	5*	n.d.	n.d.	n.d.
False negatives (%)	49	-	7*	3*	n.d.	n.d.	n.d.

* all 'out of domain'

4.7.3.2 Existing chemicals listed on Annex I

Of the structures submitted to CATABOL forecast approximately 60% is out of the structure domain (34/59 and 57/93). CATABOL can still make predictions for substances that are not in the structure domain; it is verified how good these prediction are. Of the R50-51 not R53 set, none of these substances occurs in the MITI training set. It follows that 59 'ready' substances are available for an external verification of CATABOL.
Of the R53 not R50-51 set, 17 substances also occurred in the MITI training set. It follows that 76 substances are available for an external verification.

R50 substances

For an external verification 59 substances were available which are classified in Annex I as R50 implying that these substances should be rapidly degradable (see Table 4.8). For the prediction with CATABOL, 25 of these substances were in the applicability domain and 28% (=7/25) of the substances were predicted 'not readily biodegradable' (= false negative) by CATABOL. For substances not in the applicability domain, 65% (=22/34) is predicted as 'not readily biodegradable' (= false negative). It is evident that the applicability domain significantly discriminates the predictions according to their reliability.

Via the N-class data base an attempt has been made to find the reasons for considering these substances as rapidly degradable. For 5 substances R53 was not applied because they degrade to non-classifiable compounds. Five other compounds were persistent in a Ready Test but

degrade rapidly in soil or water/sediment. For 13 substances no information could be found. Only for four compounds it was obvious that they were readily biodegradable. So in fact, the number of real false negative could still be limited.

Table 4.8: Comparison of CATABOL BOD predictions for substances 'in the domain' and 'out of domain' for Annex 1 substances that are not in the training set. n= number of substances.

	R50	R53
n total	59	76
n in structure domain	25	20
False positive (%)		20%
False negative (%)	28	
n not in structure domain	34	56
False positive (%)		3%
False negative (%)	65%	

R53 substances
For an external verification 76 substances were available classified in Annex I as 'not ready' (see Table 4.8). For the prediction with CATABOL 20 of these substances were in the applicability domain. 20% (=4/20) of the predictions as were not confirmed by the Annex I classification, and is regarded as a false positive outcome. For substances not in the applicability domain, 4% (=2/54) is predicted as 'readily biodegradable'. Because of the relative small number of substances the difference between predictions in the domain and prediction on substances not in the domain should be considered as a coincidence. Moreover there is not a plausible explanation for better predictions for substances that are not in the domain than for substances that are in the domain.

For both categories (R50 and R53 substances) there is a tendency for more false negative predictions (28-65%) average 49% than for false positive predictions (4-20%), average 8%. However, if the model applicability domain is accounted for, the performance of the model is similar for both classes, averaging 25% for false negatives and positives.

When CATABOL would be used in the classification and labelling process, for situations were experimental information is lacking, the occurrence of false positives is very undesirable. For a cautious approach, the occurrence of false negatives is acceptable for regulators. However, too many 'not ready' classifications lead to increased cost for testing and risk assessments, which is unfavourable for the industries. A number of chemicals are declassified because of primary degradation to non-classifiable substances.

4.7.3.3 New substances
The observed and predicted BOD for each of the 83 new chemicals are presented in Appendix I. Like for the existing substances, the majority of the new substances were out of structural domain (88%). All chemicals identified to belong to the applicability domain were predicted with high accuracy ($R^2 = 0.87$) and classified correctly. It should be mentioned that ignoring of the applicability domain worsened the correlation to $R^2 = 0.59$. A small number of chemicals were out of the metabolisation domain (2%). From the 83 new substances analysed, the majority was not ready biodegradable (83%). Most of them (90%) were also predicted as not ready biodegradable. With respect to the ready biodegradable substances, 9 out of 14 substances (65%) were predicted correctly by CATABOL. The number of false positive and negatives were 5 and 7, respectively. All these substances were out of the

structural domain (see Table 4.7). Three of the false positive substances had a long alkyl chain, which might be less available for degradation than predicted. A comparison between predicted and observed BOD for the new substances is given in Figure 4.5

4.7.3.4 PBT substances

Like for the previous two groups of chemicals, also a large part of the PBT substances were out of domain (57%) and 43% were in the structural domain (see Table 4.7). As the PBT substances were selected to be potential persistent substances, most of them were not ready biodegradable (52/77). For 20 substances no screening data were available. Most of the substances were also correctly predicted as not ready biodegradable by CATABOL. Five substances appear to be ready biodegradable in a further testing procedure. Three of them were also predicted to be ready biodegradable by CATABOL. In contrast to the observation, four substances were predicted ready biodegradable. In comparison to new chemicals, three of these substances (i.e. dichlorodioctylstannane, phenol, styrenated and dodecylphenol) had again a long alkyl chain, which was apparently less bioavailable for biodegradation than predicted. All incorrect predictions, except one, were done for substances out of structural domain.

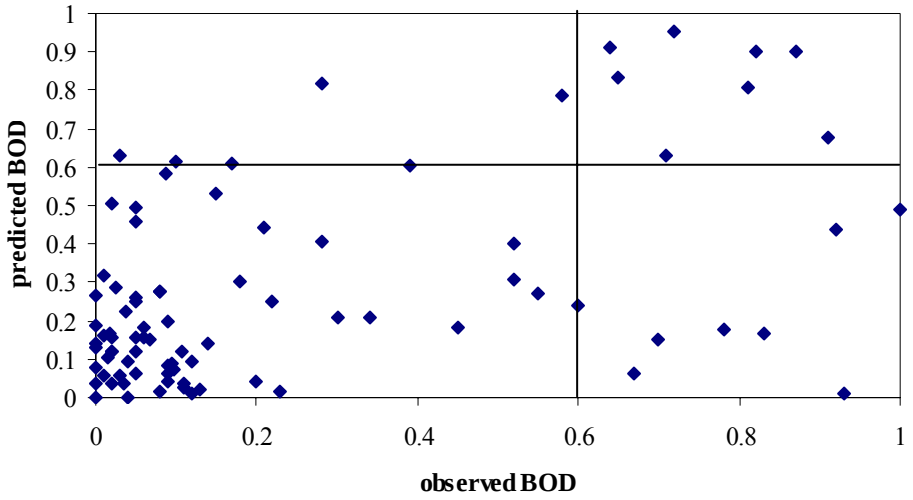


Figure 4.5: Comparison of observed and predicted BOD for new substances (correlation: 0.59). Substances in de upper left side box are incorrectly predicted to be readily biodegradable. Substances in de lower right side box are incorrectly predicted to be non-readily biodegradable.

4.7.3.5 Pesticides.

Pesticides are usually not tested in a MITI test or other ready biodegradability test. Therefore the goodness of BOD prediction can not be determined. It is striking that only 7% of the submitted pesticides are in the applicability domain, and that a large majority of the pesticide is predicted to be not readily biodegradable. For the large group of pesticides soil simulation studies and field studies are available. One of the research questions was to assess whether there is a relation between BOD predictions and half-lives obtained from soil simulation studies.

In ***Figure 4.6*** the predicted half-lives based on BOD are compared with soil degradation half-lives. The line represents the situation where predicted and measured half-lives match exactly. The dotted lines represent a variation of a factor of 2, which is done to show the natural variation of the DT50 in several soil types and between experiments. The figure shows that CATABOL predictions are on the safe side, with respect to environmental risks; DT50s are generally overestimated. This is very understandable because the CATABOL half-lives are derived from BOD which expresses complete mineralization, whereas experimental half-lives are based on disappearance of the parent compound. In the last case primary degradation contributes to the lower experimental half-lives.

The consequence for an ITS approach could be to use half-lives derived from BOD prediction with CATABOL, when experimental data are not available. In this case only in 1 out of 192 substances (=0.5%) the use of the CATABOL prediction for half-lives would lead to underestimation of environmental risks.

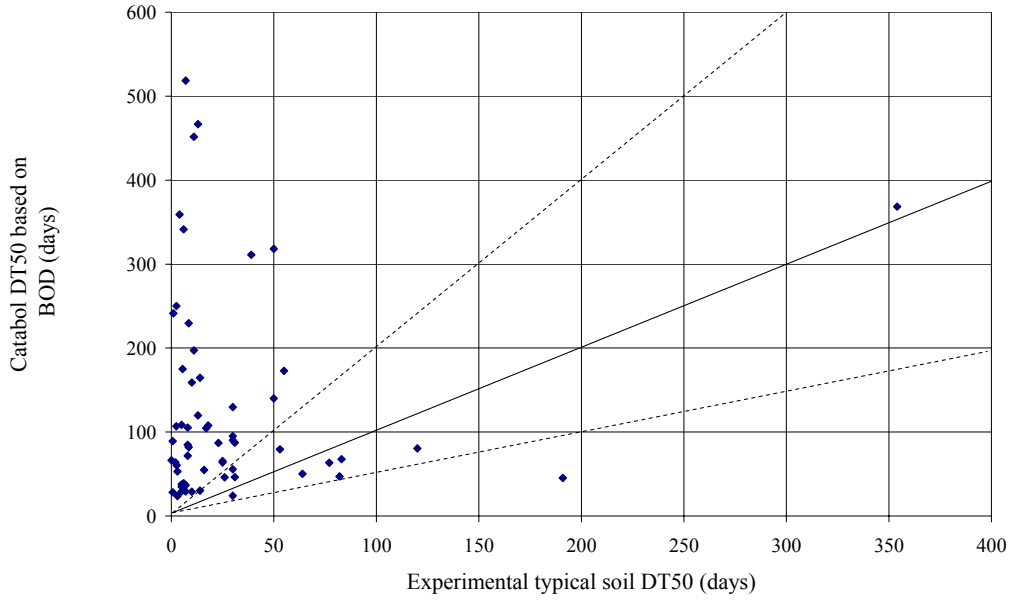


Figure 4.6: Comparison of measured soil half-lives of registered plant protection products in the Netherlands and predicted half-lifetimes based on BOD. Straight lines represent line of perfect match. Dotted lines represent natural variation of experimental data with factor of 2.

The use of remaining quantities of the parent for the estimation of the DT50, takes partial degradation into account, like is the case for experimental half-lives. One would expect a better match. This is true (see Figure 4.7), however, there are more cases where CATABOL underestimates the half-live, and consequently the environmental risks (13 out of 192 = 7% of the substances are underestimated). This makes its use in an ITS strategy less suitable.

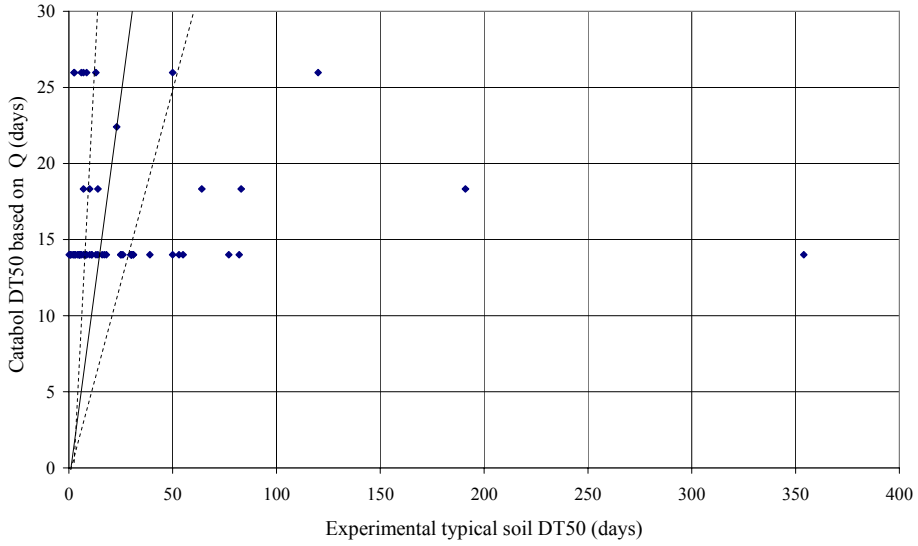


Figure 4.7: Comparison of measured soil half-lives of registered plant protection products in the Netherlands and predicted half-lifetimes based on remaining quantities. Straight line represents line of perfect match. Dotted lines represent natural variation of experimental data with factor 2.

4.7.4 Results - Metabolites

4.7.4.1 Existing substances

Ready substances are mineralized for at least 60%. This implicates that there will not accumulate a significant amount of metabolites. The production of metabolites can be expected form the 'not readily biodegradable' parent substances.

Of the 85 parent substances (see Table 4.7) that are not readily biodegradable, 52 substances (=61%) are indicated to undergo substantial transformations, however no complete mineralization. When quantities are predicted of less than 0.5 it means that 50% of the parent substance is metabolized within 28 days. This is considered a moderately rapid biodegradation. In these cases the risk evaluation should focus on persistent metabolites and not on the parent substance.

4.7.4.2 New chemicals

For the new chemicals analysed no information on the degradation products is available.

4.7.4.3 Substances discussed by WG on PBT assessment

In the absence of sufficient information, verification of the biodegradation pathway is only possible for a limited number of chemicals. Based on the available information on the metabolites formed in soil and water, there are a number of compounds for which CATABOL gives a right prediction of the biodegradation pathway, like for 2,6-di-tert-butylphenol, 4,4'-methylenedicyclohexyl_diisocyanate, octadecyl_3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate, and triphenylphosphine. The last three substances were out of structural domain.

There is also a number of substances for which the predicted degradation pathway is incorrect, as for 1-(5,6,7,8-tetrahydro-3,5,5,6,8,8-hexamethyl-2-naphthyl)ethan-1-one (AHTN), 1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethylindeno[5,6-c]pyran (HHCB), N,N-dicyclohexylbenzothiazole-2-sulphenamide and N-tert-butylbenzothiazole-2-sulphenamide. The first two chemicals are transformed at the tetrahydropyrane site to form a lactone and hydroxyl acid, whereas CATABOL predicts that degradation starts with the oxidation of the methyl on the cyclopentane part of the substance. With respect to the other two chemicals, it appears that the sulphenamide binding will be hydrolysed at first, whereas with CATABOL this binding will be stable. Consequently, none of the metabolites observed for N-tert-butylbenzothiazole-2-sulphenamide like mercaptobenzothiazole, di(benzothiazoyl-2)disulfide, t-butylamine, and benzothiazole where predicted by CATABOL. All the chemicals were out of domain.
For a large number of chemicals, CATABOL predicts that the first step in the degradation pathway has a low probability (<0.2), indicating that the parent compound is persistent. Based on the low BOD observed in the Ready Test this is confirmed for 24 of the substances (See Appendix II).

4.7.4.4 Pesticides

In the risk assessment of plant protection products ‘Ready Tests’ do not play an important role. For the registration process amongst others several degradation studies in soil are demanded as well and hydrolysis, photolysis studies.

For the verification of CATABOL the list of all registered plant protection products in the Netherlands is used. Biological pesticides (pyrethrins) are eliminated as well as polymeric structures (mancozeb, maneb, metiram), which can not be entered in CATABOL. Of pesticides that consist of several substances, all the individual components are entered. Thus, a list of 192 substances is obtained (see Appendix III).

Six substances were in the training set. 178 of the 192 substances are out of the domain (see Table 4.7), mainly due to lack of similar structures with a good BOD prediction in the training set (178 substances out of model structure domain). This affects the reliability of the BOD predictions. However, there is a considerable amount of information on metabolisation schemes, as is demonstrated by the fact that only 13 of the substances are out of the metabolisation domain.

For the complete set of 192 substances the BOD and the metabolic pathway have been computed. Only 4 of the 192 showed BOD prediction >0.6. Most of the substances would thus be considered not readily biodegradable. However, when remaining quantities of parent substances are taken into account, it is shown that many of these not readily degradable substances are partially degraded; 144 of the substances show partial degradation of at least 50%. This implies that in that case not the parent itself is persistent, but that one or more metabolites obstruct the mineralization process.

In this paragraph the metabolic products that CATABOL predicts are compared with information on major metabolites. For a verification of metabolic pathway a selection out of the set of 192 pesticides was made. The focus is on the prediction of relevant metabolites only. Relevant metabolites are defined as transformation products that can be found in quantities >10% of parent compound. Production of CO_2, H_2O, small organic molecules and inorganic fragments are not considered as major metabolites. Pesticides with predicted

metabolites in quantities >0.1 (10%) in combination with reliability of 1 are selected. This resulted in 63 pesticides for the analysis of major metabolites.

In Figure 4.8 an overview is given of the comparison of major metabolites predicted by CATABOL and reported in the Footprint database.

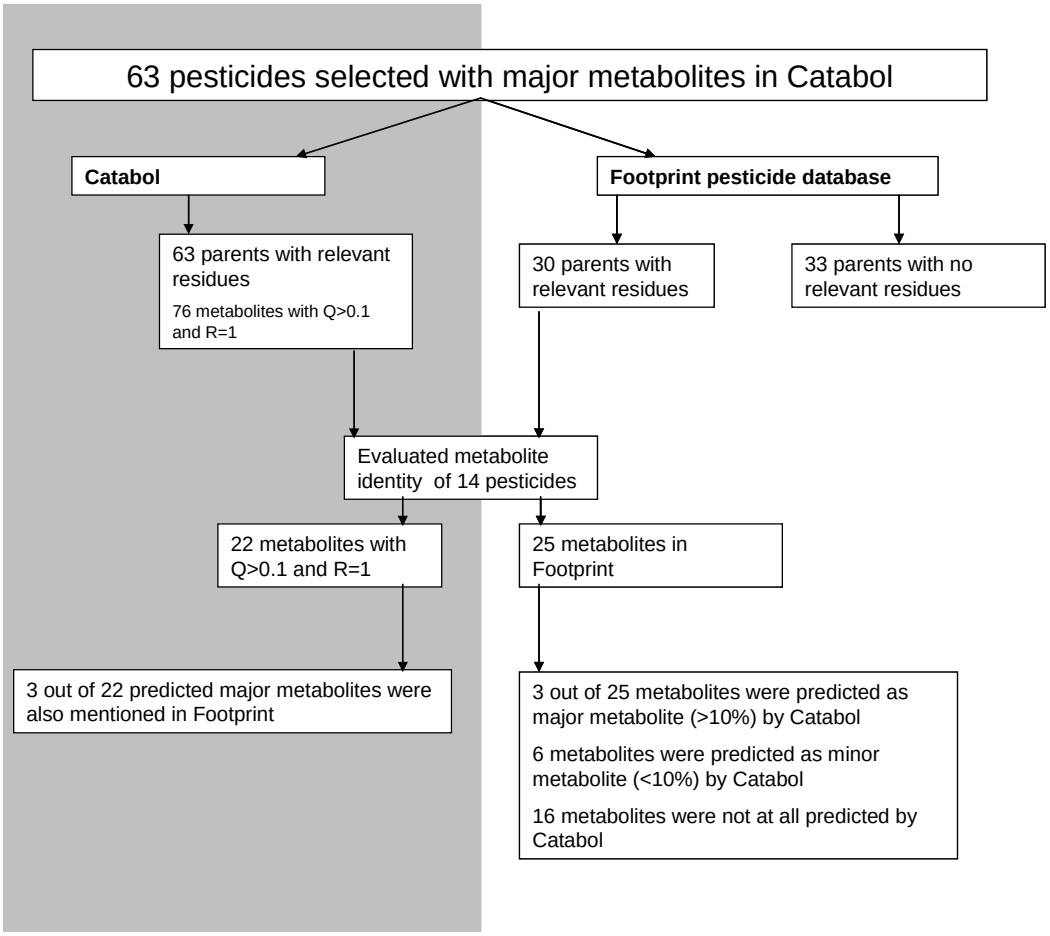


Figure 4.8: Overview of performance of metabolite prediction by CATABOL

CATABOL predicted that the 63 selected pesticides would produce 76 major metabolites (with R=1). However Footprint did not report any major metabolite for 33 of these pesticides. The remaining 30 pesticides for which Footprint did report major metabolites were selected for further comparison with predicted metabolites. Starting in alphabetical order, the analysis could be done for 14 of the 30 pesticides, because no more time was left.

The 14 pesticides for which major metabolites have been compared and the results of the comparison are given in Table 4.9. Only one of the analysed 14 pesticides was identified to belong to the applicability domain of the model and the documented metabolite in soil was predicted by CATABOL as a minor one. Evaluating the agreement between the Footprint database and the CATABOL prediction for the remaining 13 pesticides of the domain it is found that there is a good agreement for 2 pesticides: florasulam and hymexazole. The occurrence of 16 major metabolites in soil was not predicted by CATABOL. Another 5 metabolites were predicted by the pathway in CATABOL, though the amount suggests that it concerns minor metabolites. Moreover, 19 metabolites indicated by CATABOL as potential major metabolites, are not observed as major metabolites (as reported by the Footprint

database). The latter can also be caused by the fact that experiments have not paid attention to the possible formation of these metabolites, or that these metabolites are not relevant from a toxicological point of view. These results, however, should be analysed from the position that CATABOL was not trained to predict the fate of chemicals in soil.

Table 4.9: Verification results of major metabolite identity

Chem. Name	TotalDomain	Soil Metabolites in Footprint database	CATABOL
Aldicarb	Out of Domain	aldicarbsulfoxide (2-methyl-2-(methylsulfinyl)propanal O-((methylamino)carbonyl)oxime) ;	not predicted
		aldoxycarb = aldicarb-sulfon (2-mesyl-2-methylpropional O-methylcarbamoyloxime)	not predicted
Bentazone	Out of Domain	2-amino-N-isopropylbenzamide	not predicted
Bifenazate	Out of Domain	D2341-diazene (Ref: D3598);	not predicted
		4-methoxybiphenol (Ref: D1989)	predicted as minor metabolite (7.9%)
Clorpyrifos	In domain	3,5,6-trichloro-2-pyridinol (TCP) casnr 6515-38-4	predicted as minor metabolite
Famoxadone	Out of Domain	alpha-hydroxy-alpha-methyl-4-phenoxybenzene acetic acid (Ref: IN-JS940);	predicted as minor metabolite (5.9%)
		5-[4-(4-hydroxyphenoxy)phenyl-5-methyl-3-(phenylamino)-2,4-oxazolidine dione (Ref: IN-KZ007)	not predicted
Fenoxaprop-P-ethyl	Out of Domain	(D+)-2-(4-(6-chloro-2-benzoxazolyloxy)-phenoxy)-propionic acid (Ref: HOE 88406)	not predicted
Florasulam	Out of Domain	5-hydroxy florasulam;	not predicted
		(DFP-ASTCA);	predicted as minor metabolite (2.9%)
		(ASTCA)	not predicted
Fluoxastrobin	Out of Domain	HEC-5725-des-chlorophenyl (Ref: M48)	not predicted
		HEC-5725-carboxylic acid (Ref: M40)	not predicted
Folpet	Out of Domain	phthalimide	not predicted
		phthalic acid	predicted as minor metabolite (3.2%)
		phthalamic acid	not predicted
Foramsulfuron	Out of Domain	4-amino-2-[3-(4,6-dimethoxypyrimidin-2-yl)ureidosulfonyl]-N, N-dimethylbenzamide (Ref: AE F130619)	Good prediction
		dimethoxyaminopyrimidine (Ref: AE F092944)	Good prediction
Fosthiazate	Out of Domain	2-butanesulfonic acid (BSA)	not predicted
Hexythiazox	Out of Domain	5-(4-chlorophenyl)-N-(4-oxocyclohexyl)-4-methyl-2-oxothiazolidine-3-carboxamide (Ref: PT-1-9)	not predicted
		trans-5-(4-chlorophenyl)-4-methyl-2-oxothiazolidine-3-carboximide (Ref: PT-1-2)	not predicted
		trans-5-(4-chlorophenyl)-4-methyl-2-oxothiazolidine (Ref: PT-1-3)	not predicted
Hymexazole	Out of Domain	5-methyl-2-(3H)-oxazolone	Good prediction
Imidacloprid	Out of Domain	6-chloronicotinic acid	predicted as minor metabolite (0%)

4.7.4.5 Carbamates and O-P and S-P esters

Experimental data

Experimental data on metabolite formation were available for 10 out the 16 (63%) carbamates, which implies that experimental data were lacking for the remaining 6 compounds (37%). In case of organophosphates, experimental data were lacking for 12 of the 18 compounds

considered (67%), clearly highlighting the need of having available estimation methods for predicting metabolite formation.

CATABOL predictions
Twelve of the sixteen carbamates studied (75%) were out of the structure domain of the CATABOL model and hence also out of the total domain. Experimental BOD-data from the MITI test were available for three of the four carbamates that were in the structure domain. Fourteen of the eighteen organophosphorus esters (78%) were out of the structure domain of the model and hence also out of the total domain. Experimental BOD-data from the MITI test were available for three of the four organophosphates that were in the structure domain.

Comparison between CATABOL predictions and experimental data
CATABOL in general predicts the formation of a large number of stable and less stable intermediates. This is done according to a hierarchic systematic. When comparing predictions and experimental data, it is important to realize that reports on metabolites observed during biodegradation testing in soil, are not always encompassing 'all' metabolites formed. Apart from analytical constraints (detection limits), there often is bias with regard to the expected metabolites formed. The latter implies that not all metabolites predicted to be formed by CATABOL were taken into consideration at the time of deciding on which metabolites to focus on during the experimental degradation study. When comparing predicted and observed metabolites, it is important to take note of these (and additional) considerations. In this study the focus was therefore not on matching individual metabolites. Instead the focus was on deducing from the molecular structure of the metabolites reported to be formed during biodegradation in soil, whether these metabolites could be formed as the outcome of the pathways predicted by CATABOL. Or, in other words, the focus was not on individual compounds but on degradation pathways, investigating the possibility of the reported metabolites having the possibility of being formed along one of the degradation pathways predicted by CATABOL to take place. To provide an example: a common transformation step for S-P-esters containing a thiobenzene-group is the oxidation of the S-atom attached to the benzene-moiety. This transformation leaves the S=P linkage intact and yields the corresponding sulfoxide as a stable metabolite. CATABOL, however, predicts the oxidation of the P=S moiety to yield the corresponding P=O ester, with a probability of 1. In this case, the experimentally observed S-P esters, oxidized at the benzenic S-atom cannot be formed along the CATABOL-predicted pathway of S=P oxidation. The observed and predicted metabolites are presented in Appendix IV. This is graphically illustrated in Figure 4.9.

Carbamates
CATABOL predictions matched the experimental findings for 3 out of the 10 compounds for which experimental data were available: for 7 of the carbamates studied with experimental data on metabolite formation, the experimentally observed metabolites cannot be formed along the pathways predicted by CATABOL. Of the three compounds for which the CATABOL predictions were correct, one carbamate was in the structural domain and two were out of domain.

Organophosphates
CATABOL predictions matched the experimental findings for 2 out of the 6 compounds for which experimental data were available. This was not the case for 4 organophosphates studied (67%). Of the 2 compounds for which the CATABOL predictions were correct, one was in the structural domain and one was out of domain.

Overall
In Appendix III a more detailed comparison is given of the metabolites predicted to be formed by CATABOL (indicated as 'est.'), and the metabolites actually found in soil (indicated as 'exp.').

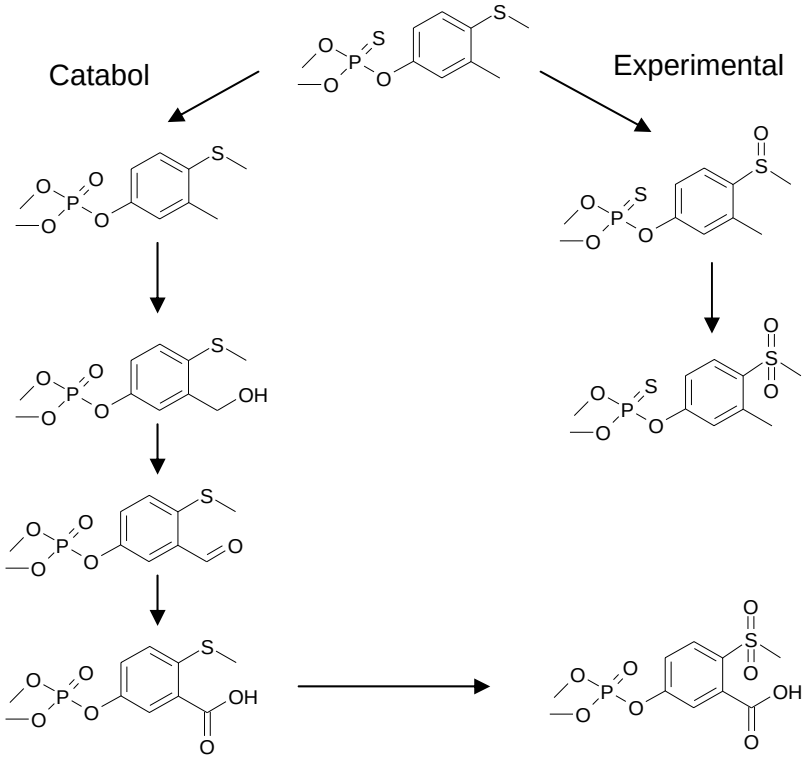


Figure 4.9: Illustration of the difference between the CATABOL-predicted metabolites of fenthion and the metabolites found in soil.

4.8 Discussion and conclusion

This study concerned the applicability of CATABOL for regulatory purposes. In this respect several aspects should be considered:

- Whether or not substances fall within the applicability domain.
- The reliability to predict biodegradability in an environment different from the training test conditions used to develop the model.
- The reliability to predict the metabolites and their relative quantities observed in an environment different from the training test conditions used to develop the model.
- Criteria to define when CATABOL outcomes are reliable enough to be used for regulatory purpose (for BOD as well as the metabolic pathway).

The majority of substances selected falls within the general parameter domain and within the metabolisation domain. This implies that CATABOL's transformation library has a wide coverage for degradation of these substances. However, as a large number of substances investigated are out of the structural domain (57-93%), the BOD predictions and in concomitant the probabilities of the metabolic pathways are not supported by the training set. Nevertheless, for the selected chemicals the false positives (predicted ready biodegradable – observed: not readily biodegradable) where limited to 5-8%.

The OECD QSAR Validation Principles recommend that the models are only used for substances that fall within the applicability domain. As the majority of the substances investigated were out of structure domain, this would limit the use of CATABOL. However, when the goodness of the prediction is related to the cut-off value of 60% BOD, CATABOL performs well in predicting 'not readily biodegradability', irrespective of whether the substance belongs to the structure domain. For ready biodegradable substance the performance is less. As the chemicals selected in this study were in majority not readily biodegradable, they are less suitable to draw a conclusion on the goodness of 'ready' predictions.

Verification of major metabolites of some existing chemicals and pesticides formed reveals that the predictive power of CATABOL for identifying major metabolites formed in soil is poor: many major metabolites are predicted that are not observed and many observed metabolites are not predicted. The majority of these substances (>90%) are 'out of the domain'. As the number of chemicals 'in the domain' is very small, it is not possible to assess whether the performance is better for substances that 'in the domain'. Our findings correspond with a comparable verification done with an earlier version of CATABOL by Sinclair et al. (2003) who found that only 24% of the experimentally derived degradation products were predicted correctly. One explanation for the mismatch in metabolite predictions could be that in reality a large number of different micro-organisms are present in sludge, water, sediment and soil. Therefore different enzymatic routes / degradation pathways can occur whereas CATABOL will only provide the user with a single degradation pathway, reducing the number of primary transformation products. Micro-organisms are generally specialised and could break down substances via routes which on forehand are not optimal. Another explanation is that the hierarchy (probabilities) of the pathway in the transformation library is not correct, or that relevant transformation steps for soil are missing. It also appeared that a number of chemicals with a long alkyl-chain was predicted as readily biodegradable whereas no degradation was observed. Apparently these compounds could be less bioavailable in the media used and as a consequence less accessible for enzymatic attack.

Although the metabolites in soil are poorly predicted, the degradation rates in soil can be reasonably estimated by deriving DT50 from BOD or remaining quantities in soil. Values predicted in this way could be used as upper limit values in risk assessments when no data are available. The chance of underestimation of degradability is small.

Despite the poor prediction of metabolites CATABOL still predicts well the 'ready biodegradability' of the substance. For a large number of chemicals (see section 4.7.3), the screening data indicate that the parent compound should be persistent which is confirmed by CATABOL. Apparently, CATABOL is able to assess whether the transformation will be blocked by a recalcitrant fragment, without being able to predict the exact identity of the recalcitrant degradation product.

CATABOL also reports reliability of the biochemical pathway, which was often below 0.5. As there seems to be no relation with the applicability domain, the situation can occur that substances that are 'out of domain' are still having a high reliability or that substances that are 'in the domain' have very low reliabilities. It is therefore still unclear how this reliability index should be used.

Despite the uncertainties discussed above and the fact that many substances are still out of domain CATABOL predictions could still be useful in a Weight-of-Evidence approach or to

target further testing.For substances for which little or no information is available, CATABOL could be used to investigate whether primary degradation is likely to occur and whether the metabolites formed are of concern by predicting e.g. their log Kow as an indication for their bioaccumulation potential. This would especially be of interest when CATABOL predicts spontaneous reactions with a high probability, where the likelihood of occurrence is assumed to be high.

For example, in the case of tetrachlorophthalic anhydride CATABOL predicts that the substance will undergo a fast alhydride hydrolysis into a degradation product with a log Kow of 3.6, which in a PBT assessment does not meet the B-criterion. For tris(2,4-di-tert-butylphenyl)phosphate, CATABOL predicts that the substance undergoes an alkylphosphinite hydrolysis, resulting in the formation of a degradation product with a log Kow of 5.3, which would be interesting for further investigation on its bioaccumulation potential. In accordance to the experimental observation, CATABOL also predicts that 4,4'-methylenedicyclohexyl-diisocyanate will hydrolyse into 4,4’-diaminodicyclohexylmethane with a log Kow <4.5 and is therefore is de-listed as a PBT candidate. Another interesting group of substances are the investigated pigments (no. 6, 7, 8 and 9 in Appendix II) which by themselves have a low octanol solubility, suggesting a low bioaccumuative potential, but based on the high probability of the first degradation steps according to CATABOL these chemicals should degrade into metabolites which could be more octanol soluble and still having a high bioaccumulation potential. However, the biodegradation of these compounds could also be hampered by the low water solubility and still result in a high persistency.

The fact that CATABOL gives information on the quantities of degradates in addition to BOD of the parent compound is a feature that can be very useful in targeting the risk assessment and testing. It is very useful to determine that not the parent but a recalcitrant degradate is responsible for the low BOD value.

CATABOL could be used to verify read across. For example, according to CATABOL prediction phenol, 4-nonyl branched compounds are much less degradable than nonylphenol itself due to the low probability of methyl group oxidation. Based on this finding read across between nonphenol and phenol, 4-nonyl branched should be considered with care.

Based on the findings discussed above, the following recommendation for further work can be given:

- Identify missing pathways for chemicals with a poor prediction of BOD.
- Develop guidance on the interpretation of the probability, reliability and applicability.
- Adding a range of separate degradation pathways reflecting the microbial diversity in soil and sediment as proposed by Boxall et al. (2004).

5. Skin and respiratory sensitisation

Disclaimer
The text below is based on the preliminary results of the REACH Implementation Project (RIP) 3.3 phase 2 endpoint working group (EWG) for skin and respiratory sensitisation, of which the report was in draft stage (awaiting the last round of comments) when issuing this report. The draft is a working document and can not in any circumstances be regarded as an official position of the European Commission. Because of the draft stage, the ITS schemes were not copied into this report. The working group was chaired by RIVM (Maaike van Zijverden), and the chair as well as several other persons participating in this group(Henk van Loveren, Etje Hulzebos, Betty Hakkert) and its management group (Betty Hakkert) were funded by DGM, within the framework of M/601200 (EOH). Other group members were: J. Arts (TNO), D. Basketter (Unilever), Carsten Goebel (Procter and Gamble), I. Jowsey (Unilever), I. Kimber (Syngenta), K. Lundberg (KemI), H. McGarry (HSE), B. Oredsson-Hagstrom (KemI), G. Patlewicz (ECB), A. Penninks (TNO), and C. Rovida (ECVAM). The extract below and its conclusions do not necessarily represent the view of the above group.

5.1 Introduction

In the context of occupational and environmental health, one of the most important allergic diseases caused by chemicals is contact hypersensitivity (CHS) or skin sensitisation. Currently, skin sensitisation testing is obligatory for new substances produced at a tonnage level >0.1 tonne/year (Annex 7B, 67/548/EEC). Under REACH, information on skin sensitising potential (for classification and labelling purposes) is required at the lowest production level (1 tonne/year). The information requirements for sensitisation are described by REACH Annexes V to XI, that specify the information that shall be submitted for registration and evaluation purposes. Column 1 of Annex VII clearly informs on the standard information requirement for skin sensitisation data for substances produced or imported in quantities of >1 tpa (tonnes per annum). It states:
'The assessment of skin sensitisation shall comprise the following consecutive steps:
1) an assessment of the available human, animal and alternative data,
2) *in vivo* testing'.

Column 2 of Annex VII lists specific rules according to which the required standard information may be omitted, replaced by other information, or adapted in another way. If the conditions are met under which column 2 of this Annex allows adaptations, the fact and the reasons for each adaptation should be clearly indicated in the registration. For skin sensitisation column 2 reads: 'Step 2 does not need to be conducted if:

- the available information indicates that the substance should be classified for skin sensitisation or corrosivity; or
- the substance is a strong acid (pH <2.0) or base (pH >11.5); or
- the substance is flammable in air at room temperature.

The Murine Local Lymph Node Assay (LLNA) is the first-choice method for *in vivo* testing. Only in exceptional circumstances should another test be used. Justification for the use of

another test shall be provided'. However, in certain cases other *in vivo* methods may be more appropriate. In such cases justification for the use of another test shall be provided.

In REACH, no information requirements are present for respiratory sensitisation. Respiratory sensitisers are indicated though as being of highest concern in REACH Article 114, and respiratory sensitisation is mentioned in Annex I and XV which deal with respectively chemical safety report and preparation of the dossiers.

In addition to these specific rules, the required standard information set may be adapted according to the general rules contained in Annex XI. In this case as well, the fact and the reasons for each adaptation should be clearly indicated in the registration.
In the following sections, first the mechanism of sensitisation is discussed, after which an overview is given on gathering and evaluating data on respectively skin and respiratory sensitisation.

5.2 Mechanism of skin and respiratory sensitisation

Among the key steps required for a chemical to induce sensitisation via skin contact are gaining access to the viable epidermis, protein binding, metabolic activation (if required), internalization and processing by Langerhans cells (LC), transport of antigen by LC to draining lymph nodes, and presentation to and recognition by T lymphocytes. For chemicals that sensitise via the respiratory tract, the relevant mechanisms are believed to be essentially similar, although gaining access to the respiratory epithelium may be somewhat easier than at skin surfaces due to the lack of a stratum corneum. Moreover, because the lining of the respiratory tract, the professional antigen presenting cells and regulatory mechanisms in the respiratory tract differ from those in the skin, they all may have an impact on the type of immune responses evoked. Although the site of induction of an adaptive immune response to a chemical allergen may be influenced by local conditions and local immunoregulatory mechanisms, the fact remains that the inherent properties of the chemical itself play a major – and possibly the major – role in determining whether an immune response is induced and determining the qualitative characteristics of that response.

Although it is sometimes assumed that immune responses induced following encounter with antigen in or on the skin are often of selective Th1-type, this is not necessarily the case. It is clear that cutaneous immune responses can be of either Th1- or Th2-type according to the nature of the antigen.

Allergic responses in the respiratory tract have a tendency to develop as Th2 type reactions (Maestrelli et al., 1997). Th2 type immune responses are characterised by the production of cytokines such as IL-4 and IL-5 and by the production of IgE antibodies. However, the mechanisms through which chemicals are able to induce sensitisation of the respiratory tract are not fully understood and there remains controversy about the roles played by IgE antibody-mediated mechanisms, and whether IgE represents a mandatory universal requirement for the induction by chemicals of allergic sensitisation of the respiratory tract. There are two, non-mutually exclusive, possibilities. The first is that IgE plays a central role but that for one or more of various reasons it is not being detected accurately in the serum of patients with occupational asthma. The second is that allergic sensitisation of the respiratory tract by chemicals can be effected through IgE antibody-independent immunological mechanisms (Kimber and Dearman, 2002; 2005). These may also include Th1-type immune responses. In this context it has been reported, for instance, that inhalation challenge of sensitised rodents with contact allergens may elicit respiratory allergic reactions that lead up

to airway hyperreactivity to non-specific stimuli (Garssen et al., 1991; Garcia et al., 1992; Buckley and Nijkamp, 1994; Zwart et al., 1994; Satoh et al., 1995; Arts et al., 1998).
It should be borne in mind that it is possible that inhalation exposure to a contact allergen might cause an adverse reaction in the respiratory tract. This comes as no surprise because it is clear that contact sensitisation is systemic in nature and that there is no reason to suppose that encounter of sensitised animals with the relevant contact allergen at respiratory epithelial surfaces will not cause an adverse immunologic reaction. However, it is important to note that in reality only a very few precedents for the elicitation of pulmonary reactions by skin sensitizing chemicals in humans have been observed, and in practice it may not represent a significant health issue.

In addition, there is a growing body of evidence that effective sensitisation of the respiratory tract by chemicals defined as respiratory allergens (such as for instance the acid anhydrides, diisocyanates and others) can and does occur in response to dermal contact (reviewed by Kimber and Dearman (2002).

There are also experimental animal data and human evidence for sensitisation by inhalation and skin effects following dermal challenge. Therefore, it is not necessarily the case that chemicals that cause allergic dermal reactions require sensitisation via the skin, or that chemicals that cause allergic airway reactions require sensitisation via the respiratory tract.

5.3 Skin sensitisation: gathering and evaluating information

5.3.1 Non-testing data on skin sensitisation

Non-testing methods for skin sensitisation cover a breadth of different approaches namely read-across/chemical categories, chemistry considerations and (Q)SARs. Read-across/chemical categories will be more generically explored elsewhere (RIP 3.3-2 cross cutting guidance on (Q)SARs).

5.3.1.1 Gathering non-testing data on skin sensitisation

A compendium of available (Q)SARs is not in existence at the present time. Work is being carried out by ECB to develop an inventory of evaluated (Q)SARs which will populate the (Q)SAR Application Toolbox, which is a larger project currently led by the OECD. This ECB inventory is being designed to help a user to determine the validity and applicability of a model for a specific chemical and purpose. This is relevant to the assessment of adequacy. The OECD principles (described in OECD, 2004) will help to characterise the validity of a given model. Preliminary practical guidance on their interpretation has been developed (Worth et al., 2005). Evaluated (Q)SARs will be documented in (Q)SAR Reporting Formats (see section 5.3.1.2). More generic information on evaluating QSARs, their predictions and reporting formats is provided elsewhere (RIP 3.3-2 cross cutting guidance on (Q)SARs).

Exploring the reaction chemistry of compounds forms the basis of most read across justifications and many of the available skin sensitisation (Q)SARs. The skin sensitisation potential of a chemical is related to its ability to react with skin proteins to form covalently linked conjugates and recognition of these by the immune system. In the vast majority of cases, this is dependent on electrophilic reactivity of the skin sensitiser or a derivative produced (usually by oxidation) *in vivo* or abiotically (Barrat et al., 1997). There are various types of electrophile-nucleophile reactions in skin sensitisation, perhaps the most frequently encountered are: Michael-type reactions; S_N2 reactions; S_NAr reactions; acylation reactions

and Schiff-base formation. These chemical reaction mechanisms can serve as a means of describing the domain of applicability (the scope) of a (Q)SAR or form the basis for grouping chemicals into chemical categories. Recent work in this area has been described in (Aptula et al., 2005; Roberts and Williams, 1982). There are relatively few (Q)SARs for skin sensitisation reported in the peer reviewed literature. Available models include local and global (Q)SARs as well as expert systems.

The majority of available **local models** have been developed for direct-acting electrophiles using the relative alkylation index (RAI) approach. This is a mathematical model derived by Roberts and Williams (1982). It is based on the concept that the degree of sensitisation produced at induction, and the magnitude of the sensitisation response at challenge, depends on the degree of covalent binding (haptenation; alkylation) to carrier protein occurring at induction and challenge. This approach has been shown to be mechanistically robust but the breadth of available models so far is still somewhat limited. These types of models assume a reasonable appreciation of chemistry. More effort is needed to encode reactivity into descriptors; this could be achieved through the systematic generation of *in vitro* peptide reactivity data as outlined in Aptula et al. (2005) and in the next section.

Global Statistical models usually involve the development of empirical QSARs by application of statistical methods to sets of biological data and structural descriptors. These are perceived to have the advantage of being able to make predictions for a wider range of chemicals. In some cases, the scope/domain of these models are well described, in most other cases a degree of judgment is required in determining whether the training set of the model is relevant for the chemical of interest. Criticism often levied at these types of models is that they lack mechanistic interpretability.

Several **Expert systems** for skin sensitisation are commercially available. Hereafter some examples are given. Statistical models include TOPKAT (marketed by Accelrys Inc (San Diego, USA), http://www.accelrys.com/products/topkat/) and CASE. The Danish Environmental Protection Agency (EPA) constructed a (Q)SAR database comprising predictions made by some 70 models for about 166,000 organic chemicals for a wide range of different endpoints. A collaborative project between the Danish EPA and the European Chemicals Bureau (ECB) has resulted in the development of an internet-accessible version of this database. (Q)SAR estimates for the MCASE skin sensitisation model are included in this version. The Danish (Q)SAR Database may be accessed from the ECB website http://ecb.jrc.it/QSAR/. Knowledge based systems include Derek for Windows (marketed and developed by LHASA Ltd (Leeds, UK), http://www.lhasalimited.org/index.php) and Hazard Expert. An example of a hybrid system is TIMES, which integrates a skin metabolism simulator with 3D-QSARs for evaluating reactivity of chemicals in order to predict their skin sensitisation potency (Dimitrov et al., 2005; Roberts et al., 2006).

Clearly there is a breadth of different (Q)SARs and expert systems available for the estimation of skin sensitisation hazard. The approaches are quite varied and each has been developed on different sets of *in vivo* data (principally GPMT and LLNA). Whilst efforts have been made to characterise a number of the literature based models in terms of the OECD principles for QSAR validation (see Roberts et al., 2007, as an example), further work is still required for some of the commercial systems. In addition, in many cases these models have been demonstrated to be reasonable for predicting skin sensitisers correctly but are limited in predicting non-sensitisers correctly (ECETOC, 2003). For this reason, careful interpretation of model predictions needs to be considered in light of other information e.g. analogue read-across (other similar chemicals with respect to their mechanistic domain). Further work

should explore encoding more knowledge/rules for non-reactive chemicals as well as those chemicals likely to undergo chemical or metabolic transformation.

Consideration of which model(s) to apply will depend on the specific chemical of interest, the underlying training set data and the applicability domain. These issues are described more fully in RIP 3.3-2 cross cutting guidance on (Q)SARs). An example is illustrated here; if the chemical falls into a chemistry reactivity domain that is well characterised, then a local (Q)SAR model developed for this domain (such as those previously described) will give rise to the most robust prediction of skin sensitisation. Where the mechanism is not understood or not known *a priori* one or more of the expert systems such as TOPKAT, Derek for Windows or the other systems already described will be best placed to provide an estimate. These systems whilst not wholly transparent do provide a reasonable amount of supporting information to enable the robustness of a prediction to be evaluated. This is discussed in more detail in section 5.3.1.2.

5.3.1.2 Evaluating non-testing data on skin sensitisation

The evaluation and assessment of a chemical using (Q)SARs is dependent on both the chemical of interest and the (Q)SAR model(s) used to make a prediction. Here it is attempted to provide some specific advice for skin sensitisation. One of the **first** steps to consider is what information already exists on chemicals 'similar' to the one of interest. Chemical similarity is a widely used concept in toxicology, and is based on the hypothesis that similar compounds have similar biological activities. This forms the underlying basis for developing (Q)SARs. In the case of skin sensitisation, the most robust means of comparing two or more chemicals is through an evaluation of their likely chemical reactivity. If the chemical reactivity is not known, or cannot be determined through experimentation, a pragmatic means of identifying similar chemicals can be done through a substructural/analogue search. There is a number of available computational tools and databases that facilitate the search and retrieval of similar analogues.

Many of currently available tools containing public data have focused on endpoints such as carcinogenicity, mutagenicity or acute toxicity. This means that an additional search is needed to identify skin sensitisation data. Much of the available skin sensitisation experimental data resides in peer reviewed publications. Cronin and Basketter (1994) published the results of over 270 *in vivo* skin sensitisation tests (mainly from the guinea pig maximisation test). All data were obtained in the same laboratory and represent one of the few occasions when large amounts of information from corporate databases were released into the open literature. A larger database of animal and human studies for 1034 compounds is described by Graham et al. (1996), the MCASE database. In addition a comparatively large number of data have been published for the local lymph node assay: full test results were published by Ashby et al. (1995) and a more extensive compilation of 211 LLNA is presented by Gerberick et al. (2005). These publications are invaluable to identify analogues with associated skin sensitisation test data.

The **second** step involves an assessment of the similarity of the analogues identified, in order to identify an available local (Q)SAR for that chemical class/mechanistic group.

If an appropriate local model can not be identified then a **third** step of evaluating a chemical using one of the available global models/expert systems is merited. Here a prediction needs to be evaluated in the context of the likely chemistry and the available similar chemicals available within the training set: i.e. is the compound of interest within the scope of the model and are similar chemicals in the training set of the model well predicted. This type of information adds weight to whether the estimate derived is meaningful and relevant.

Although the main factors driving skin sensitisation (and therefore the (Q)SARs) is the underlying premise of the electrophilicity of a chemical, other factors such as hydrophobicity encoded in the octanol-water partition coefficient (LogKow) may also be considered as playing a role in the modifying the sensitisation response observed. Within DfW, an assessment of the likely skin penetration ability is made using the algorithm by Potts and Guy. This relates the Kp value to LogKow and MW Potts and Guy (1992). It is then possible to rationalise the output in terms of bands of penetration potential. Some have been described by Howes et al. (1996).

Specific model and prediction information can be described in more detail in reporting formats ((Q)SAR Reporting Format). This summarises the pertinent information to consider for given model when evaluating an estimate as well as the estimate itself. More details are provided elsewhere (RIP 3.3-2 cross cutting guidance on (Q)SARs).

Other information such as results in other assays, for instance the Ames test (a common feature of genotoxic substances is that they can bind covalently to DNA and cause direct DNA damage (Wolfreys and Basketter, 2004)) and ecotoxicity may provide supporting information about the electrophilicity of the chemical of interest and hence its likely sensitisation ability.

5.3.2 *In vitro* data on skin sensitisation

At present, no officially adopted EU-OECD *in vitro* tests for skin sensitisation exist. However, several systems are in the course of development, based on an improved understanding of the biochemical and immunological mechanisms underlying the process (Worth and Balls, 2002).

5.3.2.1 Gathering in vitro data on skin sensitisation

Up to now *in vitro* assays to detect the sensitising properties of a chemical are under development for the following areas:

- Chemical reactivity: since the majority of chemical allergens is electrophilic and reacts with nucleophilic amino acids, peptide reactivity assays can give an indication of skin sensitisation potency (Eskes and Zuang, 2005, Gerberick et al., 2005, Aptula et al, 2006).

- Cell-based assays: the knowledge that changes occur in epidermal Langerhans cells as a result of exposure to chemical allergens (e.g. the expression of surface markers and/or cytokines) and that Langerhans cells can be replaced by blood derived dendritic-like cells was applied to design *in vitro* alternative tests (Kimber et al., 2001, Tuscl et al., 2000, Casati et al., 2005; Ryan et al., 2005; Sakaguchi et al., 2006). These systems have been shown to express various mediators and/or markers of activation following exposure to chemical sensitisers and attempts to develop robust assays have started.

- Epidermal bioavailability: Skin penetration is a prerequisite for skin sensitisation. Information about the skin penetration properties can help to evaluate the potential of a chemical to be identified as a skin sensitiser (ECVAM, 2006).

Due to the complexity of the mechanisms of skin sensitisation, a single test will probably not be able to replace the currently required animal procedures. Efforts are still needed to identify the most relevant endpoints in the optimisation of existing tests. However, a combination of several *in vitro* tests, covering the relevant mechanistic steps of skin sensitisation, into a test battery could possibly lead to replacement of *in vivo* tests (Eskes and Zuang 2005). How the outputs from these tests could be combined is not as yet determined, although a general

strategy has been presented (Jowsey et al., in press 2006). Until that date, *in vitro* tests may be used as supportive evidence in combination with other types of data for the identification of allergens.

5.3.2.2 Evaluating in vitro data on skin sensitisation

It has to be kept in mind that:

- no *in vitro* test has yet been validated;
- current *in vitro* assays only cover a (specific) part of the process of sensitisation;
- so far *in vitro* data can only be used in a Weight-of-Evidence approach, e.g. in conjunction with other data, and currently for positive identification of sensitisers only.

If *in vitro* data for skin sensitisation will be considered for the evaluation, expert judgment is needed (Hartung et al., 2004). Currently, the relevance of *in vitro* data depends on the question how it can be related to animal data. Peptide binding assays are probably most developed, but interpretation needs to be performed with caution.

5.3.3 Animal data on skin sensitisation

Animal data for skin sensitisation may be derived from either guideline-compliant tests or non-guideline-compliant tests. These are separately described below.

5.3.3.1 Gathering animal data on skin sensitisation

Guideline-compliant tests

For new *in vivo* testing of skin sensitisation potential, the murine local lymph node assay (LLNA) is the REACH Annex VII-endorsed method. This assay has been validated internationally and has been shown to have clear animal welfare benefits and scientific advantages compared with the guinea pig tests described below. The LLNA is designed to detect the potential of chemicals to induce sensitisation as a function of lymphocyte proliferative responses induced in regional lymph nodes. This method is described in OECD guideline 429.

Two further animal test methods for skin sensitisation are described in OECD guideline 406: the guinea pig maximisation test (GPMT) and the Buehler test. The GPMT is an adjuvant-type test in which the acquisition of sensitisation is potentiated by the use of Freund's Complete Adjuvant (FCA) and in which both intradermal and topical exposure are used during the induction phase. The Buehler test is a non-adjuvant method involving for the induction phase topical application only.

Both the GPMT and the Buehler test are able to detect chemicals with moderate to strong sensitisation potential, as well as those with relatively weak sensitisation potential. In such methods activity is measured as a function of challenge-induced dermal hypersensitivity reactions elicited in test animals compared with controls. Since the LLNA is the approved method for new *in vivo* testing, the use of the standard guinea pig tests to obtain new data on skin sensitisation potential will be acceptable only in exceptional circumstances and will therefore require scientific justification. However, existing data of good quality deriving from such tests will be acceptable and will, if providing clear results, preclude the need for further *in vivo* testing.

ECETOC Monograph 29 (2000) contains a useful discussion of these tests.

Non-guideline compliant tests and refinements to the standard assays
Existing data may be available from tests that do not have an OECD guideline, for example:

- Other guinea pig skin sensitisation test methods (such as the Draize test, optimisation test, split adjuvant test, open epicutaneous test);
- Additional tests (such as the mouse ear swelling test).

Information may also be available from other end points, for example, repeated dose dermal studies that show effects indicative of an allergic response, such as persistent erythema and/or oedema.

For new testing, refinements to the existing guideline methods may also be possible. In such cases, care should be taken to ensure that any modifications or deviations from standard methodologies are scientifically justified. For example, it might be feasible to conduct a reduced version of the LLNA (rLLNA) in which assessments are made on the basis of results from a vehicle control and a single (highest) concentration of the test substance (Kimber et al., 2006). In such cases, it is recommended that expert advice be sought before commencing the tests.

5.3.3.2 Evaluating animal data on skin sensitisation

Well reported studies using internationally acceptable protocols, particularly if conducted in accordance with the principles of GLP, can be used for hazard identification. Other studies (see below), not fully equivalent to OECD test protocols, can, in some circumstances, provide useful information. Particular attention should be paid to the quality of these tests and the use of appropriate positive and negative controls. The specificity and sensitivity of all animal tests should be monitored through the inclusion of appropriate positive and negative controls. In this context, positive controls are the 6-monthly sensitivity checks with an appropriate positive control substance, and negative controls are the vehicle-treated control animals included as part of each test.

Guideline-compliant tests
For the conduct and interpretation of the LLNA OECD guideline 429 provides guidance on the recommended vehicles, number of animals per group, concentrations of test chemical to be applied and substances to be used as a positive control. A preliminary study or evaluation of existing acute toxicity/dermal irritation data is normally conducted to determine the highest concentration of test substance that is soluble in the vehicle but does not cause unacceptable local or systemic toxicity. The submission of historical control data will demonstrate the ability of the test laboratory to produce consistent responses. Based on the use of radioactive labelling, chemicals that result in a stimulation index (SI) of >3 (whilst taking into account the dose response relationship) at one or more test concentrations are considered to be positive for skin sensitisation. Both positive and negative responses in the LLNA conducted as described in OECD guideline 429 meet the data requirements for classification of a substance as a skin sensitiser: no further testing is required.

The guinea pig test methods described in OECD 406, the GPMT (Magnusson and Kligman 1969, Schlede and Eppler 1995) and the Buehler, can also be used for hazard identification. Recommendations on conducting and analysing these methods are provided by Steiling et al. (2001). Particular attention should be paid to the quality of these tests with consideration given to the numbers of test and control guinea pigs, number or percentage of test and control animals displaying skin reactions, whether skin irritation was observed at the induction phase, whether the maximal non-irritating concentration was used at the challenge phase, the choice of an appropriate vehicle, whether there are signs of systemic toxicity, staining of the skin by the test material that may obscure any skin reactions, results of rechallenge treatments if performed, checking of strain sensitivity at regular intervals by using an appropriate control

substance (as specified in OECD guidelines). Currently (2006), the recommended interval is 6 months.

The investigation of doubtful reactions in guinea pig tests, particularly those associated with evidence of skin irritation following first challenge, may benefit from rechallenge of the test animals. In cases where reactions may have been masked by staining of the skin, other reliable procedures may be used to assist with interpretation; where such methods are used, the submitting laboratory should provide evidence of their value.

Non-guideline compliant tests and refinements to the standard assays
The submitted dossier should include scientific justification for conducting any new test that is a modification or deviation from guideline methods. In such cases, it would be advisable to seek appropriate expert advice on the suitability of the assay before testing is begun.

For hazard identification, it may be possible to use a reduced LLNA (rLLNA) (Kimber et al., 2006) (under evaluation by ECVAM, november 2006), which reduces the use of animals by requiring only a single (high) dose group and a concurrent negative control group. A preliminary study or evaluation of existing acute toxicity/dermal irritation data is normally conducted to determine the highest concentration of test substance that is soluble in the vehicle, but that does not cause unacceptable local or systemic toxicity. Although a concurrent positive control group is not required, registrants would be required to submit historical positive control data supportive of their competence. The rLLNA is currently not validated or accepted as a stand-alone test, and should be used only if the test laboratory can demonstrate that it has a high level of technical competence in the technique, and hazard identification is the primary objective.

As in the standard (OECD guideline-compliant) LLNA, group sizes should comprise four or five animals. A positive result in a rLLNA will suffice in circumstances where risk assessment and/or risk management is NOT required. Registrants should be aware that the rLLNA is scientifically less rigorous than the standard LLNA, with an associated increased level of uncertainty.

Alternative vehicles to those listed in OECD 429 may be used in the LLNA if sufficient scientific justification is provided. OECD 429 also states that endpoints other than radioactive labelling may be used to assess proliferation, on condition that justification and scientific support, which will include full citations and a description of the methodology, are provided.

Historically, guinea pig studies that are not fully equivalent to OECD test protocols have been conducted and can provide useful hazard information. These studies include, but are not limited to, the following: Draize test, optimisation test, split adjuvant test, open epicutaneous test and the cumulative contact enhancement test. In the case of positive results the substance may be considered as a potential skin sensitiser. If, taking into account the above quality criteria, especially the positive and negative control data, there is a clear negative result, i.e. no animals displaying any signs of sensitisation reactions, then no further animal testing is required. Where there is a low level of response, the quality of the study is questionable, or where unacceptably low concentrations of the test material have been used for induction and/or challenge, further testing may be required.

5.3.4 Human data on skin sensitisation

5.3.4.1 Gathering human data on skin sensitisation

Human data on cutaneous (allergic contact dermatitis and urticarial) reactions may come from a variety of sources:

- Consumer experience and comments, preferably followed up by professionals (e.g. diagnostic patch tests);
- Diagnostic clinical studies (e.g. patch tests, repeated open application tests)
- Records of workers' experience, accidents, and exposure studies including medical surveillance;
- Case reports in the general scientific and medical literature;
- Consumer tests (monitoring by questionnaire and/or medical surveillance);
- epidemiological studies;
- Human experimental studies such as the human repeat insult patch test (HRIPT, Stotts 1980) and the human maximization test (HMT, Kligman 1966), although it should be noted that *new* experimental testing for hazard identification in humans, including HRIPT and HMT, is not acceptable for ethical reasons).

5.3.4.2 Evaluating human data on skin sensitisation

When reliable and relevant human data are available, they can be useful for hazard identification and may even be preferable over animal data. However, lack of positive findings in humans does not necessarily overrule positive and good quality animal data.

Well conducted human studies can provide very valuable information on skin sensitisation. However, in some instances (due to lack of information on exposure, a small number of subjects, concomitant exposure to other substances, local or regional differences in patient referral etcetera) there may be a significant level of uncertainty associated with human data. Moreover, diagnostic tests are carried out to see if an individual is sensitised to a specific agent, and not to determine whether the agent can cause sensitisation.

For evaluation purposes, existing human experience data for skin sensitisation should contain sufficient information about:

- The test protocol used (study design, controls);
- The substance or preparation studied (should be the main, and ideally, the only substance or preparation present which may possess the hazard under investigation);
- The extent of exposure (magnitude, frequency and duration);
- The frequency of effects (versus number of persons exposed);
- The persistence or absence of health effects (objective description and evaluation);
- The presence of confounding factors (e.g. pre-existing dermal health effects, medication, presence of other skin sensitisers);
- The relevance with respect to the group size, statistics, documentation;
- The 'healthy worker' effect.

Human experimental studies on skin sensitisation are not normally conducted and are generally discouraged. Where human data are available, then quality criteria and ethical considerations are presented in ECETOC monograph no. 32 (ECETOC, 2002)

Ultimately, where a very large number of individuals (eg.10^5) have frequent (daily) skin exposure for at least two years and there is an active system in place to pick up complaints and adverse reaction reports (including via dermatology clinics), and where no or

only a very few isolated cases of allergic contact dermatitis are observed then the substance is unlikely to be a significant skin sensitiser. However, information from other sources should also be considered in making a judgment on the substance's ability to induce skin sensitisation.

It is emphasised that testing with human volunteers is strongly discouraged, but when there are good quality data already available they should be used as appropriate in well justified cases.

5.3.5 Concluding on suitability for classification and labelling for skin sensitisation

Details about Classification and Labelling are reported in REACH, Section 4 of Annex VI. According to this section, classification and labelling should comply with Council Directive 67/548/EEC. In addition to Directive 67/548/EEC, REACH demands that all available information must be reported, including when no relevant data are present. Standard information required for skin sensitisation is described in Annex VII of REACH, i.e. for any substance manufactured or imported in quantity of 1 tonne or more.

A substance can be classified as ‘skin sensitiser’ following a flow chart for Integrated Testing Strategy (ITS) (not included). At the moment, labelling for skin sensitisation is with symbol Xi, the indication of danger ‘Irritant’ and the risk phrase R43 (R43: May cause sensitisation by skin contact). The labelling may change when the Global Harmonised System will come into force in Europe. An expert group on skin sensitisation was nominated by the European Commission to investigate the possibility to apply classification criteria for the definition of skin sensitiser potency. The outcome was to assign sensitisers to 1 to 3 categories according to potency (moderate, strong, extreme). However this convention has still no endorsement from competent Authorities.

5.4 Respiratory sensitisation: gathering and evaluating information

5.4.1 Non-testing data on respiratory sensitisation

Attempts to model respiratory sensitisation have been hampered by a lack of a predictive test protocol for assessing chemical respiratory sensitisation. (Q)SAR models are available but these have largely been based on data for chemicals reported to cause respiratory hypersensitivity in humans.

Agius et al. (1991) made qualitative observations concerning the chemical structure of chemicals causing occupational asthma. This work drew attention to the large proportion of chemical asthmagens with at least two reactive groups, e.g., ethylene diamine and toluene diisocyanate. The earlier work was followed up by a simple statistical analysis of the occurrence of structural fragments associated with activity, with similar conclusions (Agius et al., 1994; Agius, 2000).

The MCASE group has developed three models for respiratory hypersensitivity (Karol et al., 1996; Graham et al., 1997; Cunningham et al., 2005). The Danish (Q)SAR Database has an in-house model for respiratory hypersensitivity for which estimates can be extracted from the on-line database (available at http://ecb.jrc.it/QSAR). Derek for Windows contains several alerts derived from a set of respiratory sensitisers/asthmogens (Payne and Walsh, 1995).

Whilst the available structural alerts (SAR) are transparent and easily to apply (Agius et al., 1991; Agius et al., 1994; Agius, 2000; Payne and Walsh, 1995), it should be stressed that these are derived on the basis of chemical asthmagens not specifically chemical respiratory allergens. A need therefore remains to develop new (Q)SARs as and when a robust predictive test method becomes available.

Given lack of available (Q)SARs for respiratory sensitisation, it is not possible to provide any additional guidance on the evaluation of non-testing data for respiratory sensitisation.

5.4.2 *In vitro* data on respiratory sensitisation

No *in vitro* tests specific for respiratory sensitisation are available yet (2006), due to the complexity of the mechanisms of the sensitisation process. If such a method were to become available then it would need to be assessed for its relevance and reliability (Hartung et al., 2004). Efforts are still needed to identify the most relevant endpoints in the optimisation of existing tests. However, a combination of several *in vitro* tests, covering the relevant mechanistic steps of respiratory sensitisation, into a test battery could eventually lead to replacement of the *in vivo* tests.

Given lack of available *in vitro* tests for respiratory sensitisation, it is not possible to provide any additional guidance on the evaluation of non-testing data for respiratory sensitisation.

5.4.3 Animal data on respiratory sensitisation

5.4.3.1 Gathering animal data on respiratory sensitisation

At present, although a number of test protocols has been published to detect respiratory allergenicity of low molecular weight compounds, none of these are validated nor are these widely accepted. One approach that might be of some value in characterizing the likely respiratory sensitizing activity of chemicals is application of the LLNA, or of other tests for measuring skin sensitisation potential. Although the LLNA was developed and validated for the identification of contact allergens, there is evidence that chemical respiratory allergens will also elicit positive responses in this assay (Kimber, 1995). That is, chemicals known to cause respiratory allergy and occupational asthma have been shown to test positive in the LLNA. Among such chemicals are acid anhydrides (such as trimellitic anhydride and phthalic anhydride), diisocyanates (including diphenylmethane diisocyanate and hexamethylene diisocyanate) and certain reactive dyes. In fact, the view currently is that most, if not all, chemical respiratory allergens are able to elicit positive responses in the LLNA, or in other tests for skin sensitisation, such as the GPMT test. This is true even of those chemical respiratory allergens, such as phthalic anhydride, for instance, that are implicated virtually exclusively with the induction of chemical respiratory allergy and have rarely, if ever, been shown to cause allergic contact dermatitis. Against this background and in combination with other data it might be possible to conclude in a Weight-of-Evidence assessment that chemicals that (at an appropriate test concentration and test conditions, i.e. skin penetration should have occurred) are negative in the LLNA, as well as being considered as not being skin sensitisers, can also be regarded as lacking the potential to cause allergic sensitisation of the respiratory tract.

One approach that has been proposed for the identification of chemicals that have the potential to cause allergic sensitisation of the respiratory tract is one in which activity is measured as a function of the profiles of cytokines produced by draining lymph node cells in mice exposed more chronically (over a 2 week period) to the test chemical (Dearman et al., 2002). This method is predicated on an understanding that allergic sensitisation of the

respiratory tract is favoured by selective Th2-type immune responses and that in many instances chemical respiratory allergy and occupational asthma are associated with IgE antibody. Using this approach chemical respiratory allergens are identified as a function of their ability to stimulate in mice the selective development of preferential Th2-type immune responses associated with a predominance of type 2 cytokine secretion by draining lymph node cells (Dearman et al., 2002; 2003). Specifically, chemical contact allergens promote Th1 responses characterised by an enhanced production of IFN-gamma, whereas chemical respiratory allergens promote Th2 responses characterised by enhanced production of IL-4, IL-5 and IL-13. Many variables other than the compound itself, such as concentration used to induce sensitisation, duration of the sensitisation period, and presence or absence of mitogens to reveal differences in cytokine expression have all been noted to have impact on the outcome (Van Och et al., 2002). There are general guidelines now available for the conduct of the method (Dearman et al., 2003), however, this method has not yet been formally validated and nor is it widely accepted.

Another, relatively simple approach may serve the purpose to specifically predict sensitisation of the respiratory tract: i.e. increases in total serum IgE antibodies after induction. This method is based on statistically significant increases in total serum IgE (see review by Arts and Kuper, 2006).

Methods that use both an induction and an inhalation elicitation or challenge phase and which include different parameters such as total and/or specific IgE antibody determinations, lung function testing, tests for aspecific hyperreactivity (e.g. methacholine challenges), bronchoalveolar lavage measurements, and histopathological examination of the entire respiratory tract, may provide (additional) information on the potential of chemicals to cause respiratory sensitisation. These methods usually use high IgE-responding animal strains; to test for Th1-mediated responses low IgE-responding strains should typically be used. Several of these models have been reviewed recently (Arts and Kuper, 2006).

There are currently no predictive methods to identify chemicals that induce asthma through non-immunological mechanisms, however, when performing challenge tests including non-sensitised but challenged controls information can be obtained on non-immunological effects of these chemicals.

5.4.3.2 Evaluating animal data on respiratory sensitisation

Although the LLNA does not represent a method for the specific identification of chemical respiratory allergens, there is evidence that chemical respiratory allergens will also elicit positive responses in this assay (Kimber, 1995). The interpretation is therefore that a chemical which fails to induce a positive response in the LLNA (at an appropriate test concentration) most probably lacks the potential for respiratory allergy. Conversely, it cannot be wholly excluded that a chemical that induces a positive response in the LLNA, might sensitise the respiratory tract upon inhalation or via dermal exposure. Any potential hazard for respiratory sensitisation could only be positively identified by further testing, although such testing is neither validated nor widely accepted.

One further approach to the identification of chemicals that have the potential to induce allergic sensitisation of the respiratory tract is ‘cytokine fingerprinting’ (Dearman et al., 2002; see section 1.4.4.1). In addition, there are other approaches that have been proposed and these have been reviewed recently (Arts and Kuper, 2006) - although again it is important to emphasize that there are currently no fully evaluated or validated animal models for the predictive identification of chemical respiratory allergens available.

As indicated previously, some chemicals may have the potential to induce pulmonary reactions via Th1-type immune responses. Studies with typical skin allergens such as DNCB, DNFB and picryl chloride (trinitrochlorobenzene) in BALB/c mice, guinea pigs or Wistar rats have shown the potential of these chemicals to induce allergic reactions in the lungs that are independent of IgE (Garssen et al., 1991; Garcia et al., 1992; Buckley and Nijkamp, 1994; Zwart et al., 1994; Satoh et al., 1995; and see for a review Arts and Kuper, 2006). Sensitisation and challenge with DNCB resulted in laryngitis in low IgE-responding Wistar rats (Arts et al., 1998). In addition, cellular immune responses to these sensitisers were shown to be associated with hyperreactivity of the airways to non-specific stimuli (Garssen et al., 1991). For these reasons, it might be the case that people who are sensitised via the skin might suffer adverse pulmonary reactions if they were to inhale sufficient amounts of the contact allergen to which they were sensitised. As indicated previously, very few precedents for the elicitation of pulmonary reactions by skin sensitizing chemicals in humans have been observed, and in practice it may appear not to represent a health issue.

5.4.4 Human data on respiratory sensitisation

5.4.4.1 Gathering human data on respiratory sensitisation

Human data on respiratory reactions (asthma, rhinitis, alveolitis) may come from a variety of sources:

- consumer experience and comments, preferably followed up by professionals (e.g. bronchial provocation tests, skin prick tests and measurements of specific IgE serum levels);
- records of workers' experience, accidents, and exposure studies including medical surveillance;
- case reports in the general scientific and medical literature;
- consumer tests (monitoring by questionnaire and/or medical surveillance);
- epidemiological studies.

5.4.4.2 Evaluating human data on respiratory sensitisation

Although human studies may provide some information on respiratory hypersensitivity, the data are frequently limited and subject to the same constraints as human skin sensitisation data. For evaluation purposes, existing human experience data for respiratory sensitisation should contain sufficient information about:

- The test protocol used (study design, controls)
- The substance or preparation studied (should be the main, and ideally, the only substance or preparation present which may possess the hazard under investigation)
- The extent of exposure (magnitude, frequency and duration)
- TRhe frequency of effects (versus number of persons exposed)
- The persistence or absence of health effects (objective description and evaluation)
- The presence of confounding factors (e.g. pre-existing respiratory health effects, medication; presence of other respiratory sensitisers)
- The relevance with respect to the group size, statistics, documentation
- The 'healthy worker' effect.

For respiratory sensitisation, no clinical test protocols for experimental studies exist but tests may have been conducted for diagnostic purposes, e.g. bronchial provocation test. The test should meet the above general criteria, e.g. be conducted according to a relevant design

including appropriate controls, address confounding factors such as medication, smoking or exposure to other substances, etcetera. Furthermore, the differentiation between the symptoms of respiratory irritancy and allergy can be very difficult. Thus, expert judgment is required to determine the usefulness of such data for the evaluation on a case-by-case basis.

Although predictive models are under validation, there is as yet no internationally recognized animal method for identification of respiratory sensitisation. Thus human data are usually evidence for hazard identification.

Where there is evidence that significant occupational inhalation exposure to a chemical has not resulted in the development of respiratory allergy, or related symptoms, then it may be possible to draw the conclusion that the chemical lacks the potential for sensitisation of the respiratory tract. Thus, for instance, where there is evidence that a large cohort of subjects have had opportunity for regular inhalation exposure to a chemical for a sustained period of time in the absence of respiratory symptoms, or related health complaints, then this will provide reassurance regarding the absence of a respiratory sensitisation hazard.

5.4.5 Concluding on suitability for classification and labelling for respiratory sensitisation

Details about classification and labelling are reported in REACH, Section 4 of Annex VI. According to this section, classification and labelling should comply with Council Directive 67/548/EEC. In addition to Directive 67/548/EEC, REACH demands that all available information must be reported, including when no relevant data are present.

In REACH, respiratory sensitisers are included among the substances of higher concern with CMRs (Carcinogenic, Mutagenic, toxic for Reproduction) and regulated in Annex I of Directive 67/548/EEC. Annex I contains a list of dangerous substances, including respiratory sensitisers. Annex XV in REACH lays down general principles for preparing dossiers to propose and justify harmonised classification and labelling of CMRs and respiratory sensitisers.

Potential hazard for respiratory sensitisation cannot be easily addressed, as validated testing methods are currently not available. A probable hazard for respiratory sensitisation should be mentioned in the Safety Data Sheet. A substance can be classified as ‘respiratory sensitiser’ following a flow chart for Integrated Evaluation Strategy (IES) (not included). At the moment, labelling for ‘respiratory sensitisers’ is with symbol Xn, the indication of danger ‘Harmful’ and the risk phrase R42 (R42: May cause sensitisation by inhalation). The labelling may change when a Global Harmonised System will come into force in Europe.

5.5 Remaining uncertainty

Reliable data can be generated on skin sensitisation from well designed and well conducted studies in animals. The use of adjuvant in the GPMT may lower the threshold for irritation and so lead to false positive reactions, which can therefore complicate interpretation (running a pre-test with FCA treated animals can provide helpful information). In international trials, the LLNA has been shown to be reliable, but like the guinea pig tests is dependent on the vehicle used, and it can occasionally give false positive results with irritants. Where tests (guinea pig/mouse) rely on topical exposure rather than intradermal injection, false negatives may occur where the substance is poorly absorbed into the skin. Therefore, careful consideration should be given to the vehicle used and the type of test performed. In some circumstances inconsistent results from similar guinea pig studies, or between guinea pig and

LLNA studies, might increase the uncertainty of making a correct interpretation. Finally, where data is derived from an experimental study in human volunteers, then consideration must be given to whether individual variability is such that it is not scientifically sound to generalize from a limited test panel.

When considering whether or not a substance is a respiratory sensitiser, observations of idiosyncratic reactions in only a few individuals with hyper-reactive airways are not sufficient to indicate the need for classification.

5.6 Dose response assessment and potency

There is evidence that for both skin sensitisation and respiratory hypersensitivity dose-response relationships exist (although these are frequently less well defined in the case of respiratory hypersensitivity). The dose of agent required to induce sensitisation in a previously naïve subject or animal is usually greater than that required to elicit a reaction in a previously sensitised subject or animal; therefore the dose-response relationship for the two phases will differ. Little or nothing is known about dose-response relationships in the development of respiratory hypersensitivity by non-immunological mechanisms.
It is frequently difficult to obtain dose-response information from either existing human or guinea pig data where only a single concentration of the test material has been examined. With human data, exposure measurements may not have been taken at the same time as the disease was evaluated, adding to the difficulty of determining a dose response.

Dose-response data can, however, be generated from local lymph node assays or, in exceptional cases, using specially designed guinea pig test methods. Such types of data can give data on induction and elicitation thresholds in these models, but it must be remembered these cannot be translated directly to human thresholds.

Measurement of potency
Appropriate dose-response data can provide important information on the potency of the material being tested. This can facilitate the development of more accurate risk assessments. This section refers to potency in the induction phase of sensitisation.

Neither the standard LLNA nor the GPMT/Buehler is specifically designed to evaluate the skin sensitising potency of test compounds, instead they are used to identify sensitisation potential for classification purposes. However, all could be used for some estimate of potency. The relative potency of compounds may be indicated by the percentage of positive animals in the guinea pig studies in relation to the concentrations tested. Likewise, in the LLNA, the EC3 value (the dose estimated to cause a 3-fold increase in local lymph node proliferative activity) can be used as a measure of relative potency (ECETOC, 2000). Often linear interpolation of a critical effect dose from the EC3 is proposed (ECETOC), but more advanced statistical approaches basing conclusions on the characteristic of the dose-response curve and variability of the results is also used (Basketter et al., 1999, Van Och et al., 2000). The dose-response data generated by the LLNA makes this test more informative than guinea pig assays for the assessment of skin sensitising potency. EC3 data correlate well with human skin sensitisation induction thresholds derived from historical predictive testing (Schneider and Akkan, 2004; Griem, 2003; Basketter et al., 2005). Accordingly, there are proposals for how this information may be used in a regulatory sense (Basketter et al., 2005) and for risk assessment.

5.7 Additional considerations

Chemical allergy is commonly designated as being associated with skin sensitisation (allergic contact dermatitis), or with sensitisation of the respiratory tract (asthma and rhinitis). In view of this it is sometimes assumed that allergic sensitisation of the respiratory tract will result only from inhalation exposure to the causative chemical, and that skin sensitisation necessarily results only from dermal exposure. This is misleading, and it is important for the purposes of risk management to acknowledge that sensitisation may be acquired by other routes of exposure. Since adaptive immune responses are essentially systemic in nature, sensitisation of skin surfaces may theoretically develop from encounter with contact allergens via routes of exposure other than dermal contact (although in practice this appears to be uncommon). Similarly, there is evidence from both experimental and human studies which indicate that effective sensitisation of the respiratory tract can result from dermal contact with a chemical respiratory allergen. Thus, in this case, it appears that the quality of immune response necessary for acquisition of sensitisation of the respiratory tract can be skin contact with chemical respiratory allergens (Kimber and Dearman 2002). Such considerations have important implications for risk management. Thus, for instance, there is a growing view that effective prevention of respiratory sensitisation requires protection of both skin and respiratory tracts. This includes the cautious use of known contact allergens in products to which consumers are (or may be) exposed via inhalation, such as sprays. The generic advice is that appropriate strategies to minimise the risk of sensitisation to chemical allergens will require consideration of providing protection of all relevant routes of exposure.

5.8 RIVM discussion, possibilities for further research

The text in the sections 5.1-5.7 was taken from the final draft report from the RIP 3.3-2 EWG on sensitisation. The ITS for skin sensitisation and IES for respiratory sensitisation as proposed by the EWG were not included in this report because the RIP project was still in draft phase. In the first part of the present project in 2005, however, ITS elements for skin sensitisation were identified (see Figure 5.1). In 2005 already, it was concluded that due to the complexity of the mechanisms of skin sensitisation, a single non-animal test would not be able to replace a complete *in vivo* experiment. However, replacement should not be the only goal. A tiered non-animal approach as suggested by Jowsey et al., (2006) might be more realistic. These authors state that on the long run a combination of different alternatives may be the best option for a replacement of *in vivo* testing. Their proposal comprises five modules covering the presence of structural alerts and subsequently processes of bioavailability, protein reactivity, dendritic cell maturation and T cell proliferation. Although such a tiered approach consisting of non-animal methods, and possibly including *in vivo* testing as a last resort for a very small percentage of the total amount of chemicals, may be realistic for the near future, this will not be implemented before the start of REACH in June 2007.

Besides the fact that a non-animal strategy is scientifically premature, a major hurdle might be the willingness of industry to perform such a battery of tests, especially when they are more expensive and laborious than the *in vivo* alternative. Another threat to this alternative approach might be the upcoming introduction of GHS (globally harmonized system), an international system that soon will be adopted to provide a basis for the classification and labelling of substances and mixtures. GHS will revise symbols and pictograms to address chemical hazard, and in the field of sensitisation it is under discussion whether or not to introduce the concept of potency in terms of both required threshold to induce sensitisation in

humans as well as subsequent threshold to challenge allergic reactions. Therefore in GHS the distinction is not so much whether or not a substance is a sensitiser, but whether or not a substance is a weak, moderate or strong sensitiser. So far, this so called potency ranking can only be performed by means of high quality *in vivo* data, and more specifically LLNA data, from which dose-response relationships can be derived. RIVM is already involved in national and international discussions on the possible introduction of potency ranking for sensitisation, but more work could be performed on estimating possible implications for risk assessment strategies. Besides, most non-animal methods, also the reduced LLNA is not suited for this purpose. This *in vivo* method, in which only a high dose group instead of three dose groups is tested, will probably play an important role in a testing strategy aiming at less animal use under REACH. But before any rLLNA is performed it should be clear that for the substance in question no potency data are wanted.

The performance and applicability domains of the different guinea pig and mice tests for determining skin sensitisation potential is still under a lot of discussion and could also be an interesting area of future work for RIVM. In REACH the LLNA is mentioned as '…the first-choice method for *in vivo* testing. Only in exceptional circumstances should another test be used. Justification for the use of another test shall be provided'. Industry has major objections to this phrasing. One of their main points is the fact that the LLNA has not been validated for all types of substances, and that this may result in overclassification. Until this date, guinea pig tests have not been validated either, but they do have a use history of several decades. Real proof, however, for the LLNA being too sensitive is not available, although this may become available shortly. Recently a report was sent to the RIP 3.3 group in which substances were concurrently tested in the GPMT and the LLNA (prof Dr A. Wendel, STZ In Vitro Pharmakologie und Toxikologie, University of Kostanz, Germany). After critical review of this kind of studies it might be concluded that indeed there is a need for other *in vivo* possibilities. To be prepared for this situation, an opening for performing other *in vivo* studies next to the LLNA should be included in the RIP 3.3 guidance for industry.

REACH does not include testing requirements for respiratory sensitisation. RIVM, TNO and some other EWG participants did however want to raise awareness about the fact that sensitisation is a systemic process, and inhalatory exposure to any allergen should be prevented as much as possible. So far, it is unclear whether respiratory sensitisation to contact allergens really does not represent a health issue, or that it is simply not diagnosed as such. But whereas this risk is still unclear, animal experiments show that there is indeed a hazard, since it is possible to sensitise animals to contact allergens administered via the respiratory tract. Careful application of contact allergens in products to which consumers may be exposed via inhalation is therefore warranted. The development of models and testing guidance for respiratory sensitisation should therefore also be stimulated, and RIVM might play a role in this area.

RIVM contributes to alternatives in different areas, for example in the area of ex-vivo or non-radioactive LLNA testing (eg. Van Och et al., 2000, 2002). In the non-testing area, mainly consisting of (Q)SARs, RIVM contributes to the scientific validation of the (non-commercial) models, which is of key importance. Formal validation is probably less relevant for those models, and is also not required within REACH (Annex XI). Moreover, in order to be able to improve current models and/or build new ones, there is also the need for additional data bases filled with good quality data. RIVM will participate in EU projects in which this area is covered. RIVM is also involved in the development of toxicogenomics based *in vitro* alternatives for sensitisation tests.

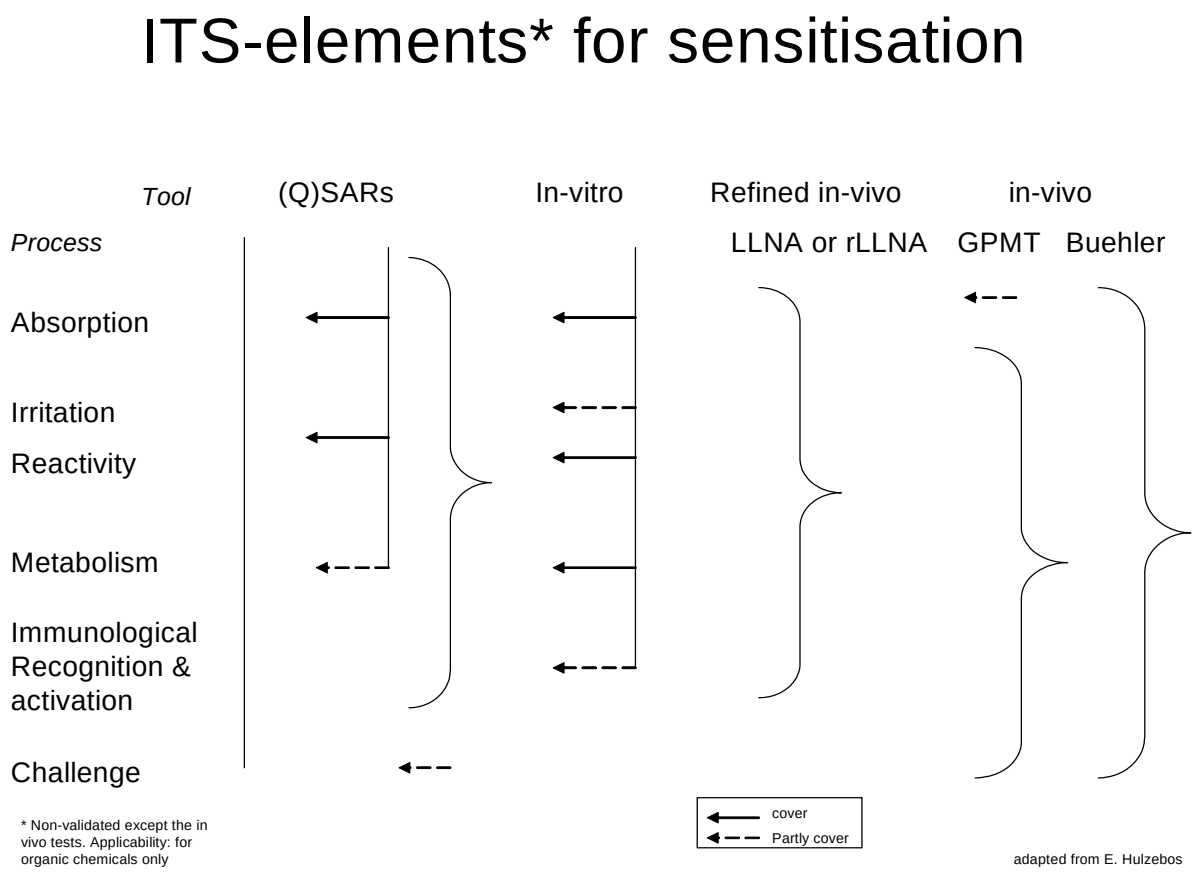


Figure 5.1: ITS elements for skin sensitisation: tools that apply to the different processes

6. ITS on reproductive toxicity

6.1 Introduction

Reproductive toxicity testing according to most current international guidelines involves a prenatal developmental toxicity study (OECD TG 414) in a rodent and a non-rodent species and a one- or two-generation reproduction toxicity study (OECD TG 415 and 416, respectively). Two generations studies are very cost- and time-intensive and require relatively large numbers of animals.

REACH requires at least a reproductive toxicity screening test or comparable information at the 10 and 100 tonnes per annum level. A developmental toxicity study in one species is required for substances of concern manufactured or imported in quantities of 10 tonnes or more. The outcome of this test should be the basis to decide on the need to perform a developmental toxicity study in a second species. A first developmental toxicity study is mandatory for all compounds at the 100 tonnes level and higher. Starting from 10 tonnes per annum, the two-generation test is required if there are indications of potential reproductive toxicity from a repeated dose toxicity study or the substance has a close structural relationship with a known reproductive toxicant. In the absence of indications of potential reproductive toxicity, the two-generation study is only required for substances that are manufactured or imported in quantities of 1000 tonnes or more (EC, 2006).

It has been estimated that the two-generation test will be required for 7.5% of all substances and will use nearly 40% of the laboratory animals under REACH. Reproductive and developmental toxicity testing together may require around 70% of all experimental animals used in REACH (Van der Jagt et al., 2004). Concern has already been expressed on the difficulties to carry out a complete reproductive toxicity test program for all such chemicals (Combes et al., 2003) and on the need to develop simplified reproductive toxicity testing.

In the view of the upcoming changes in regulation, there is a need to retrospectively evaluate the efficiency of the testing strategy followed to the present. After more than twenty years of chemical testing, considerable amounts of toxicological data have been generated. This toxicological data can be used for retrospective analyses. In the present report, the aim is to assess:

- The added value to risk assessment of the two-generation reproductive toxicity study when a subchronic study is available.
- The impact on risk assessment of the second generation within the two-generation reproductive toxicity study.
- The added value to risk assessment of the rabbit developmental study when a rat developmental study is available.

For all three analyses, the impact on both the derived NOAEL and the classification for toxicity to fertility will be assessed. The lowest NOAEL among the toxicological studies for a substance will constitute the point of departure to derive a human safe dose. Thereafter, this human safe dose and the information on exposure will be used for the risk assessment of the substance. On the other hand, labelling can lead to restrictions on the use of a substance or mixtures containing it. In addition, classification of a substance can have important consequences within REACH. First, the classification of a substance in Directive 92/32/EEC

Annex 1 is considered a trigger to perform a risk assessment. And second, REACH will establish an authorization system for substances of very high concern, which include persistent, bioaccumulative and toxic substances, very persistent and very bioaccumulative substances, and substances classified as Category 1 and 2 mutagenic, carcinogen or reproductive toxicant. For all these substances, a complete risk assessment will be performed for these substances prior to their actual use (http://ecb.jrc.it/REACH/).

6.2 Background

6.2.1 Two-generation study versus repeated dose toxicity studies

Previous publications have compared the sensitivity of different endpoints to evaluate reproductive toxicity (fertility, organ weight, organ histopathology…) and have suggested that the reproductive toxicity studies are not more sensitive than repeated dose toxicity studies. Mangelsdorf et al. (2003) reviewed previous reports (e.g. Takayama et al., 1995; Gray et al., 1989; Ulbrich and Palmer, 1995; Chapin and Sloane, 1997) and concluded that testes histopathology was the most sensitive endpoint to detect adverse effects on male reproduction. Other parameters, such as reproductive organ weights and sperm parameters, showed also higher sensitivity than fertility parameters (e.g. number of implantations per female). Testes weight and histopathology evaluations are required in the rat oral subacute study (OECD Test Guidelines 407); and testes, epididymides, uterus, and ovaries weight and histopathology evaluations are required in the rat subchronic study (OECD Test Guidelines 408, 411, and 413). Sakai et al. (2000) showed that lesions in male reproductive organs can be detected in most cases after two weeks of treatment. Therefore, the length of these repeated dose tests (the subacute and the subchronic study mentioned above) should be sufficient in most cases to detect effects on testes weight and histopathology of a male reproductively toxic substance, if tested at an appropriate dose. In accordance with this conclusion, Chapin et al. (1988) showed a high degree of concordance in terms of reproductive toxicity between 90-day general toxicity studies and continuous breeding studies in mice. In addition, a similar sensitivity of these tests is suggested, because similar doses were selected in both tests (Chapin et al., 1998).

A recent report by the TERA group (Toxicology Excellence for Risk Assessment; Gadagbui et al., 2005) compared the NOAELs obtained in a rat chronic and a rat two-generation study for a series of chemicals. The analysis showed that, in general, lower NOAELs were obtained in rat chronic than in rat reproductive toxicity studies (Gadagbui et al., 2005). However, in 10 out of 128 pesticides, the NOAEL obtained in the rat reproductive toxicity study was reported as being more than 10-fold lower than that obtained in the rat chronic study (Gadagbui et al., 2005).

Most of the evidence presented above supports the similarity in terms of the sensitivity between repeated dose toxicity studies (if including parameters such as testes histopathology) and two-generation studies. Nevertheless, it could be argued that the distribution of NOAEL ratios reported in the TERA report showed the chronic test as more sensitive because most of the chemicals included in the database are not toxic to reproduction (i.e. only 9 out of 128 substances were classified as toxic to reproduction by 92/32/EEC Annex 1 or U.S. California E.P.A.). On the other hand, the the chemicals in the TERA report for which reproductive toxicity appeared most sensitive require further scrutiny. In principle, the two-generation test is still useful to detect effects on endpoints and mechanisms of reproductive toxicity other than testicular toxicity, and to detect a possibly higher response due to lifetime

exposure. Furthermore, specific knowledge on reproductive endpoints in relation to general toxicity is informative for classification and labelling.

6.2.2 Rabbit developmental toxicity study versus rat developmental toxicity study

The developmental studies aim to detect abnormal prenatal development and congenital malformations induced by exogenous chemicals or physical agents. In contrast to most toxicological tests, developmental studies are usually required in two species, a rodent and a non-rodent. One of the reasons for this requirement is the thalidomide catastrophe in 1961. Thalidomide, a sedative that was used in pregnancy, caused malformations in about 20% of newborns of mothers who ingested the drug during the sensitive period (Shepard, 1998). Thalidomide shows species differences in both effective doses and in pattern of effects observed. The teratogenic doses in rabbits were about 50-fold higher than those in humans, but the type of malformations observed (limb defects) were similar between both species (Shepard, 1998). In contrast, most rat strains did not show teratogenicity but embryo- or fetal-lethality (Schardein, 2000).

For many teratogenic agents, death, malformation, or growth retardation can be observed depending on the time of embryonic or fetal exposure and the amount or dose encountered. Animal studies are used to identify whether a hazard for developmental toxicity exist and to determine a NOAEL. Nevertheless, it is not realistic to expect a teratogenic manifestation similar to the one expected in humans. In addition, in order to classify an industrial chemical as toxic to development, the observed effects should not be secondary to maternal toxicity (Directive 92/32/EEC). This consideration may not be relevant for medicinal drugs, where the pharmacological action may outweigh toxic effects for mother and child in a risk benefit analysis.

Several reviews exist on the comparability of developmental toxicity among different test species (Frankos., 1985; Schardein and Keller, 1989; Jelovsek et al., 1989; Schardein, 2000; Hurt et al., 2003). These reviews have led to the general opinion that a developmental study in a non-rodent species is necessary for the toxicological evaluation of most substances, including industrial chemicals. However, these reviews include no consideration of whether developmental toxicity is or is not a consequence of maternal toxicity, and they interpret the differences in developmental toxic manifestations observed in different species (e.g. resorptions and malformations) by default as relevant for risk assessment.

6.2.3 Second generation versus first generation in a two-generation reproductive toxicity study

Increasing pressure is exerted by some stakeholders to reduce the two-generation study into a one-generation study. A general requirement of a one-generation instead of a two-generation study would considerably reduce the number of animals and other costs involved in these lengthy studies.

A proposal was recently presented by the Agricultural Chemical Safety Assessment (ACSA) Technical Committee of the ILSI Health and Environmental Sciences Institute (HESI) for an improved approach to assessing the safety of crop protection chemicals. In the tiered approach proposed, an enhanced one-generation study was introduced. This test reduces the pre-mating period to 4 weeks for males and 2 weeks for females and doses pups until PND70,

which would be divided in three subsets, one for clinical pathology and developmental neurotoxicity, one for immunotoxicity, and one for estrous cycles and possible continuation towards a second generation. The extension of the one-generation study to additional litters or to a second generation would depend on special triggers (Cooper et al., 2006). Adverse effects on fertility or fecundity of the parental generation, abnormal sexual development of the F1 pups, or deaths or evidence of toxicity to the F1 preweaning, as well as equivocal effects on those parameters, were considered triggers for the extension of the study to a second generation.

The validity of this enhanced 1-gen study was supported by a preliminary analysis performed by the USEPA Office of Pesticide Programs Health Effects Division and summarized in Cooper et al. (2006). In that analysis a data set of approximately 350 pesticide chemicals indicated that, with two possible exceptions, adverse effects would have been detected from first generation results and other available toxicology data. However, only the preliminary conclusions, without the underlying data, have been published. In addition, the question remains on whether the conclusions reached for pesticides would as well apply to other types of chemicals.

6.3 Data collection

A database has been constructed with rat subchronic (90-days), rat reproductive (two-generations), rat developmental toxicity, and rabbit developmental toxicity tests for substances that:

- are classified as toxic to fertility (R60 or R62) or development (R61or R63), or R64 up to and including the 30th ATP of Annex 1, or
- are known to cause male and/or female reproductive according to the US State of California Environmental Protection Agency toxicity, or
- are known to cause toxicity to development or to fertility according to the Dutch Health Council.

Publicly available, peer-reviewed national or international documentations were preferably used (Table 6.1). When no data or insufficient data for a specific substance were found in these sources, a search was done in the IUCLID files, the open literature (Medline), in the publicly available ECB meeting reports and working documents, and in the confidential evaluation files used for classification and labelling within RIVM. In order to keep the confidentiality of the data, no identification of those substances will be provided in the dissemination of results.

A comparable database was created for substances not classified for either toxicity to fertility or development. Data for these substances were obtained from the same sources as for classified substances, except that no confidential information was used.

Additionally, data were collected on the chronic (and subchronic, if available) and the two-generation study for substances that show differences higher than 10-fold between these two studies in the TERA report.

It should be noted that some of the included studies did not completely follow the OECD guidelines (408 or 413 for the subchronic study; 416 for the two-generation study). For example, this might regard the number of animals included or type of endpoints assessed in

Table 6.1. Peer-reviewed national or international data sources used.

Source	*Webpage*
EU Risk Assessment Reports for Existing Substances	*http://ecb.jrc.it/esis/*
OECD Screening Information Dataset for High Production Volume Substances	*http://www.chem.unep.ch/irptc/sids/OECDSIDS/sidspub.html*
Joint Meeting on Pesticides Residues	*http://www.inchem.org/pages/jmpr.html*
Environmental Health Criteria Monographs	*http://www.inchem.org/pages/ehc.html*
Concise International Chemical Assessments Documents	*http://www.inchem.org/pages/cicads.html*
US Agency for Toxic Substances and Disease Registry	*http://www.atsdr.cdc.gov/toxpro2.html#bookmark05*
US National Toxicology Program- Center for the Evaluation of Risks to Human Reproduction	*http://cerhr.niehs.nih.gov/reports/index.html*
Canadian Pest Management Regulatory Agency	*http://www.pmra-arla.gc.ca/english/pubs/reg-e.html*
Health Council of the Netherlands	*http://www.gr.nl/*
German Advisory Committee on Existing Chemicals of Environmental Relevance	*http://www.hirzel.de/bua-report/download_53_99.html*
California EPA evaluation reports	*http://www.oehha.org/prop65/hazard_ident/hazard_id.html*

the study. In order to provide some information on the quality of the studies, these were rated as follows: 1 and 2 for studies that certainly or probably followed OECD guidelines, respectively; 3 and 4 for studies that probably and certainly did not follow OECD guidelines, respectively. In addition, some studies found effects at the lowest tested level and did not allow the derivation of a NOAEL. In those cases, the lowest tested dose divided by 3 was used as an estimation of the NOAEL, except when effects were observed at lower doses in other studies for the same substance. Finally, no subchronic study was available for some reproductive toxicants. For these substances, the NOAEL from other repeated dose toxicity studies or the NOAEL for systemic toxicity in the P_0 generation of the two-generation study were used as an estimation of the subchronic NOAEL. In order to have a representative and relevant number of substances, strict compliance with OECD guidelines and inclusion of a NOAEL could not be required as inclusion criteria. However, these studies were reported in the databases used and therefore considered to have sufficient quality.

The highest tested dose, NOAELs and LOAELs (Lowest-Observed-Adverse-Effect-Level) for the critical effect and for reported effects on reproductive organs or associated glands, as well as details on the nature of these effects, were collected for subchronic studies. NOAELs and LOAELs for the effects on the parental generation (P_0), on fertility and on pup

development (F_1 or F_2) or in the first generation adults (F_1) were collected for two-generation studies. When several studies were available for the same substance, those that did not follow OECD guidelines were excluded, and the geometric mean of the NOAELs was calculated for the remaining studies. Additionally, the effects found at all tested doses and at all generations were recorded in a separate database. This database was used to compare the first and the second generation in the two-generation study.

In addition to the substances classified for developmental toxicity in Annex 1, two frequently quoted reviews on comparative developmental toxicity between species were used to identify substances where the rabbit study might add critical information to the rat study. One of these reviews was the book 'Chemically induced birth defects' (Schardein, 2000). Substances that were reported as being teratogenic in rabbits, but not in rats were selected. Apart from the information contained in the book, additional data for the rat and rabbit studies on those substances were searched for in the book 'Catalog of teratogenic agents' (Shepard, 1998), and the open literature. The second review, was the paper 'Proposal for a tiered approach to developmental toxicity testing for veterinary pharmaceutical products for food-producing animals' (Hurtt et al., 2003). Similarly as in the previous review, substances that were reported as being developmental toxicants in rabbits, but not in rats were selected. Additional information on these substances was obtained from the European Medicine Agency Reports, from the JEFCA Monographs and Evaluations, and from the open literature.

The data collected were used to compare the effects observed in the rat and rabbit study, in terms of occurrence of developmental toxicity and dependence/independence of maternal toxicity. In addition, the ratio was calculated between the NOAELs for developmental toxicity and those for maternal toxicity from the rat and the rabbit studies.

6.3.1 Data availability

Annex 1 contains 182 substances classified for reproductive or developmental toxicity. Some of these substances are very closely related (e.g. cadmium chloride and cadmium fluoride). Therefore, the list in annex 1 was reduced to 141 substances by grouping compounds with these two criteria: they were different salts of the same chemical, or they were active metabolite and parent compound. Of these 141 substances, 76 were classified for reproductive toxicity and 89 were classified for developmental toxicity. Despite the number of sources screened for the analysis, data could not be found for 11 substances out of the 141 substances included in Annex 1. These data are likely to be held by EU members other than the Netherlands in confidential files for classification and labelling and for new substances notifications.

The list of substances known to cause reproductive toxicity by California EPA included 56 substances that were not present in the EU Annex 1. Thirty five of these substances are drugs (medicinal or abuse). Toxicological data were found for all other substances. The list of substances classified as toxic to reproduction (fertility or development) by the Dutch Health Council included 17 substances that were not present in the EU Annex 1. Toxicological data were found for all these substances.

6.3.2 Studies included in the two-generation study versus subchronic study comparison

Within the selected substances for which toxicological data were found, approximately a third of the substances classified as toxic to fertility and a third of the substances classified as

toxic to development had been tested in a two-generation rat study (25 out of 76 substances with R60 or R62, 37 out of 89 substances with R61 or R63, 5 out of 21 substances in the California EPA list, none of the 17 substances in the Dutch Health Council list). The classification and labelling working group had classified most of the remaining substances as toxic to fertility based on reproductive toxicity observed in mouse studies, other rat fertility studies, and/or subacute or (sub)chronic studies. The classification as toxic to development was, for most of the substances, based on developmental toxicity studies. Finally, in a few cases, classification was based on structural similarities with other classified substances.

Altogether a two-generation study was found for 48 substances considered as reproductive toxicants. No subchronic study was available for 7 of these substances. In the latter cases, the NOAEL from other repeated dose toxicity study (a subacute study, a 6-month study, and a chronic study were used for three substances) or the NOAEL for systemic toxicity in the P_0 generation of the two-generation study (4 substances) were used as an estimation of the subchronic NOAEL.

Additionally, data from a two-generation study and a subchronic study were collected for 76 substances not considered reproductive toxicants.

6.3.3 Studies included in the second generation versus first generation comparison

Some of the studies used for the comparison between the subchronic and the two-generation study could not be used for the comparison between the first and the second generation. The reason is that the summary reports available for those substances did not include details on the effects observed in the two-generation study. As a consequence, two- or multi- generation studies were available for a total of 47 substances classified for reproductive toxicity up to and including the 30th ATP of Annex 1 or considered as toxic to fertility by the California EPA. In addition, a two- or multi-generation study was collected for a total of 82 substances that were not classified for reproductive toxicity. The information for the studies was obtained from the sources mentioned in Table 1. For some substances more than one study was found. In total, 162 multi-generation studies for 129 substances were included in the analyses.

6.3.4 Studies included in the rabbit developmental toxicity study versus the rat developmental toxicity study comparison

Despite the number of sources screened for the analysis, data could not be found for 7 substances out of the 89 substances included in Annex 1. These data are likely to be held in the confidential classification and labelling files and/or in the new substances dossiers of EU members other than the Netherlands. In addition, toxicological data were found for the 15 substances classified as developmental toxicants by the Dutch Health Council that are not present in the Annex 1.

6.4 Completeness and quality of the database

The number of substances classified for reproductive toxicity is limited and sometimes they were tested in studies that did not completely follow OECD guidelines. Nevertheless, most of the substances currently classified were included in the analyses. In addition, data on

developmental toxicity were included for medicine and veterinary drugs, increasing considerably the number of evaluated substances.

The conclusions will be based on the substances evaluated to the present with information that could be accessed. The concern will always remain that a future substance might have different characteristics than the ones tested to the moment, and that for such hypothetical substance, the second generation in the two-generation test and/or the rabbit developmental toxicity test might be necessary to identify its hazard.

6.5 Expected results

The results will describe the impact of the two-generation study, the second generation within this study, and the rabbit developmental study to risk assessment and risk management. For example, the study will provide information on the proportion of reproductive toxic substances that would not be classified if a two-generation study would be missed, or information on the differences between the NOAELs obtained in a two-generation study versus a subchronic study. In addition, the implications of the results for an ITS under REACH will be discussed.

The results obtained from these analyses will be disseminated by publishing them in peer-reviewed journals and presenting them in international conferences.

7. Discussion and conclusions

The need for ITS
The aim of this report was to investigate and analyse existing alternative methods and integrated testing strategies. Secondly, selected topics relevant for the subject of ITS were investigated: the Weight-Of-Evidence (WoE) process as a general subject and subsequently components of ITS for the endpoints environmental degradation, sensitisation and reproduction toxicology.

The consequence of REACH is that in a relative short time period the risk of a large group of chemicals has to be assessed. This implies that a large amount of information on the fate and effects of chemicals will become available. For reasons of animal welfare, costs and logistics, it is important to limit the number of tests to be conducted. This means that alternative methods (non-testing methods and *in vitro* tests) have to be developed and validated. However, as concluded by the European Chemicals Bureau, it will require many years before an extensive program aimed at the development and validation of alternative methods could bear fruit, whereas the full replacement of animal testing would not be possible at all for some endpoints.

It is therefore that Integrated Testing Strategies (ITS) are widely proposed as a solution towards the increased testing requirements for classification and labelling and risk assessment under REACH. ITS combine non-testing, *in vitro, in vivo* and exposure information. They are hierarchical in nature, starting by making maximum use of existing effects and exposure data. Key to the resulting decision schemes is the WoE process to be followed which should be as explicit as possible in order to determine the uncertainty in their outcome and to allow consistent and transparent decision making.

ITS are nothing new: such strategies have been developed for both classification and labelling and risk assessment in various regulatory frameworks to a varying degree and in different ways. However, so far the integration of alternative methods in these ITS is very limited. Given the conclusions above, the transition of the current strategies for most endpoints into strategies encompassing more – let alone exclusively - alternatives will require a huge research effort and will take a considerable period of time. Fortunately, such efforts are underway within the scope of research programs of OECD and EU. For instance, a 10 M€ research program on ITS has been initiated within the scope of the EU 6th Framework Program. Updating of guidance on ITS to be able to absorb the results of such research needs to be performed on a continuous basis.

Components of ITS
Under the current EU legislation for new and existing chemicals, the regulatory use of *non-testing methods* ((Q)SARs, grouping methods) and *in vitro* methods is limited and varies considerably among the Member States. This is due to the fact that there is no agreement in the scientific and regulatory communities over their application and the extent to which estimates can be relied on. Validation, limited applicability domain, and poor availability of guidance are the major limitations of these approaches for regulatory testing. Nevertheless, some methods are already part of the EU legislation on chemicals. So there is already some regulatory implementation and acceptance in the EU Member States, but further work needs to be done.

Optimization of in vivo tests by a) only performing tests that provide relevant data; b) eliminating redundant tests; c) using one sex; d) applying some tests simultaneously to the same animals; and e) making greater use of screens and preliminary testing also are widely regarded to be a useful option which warrants further research.
Possibilities for *exposure based waiving* of tests are promising, especially if combined with (Q)SAR or read-across, but it requires further investment in the development of exposure models and it also needs precise information on the use pattern of the chemicals (e.g. downstream use information) and on risk reduction measures in place, which are current bottlenecks. Under REACH the operationalisation of the concept of exposure scenario should go a long way in solving this. An exposure scenario is a description of a control strategy for substances, giving realistic operational conditions for manufacture of a substance or identified use(s) of a substance, a group of substances or a preparation.

Toxicogenomic approaches are recognised as not yet developed enough for direct replacement of existing approaches, but it could give supportive evidence on a case-by-case basis. These techniques can provide additional evidence of both exposure to and effects of pollutants.

This report discusses one overarching element of ITS, WoE, and aspects of ITS for three endpoints: environmental degradation, sensitisation and reproductive toxicology. In the following, main discussion points and conclusions are presented

WoE
The discussion on WoE in ITS is still rather young. The WoE is often mentioned in the risk assessment literature, without adequate documentation. In many cases, it is not clear which methods were used, how they were applied to the scientific evidence, what the results were and how these were used to make decisions in a specific risk assessment. Both qualitative and quantitative weighting methods are in use. An important issue in WoE is the influence of expert judgment that needs to be recognized and made explicit as far as possible. It should be documented what type of information is assessed, why it is assessed and which quality criteria are used. The interpretative methods as well as the weighting procedure should be transparent and clear separation between scientific evidence and (value driven) expert judgment is needed.

Prior knowledge about chemicals can be used in a WoE approach in a Bayesian framework and this approach has been explained in this report and demonstrated for biodegradation data. Such prior knowledge could be derived e.g. from experimental results for a class of similar chemicals, or information from model predictions. The advantage of the Bayesian analysis is that, once familiar with the terminology and notation, relatively simple calculations can demonstrate the influence of the prior knowledge on the outcome of the model predictions. The consequences of the additional information can then be evaluated for increased certainty on the outcome of the test. This, however, emphasizes that the user of the information needs to decide which certainty is acceptable or not, since the outcome of the analysis is expressed as probabilities.

The current analysis shows that the use of Bayesian statistics allows quantifying the information value of additional information in sequential steps of an ITS. Although the predictive value can be improved by applying a model battery, as demonstrated in the biodegradation example, it still needs to be decided whether one is satisfied with a model outcome, or needs to conduct further testing. This decision is not only a matter of agreeing on

cut-off probabilities for the uncertainty one is willing to accept. Additional considerations on costs and magnitude of the potential risk are important as well. It is therefore needed to expand the statistical framework towards decision making. The current application as demonstrated for biodegradation data needs to be expanded towards different assessment endpoints. Although the updating process will in theory always be the same, variations are needed to deal with different types of information such as discrete or continuous data and expert judgement.

Biodegradation
This report reviews briefly the current status of QSAR applications for environmental degradation with regard to classification and labelling, PBT assessment and risk assessment. It further concentrates on the usefulness of CATABOL as a tool for the assessment of biodegradation and biodegradation products within the scope of an ITS.

Under the current EU legislation for new and existing chemicals, the regulatory use of estimation models or (Q)SARs is limited and preliminary of nature. It takes account of the fact that the models show better performance in their predictions of not-ready biodegradability. The application of (Q)SARs in exposure assessment is only possible if a substance can be judged as being readily degradable or not readily degradable and if a degradation rate constant can be predicted with sufficient certainty.

Although, the EU TGD highlights that, where degradation occurs, consideration should be given to the properties (including toxic effects) of the products that might arise, that information does not exist for many compounds. Guidance is needed to establish the criteria upon which metabolites of concern may be identified and to determine when a metabolite would not be of concern. Degradation route studies are, however, complex and costly, and it is often very difficult to identify the minor degradation products in a system. A promising model which can be used for quantitative assessment of biodegradability in biodegradation pathways of chemicals is CATABOL. This system generates most plausible biodegradation products and provides quantitative assessment for their physicochemical properties and toxic endpoints.

The main conclusions of the analysis of CATABOL are:
1. When the goodness of the prediction is related to the cut-off value of 60% BOD, CATABOL performs well in predicting ‘not readily biodegradability’. The verification study showed that BOD predictions for substances that are ‘out of the domain’ are approximately as good as predictions for substances that are ‘in the domain’. For ready biodegradable substances the performance is less. As the chemicals selected in this verification study were in majority not readily biodegradable, they are less suitable to draw a conclusion on the goodness of ‘ready’ predictions.
2. Verification of major metabolites of some existing chemicals and pesticides formed reveals that the predictive power of CATABOL for identifying major metabolites formed in soil is poor: many major metabolites are predicted that are not observed and many observed metabolites are not predicted. The majority of these substances (>90%) are ‘out of the domain’. As the number of chemicals ‘in the domain’ is very small, it is not possible to assess whether the performance is better for substances that are ‘in the domain’.
3. Nevertheless, when no data are available, the degradation rates in soil can be estimated by deriving DT50 from BOD or remaining quantities in soil. Values predicted in this way could be used as (worst case) upper limit values in risk assessments.

4. Despite the poor prediction of metabolites formed in soil CATABOL still predicts well the biodegradability of the substance. For a large number of chemicals (see section 4.7.3), the screening data indicate that the parent compound should be persistent which is confirmed by CATABOL. Apparently, CATABOL is able to assess whether the transformation will be blocked by a recalcitrant fragment, without being able to predict the exact identity of the recalcitrant degradation product.

Overall, despite the uncertainties discussed above and the fact that many substances investigated are out of domain, CATABOL predictions could still be useful in a WoE approach or to target further testing. For substances for which little or no information is available, CATABOL could be used to investigate whether primary degradation is likely to occur and whether the metabolites formed are of concern by predicting e.g. their the log Kow as an indication for their bioaccumulation potential. The fact that CATABOL gives information on the quantities of degradation products in addition to BOD of the parent compound is a feature that can be very useful in targeting the risk assessment and testing. CATABOL could also be used to verify read across.

Sensitisation
The section of sensitisation reflects the discussions in the RIP 3.3-2 Endpoint Working Group on ITS for this endpoint up to the end of 2006. It includes the contributions from the Netherlands' delegation to this Working Group which was supported by the two projects which made this report possible.

There are many different (Q)SARs and expert systems available for the estimation of skin sensitisation hazard. The approaches are quite varied and each has been developed on different sets of *in vivo* data. In many cases these models have been demonstrated to be reasonable for predicting skin sensitisers correctly but are limited in predicting non-sensitisers correctly. For this reason, careful interpretation of model predictions needs to be considered in light of other information e.g. analogue read-across. Further work should explore encoding more knowledge/rules for non-reactive chemicals as well as those chemicals likely to undergo chemical or metabolic transformation.

At present, no officially adopted EU-OECD *in vitro* tests for skin sensitisation exist. However, several systems are in the course of development, based on an improved understanding of the biochemical and immunological mechanisms underlying the process. Current *in vitro* assays only cover a (specific) part of the process of sensitisation and so far *in vitro* data can only be used in a WoE approach in conjunction with other data, and currently for positive identification of sensitisers only.

Due to the complexity of the mechanisms of skin sensitisation, a single non-animal test would not be able to replace a complete *in vivo* experiment. However, this probably should not be the ultimate goal. A tiered approach might be more realistic based on both non-testing and testing approaches. A tiered ITS approach consisting of non-animal methods, and possibly including *in vivo* testing as a last resort for a very small percentage of the total amount of chemicals, may be realistic for the near future. However, this is not expected to be implemented before the start of REACH in June 2007.

Besides the fact that a non-animal ITS is scientifically premature, a major hurdle might be the willingness of industry to perform such a battery of tests, especially when they are more expensive and laborious than the *in vivo alternative*. Another threat to this alternative

approach might be the upcoming introduction of GHS (Globally Harmonised System) possibly introducing the concept of potency. RIVM is already involved in national and international discussions on the possible introduction of potency ranking for sensitisation, but more work could be performed on estimating possible implications for risk assessment strategies.

The performance and applicability domains of the different guinea pig and mice *in vivo* tests for determining skin sensitisation potential is still under a lot of discussion. In REACH, the LLNA is mentioned as the first-choice method for *in vivo* testing. After critical review of this kind of studies it might be concluded that there is a need for other *in vivo* possibilities. REACH does not include testing requirements for respiratory sensitisation. A number of test protocols has been published to detect respiratory allergenicity of low molecular weight compounds, but none of these are validated nor are these widely accepted. Given lack of available (Q)SARs and *in vitro* tests for respiratory sensitisation, it is not possible to provide any additional guidance on the evaluation of non-testing data for respiratory sensitisation. Therefore, the development of models and testing guidance for respiratory sensitisation should be stimulated. It might be possible to conclude in a WoE assessment that chemicals that are negative in the LLNA, as well as being considered as not being skin sensitisers, can also be regarded as lacking the potential to cause allergic sensitisation of the respiratory tract.

Reproductive toxicity
In the view of the upcoming changes in regulation, there is a need to retrospectively evaluate the efficiency of the ITS followed to the present. After more than twenty years of chemical testing, considerable amounts of toxicological data have been generated. This toxicological data can be used for retrospective analyses. In the present report such a study is described focusing on the added value to risk assessment of the two-generation reproductive toxicity study when a subchronic study is available, the impact on risk assessment of the second generation within the two-generation reproductive toxicity study and the added value to risk assessment of the rabbit developmental study when a rat developmental study is available. For all three analyses, the impact on both the derived NOAEL and the classification for toxicity to fertility will be assessed. Most of the substances currently classified for reproductive toxicity were included in the analyses. In addition, data on developmental toxicity was included for medicine and veterinary drugs, increasing considerably the number of evaluated substances. The study will, among others, provide information on the proportion of reproductive toxic substances that would not be classified if a two-generation study would be missed, or information on the differences between the NOAELs obtained in a two-generation study versus a subchronic study. In addition, the implications of the results for an ITS under REACH will be discussed.

The results obtained from these analyses will be disseminated by publishing them in peer-reviewed journals and presenting them in international conferences.

Acknowledgments

The authors thank Etje Hulzebos, Wim Mennes Jean-Paul Rila, E. Rorije, Wout Slob, all of RIVM, Inge Mangelsdorf of Fraunhofer Insitute for Toxicology and Experimental Medicin in Germany, Vicki Dellarco of theUS-EPA Office of Pesticide Program and David Dix of US-EPA National Center for Computational Toxicology, Ovanes Mekenyan and Saby Dimitrov of the Laboratory of Mathematical Chemistry of Bourgas University in Bulgaria, and Mark Bonnell of Environment Canada for their valuable contributions to this project.

The authors also thank the Director General of RIVM and the Directorate-General for Environmental Protection, Directorate for Chemicals, Waste and Radiation Protection for making this project possible.

References

Chapter 1

Bradbury, S., Feijtel, T.,van Leeuwen, K. 2004. Meeting the scientific needs of ecological risk assessment in a regulatory context. Environ.Sci.Technol. 38/23, 463a-470a.

Combes, R., Barratt, M., Balls, M. 2003. An overall strategy for the testing of chemicals for human hazard and risk assessment under the EU REACH system. ATLA 31: 7-19

EC. 2003. Technical Guidance Document in support of the Commission Directive 93/67/EEC on risk assessment for new substances, Commission Regulation (EEC) No.1488/94 on risk assessment for exisiting substances, and Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market, Part 3. Ispra, Italy, European Chemicals Bureau, Joint research Centre, EUR 20416 EN/1

EC. 2004. Commission Directive 2004/73/EC of 29 April 2004 adapting to technical progress for the twenty-ninth time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. O.J. No L152 2004. (2004)

EC. 2006. Proposal for a regulation of the European Parliament and of the Council concerning the Registration, Evaluation, Authorisation and restriction of Chemicals (REACH), establishing a European Agency and amending Directive 1999/45/EC and Regulation (EC) {on persitent organic pollutants}. European Commission, Brussels, Belgium, 2003/0256 (COD).

JRC. 2005. REACH and the need for Intelligent Testing Strategies. Ispra, Italy, Joint Research Centre, Institute of Health and Consumer Protection. EUR 2154 EN.

Pedersen, F., de Bruijn, J., Munn, S., van Leeuwen, K. 2003. Assessment of addiitonal testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives. Ispra, Italy, Institute of Health and Consumer Protection, Joint Research Centre, EUR 20863.

TAPIR. 2005. Scoping study on the Development of a Technical Guidance Document on Information requirements on Intrinsic Properties of Substances (RIP 3.3-1). TAPIR Three point three – A Project for the Information Requirements of REACH Final Report – August 2nd 2005. Commission contract number – 22553-2004-12 F1SC ISPBE.

Van der Jagt, K., Munn, S., Tørslov, J., de Bruijn, J. 2004. Alternative approaches can reduce the use of test animals under REACH. Addendum to the report: Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives. Ispra, Italy, Institute of Health and Consumer Protection, Joint Research Centre, EUR 21405 EN.

Worth, A. and Balls, M. 2002. Alternative (non-animal) methods for chemical testing: current status and future testing. A report prepared by ECVAM and the ECVAM Working

Group on Chemicals. Ispra, Italy, European Centre for the Validation of Alternative Methods. Institute of Health and Consumer Protection, Joint Research Centre. Also published in ATLA 30, Supplement 1: 1-125.

Chapter 2

Barlow, S. 2005. Threshold of toxicological concern (TTC). A tool for assessing substances of unknown toxicity present at low levels in the diet. Brussels, Belgium, International Life Sciences Institute Europe. ILSI Europes Concise Monograph Series 2005:1-31.

Blaauboer, B.J., Barrett, M.D., Houston, J.B. 1999. The integrated use of alternative methods in toxicological risk evaluation. ATLA 27, 229-237.

Botham, P.A. 2004. Acute systemic toxicity - prospects for tiered testing strategies . Toxicology in invitro 18, 227-230.

Combes, R. and Balls, M. 2005. Intelligent testing strategies for chemicals testing - a case of more haste, less speed? ATLA 33, 289-297.

EC. 2006. Proposal for a regulation of the European Parliament and of the Council concerning the Registration, Evaluation, Authorisation and restriction of Chemicals (REACH), establishing a European Agency and amending Directive 1999/45/EC and Regulation (EC) {on persitent organic pollutants}. European Commission, Brussels, Belgium, 2003/0256 (COD).

ECETOC. 2003. (Q)SARs: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications. Brussels, Belgium, European Centre for Ecotoxicology and Toxicology of Chemicals, ECETOC Technical Report No. 89. ISSN-0773-8072-89.

Gant, T.W. and Zhang, S. 2005. In pursuit of effective toxicogenomics. Mutat. Res. 575, 4-16.

Groen, K., Rila, J.-P., Pelgrom, S.M.G.J., Sips, A.J.A.M., Hakkert, B.C. 2005a. Alternative in vitro methods for systemic toxicity studies, ready for REACH? Bilthoven, RIVM, SEC-report No. 09446A00. Available from RIVM Expertise Centre for Substances.

Groen, K., Slob, W., Van Eijkeren, J.C.H., Piersma, A.H., Freidig, A., Verwei, M., Hakkert, B., Sips, A.J.A.M. 2005b. Quantitative translation of in vitro concentration into in vivo doses. Bilthoven, RIVM, RIVM report 340720001. Available from RIVM Expertise Centre for Substances.

Hanway, R.H. 2002a. The use of toxicological read-across data in the notification of new chemicals. Industrial Chemicals Unit, HSE, Magdalen House, Stanley Precinct, Bootle, Merseyside, L20 3QZ, UK.

Hanway, R.H. 2002b. The use of read-across for environmental data in the notification of new chemicals. Industrial Chemicals Unit, HSE, Magdalen House, Stanley Precinct, Bootle, Merseyside, L20 3QZ, UK.

Jaworska, J., Aldenberg, T., Nikolova, N. 2005. Review of methods for assessing the applicability domains of SARs and QSARs. Final report to the Joint Research Centre

(Contract No. ECVA-CCR.496575-Z).

Jeram, S., Sintes, J.M.R., Halder, M., Fentanes, J.B., 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. Regul. Toxicol. Pharmacol. 42, 218-224.

JRC. 2005. REACH and the need for Intelligent Testing Strategies. Ispra, Italy, Joint Research Centre, Institute of Health and Consumer Protection. EUR 2154 EN.

Nendza, M. and Hermens, J. 1995. Properties of chemicals and estimation methodologies. In: Van Leeuwen, C.J. and Hermens, J. (eds) Risk assessment of chemicals. An introduction, pp. 239-292. Dordrecht, The Netherlands, Kluwer Academic Press.

Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D. , Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, D.T., van de Sandt, J.J.M., Tong, W., Veith, G., Yang, C. 2005. Current status of methods for defining the applicability domain of (Quantitative) Structure- Activity Relationships. ATLA 33: 155-173.

OECD. 1994. US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relationships. Paris, France, Organisation for Economic Cooperation and Development. OECD Environment Monograph No. 88. OCDE/GD(94)28. 367 pp. (http://www.epa.gov/opptintr/MPD-SAR).

OECD. 2004a. The report from the expert group on (quantitative) structure-activity relationships [(Q)SARs] on the principles for the validation of (Q)SARs. Paris, France, Organisation of Economic Cooperation and Development, Environment Directorate, 37th Joint Meeting of the Chemical Committee of the Working Party on Chemicals, Pesticides and Biotechnology, ENV/JM/MONO(2004)24, OECD Series on Testing and Assessment No. 49.

OECD 2004b Manual for investigation of HPV chemicals. Paris, France, Organization for Economic Co-operation and Development Available form www. oecd.org.

OECD/IPCS. 2005. Report of the OECD/IPCS Workshop on Toxicogenomics, Kyoto, 13-15 Oct. 2004. Paris, France, Organisation of Economic Cooperation and Development, Environment Directorate, Joint Meeting of the Chemical Committee of the Working Party on Chemicals, Pesticides and Biotechnology, ENV/JM/MONO(2005)10, OECD Series on Testing and Assessment No. 50. http://www.oecd.org/document/29/0,2340,en_2649_34377_34704669_1_1_1_1,00.html.

OECD. 2006. Report on the regulatory uses and applications in OECD member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. Paris, France, Organisation of Economic Cooperation and Development, Environment Directorate, Joint Meeting of the Chemical Committee of the Working Party on Chemicals, Pesticides and Biotechnology, ENV/JM/MONO(2006)25, OECD Series on Testing and Assessment No. 58.

Rila J.-P., Bos P., van Hoeven-Arentzen, P.H., Maslankiewicz, L., Hakkert, B. 2005.

Applicability of (non)-validated in vitro methods for regulatory requirements. Bilthoven, RIVM, SEC report No. EOH/I 09445. Available from RIVM Expertise Centre for Substances.

Rila, J.-P., Bos, P., Hulzebos, E., Hakkert, B. 2006. Workability of guidance on category/read across approaches for selected groups of chemicals. Bilthoven, RIVM, Report No 601200009. Available from RIVM Expertise Centre for Substances.

TAPIR. 2005. Scoping study on the Development of a Technical Guidance Document on Information requirements on Intrinsic Properties of Substances (RIP 3.3-1). TAPIR Three point three – A Project for the Information Requirements of REACH Final Report – August 2nd 2005. Commission contract number – 22553-2004-12 F1SC ISPBE.

TERA. 2005. Record for TERA peer consultation on Health Canada's Complex Hazard Tool (ComHaz) preliminary weight of evidence framework for genotoxic carcinogenicity. Cincinnati, USA, Toxicology Exellence for Risk Assessment.

Waters, M.D. and Fostel, J.M. 2004. Toxicogenomics and systems toxicology: aims and prospects. Nature Rev. - Genetics 5: 936-948.

Worth, A.P., Bassan, A., Gallegos, A., Netzeva, T.I., Patlewicz, G., Pavan, M., Tsakovska, I. and Vracko, M. 2005. The characterization of (Quantitative)Structure-Activity Relationships: preliminary guidance. Ispra, Italy, Joint Research Centre, European Chemicals Bureau, EUR 21866 EN.

Chapter 3

Boethling, R.S., Lynch, D.G., Jaworska, J.S., Tunkel, J.L., Thom, G.C., Webb, S. 2004. Using Biowin^{TM}, Bayes, and batteries to predict ready biodegradability. Environ. Toxicol. Chem., 23, 911-920.

Calabrese, E.J, Baldwin, L.A., Kostecki, P.T. and Potter, T.L.1997. A Toxicologically Based Weight-of-Evidence Methodology for the Relative Ranking of Chemicals of Endocrine Disruption Potential Reg. Toxicol. Pharmacol. 26:36-40

Campbell, M.J and Machin, D. 1993. Medical statistics. A commonsense approach (2nd ed.). Chichester, England, John Wiley,

Carlin, B.P., and Louis, T.A. 2000. Bayes and empirical bayes methods for data analysis. 2nd Ed. Chapman & Hall/CRC.

Denison, R.A. and Balbus, J.M. 2006. Experiences using integrated testing approaches to fulfil information requirements for testing and assessment.39th Joint Meeting of Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology, 15-17 feb. 2006.

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

Gelman, A.,Carlin, J.B., Stern, H.S., Rubin, D.B. 2004. Bayesian Data Analysis. 2nd Ed., Chapman & Hall/CRC.

Hulzebos, E., Sijm, D., Traas, T., Posthumus, R. and Maslankiewicz, L. 2005. Validity and validation of expert (Q)SAR systems. SAR QSAR Environ. Res. 16, 385-401.

Klimisch, H.J., Andreae, M., Tillmann, U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25:1-5.

Menzie C., Henning, M.H., Cura, J., Finkelstein, K., Gentile, J., Maughan, J., Mitchell, D., Petron, S., Potocki, B., Svirsky, S., Tyler, P. 1996. A weight-of-evidence approach for evaluating ecological risks: Report of the Massachusetts Weight-of-Evidence Workgroup. Human Ecol. Risk Assessment 2: 277-304.

Posthumus, R, Traas, T.P., Peijnenburg, W.J.G.M., Hulzebos, E.M. 2005. External validation of Epiwin biodegradation models. SAR QSAR Environ. Res. 16:135-148.

Press, S.J. 2003. Subjective and objective Bayesian statistics: principles, models, and application.Wiley, Hoboken NJ, USA.

Proctor, D.M., Otani, J.M., Finley, B.L., Paustenbach, D., Bland, J.A., Spezier, N., Sargent, E.V. 2002. Is hexavalent chromium carcinogenic via ingestions? A weight-of-evidence review. J. Toxicol. Environ. Health 65:701-746.

TAPIR. 2005. Scoping study on the Development of a Technical Guidance Document on Information requirements on Intrinsic Properties of Substances (RIP 3.3-1). TAPIR Three point three – A Project for the Information Requirements of REACH Final Report – August 2nd 2005. Commission contract number – 22553-2004-12 F1SC ISPBE.

US EPA. 2000. Estimation program interface for windows. [http:www.eap.gov/oppt/exposure/docs/episuitedl.htm], December 15, 2006

Weed, D.L. 2005. Weight of evidence: a review of concepts and methods. Risk Anal. 25: 1545-1557.

Winkler, R.L. 2003. An introduction to Bayesian inference and decision (2nd ed.). Probabilistic publishing, Gainsville FL, USA.

Chapter 4

Boethling, R.S., Lynch, D.G., Thom, G.C. 2003. Predicting Ready Biodegradability of Premanufacture Notice Chemicals. Environ. Toxicol. Chem. 22, 837-844.

Boxall, A.B.A., Sinclair, C.J., Fenner, K., Kolpin, D., Maund, S.J. 2004. Environ. Sci.Technol. October 1: 369a-375a.

Dimitrov, S., Kamenska, V., Walker, J.D., Windle, W., Purdy, R., Lewis, M., Mekenyan, O. 2004. Predicting the biodegradation products of perfluorinated chemicals using Catabol. SAR QSAR Environ. Res. 15 (1): 69-82.

Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., Mekenyan, O. 2005. A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models. J. Chem. Inf. Model 45: 839-849.

EC. 2003. Technical Guidance Document in support of the Commission Directive 93/67/EEC on risk assessment for new substances, Commission Regulation (EEC) No.1488/94 on risk assessment for exisiting substances, and Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market, Part 3. Ispra, Italy, European Chemicals Bureau, Joint research Centre, EUR 20416 EN/1

Hulzebos, E., Sijm, D., Traas, T., Posthumus, R., Maslankiewicz, L. 2005. Validity and validation of expert (Q)SAR systems. SAR QSAR Environ. Res. 16: 385-401.

Hulzebos, E.M. and Posthumus, R. 2003. (Q)SARs: gatekeepers against risk on chemicals? SAR QSAR Environ. Res. 14: 285-316.

Jaworska, J.S., Boethling, R.S., Howard, P.H. 2003. Recent developments in broadly applicable structure-biodegradability, Environ. Toxicol. Chem. 22: 1710-1723.

Jaworska, J., Dimitrov, S., Nikolova, N., Mekenyan, O. 2002. Probabilistic assessment of biodegradability based on metabolic pathways: Catabol system. SAR QSAR Environ. Res. 13: 307-323.

Klopman, G. and Tu, M. 1997. Structure-biodegradability study and computer-automated prediction of aerobic biodegradation of chemicals. Environ. Toxicol. Chem. 16: 1829-1835.

Nendza, M. 2004. Prediction of Persistence. In: Cronin, M.T.D. and Livingstone D.J. Predicting Chemical Toxicity and Fate. CRC Press 315- 331.

OECD. 1981-2006. Guidelines for testing of chemicals. Paris, France. Organization for Economic Co-operation and Development.

Posthumus, R., Traas, T.P., Peijnenburg, W.J., Hulzebos, E.M. 2005. External validation of EPIWIN biodegradation models. SAR QSAR Environ. Res. 16: 135-48.

Rorije, E., Peijenburg, W.J.G.M., Klopman, G. 1998. Structural requirements for anaerobic biodegradation of organic chemicals: a fragment model analysis. Environ. Toxicol. Chem. 17: 1943-1950.

Rorije, E., Loonen H., Muller, M., Klopman, G., Peijenburg, W.J.G.M. 1999. Evaluation and application of models for the prediction of ready biodegradability in the MITI-test. Chemosphere 38: 1409-1417.

Sakuratani, Y, Yamada, J., Kasai, K., Noguchi, Y., Nishihara, T. 2005. External validation of the biodegradability prediction model CATABOL using data sets of existing and new chemicals under the Japanese Chemical Substances Control Law. SAR QSAR Environ. Res. 16 (5): 403–431.

Sinclair, C.J., Dimitrov, S., Hetheridge, M.J., Boxall, A.B.A., Mekenyan, O., Comber, M. 2003. Assessment of the expert system CATABOL for predicting transformation products and pathways in soils. In: Proceedings of the 13th SETAC Europe Annual Meeting, 27 April–1 May, Hamburg, Germany. SETAC Europe, Brussels, Belgium.

Struijs, J. and Stoltenkamp, J. 1994. Testing surfactants for ultimate biodegradability. Chemosphere 28,: 1503-1523.

TAPIR. 2005. Scoping study on the Development of a Technical Guidance Document on Information requirements on Intrinsic Properties of Substances (RIP 3.3-1). TAPIR Three point three – A Project for the Information Requirements of REACH Final Report – August 2nd 2005. Commission contract number – 22553-2004-12 F1SC ISPBE.

United Nations. 2003. Globally Harmonized System for Classification and Labelling of chemicals (GHS). United Nations, Geneva, ST/SG/AC.10/30, ISBN 92-1-116840-6.

Chapter 5

Agius, R.M. 2000. Why are some low molecular weight agents asthmagenic? Occup. Med.-State of the Art Rev. 50:369–384.

Agius, R.M., Nee, J., McGovern, B., Robertson, A. 1991. Structure activity hypotheses in occupational asthma caused by low molecular weight substances. Ann. Occup. Hyg. 35:129–137.

Agius, R.M., Elton, R.A., Sawyer, L., Taylor, P. 1994. Occupational asthma and the chemical properties of low molecular weight organic substances. Occup. Med.-State of the Art Rev. 44: 34–36.

Aptula, A.O. and Roberts, D.W. 2006.Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. Chem. Res. Toxicol. 19(8):1097-105.

Aptula, A.O., Patlewicz, G., Roberts, D.W. 2005. Skin Sensitisation: Reaction Mechanistic Applicability Domains for Structure-Activity Relationships. Chem. Res. Toxicol. 18: 1420-1426.

Aptula, A.O., Patlewicz, G., Roberts, D.W., Schultz, T.W. 2006. Non-enzymatic glutathione reactivity and in vitro toxicity: a non-animal approach to skin sensitisation. Toxicol. In Vitro. 20(2):239-47.

Arts, J.H.E. and Kuper, C.F. 2006. Animal models to test respiratory allergy of low molecular weight chemicals. Methods (in press).

Arts, J.H.E., Kuper, C.F., Spoor, S.M., Bloksma, N. 1998. Airway morphology and function of rats following dermal sensitization and respiratory challenge with low molecular weight chemicals. Toxicol. Appl. Pharmacol. 152: 66-76.

Arts, J.H.E., Bloksma, N., Leusink-Muis, A., Kuper, C.F. 2003. Respiratory allergy and pulmonary irritation to trimellitic anhydride in Brown Norway rats. Toxicol. Appl. Pharmacol.,187: 38-49.

Ashby, J., Basketter, D.A., Paton, D., Kimber, I. 1995. Structure activity relationships in skin sensitisation using the murine local lymph node assay. Toxicology 103: 177-194.

Barratt, M.D., Basketter, D.A., Roberts, D.W. 1997. Quantitative structure activity relationships. In Lepoittevin J-P, Basketter DA, Dooms-Goossens A, Karlberg A-T, eds, The Molecular Basis of Allergic Contact Dermatitis. Springer-Verlag Heidelberg. pp. 129-154.

Basketter, D.A., Lea, L.J., Dickens, A., Briggs, D., Pate, I., Dearman, R.J., Kimber, I. 1999. A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. J. Appl. Toxicol. 19: 261-266.

Basketter, D.A., Andersen, K.E., Lidén, C., van Loveren, H., Boman, A., Kimber, I., Alanko, K., Berggren, E. 2005. Evaluation of the skin sensitising potency of chemicals using existing methods and considerations of relevance for elicitation. Contact Dermatitis, 52: 39 - 43.

Buckley, T.L. and Nijkamp, F.P. 1994. Mucosal exudation associated with a pulmonary delayed-type hypersensitivity reaction in the mouse. Role for the tachykinins. J. Immunol. 153: 4169-4178.

Casati, S., Aeby, P., Basketter, D.A., Cavani, A., Gennari, A. Gerberick, G.F. Griem, P. Hartung, T., Kimber, I., Lepoittevin, J.P., Meade, B.J., Pallardy, M., Rougier, N., Rousset, F., Rubinstenn, G., Sallusto, F., Verheyen, G.R., Zuang,V. 2005. Dendritic cells as a tool for the predictive identification of skin sensitisation hazard. ATLA 33: 47-62.

Cronin, M.T. and Basketter, D.A.1994. Multivariate QSAR analysis of a skin sensitisation database. SAR QSAR Environ. Res. 2(3):159-79.

Cunningham, A.R., Cunningham, S.L., Consoer, D.M., Moss, S.T., Karol, M.H. 2005. Development of an information-intensive structure-activity relationship model and its application to human respiratory chemical sensitisers. SAR QSAR Environ. Res. 16(3):273-85.

Dearman, R.J., Warbrick, E.V., Skinner, R., Kimber, I. 2002. Cytokine fingerprinting of chemical allergens: species comparisons and statistical analyses. Food Chem. Toxicol. 40: 1881-1892

Dearman, R.J., Betts, C.J., Humphreys, N., Flanagan, B.F., Gilmour, N.J., Basketter, D.A., Kimber, I. 2003. Chemical allergy: considerations for the practical application of cytokine profiling. Toxicol. Sci. 71, 137-145.

Dimitrov, S.D., Low, L.K., Patlewicz, G.Y., Kern, P.S., Dimitrova, G.D., Comber, M.H., Phillips, R.D., Niemela, J., Bailey, P.T., Mekenyan, O.G. 2005. Skin sensitisation: modeling based on skin metabolism simulation and formation of protein conjugates. Int. J. Toxicol. 24(4):189-204.

ECETOC. 2002. Use of human data in hazard classification for irritation and sensitization. Brussels, European Centre for Toxicology and Ecotoxicology of Chemicals, ECETOC Monograph No.32.

ECETOC. 2003. (Q)SARs: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management

applications. Brussels, European Centre for Toxicology and Ecotoxicology of Chemicals, ECETOC Technical Report No. 89.

ECVAM. 2006. Workshop skin sensitisation and epidermal disposition. The relevance of epidermal bioavailability for sensitisation hazard identification/risk assessment, Jan. 2006. Ispra, Italy, European Centre for the Validation of Alternative Methods, Institute of Health and Consumer Protection, Joint Research Centre.

Eskes, C. and Zuang, V.,eds. 2005. Alternative (Non Animal) Methods for Cosmetic Testing: Current Status and future Prospects. ATLA 33, Supplement 1: 83-103.

Garcia, H., Salter-Cid, L., Stein-Streilein, J. 1992. Persistent interleukin-2 activity and molecular evidence for expression of lymphotoxin in the hapten-immune model for pulmonary interstitial fibrosis. Am. J. Respir. Cell. Mol. Biol. 6: 22-28.

Garssen, J., Nijkamp, F.P., Vliet van der, H., Loveren van, H. 1991. T-cell mediated induction of airway hyperreactivity in mice. Am. Rev. Resp. Dis. 144: 931-938.

Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. 2005. Compilation of historical local lymph node data for evaluation of skin sensitisation alternative methods. Dermatitis 16(4):157-202.

Graham, C., Gealy, R., Macina, O.T., Karol, M.H., Rosenkranz, H.S. 1996. QSAR for allergic contact dermatitis. Quant. Struct.-Act. Relat. 15: 224-229.

Graham, C., Rosenkranz, H.S., Karol, M.H. 1997. Structure-activity model of chemicals that cause human respiratory sensitisation. Regul. Toxicol. Pharmacol. 26:296–306.

Griem, P., Goebel, C., Scheffler, H. 2003. Proposal for a risk assessment methodology for skin sensitisation potency data. Regul. Toxicol. Pharmacol., 38: 269-290.

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Janusch Roi, A., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V. 2004. A modular approach to the ECVAM Principles on Test Validity ATLA 32: 467-472.

Howes, R.G., Hadgraft, J., Heylings, J., Hoeck, U., Kemper, F., Maiback, H., Marty, J.-P., Merk, H., Parra, J., Rekkas, D., Rondelli, I., Schaefer, H., Täuber U., Verbiese, N.1996. Methods for assessing percutaneous absorption. ATLA 24: 81-106.

Jowsey, I.R., Basketter, D.A., Westmoreland, C., Kimber, I. 2006. A future approach to measuring relative skin sensitising potency: a proposal. J. Appl. Toxicol. (in press).

Karol, M.H., Graham, C., Gealy, R., Macina, O.T., Sussman, N., Rosenkranz, H.S. 1996. Structure-activity relationships and computer-assisted analysis of respiratory sensitisation potential. Toxicol. Lett. 86:187-91

Kimber, I. 1995. Contact and respiratory sensitization by chemical allergens: uneasy relationhips. Am. J. Contact Derm. 6: 34-39.

Kimber, I. and Dearman, R.J. 2002. Chemical respiratory allergy: role of IgE antibody and relevance of route of exposure. Toxicology 181-182: 311-315.

Kimber, I. and Dearman, R.J. 2005. What makes a chemical a respiratory sensitiser. Current Opinion Allergy Clin. Immunol. 5: 119-124.

Kimber, I., Pichowski, J.S., Betts, C.J., Cumberbatch, M., Basketter, D., Dearman, R. 2001. Alternative Approaches to the Identification and Characterisaton of Chemical Allergens. Toxicol. In Vitro 15: 307-312.

Kimber, I., Dearman, R.J., Betts, C.J., Gerberick, G.F., Ryan, C.A., Kern, P.S., Patlewicz, G.Y., Basketter, D.A. 2006.The local lymph node assay and skin sensitisation: a cut-down screen to reduce animal requirements? Contact Dermatitis 54: 181-185.

Kligman, A.M. 1966. The identification of contact allergens by human assay. III. The maximization test: a procedure for screening and rating contact sensitisers. J. Invest. Dermatol. 47: 393-409.

Maestrelli, P., Occari, P., Turato, G., Papiris, S.A., Di Stefano, A., Mapp, C.E., Milani, G.F., Fabbri, L.M., Saetta, M. 1997. Expression of interleukin (IL)-4 and IL-5 proteins in asthma induced by toluene isocyanate. Clin. Exp. Allergy 27: 1292-1298.

Magnusson, B. and Kligman, A.M. 1969. The identification of contact allergens by animal assay. The guinea pig maximisation test. J. Invest. Dermatol. 52: 268-276.

OECD. 2004. The Report from the Expert Group on (Quantitative) Structure Activity Relationship ([Q]SARs) on the Principles for the Validation of (Q)SARs. Paris, France, Organization for Economic Co-operation and Development. ENV/JM/MONO(2004)24. OECD Series on Testing and Assessment No. 49. http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html. December 15, 2006.

Payne, M.P. and Walsh, P.T. 1995. Structure-activity relationships for respiratory sensitisation. Proceedings, British Toxicology Society Meeting York, Yorkshire, UK, March/April 1995.

Potts, R.O. and Guy, R.H. 1992. Predicting skin permeability. Pharmaceutical Res. 9(5): 663-669.

Roberts, D.W., Williams, D.L. 1982. The derivation of quantitative correlations between skin sensitisation and physico-chemical parameters for alkylating agents and their application to experimental data for sultones. J. Theor. Biol. 99: 807-825.

Roberts, D.W., Aptula, A.O., Patlewicz, G.Y. 2006.Mechanistic Applicability Domains for Non-Animal Based Toxicological Endpoints. QSAR Analysis of the Schiff Base Applicability Domain for Skin Sensitisation. Chem. Res. Toxicol. 19(9):1228-1233.

Roberts, D.W., Aptula, A.O., Cronin, M.T., Hulzebos, E., Patlewicz, G. 2007. Global (Q)SARs for Skin Sensitisation – Assessment Against OECD Principles. SAR QSAR Environ. Res. (in press)

Ryan, C.A., Gerberick, G.F., Gildea, L.A., Hulette, B.C., Betts, C.J., Cumberbatch, M., Dearman, R.J., Kimber, I. 2005. Interactions of contact allergens with dendritic cells: opportunities and challenges for the development of novel approaches to hazard assessment. Toxicol. Sci. 88: 4-11.

Sakaguchi, H., Ashikaga, T., Miyazawa, M., Yoshida, Y., Ito, Y., Yoneyama, K., Hirota, M., Itagaki., H., Toyoda, H., Suzuki., H. 2006. Development of an in vitro skin sensitisation test using human cell lines; human Cell Activation Test (h-CLAT), I and II. Toxicol. In Vitro 20: 767-773 and 774-784.

Satoh, T., Kramarik, J.A., Tollerud, D.J, Karol, M.H. 1995. A murine model for assessing the respiratory hypersensitivity potential of chemical allergens. Toxicol. Lett. 78: 57-66.

Schlede, E. and Eppler, R. 1995. Testing for skin sensitisation according to the notification procedure for new chemicals: the Magnusson and Kligman test. Contact Dermatitis 32: 1-4.

Schneider, K. and Akkan, Z. 2004. Quantitative relationship between the local lymph node assay and human skin sensitisation assays. Regul. Toxicol. Pharmacol. 39: 245-255.

Steiling, W., Basketter, D.A., Berthold, K., Butler, M., Garrigue, J.-L., Kimber, I., Lea, L., Newsome, C., Roggeband, R., Stropp, G., Waterman, S., Wieman, C. 2001. Skin sensitisation testing - new perspectives and recommendations. Food Chem. Toxicol. 39: 293-301.

Stotts, J. 1980. Planning, conduct and interpretation of human predictive sensitisation patch tests. In: Current concepts in cutaneous toxicity. Academic Press, p. 41-53.

Tuscl, H., Kovac, R., Weber, E. 2000. The Expression of Surface Markers on Dendritic Cells as Indicators for the Sensitising Potential of Chemicals. Toxicol. In Vitro 14: 541-549.

Van Och, F.M.M., van Loveren, H., de Jong, W.H., Vandebriel, R.J. 2002. Cytokine production induced by low-molecular-weight chemicals as a function of the stimulation index in a modified local lymph node assay: an approach to discriminate contact sensitizers from respiratory sensitizers. Toxicol. Appl.Pharmacol. 184: 46-56.

Van Och, F.M., Slob, W., de Jong, W.H., Vandebriel, R.J., van Loveren, H. 2000. A quantitative method for assessing the sensitizing potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of the uncertainty margins. Toxicology 20;146(1):49-59.

Wolfreys, A. and Basketter, D.A. 2004. Mutagens and Sensitisers – an unequal relationship? J. Toxicol. Cutaneous Ocular Toxicol. 23(3): 197-205.

Worth, A. and Balls, M., eds. 2002. Alternative (non-animal) Methods for Chemical Testing: Current Status and Future Prospects. ATLA Vol. 30 Supplement 1: 49-53.

Worth, A.P., Bassan, A., Gallegos, A., Netzeva, T.I., Patlewicz, G., Pavan, M., Tsakovska, I., Vracko, M. 2005. The characterisation of (Quantitative) Structure-Activity Relationships: Preliminary Guidance. ECB Report EUR 21866 EN Ispra, Italy: European Commission, Joint Research Centre.

Zwart, A., Arts J.H.E., Kuper, C.F. 1994. Wave propagation: a new parameter in the description of mechanical airway impedance. Eur.Respir. Rev. 4: 203-209.

Chapter 6

Chapin, R.E. and Sloane, R.A. 1997. Reproductive assessment by continuous breeding: evolving study design and summaries of ninety studies. Environ. Health Perspect. 105 (Suppl 1): 199-205.

Chapin, R.E., Sloane, R.A., Hasemann, J.K. 1988. Reproductive endpoints in general toxicity studies: are they predictive? Reprod. Toxicol. 12(4): 489-494.

Combes, R., Barratt, M., Balls, M. 2003. An overall strategy for the testing of chemicals for human hazard and risk assessment under the EU REACH system. Altern. Lab. Anim. 31(1): 7-19.

Cooper, R.L., Lamb, J.C., Barlow, S.M., Bentley, K., Brady, A.M., Doerrer, N.G., Eisenbrandt, D.L., Fenner-Crisp, P.A., Hines, R.N., Irvine, L.F., Kimmel, C.A., Koeter, H., Li, A.A., Makris, S.L., Sheets, L.P., Speijers, G., Whitby, K.E. 2006. A tiered approach to life stages testing for agricultural chemical safety assessment. Crit. Rev. Toxicol. 36(1):69-98.

EC. 2006. Proposal for a regulation of the European Parliament and of the Council concerning the Registration, Evaluation, Authorisation and restriction of Chemicals (REACH), establishing a European Agency and amending Directive 1999/45/EC and Regulation (EC) {on persitent organic pollutants}. European Commission, Brussels, Belgium, 2003/0256 (COD).

Frankos, V.H. 1985. FDA perspectives on the use of teratology data for human risk assessment. Fundam. Appl. Toxicol. 5(4): 615-25.

Gadagbui, B., Zhao, J., Maier, A., Dourson, M. 2005. The scientific rationale for deriving database and toxicodynamic uncertainty factors for reproductive or developmental toxicants. Cincinnati, USA, Toxicology Exellence for Risk Assessment.

Gray, L., Ostby, J., Ferrel, J., Sigmno, R., Cooper, R., Linder, R., Rehnberg, G., Goldman, J., Laskey, J. 1989. Correlation of sperm and endocrine measures with reproductive success in rodents. In: Burger, E.J. (Ed.), Progress in Clinical and Biological Reseach, vol. 302. Liss, New York, pp. 193-209.

Hurtt, M.E., Cappon, G.D., Browning, A. 2003. Proposal for a tiered approach to developmental toxicity testing for veterinary pharmaceutical products for food-producing animals. Food Chem. Toxicol. 41(5): 611-9.

Jelovsek, F.R., Mattison, D.R., Chen, J.J. 1989. Prediction of risk for human developmental toxicity: how important are animal studies for hazard identification? Obstet. Gynecol. 74(4): 624-36.

Mangelsdorf, I., Buschmann, J., Orthen, B. 2003. Some aspects relating to the evaluation of the effects of chemicals on male fertility. Regul. Toxicol. Pharmacol. 37(3): 356-69.

Sakai, T., Takahashi, M., Mitsumori, K., Yasuhara, K., Kawashima, K., Mayahara, H., Ohno, Y. 2000. Collaborative work to evaluate toxicity on male reproductive organs by repeated dose studies in rats--overview of the studies. J. Toxicol. Sci. 25: 1-21.

Schardein, J. 2000. Chemically Induced Birth Defects. 3rd Edition. New York, Marcel Dekker, Inc.

Shepard, T.H. 1998. Catalog of teratogenic agents. 9th Edition. Baltimore (MD), Johns Hopkins University Press.

Takayama, S., Akaike, M., Kawashima, K., Takahashi, M., Kurokawa, Y. 1995. Studies on the optimal treatment period and parameters for detection of male fertility disorder in rats--introductory summary. J. Toxicol. Sci. 20(3): 173-82.

Ulbrich, B. and Palmer, A.K. 1995. Detection of effects on male reproduction—a literature survey. J. Am. Coll. Toxicol. 14(4): 293-327.

Van der Jagt, K., Munn, S., Tørsløv, J., de Bruijn, J. 2004. Alternative approaches can reduce the use of test animals under REACH. European Commission Report EUR 21405 EN.

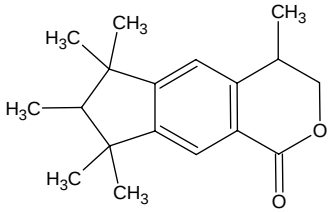
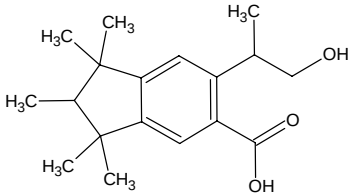
Appendices

Appendix I: CATABOL prediction of the BOD for 83 new chemicals notified in the Netherlands in comparison with the observed BOD.

Chem.	Reliability	BOD Observed	BOD predicted	General parametric requirements	Model structure domain	Unable to metabolized	Total Domain
1	0.1105	0.93	0.008	In domain	Out of domain	In domain	Out of Domain
2	0.5601	0.87	0.902	In domain	In domain	In domain	In domain
3	0.0928	0.34	0.208	In domain	Out of domain	In domain	Out of Domain
4	0.0597	0.017	0.167	In domain	In domain	In domain	In domain
5	0.2121	0.58	0.785	In domain	Out of domain	In domain	Out of Domain
6	0.1349	0.83	0.166	In domain	Out of domain	In domain	Out of Domain
7	0.0472	0.21	0.441	In domain	In domain	In domain	In domain
8	0.0471	0.18	0.3	In domain	Out of domain	In domain	Out of Domain
9	0.1034	0.106	0.12	In domain	Out of domain	In domain	Out of Domain
10	0.5681	0.82	0.903	In domain	In domain	In domain	In domain
11	0	0.13	0.023	In domain	Out of domain	In domain	Out of Domain
12	0.3053	0.15	0.53	In domain	Out of domain	In domain	Out of Domain
13	0.0375	0.02	0.122	In domain	Out of domain	In domain	Out of Domain
14	0	0.04	0	In domain	Out of domain	In domain	Out of Domain
15	0.1104	0.087	0.584	In domain	Out of domain	In domain	Out of Domain
16	0.5686	0.65	0.831	In domain	Out of domain	In domain	Out of Domain
17	0.0724	0.095	0.087	In domain	Out of domain	In domain	Out of Domain
18	0	0.067	0.149	In domain	Out of domain	In domain	Out of Domain
19	0.4666	0.05	0.493	In domain	Out of domain	In domain	Out of Domain
20	0.7766	0.72	0.952	In domain	In domain	In domain	In domain
21	0.4886	0.71	0.631	In domain	Out of domain	In domain	Out of Domain
22	0.0357	0.02	0.119	In domain	Out of domain	In domain	Out of Domain
23	0.0069	0.06	0.18	In domain	In domain	In domain	In domain
24	0.4432	0	0.131	In domain	Out of domain	In domain	Out of Domain
25	0.1579	0.7	0.152	In domain	Out of domain	In domain	Out of Domain
26	0.0574	0	0.265	In domain	In domain	In domain	In domain
27	0.4388	0.92	0.437	In domain	Out of domain	In domain	Out of Domain
28	0.0414	0	0.143	In domain	Out of domain	In domain	Out of Domain
29	0	0.2	0.044	In domain	Out of domain	In domain	Out of Domain
30	0.4611	0.1	0.615	In domain	Out of domain	In domain	Out of Domain
31	0.4813	0.28	0.816	In domain	Out of domain	In domain	Out of Domain
32	0.2746	0.01	0.318	In domain	In domain	In domain	In domain
33	0.1505	0.037	0.222	In domain	Out of domain	In domain	Out of Domain
34	0.2155	0.05	0.458	Out of domain	Out of domain	In domain	Out of Domain
35	0.0167	0	0.038	In domain	Out of domain	In domain	Out of Domain
36	0.1804	0.55	0.27	In domain	Out of domain	In domain	Out of Domain
37	0.3362	1	0.491	In domain	Out of domain	In domain	Out of Domain
38	0.1525	0.52	0.4	In domain	Out of domain	In domain	Out of Domain
39	0.1387	0.91	0.677	In domain	In domain	In domain	In domain
40	0.0565	0.78	0.179	In domain	Out of domain	In domain	Out of Domain
41	0.0998	0.52	0.305	In domain	Out of domain	In domain	Out of Domain
42	0	0	0.002	In domain	Out of domain	In domain	Out of Domain
43	0.1207	0.3	0.209	In domain	Out of domain	In domain	Out of Domain
44	0.056	0.09	0.083	In domain	Out of domain	In domain	Out of Domain
45	0.1843	0.39	0.602	In domain	Out of domain	In domain	Out of Domain

46	0.0467	0.01	0.16	In domain	Out of domain	In domain	Out of Domain
47	0	0.02	0.503	In domain	Out of domain	In domain	Out of Domain
48	0	0.02	0.154	In domain	Out of domain	In domain	Out of Domain
49	0.0824	0.08	0.016	In domain	Out of domain	In domain	Out of Domain
50	0.3521	0.64	0.91	In domain	In domain	In domain	In domain
51	0.0137	0.14	0.143	In domain	Out of domain	In domain	Out of Domain
52	0.092	0.05	0.248	In domain	Out of domain	In domain	Out of Domain
53	0.0127	0.05	0.064	Out of domain	Out of domain	In domain	Out of Domain
54	0.3772	0.17	0.609	In domain	Out of domain	In domain	Out of Domain
55	0.2234	0.45	0.182	In domain	Out of domain	In domain	Out of Domain
56	0	0.015	0.106	Out of domain	Out of domain	In domain	Out of Domain
57	0.5394	0.81	0.808	In domain	Out of domain	In domain	Out of Domain
58	0	0.11	0.028	In domain	Out of domain	In domain	Out of Domain
59	0	0.23	0.018	In domain	Out of domain	In domain	Out of Domain
60	0	0	0.078	In domain	Out of domain	In domain	Out of Domain
61	0.0461	0	0.188	In domain	Out of domain	In domain	Out of Domain
62	0.0357	0.05	0.119	In domain	Out of domain	In domain	Out of Domain
63	0.0727	0.12	0.095	Out of domain	Out of domain	Out of domain	Out of Domain
64	0.0491	0.05	0.158	In domain	Out of domain	In domain	Out of Domain
65	0.0188	0.09	0.06	In domain	Out of domain	In domain	Out of Domain
66	0.0206	0.06	0.154	In domain	Out of domain	In domain	Out of Domain
67	0.1496	0.08	0.276	In domain	Out of domain	In domain	Out of Domain
68	0.077	0.03	0.057	In domain	Out of domain	In domain	Out of Domain
69	0.275	0.22	0.25	Out of domain	Out of domain	In domain	Out of Domain
70	0.0932	0.09	0.196	In domain	Out of domain	In domain	Out of Domain
71	0.156	0.6	0.24	In domain	Out of domain	In domain	Out of Domain
72	0	0.11	0.037	In domain	Out of domain	In domain	Out of Domain
73	0	0.09	0.042	In domain	Out of domain	Out of domain	Out of Domain
74	0.0823	0.04	0.096	In domain	Out of domain	In domain	Out of Domain
75	0.092	0.12	0.01	In domain	Out of domain	In domain	Out of Domain
76	0.0734	0.05	0.259	In domain	Out of domain	In domain	Out of Domain
77	0.2813	0.08	0.277	In domain	Out of domain	In domain	Out of Domain
78	0.0808	0.03	0.63	In domain	Out of domain	In domain	Out of Domain
79	0.1538	0.02	0.038	In domain	Out of domain	In domain	Out of Domain
80	0.0243	0.67	0.06	In domain	Out of domain	In domain	Out of Domain
81	0.1773	0.025	0.288	In domain	Out of domain	In domain	Out of Domain
82	0.09	0.097	0.073	In domain	Out of domain	In domain	Out of Domain
83	0.0373	0.01	0.057	In domain	Out of domain	In domain	Out of Domain

Appendix II: Experimental data and CATABOL prediction on the biodegradability of 76 substances discussed by the working group on PBT assessment

No	Chem. Name	
1	1-(5,6,7,8-tetrahydro-3,5,5,6,8,8-hexamethyl-2-naphthyl)ethan-1-one (AHTN)	Data available EU RAR : An extensive study was performed into the biotransformation of ^{14}C-AHTN in activated sludge. After 3 days, a variety of more polar metabolites were detected. The half-life of the parent AHTN was 12-24 h. AHTN was largely biotransformed to polar metabolites within 20 days. Catabol prediction, out of structural domain (BOD: 0.013) First step is methyl group oxidation with low probability (pathway contains more methyl group oxidation steps. The metabolites are not formed in the CATABOL predictions.
2	1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethylindeno[5,6-c]pyran (HHCB)	Data available Experimental derived DT50 values for fresh water are 2 d for water and 79 d for sediment. DT50 values determined in sludge and soil are in the same range. The highest DT50 value is 105 d for sludge amended soil. The half-life time for the parent in activated sludge was 21 hours (so a first order rate constant 0.033 h^{-1}) and about 85% disappeared in 150 hours. Initially a first metabolite with TLC elution time similar to lactone appeared and accounted for about 40% of the original radioactivity between day 1 and 8. Gradually a second metabolite increased, to up to 45% of the radioactivity after 650 hours. This metabolite had a similar elution time as the hydroxycarboxylic acid. The structures are shown in below. A third, highly polar metabolite made up to 15% of the radioactivity after 150 hours.  *HHCB-lactone or Galaxolidone*  *Hydroxy acid* CATABOL prediction, out of structural domain (BOD: 0.014)

		First step is methyl group oxidation with low probability (pathway contains more methyl group oxidation steps). The metabolites are not formed in the CATABOL predictions.
3	1,2,3-trichlorobenzene in training set	Data available No info on biodegradable available CATABOL prediction, in of structural domain (BOD: 0.025) Parent compound has a log Kow of 3.9? Parent compound is persistent because the first steps aromatic ring oxidations has low probability (P=0.3), followed by aromatic ring cleavage (P=0)
4	1,2,4-trichlorobenzene in training set	Data available EU RAR: The ready biodegradability was studied with a method corresponding to OECD TG 301C, Modified MITI (I) test. In the aerobic study, the degradation measured as BOD was 0% after 14 days (MITI 1992). However, the high concentration of 1,2,4-TCB employed in the test may have resulted in toxicity to the micro-organisms. CATABOL prediction, in of structural domain (BOD: 0.025) See above
5	2,2',6,6'-tetra-tert-butyl-4,4'-methylenediphenol	Data available Not readily biodegradable; 0% degradation at day 28 (OECD 301) CATABOL prediction, out of structural domain (BOD: 0.009) The substance is persistent, first step is methyl group oxidation with low probability, followed by several comparable degradation steps
6	2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2,4-dimethylphenyl)-3-oxobutyramide]	Data available No exp data CATABOL prediction, out of structural domain (BOD: 0.128) After two rapid diketone and unsaturated ketone oxidation and nitrile and amide hydrolysis two compounds are formed: c1(N)c(C)cc(C)cc1 and c1(-c2cc(Cl)c(N=NCC(=O)O)cc2)cc(Cl)c(N=NCC(=O)O)cc1 both are persistent and the last one

		has a log Kow of 4.8!!!!!
7	2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2-methylphenyl)-3-oxobutyramide]	Data available See above CATABOL prediction, out of structural domain (BOD: 0.388) See above
8	2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(4-chloro-2,5-dimethoxyphenyl)-3-oxobutyramide]	Data available See above CATABOL prediction, out of structural domain (BOD: 0.244) See above
9	4,4'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[2,4-dihydro-5-methyl-2-phenyl-3H-pyrazol-3-one]	Data available See above CATABOL prediction, out of structural domain (BOD: 0.046) The substance will degrade via geminal derivatives decomposition. The formed metabolite c1(-c2cc(Cl)c(N=NC3C(C)(O)NN(c4ccccc4)C3=O)cc2)cc(Cl)c(N=NC2C(C)(O)NN(c3ccccc3)C2=O)cc1 might have different properties???
10	2,4,6-trichlorophenol (in training set)	Data available 1) Readily biodegradable in S Trenton Delaware River Water with suspended sediment; avg. half-life 6,3 days; range from 2,75 to 70 days (10g to 1g suspended sediment). (Data also included in IUCLID) HSDB Database. 2) Water from 11 sites on the Delaware River ; 100% biodeg. in filtered water, in 10-14 days (8days lag time) and 100% degradation in water with sediment in 8-10 days (2 days lag time). HSDB Database. 3) Aerobic aquatic half-life: 7 - 70 days (Howard et al. 1991). CATABOL prediction, in of structural domain (BOD: 0.033) The substance degrades as follows: dehalogenation (P=0.3), aromatic ring oxidation (P=0.3) and aromatic ring cleavage

		(P=0).
11	2,4-dinitrotoluene (in training set)	Data available P-screening criterion is fulfilled, because the Zahn-Wellens test with adapted inocula does not fulfil the specific criteria. CATABOL prediction, in of structural domain (BOD: 0.000) First steps are nitro groups reduction with low probability (P=0) under aerobic conditions.
12	2,4-di-tert-butylphenol in training set !!!!	Data available Test with aerobic, activated sludge shows only 2% biodegradation after 28 days The result is consistent with the data for 2,6-di-tert-butylphenol. CATABOL prediction, in of structural domain (BOD: 0.016) Substance is persistent because the first step methyl group oxidation has a low probability (P=0.2), followed by several methyl group oxidations.
13	2,6-di-tert-butylphenol	Data available BHT is unstable in water and forms several products e.g. BHT-COOH and BHT-OH. It does however not seem to be ready biodegradable. IUCLID lists results from 2 ready biodegradation tests, one giving <10% degradation in 20 days (OECD 301 D) and the other giving 4.5% degradation in 28 days (OECD 301C). CATABOL prediction, in of structural domain (BOD: 0.016) Substance is likely to be persistent because the first step methyl group oxidation has a low probability (P=0.2). The formation of BTH-OH and BHT-COOH is also predicted by CATABOL. Although the former precursor is rapidly transformed in an aldehyde which is rapidly transformed in the carboxyl group. This substance is relatively persistent (next step is methyl group oxidation).
14	4,6-di-tert-butyl-m-cresol	Data available Not biodegradable, 0 % after 28 d Directive 84/449/EEC. CATABOL prediction, in of structural domain (BOD: 0.016) Substance is persistent because the first step methyl group oxidation has a low probability (P=0.2), followed by several

		methyl group oxidations.
15	2-ethylhexyl_10-ethyl-4,4-dioctyl-7-oxo-8-oxa-3,5-dithia-4-stannatetradecanoate	Data available Significant levels of degradation (up to 40%) were observed in a standard degradation test. A hydrolysis test indicates that hydrolysis does occur, and that dioctyl tin oxide/hydroxide can be produced. CATABOL prediction, out of structural domain (BOD: 0.606) According to CATABOL the octyl group are oxidized at first, after which the other chains are degraded.
16	Dichlorodioctylstannane	Data available The substance is not readily degradable in standard Ready Tests, although some degradation is achieved. Rapid hydrolysis is anticipated by analogy with dibutyl tin compounds and this is being investigated. Data provided appears to indicate that dichlorodioctylstannane react with water to produce the relevant oxide/hydroxide. The chemistry is complicated by the relative insolubility of each of the dioctyl species and by the potential for the oxide to polymerise and precipitate. Indications of the behavioir in water appear to contradict earlier data that showed rapid reaction in water to produce an insoluble oxide, although losses of approximtely 50% were observed during the 24 hours renewal periods. Measured concentrations of dioctyltin dichloride ranged from 0.27-0.42 mg/L. Data from these ecotoxicity tests give indications of hydrolysis in water during the test but does not support more recent observations made by industry of rapid precipitation with only very low dissolved tin levels.. CATABOL prediction, out of structural domain (BOD: 0.776) The different alkyl chains are degraded via several oxidation steps (e.g. beta oxidation), the remaining persistent metabolite is C[Sn](C)(Cl)Cl.
17	3-methyl-1-(2,6,6-trimethylcyclohex-1-en-1-yl)penta-1,4-dien-3-ol	Data available Not readily biodegradable, 52 % after 28 d (OECD 301 C). A ready biodegradability test according to OECD guideline 301 B has been performed by BASF 2003 in compliance with GLP. The degree of degradation is 63 & 64 % (two replicates) after 28 days of incubation. CATABOL prediction, out of structural domain (BOD: 0.053) According to CATABOL the biodegradability is slow because the first steps are epoxidations with a relative low probability (P=0.38).

18	4,4'-methylenedicyclohexyl_diisocyanate	Data available The substance is not readily biodegradable based on experimental results (0%). There is also a MITI test for the reaction product which shows it is ready biodegradable. The foreseen hydrolysis product (4,4'-diaminodicyclohexylmethane) is not B (log Kow (OECD 107) = 2.03 and log Kow (QSAR) = 2.55). Considering that the hydrolysed product (4,4'-diaminodicyclohexylmethane) is not a PBT CATABOL prediction, out of structural domain (BOD: 0.372) According to CATABOL the assumed hydrolysis product is indeed formed (23%). Other major metabolites are C(=O)(O)C(CC(=O)O)CC1CCC(N)CC1 (16%) and C(=O)(O)C(CC(=O)O)CC(C(=O)O)CC(=O)O (41%)
19	4-chloro-1-(2,4-dichlorophenoxy)-2-nitrobenzene	Data available Based on a (C4-C EEC) test substance should be considered as not readily biodegradable (<1% in 28 days). CATABOL prediction, in of structural domain (BOD: 0.000) Parent compound is persistent, the first step is a nitro group reduction.
20	5-nonylsalicylaldehyde_oxime	Data available The available exp. data indicate that no biodegradadtion occur. CATABOL prediction, out of structural domain (BOD: 0.479) According to CATABOL the alkyl will rapidly degrade, forming a metabolite c1(O)c(C=NO)cc(CCC(=O)O)cc1 in 85% with a log Kow of 1.7.
21	alpha,alpha,alpha,4-tetrachlorotoluene	Data available At specific pH (demineralized water), the chemical is hydrolized. CATABOL prediction, out of structural domain (BOD: 0.020) First step is dehalogenation with a low probability (P=0.06). Hydrolysis is not predicted.
22	anthracene,_pure	Data available

	in training set !!!!	Anthracene is not readily biodegradable according to the MITI I test. CATABOL prediction, in of structural domain (BOD: 0.011) The substance is persistent because the first step is an aromatic ring oxidation with a low probability (P=0.0.03). After that another aromatic ring oxidation have to take place.
23	barium_bis[2-[(2-hydroxynaphthyl)azo]naphthalenesulphonate]	Data available No experimental data. CATABOL prediction, out of structural domain (BOD: 0.019) After hydrolysis (splitting off Barium) the substance is persistent: first step is an azo compound reduction with low probability (P=0). The metabolites formed are also persistent.
24	bis(2,4-dichloro-5-nitrophenyl)_carbonate	Data available No experimental data available. CATABOL prediction, out of structural domain (BOD: 0.057) According to Cat the first two steps are ester hydrolysis of the COC binding. It is this part of the molecule which is out of domain (relaibility?). The metabolites formed have a log Kow of 3.2 en 2.4. Based on that the substance will not be a PBT.
25	Clofenotane_(=_p,p-DDT) in training set	Data available Experimental data indicate 0% biodegradability. CATABOL prediction, in of structural domain (BOD: 0.001) First step is a dehalogenation with a low probability.
26	Cyclododeca-1,5,9-triene in training set !!!!	Data available Biodegradation studies (according to 301) indicate that the substance is not ready biodegradable, but that degradation does occur. CO2 production after 28 days: 8%, 63 days 32% and 77 days 68%. The parent compound decline to 83% after 28 days, 45% after 63 days and not detectable after 77 days. CATABOL prediction in of structural domain (BOD: 0.036)

		According to CATABOL the substance degrades via epoxidation, which after slow processes (P=o.38). Followed by subterminal oxidation and Baeyer Villiger oxidation with also a low probability (P=0.1). After the ring is open the degradation goes very rapidly.
27	Cyclododecane in training set !!!!	Data available Based on a MITI test the biodegradation is 0-12%.. CATABOL prediction in of structural domain (BOD: 0.008) Parent compound is persistent. The first steps baeyer villiger oxidation and subterminal oxidation has a low probability (P= 0.08 and 0.1)
28	Decanoic_acid,_ester_with_2-ethyl-2-(hydroxymethyl)-1,3-propanediol_octanoate	Data available Experimental data indicate that the substance is ready biodegradable. CATABOL prediction, out of structural domain (BOD: 0.691) This is in accordance with the prediction of CATABOL.
29	Di(tert-dodecyl)_pentasulphide	Data available New ready biodegradation study shows 0% degradation confirming that the screening criteria for P/vP are met. CATABOL prediction, out of structural domain (BOD: 0.009) The substance degrades via several methyl group oxidation with a low probability (P=0.2).
30	Dibenzyltoluene	CATABOL prediction in of structural domain (BOD: 0.008) CATABOL predicted that the the substance should be persistent. Parent chemical: Predicted BOD: 0.8%, non metabolised quantity: 78% (mol), probability to be metabolised: 22%. One stable metabolite: probability to be obtained: 22%, quantity: 22% (mol per mol parent), probability to be metabolised: 3%. Conclusion: CATABOL simulation confirmed that the parent substance is not readily biodegradable and provided structure of possible metabolites. QSAR on these metabolites indicated that they cannot be considerd as persistent. However, CATABOL is not in agreement with the experimental results obtained in the biodegradability test (disappearance of the aromatic cycles: 13% in 20 days, 58% in 62 days, 50% in 105 days and 67% in 149 days) as CATABOL estimated as very low the probability to obtain metabolites with less than 3 aromatic cycles (<1%).

31	Dicofol	Data available
	in training set !!!!	In a water/sediment study DCBA, DCBP and CBA are the most important metabolites. The diol form is however not observed, 3 en 4-OH DCBPA in stead. CATABOL prediction, in of structural domain (BOD: 0.004) According to CATABOL the first step in the biodegradation pathway is a dehalogenation (slow process), after which is substance DCBA is formed (P=0.7), together with DCBP (persistent; P=0.3) and diol DCBP (very persistent P=0). Subsequently further rapid degradation occurs via aromatic ring oxidation until the persistent metabolite CBA is formed (P= 0.3).
32	Diethyldimethylplumbane	Data available No standard studies are available in IUCLID or NSDB. Diethyldimethylplumbane is considered to degrade in the simular manner as tetraethyllead which appears to degrade to non PBT substances and ultimately to inorganic lead. The final degradation product inorganic lead is in this context considered not to be further discussed. CATABOL prediction, out of structural domain (BOD: 0.000) The substance is for all cases out of domain. No predictions are made.
33	Diisodecyl_phenyl_phosphite	Data available According to a Sturm study the biodegradation is 10%, which is in line with the CATABOL prediction. CATABOL prediction, out of structural domain (BOD: 0.051) The first step predicted by CATABOL is a methyl group oxidation with a low probability (P=0.22). The substance is, however, de-listed by the PBT WG because it was assumed that the biodegradation was hampered by the low water solubility. It was expected that the substance will hydrolyze rapidly as in aquatic tox studies a significant loss was observed.
34	Dioxobis(stearato)trilead	Data available A screening study gives a biodegradation of 87% in 35 days and 73% after 28 days, but does not meet the 10 day window criterion for readily biodegradable. CATABOL prediction, out of structural domain (BOD: 0.832)

		CATABOL predictions confirms that the substance should degrade rapidly.
35	bis(pentabromophenyl)_ether in training set !!!!	Data available EU RAR: Overall, decabromodiphenyl ether was found to be stable under the conditions used in the test, and so this type of process is not expected to lead to the formation of significant amounts of lower brominated congeners. CATABOL prediction, in of structural domain (BOD: 0.004) According to the predictions of CATABOL. The substance degrades slowly via dehalogenation fllowed by aromatic ring oxidation.
36	Diphenyl_ether,_octabromo_derivative	Data available The substance shows little indication of degradability. CATABOL prediction, in of structural domain (BOD: 0.016) According to CATABOL the substance will breakdown via aromatic ring oxidations and dehalogenation, which are slow processes (P=0.3), after which a very persistent metabolites is formed with the structure c1(Br)c(Br)c(Br)c(Br)c(Br)c1Oc1c(Br)c(O)c(Br)c(O).
37	Dodecylphenol	Data available Only 10% degradation was observed in the inherent test after 56 days (based on carbon dioxide evolution), and the results of compound-specific analysis indicated that no significant degradation of the test material had occurred. CATABOL prediction, in of structural domain (BOD: 0.692) CATABOL predicts readily biodegradability, fast biotransformation via omega oxidation, primary hydroxyl group oxidatio, aldehyde oxidation and beta-oxidation the substance will degrade to phenolacetic acid.
38	Endosulfan	Data available Endosulfan is volatile and long-range transportable via air (reported atmospheric half-lives vary from >2 days to 27 days). Hydrolysis of endosulfan is very pH dependent: measured DT50 at pH 5 is >200 d. Half-lives in soil for endosulfan (α + β) reported from field studies carried out in southern conditions vary between 16.5 and 167.1 d. Laboratory soil-studies, conducted at 21 to 28 ˚C, indicate a considerable difference in half-lives of α- and β-isomers. For β-isomer alone half-lives between 108 and 264 d have been reported, whereas for α -isomer alone half-lives

		between 12 and 39 d have been reported from these studies. The technical product contains ca. 30 % β-isomer and ca. 65 % α –isomer. The lower half-life of α –isomer has thus a significant influence on the overall-half-life -results of the field studies. CATABOL prediction, out of structural domain (BOD: 0.003) CATABOL predicts that the dehalogenation takes place very slowly, resulting in a very low readily biodegradability.
39	Hexabromocyclododecane in training set !!!!	Data available No degradation was observed in Ready Test. Reliable half-lives have been determined in valid simulation studies including sediment. In these studies, HBCD has been added at environmentally relevant concentrations. The half-life times for biotransformation for the sediment compartment range from 2 to 56 days. The metabolites were identified as tetrabromocyclododecene, dibromocyclododecadiene, and cyclododecatriene (CDT). These degradation products suggest that HBCD is sequentially debrominated via a series of dehaloelimination steps as a major pathway for the degradation of HBCD in the environment mediated by naturally occurring microorganisms in the wastewater sludge and aquatic sediments. CATABOL prediction, in of structural domain (BOD: 0.002) CATABOL only consideres an aerobic transformation where dehalogenation is slow via substitution of Br to a hydroxylgroup. These are very slow processes.
40	Hexachlorobenzene	Data available No data. Considered as POP. CATABOL prediction, in of structural domain (BOD: 0.009) According to CATABOL slow degradation because of the slow dehalogenation.
41	hexachlorobuta-1,3-diene	Data available Biodegradation in water: Tabak (1981) found complete disappearance (adsorption/volatilisation/degradation) within 7 days with adapted microorganisms under aerobic conditions (conc. 5-10mg/l, semistatic shake flask). 70 % adsorption and 10 % degradation in 8 days pilot low loaded STP test. Current data not evidence for ready biodegradability! CATABOL prediction, out of structural domain (BOD: 0.022) According to the CATABOL prediction the first step in the biotransformation process is an epoxidation, which has a low

probability (0.05). The metabolite formed has a log Kow of 3.4.

42 methyl_2-(4-(2,4-dichlorophenoxy)phenoxy)propionate

Data available

Substance is active ingredient in plant protection products, under re-registration under Directive 91/414/CE. DT50 in water = 363 d (pH 5); 31.7 d (pH 7); 0.52 d (pH 9) => vP in acidic conditions.

CATABOL prediction, out of structural domain (BOD: 0.098)

According to the CATABOL prediction the substance will degrade via ester hydrolysis, decarboxylation, aromatic ring oxidation to persistent metabolite starting with 5 to 4.1.

43 methyl_3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate

Data available

Data indicates that Metilox and its hydrolysis products are not degradable and are expected to end up in sediment.

CATABOL prediction, out of structural domain (BOD: 0.099)

According to the prediction of CATABOL, the ester hydrolysis has a high probability (P=0.9). The formed hydrolysis product is like to be persistent (P=0.15), with a calculated Log Kow of 4.8.

44 N,N-dicyclohexylbenzothiazole-2-sulphenamide

Data available

Two tests according to OECD 301F show that the mixture is not readily biodegradable (biodegradation 0 % and 0.64 %).

The results of the hydrolysis study indicate that the substance is hydrolytic degradable. The half lives determined at 20°C and pH 7 were between 9 and 16 days. However, the metabolites were not identified. In addition to that, the Rapporteur is of the opinion that especially for substances showing highly adsorptive behaviour, hydrolytic degradation is of minor relevance in the environment. It is expected that once released into aquatic systems, the substance might persist adsorbed to sediment and particles.

The study on field dissipation of the radio-labelled ditolyl-component indicates that degradation in soil is very slow. An initial degradation-phase of 32 days with a half-life of 11 days was followed by a second phase of slow dissipation of 88 days, The half-life calculated for the 362 days of the study was 67 days integrating binding to soil as non-extractable residues (42 % in 362 days).

CATABOL prediction, out of structural domain (BOD: 0.005)

According to the CATABOL prediction, the substance is very persistent. The hydrolysis is not predicted. The first step is an oxidative deamination and N-dealkylation (P= 0.01).

45	4-(2,4-dichlorophenoxy)aniline	Data available Aminofen is not ready biodegradable according to IUCLID where <20%ThOD in a OECD 301 D is reported. Thus, the screening criterion for P and vP is met. CATABOL prediction, in of structural domain (BOD: 0.040) First step predicted by CATABOL is an oxidative deamination followed by two ring oxidations. First step has a P of 0.85. Rong oxidation a P of 0.3. The formed metabolite is persistent (not B; log Kow 3.1).
46	Nitrofen	Data available Biodegradation studies reported in IUCLID did not fulfill the criteria for readily or inherently biodegradability. Rapid degradation was seen in water and soil in a test with a model ecosystem (Kale & Raghu, 1989). This study is not suitable to assess the biodegradability of nitrofen. CATABOL prediction, in of structural domain (BOD: 0.000) First step is nitrogroup reduction (P=0), the formed metabolite has a log Kow of 3.6, followed by a oxidative deamination (P=0.9), formed metabolite has a log Kow of 4.1. After an aromatic ring oxidationa persistent metabolite is formed.
47	Nonylphenol	Data available There are variable results from ready biodegradability tests, but two show significant degradation (53% and 62%) over 28 days although they did not meet the 10-day criterion for ready degradability. According to the PBT strategy such substances should be considered as non-persistent. (Inherent degradability was assumed in the risk assessment. P is not met. CATABOL prediction, in of structural domain (BOD: 0.901) According to CATABOL all degradation steps has a high probability.
48	Phenol,_4-nonyl-,_branched	Data available This substance is also considered ready biodegradable via read across with nonylphenol, CATABOL prediction, in of structural domain (BOD: 0.128) However, according to CATABOL calculation this substance is less degradable because the first step is a methyl oxidation which occur less rapidly than the fist degradation step of nonylphenol which is an omega oxidation.

49	N-tert-butylbenzothiazole-2-sulphenamide	Data available The chemical is not ionised at environmental pHs (pKa(1) 1.75, pKa(2) –3.43). The chemical is not readily biodegradable (0%) but it does hydrolyse in less than 1 day at pH 9 or less (METI 1996, OECD 111, t1/2 at pH 7 = 1.8 h, 25 °C). The hydrolysis products are mercaptobenzothiazole (identified), di(benzothiazoyl-2)disulfide (identified), t-butylamine (identified), 2-sulfo(sulfino)benzothiazole (predicted) and benzothiazole (identified). These hydrolysis products excluding 2-sulfo(sulfino)benzothiazole have been tested and shown to have low potential for bioaccumulation. Based on these findings, the chemical has a low potential for bioaccumulation also. CATABOL prediction, out of structural domain (BOD: 0.011) First steps are methyl groups oxidations which have a low probability (P=0.2), followed by aromatic ring oxidation which has an even lower probability (P=0.05)
50	Octabenzone in training set !!!!	Data available The rapporteur indicated that the substance was not readily biodegradable (6% in 28 days). CATABOL prediction, in of structural domain (BOD: 0.390) The alkyl group will be oxidized after which a persistent metabolite is formed with the structure c1(C(=O)c2c(O)cc(O)cc2)ccccc1.
51	octadecyl_3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate	Data available A modified MITI test (OECD 301C) showed 21-39% ultimate biodegradation (BOD) over 28 days. Primary degradation to metilox acid was observed (by HPLC). The substance was converted to metilox acid for 61-93%. 1-Octadecanol was not observed, indicating that this compound is ultimately biodegraded. In an old modified Zahn Wellens test reported in the EPA HPV Challenge Program the degradation was studied for 35 days. In this test Tween 80 was used as an emulsifier. 47 and 21% degradation (CO2 evolution) of the substance was observed at 13.3 and 25.9 mg/l. CATABOL prediction, out of structural domain (BOD: 0.563) Via several oxidation steps the alkylgroup will be oxidized. Finally a persistent metabolite will be formed (C(C)(C)(C)c1c(O)c(C(C)(C)C)cc(CCC(=O)O)c1) which is metilox acid which is also observed in the biodegradation study.
52	Octamethylcyclotetrasiloxane	Data available The hydrolysis product is reported to be the opened ring structure, a diol octamethyltetrasiloxanediol. The diol is then further hydrolysed/catalysed to shorter chain linear siloxanes, finally to dimethylsilanediol (DMSD) monomer. Depending

		on circumstances the ring can be reformed (dry conditions), but normally this process is of minor importance compared to further hydrolysis/catalysis towards shorter chain siloxanes. Nordic study results support the conclusion that relative rapid abiotic degradation of D4 in the environment takes place. The hydrolysis half-life observed in tests for D4 at 25 ºC is close to the half-life of a substance passing a standard OECD 301 'Ready biodegradability' test (t½ several days). CATABOL prediction, out of structural domain (BOD: 0.029) Degradation via methyl group oxidation. the ring cleavage is not predicted.
53	Pentachlorobenzenethiol	Data available Not biodegradable, 0% after 28d OECD Guide-line 301 D, Reference: Bayer AG data, cited in IUCLID CATABOL prediction, out of structural domain (BOD: 0.009) Degradation via dehalogenation which has a low probability (P=0.3).
54	pentaerythritol_tetrakis(3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate)	Data available The substance might be hydrolysed. Suggested degradation products are: Pentaerythritol and Metilox CATABOL prediction, out of structural domain (BOD: 0.039)
55	perylene-3,4:9,10-tetracarboxylic_dianhydride	Data available No experimental data information. CATABOL prediction, out of structural domain (BOD: 0.238) The substance seems to degrade easily upto C(=O)(O)c1c(O)c(CC(=O)O)c2c(c1CC(=O)O)cc(C(=O)O)cc2 which is persistent with log Kow of 0.8.
56	Phenol,_styrenated	Data available 7% Degradation after 28 days, OECD 301. CATABOL prediction, out of structural domain (BOD: 0.855)
57	Terphenyl	Data available 7-10% Degradation after 50 days, CO2 evolution, Acclimated inoculum, 50 % loss in 16-28 day, River die away test,

comparative study, In this test with a mixture of terphenyls, 80% degradation was observed for the o- and m-terphenyl within 45 days, with a T1/2 of 16-28 days.

CATABOL prediction, in of structural domain (BOD: 0.091)

Parent compound has low probability to degrade (P=0.2).

58 Tetrabutyltin

in training set !!!!

Data available

PBT group: In essence the molecule does not biodegrade on a ready biodegradability test.

CATABOL prediction, in of structural domain (BOD: 0.000)

Parent compound is persistent. Degradation via formation of lower alkylated tins be formed, such as tri- , di- or monobutyltins through organotin compound oxidation.

59 Tetraoctyltin

Data available

No experimental data available.

CATABOL prediction, in of structural domain (BOD: 0.761)

According to Catabol tetrabutyltin is degraded via organotin compound oxidation and tetraoctyltin not. In the present prediction the alkylchains are degraded via several oxidation steps (beta, aldehyde)

60 Tetrachlorophthalic_anhydride

Data available

A fast hydrolysis of the substance is observed. Data indicate that the substance was decarbonated in the environment and would not accumulate.

CATABOL prediction, out of structural domain (BOD: 0.065)

Confirmation that the substance will undergo a fast alhydride hydrolysis into C(=O)(O)c1c(C(=O)O)c(Cl)c(Cl)c(Cl)c1Cl (log Kow of 3.6).

61 Tetraethyllead

Data available

Number of studies confirm hydrolysis to triethyl lead which is not a PBT (low log Kow). Industry have submitted a further review of data available on wet soils showing rapid degradation of TEL to triethyl leads and ultimately inorganic lead.

CATABOL prediction, out of structural domain (BOD: 0.000)

No prediction

62 Tetramethyllead

Data available

Standard degradation tests are not available in the IUCLID, but investigation indicates that tetramethyllead in short time degrades both by abiotic and biotic pathways to inorganic lead. A 100 % decomposition in rainwater within 2 days decompose in water to form Pb2+.

CATABOL prediction, out of structural domain (BOD: 0.000)

No prediction

63 Triphenylphosphine

Data available

In a stability test in aqueous layer, the measured concentration decreased from ca. 24 µg/l at the start of the experiment to a level not detected in samples analysed at 3 h and 5 h. The reaction product was confirmed by HPLC to be triphenylphosphine oxide. Because the oxidation occurred in the test within few hours in the dark, in octanol saturated water and under argon (test temperature was 20 ºC), it is expected that triphenylphosphine oxidizes fast also under environmental conditions.

CATABOL prediction, out of structural domain (BOD: 0.007)

In accordance to the experimental data the first step in the degradation of the substance is the oxidation to triphenylphosphine oxide, which is also predicted to be a persistent metabolite (P= 0.05). The probability of this oxidation step is 0.3, which contradict the fast oxidation observed in the test described above.

64 Bis(tributyltin)oxide_(TBTO)

in training set !!!!

Data available

The C-Sn bond of TBTO is stable against hydrolysis under environmental conditions. In marine water TBTO forms mainly TBTCl, TBTOH, an aqueous complex (TBTOH2+) and calcareous compounds ((TBT)2 CO3). TBTO has been assessed slight to moderat persistent in water. Biodegradation half-lives in the literature under aerobic conditions in aquatic systems are found to 4 - 225 days (mainly primary degradation and not complete mineralisation). TBTO is more persistent in sediment and studies have demonstrated half-lifes from 1-15 years (BUA Report 238, 2003).

CATABOL prediction, in of structural domain (BOD: 0.000)

TBTO degrades via organotin compound oxidation where the different butyl chain are splitted of. The probability of occurrence is expected to be low (P=0)

65 Dioctadecyl_3,3'-thiodipropionate

Data available

	in training set !!!!	The degradability of DSTDP was studied in a slightly modified closed bottle test. Test substance dissolved in dichloromethane was added to 250-300 ml BOD bottles. Thereafter, the dichloromethane was evaporated during 24 hours on a roller bank before adding mineral medium and inoculum. The biodegradation of DSTDP was 71% in 28 days based on the BOD/ThOD-ratio. CATABOL prediction, in of structural domain (BOD: 0.836) In accordance with the experimental observation the substance will degrade rapidly via several oxidation steps. Finally a persistent metabolite is formed with the structure C(C)SCC.
66	N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine_(6PPD)	Data available In a hydrolysis study the substance showed to be unstable, abiotic degradation of 60% within 24 h. While the substance does degrade hydrolytically, there is little information on the fate and effects of the hydrolysis products. CATABOL prediction, out of structural domain (BOD: 0.035) First step is a methyl group oxidation, which probability of occurrence is expected to be low (P=0.2). Hydrolysis is not predicted.
67	tert.dodecanethiol	Data available A study of the abiotic degradation of TDM in aerated solution has been conducted. The study showed slow degradation at 20 degrees C, with a half life of approximately 150 days in algal medium and pH 7 buffer. However, slow degradation was also evident in the nitrogen-purged solutions. Low concentrations of di-t-dodecyl disulphide were found in the nitrogen-purged samples but were below the detection limit in the oxygenated samples. CATABOL prediction, out of structural domain (BOD: 0.021) First step is a methyl group oxidation, which probability of occurence is expected to be low (P=0.2), follwed by primary hydroxyl group oxidation (P=1), aldehyde oxidation (P=1) and decarboxylation (P=0.68). The next step will again be a methyl group oxidation with low P (0.2).
68	2-Ethylhexyl_10-ethyl-4-[[2-[(2-ethylhexyl)oxy]-2-oxoethyl]-thio]-4-octyl-7-oxo-8-oxa-3,5-dithia-4-stannatetradecanoate	Data available A related substance is monooctyltin trichloride, and this will have similar degradation products so should be considered with this. 2-Ethylhexyl 10-ethyl-4,4-dibutyl-7-oxo-8-oxa-3,5-dithia-4-stannatetradecanoate (CAS no. 10584-98-2) is another related substance that is also not listed. It degraded significantly (by 35%) in a ready biodegradation test and on chemical

		structural grounds it seems reasonable to assume that it would biodegrade slightly faster than the other two stannate compounds since it has shorter alkyl chains (the main hydrolysis product will be dibutyltin hydroxide). CATABOL prediction, out of structural domain (BOD: 0.524) The substance will degrade first via several oxidation steps into C(=O)(O)CC(=O)C[Sn](SCC(=O)OCC(CCCC)CC)(SCC(=O)OCC(CCCC)CC)SCC(=O)OCC(CCCC)CC, which will then further degrade via ester hydrolysis into C(=O)(O)CS[Sn](CC(=O)O)(SCC(=O)O)SCC(=O)O, which has a low bioaccumulation potential.
69	Tris(2,4-di-tert-butylphenyl)phosphate	Data available No ready biodegradability test available. It could reasonably be expected that hydolysis of the di-(t-butyl) phenyl derivative would also occur, although the rate would be uncertain. CATABOL prediction, out of structural domain (BOD: 0.038) First two steps is are alkylphosphinite hydrolysis, resulting on the formation of C(C)(C)(C)c1c(O)ccc(C(C)(C)C)c1 (log Kow of 5.3). The degradation of this metabolite via methyl group oxidation is not fast (P=0.22).
70	Ethylene-bistetrabromophthalimide	Data available Little experimental data are available; not readily biodegradable (301 C, CITI, 1981) It is speculated that under aerobic conditions, a probable degradation reaction is cleavage of the ethylene bridge With formation of either carboxylic acids or an imide. Neither the carboxylic acid nor the imide are expected to be readily biodegradable. Both are expected to have a short atmospheric half-life, to be non-bioconcentrating, and to lack aquatic toxicity. CATABOL prediction, out of structural domain (BOD: 0.040) The above statement is probably based on cat prediction. The following steps are predicted two nitrile and amide hydrolysis steps, resulting into C(=O)(O)c1c(C(=O)O)c(Br)c(Br)c(Br)c1Br and C1(=O)c2c(C(=O)N1CCN)c(Br)c(Br)c(Br)c2Br. The former is relative persistent (P=0.3) and has a log Kow of 4.6. The following steps are all dehalogenations.
71	1H-3a,7-Methanoazulene,_2,3,4,7,8,8a-hexahydro-3,6,8,8-tetramethyl-,__3R-(3.alpha.,3a.beta.,7.beta.,8a.alpha.)_-	Data available Degradation starts after a lag phase of 3 day and goes on steadily during the test period. After 28 days, mineralisation has reached 78%. The biodegradability curve does not meet the 10-day window criterion. The test is valid and therefore

		CEDRENE is considered readily biodegradable but failing the 10-day window. CATABOL prediction, out of structural domain (BOD: 0.296) Several oxidation, hydrolysis steps with high probability are followed by steps with low probability methyl group.
72	Lindane in training set !!!!	Data available Taking into account, the fact that this substance has been included in the UNECE POP protocol, the rapporteur proposes to introduce it in the list of potential PBTs. CATABOL prediction, in of structural domain (BOD: 0.000) Substance will be degraded via dehalogenation with a low probability.
73	Methylnaphthalene	Data available According to results following OECD 301F test guidelines 2-methylnaphthalene is expected to degrade at a moderate rate (50% in 28 days), and is considered inherently but not readily biodegradable. Although 2-methylenenaphtalene is according to several studies to some extent biodegradable, the available mesocosm and microcosm studies clearly indicate persistence in water and sediment. CATABOL prediction, in of structural domain (BOD: 0.179) The first step methyl group oxidation has a low probability (P=0.2), low by rapid degradation steps
74	Hexachlorocyclopentadiene	Data available EU-RAR: When activated sludge from a municipal sewage works (1 g dw/l) was exposed at 25°C to a concentration of 50 mg/l, more than 75% of HCCP applied was degraded after five days. About 49% of the HCCP applied was found in the activated sludge in the form of metabolites and 26.4% was converted to products that were soluble in water. The vast majority (89.1%) of the conversion products contained in the activated sludge was not extractable. The carbon dioxide formed was less than 0.1% of the amount of HCCP applied. CATABOL prediction, out of structural domain (BOD: 0.006) The degradation pathway contains three steps with low probability: (1) dehalogenation and (2,3) epoxidation.

75	1H-Indene-5-ethanol,_2,3-dihydro-beta.,1,1,2,3,3-hexamethyl-	Data available No experimental data available. CATABOL prediction, out of structural domain (BOD: 0.040) The degradation pathway contains a number of steps with low probability, the first ones are all methyl group oxidations and decarboxylations. The fist persistent metabolite is C1(C)(C)c2c(C(C)(C)C1C)cc(C(C)C(=O)O)cc2 and has a log Kow of 5.5.
76	2,6-di-tert-butyl-p-cresol_(BHT) in training set !!!!	Data available BHT is unstable in water and forms several products e.g. BHT-COOH and BHT-OH. It does however not seem to be ready biodegradable. IUCLID lists results from 2 ready biodegradation tests, one giving <10% degradation in 20 days (OECD 301 D) and the other giving 4.5% degradation in 28 days (OECD 301C). QSAR: BIOWIN 2 & 6 <0.5 (not ready) BIOWIN 3 = 2.3. (just above trigger of 2.2). CATABOL prediction, in of structural domain (BOD: 0.016) The degradation pathway contains several steps with low probability. The substance itself because first step methyl group oxidation has a low probability (P=02). The first steps are all methyl groups oxidations and decarboxylations with low probability (P=0.2 and 0, respectively).

Appendix III: Substances with indication of substantial primary degradation (predicted quantities <0.5).

Id#	CAS #	Chem. Name	Smiles	BOD	Quantity	Rel.
5	003173-72-6	1,5-naphthylene_diisocyanate	c1(N=C=O)c2c(c(N=C=O)ccc2)ccc1	0.077	0	0.2522
5.4			c1(N)c2c(c(N)ccc2)ccc1	0.003	0.967	1
7	025057-89-0	bentazone_(ISO)	C1(=O)c2c(cccc2)NS(=O)(=O)N1C(C)C	0.03	0.461	0.027
7.1			C(=O)(O)c1c(NS(=O)(=O)NC(C)C)cccc1	0.013	0.4186	1
8	000081-82-3	coumachlor_(ISO)	C1(O)c2c(cccc2)OC(=O)C=1C(c1ccc(Cl)cc1)CC(C)=O	0.386	0	0.0765
8.2			C1(=O)c2c(cccc2)OC(=O)C1C(c1ccc(Cl)cc1)COC(C)=O	0.372	0.1005	1
8.7			C(=O)(O)C(c1ccc(Cl)cc1)CC(=O)O	0	0.2316	0.0136
8.8			C(=O)(O)C(C)c1ccc(Cl)cc1	0	0.4868	0.0006
9	000117-52-2	coumafuryl_(ISO)	C1(O)c2c(cccc2)OC(=O)C=1C(C1=CC=CO1)CC(C)=O	0.453	0	0.0926
9.2			C1(=O)c2c(cccc2)OC(=O)C1C(C1=CC=CO1)COC(C)=O	0.44	0.1005	1
9.7			C(=O)(O)C(C1=CC=CO1)CC(=O)O	0.039	0.2316	0.0136
9.8			C(=O)(O)C(C)C1=CC=CO1	0.058	0.1569	0.0006
9.9			C1(CC)=CC=CO1	0.085	0.2961	0
10	005836-29-3	coumatetralyl	C1(O)c2c(cccc2)OC(=O)C=1C1c2c(cccc2)CCC1	0.276	0	0.0236
10.1			C1(=O)c2c(cccc2)OC(=O)C1C1c2c(cccc2)CCC1	0.276	0.1005	1
10.2			C(=O)(c1c(O)cccc1)C(C(=O)O)C1c2c(cccc2)CCC1	0.292	0.1008	0.383
10.3			c12c(C(CC(=O)O)CCC1)cccc2	0.007	0.6941	0.0356
12	070693-57-1	N-[3-[(2-acetyloxy)ethyl](phenyl-methyl)amino]-4-methoxyphenylacetamide	c1(OC)c(N(Cc2ccccc2)CCOC(C)=O)cc(NC(C)=O)cc1	0.347	0.1005	0.0333
12.1			c1(OC)c(N(Cc2ccccc2)CCO)cc(NC(C)=O)cc1	0.319	0.1489	0.383
12.2			c1(OC)c(N(Cc2ccccc2)CCO)cc(N)cc1	0.307	0.123	0
12.3			c1(O)c(N(Cc2ccccc2)CCO)cc(N)cc1	0.298	0.148	0
12.4			c1(O)c(NCCO)cc(N)cc1	0.003	0.474	0
12.3			C(=O)(O)C(C(=O)O)C(=O)O	0.678	0.1299	0
13	001663-39-4	tert-butyl_acrylate	C(=O)(C=C)OC(C)(C)C	0.328	0.375	0.6944
13.1			C(C)(C)(C)O	0.223	0.4854	1
16	000134-62-3	N,N-diethyl-m-toluamide	C(=O)(c1cc(C)ccc1)N(CC)CC	0.439	0.461	0.4944
16.12			C(C)NCC	0.674	0.1271	1
17	002275-18-5	prothoate_(ISO)	C(=O)(CSP(=S)(OCC)OCC)NC(C)C	0.05	0	0.0244
17.1			C(=O)(CSP(=O)(OCC)OCC)NC(C)C	0.012	0.7766	1
17.2			C(=O)(CSP(=O)(OCC)OCC)NC(C)CO	0.012	0.1735	0
24	000078-88-6	2,3-dichloropropene	C(=C)(Cl)CCl	0.178	0.0649	0.1421
24.1			C(=C)(Cl)CO	0.055	0.8776	0.75
25	000957-51-7	diphenamid_(ISO)	C(=O)(C(c1ccccc1)c1ccccc1)N(C)C	0.067	0.461	0.1075
25.1			C(=O)(O)C(c1ccccc1)c1ccccc1	0.002	0.5212	1
25.26			CNC	0.674	0.1271	1
26	000103-83-3	N,N-dimethylbenzylamine	C(=O)(c1ccccc1)N(C)C	0.409	0.461	0.2097
26.9			C(=O)(O)C(C(=O)O)C(=O)O	0.678	0.146	0.0054
26.14			CNC	0.674	0.1271	1
27	057837-19-1	metalaxyl_(ISO)	c1(C)c(N(C(=O)COC)C(C)C(=O)OC)c(C)ccc1	0.18	0.1005	0.1739
27.1			c1(C)c(N(C(=O)COC)C(C)C(=O)O)c(C)ccc1	0.105	0.2099	0.383
27.2			c1(C)c(N(C(=O)CO)C(C)C(=O)O)c(C)ccc1	0.054	0.3179	0.383
27.3			c1(C)c(NC(C)C(=O)O)c(C)ccc1	0.004	0.3673	0.383
28	000078-67-1	2,2'-dimethyl-2,2'-azodipropiononitrile	C(#N)C(C)(C)N=NC(C)(C)C#N	0.088	0.1656	0.0805
28.1			C(N)(=O)C(C)(C)N=NC(C)(C)C#N	0.066	0.1382	0.6667
28.2			C(N)(=O)C(C)(C)N=NC(C)(C)C(N)=O	0.038	0.321	0.4445

28.3			C(=O)(O)C(C)(C)N=NC(C)(C)C(N)=O	0.026	0.173	0.2166
Id#	CAS #	Chem. Name	Smiles	BOD	Quantity	Rel.
28.4			C(=O)(O)C(C)(C)N=NC(C)(C)C(=O)O	0	0.2023	0.1055
32	002451-62-9	1,3,5-tris(oxiranylmethyl)-1,3,5-triazine-2,4,6(1H,3H,5H)-trione	C1(=O)N(CC2CO2)C(=O)N(CC2CO2)C(=O)N1CC1CO1	0.339	0	0.1931
32.9			C(=O)(O)C(O)CN1C(=O)N(CC(O)C(=O)O)C(=O)N(CC(O)C(=O)O)C1=O	0.091	0.3224	1
32.11			C(=O)(O)C(O)CN1C(=O)N(CC(O)C(=O)O)C(=O)N(CC(=O)O)C1=O	0.077	0.2184	0.0244
32.13			C(=O)(O)C(O)CN1C(=O)N(CC(=O)O)C(=O)N(CC(=O)O)C1=O	0.05	0.148	0.0006
32.15			C(=O)(O)CN1C(=O)N(CC(=O)O)C(=O)N(CC(=O)O)C1=O	0.003	0.3075	0
33	010004-44-1	3-hydroxy-5-methylisoxazole	C1(O)C=C(C)ON=1	0.581	0	0.5777
33.2			C1(=O)C=C(C)ON1	0.539	0.461	1
34	016063-70-0	2,3,5-trichloropyridine	c1(Cl)c(Cl)cc(Cl)cn1	0.025	0.6956	0
34.1			c1(Cl)c(Cl)cc(Cl)c(O)n1	0	0.3041	0
35	000091-76-9	6-phenyl-1,3,5-triazine-2,4-diyldiamine	c1(-c2ccccc2)nc(N)nc(N)n1	0.181	0.15	0.1689
35.1			c1(-c2ccccc2)nc(O)nc(N)n1	0.177	0.1275	0.5
35.2			c1(-c2ccccc2)nc(O)nc(O)n1	0.171	0.5691	0.25
35.8			C1(=O)NC(=O)NC(=O)N1	0	0.1362	0
36	000078-57-9	menazon	c1(CSP(=S)(OC)OC)nc(N)nc(N)n1	0.138	0	0.0735
36.1			c1(CSP(=O)(OC)OC)nc(N)nc(N)n1	0.09	0.15	1
36.2			c1(CSP(=O)(OC)OC)nc(O)nc(N)n1	0.054	0.1275	0.5
36.3			c1(CSP(=O)(OC)OC)nc(O)nc(O)n1	0.005	0.6839	0.25
37	000140-31-8	2-piperazin-1-ylethylamine	C(N)CN1CCNCC1	0.104	0.2359	0.1043
37.2			C(=O)(O)CN1CCNCC1	0.004	0.7552	0.2692
41.1			c1(N)c(N(=O)=O)cc(O)cc1	0	0.8362	0.375
44	057646-30-7	methyl_N-(2,6-dimethylphenyl)-N-(2-furylcarbonyl)-DL-alaninate	C(=O)(C1=CC=CO1)N(c1c(C)cccc1C)C(C)C(=O)OC	0.168	0.1005	0.1192
44.1			C(=O)(C1=CC=CO1)N(c1c(C)cccc1C)C(C)C(=O)O	0.103	0.4147	0.383
44.2			c1(C)c(NC(C)C(=O)O)c(C)ccc1	0.004	0.4791	0.383
44.24			C(=O)(O)CCC(=O)O	0.678	0.1263	0.1467
45	024691-76-7	pyracarbolid	C(=O)(C1=C(C)OCCC1)Nc1ccccc1	0.241	0.461	0.0617
45.1			C(=O)(O)C1=C(C)OCCC1	0.068	0.1737	0.4872
45.2			C1(C)=CCCCO1	0.1	0.2275	0.0207
45.4			C1(C)(O)C(O)CCCO1	0.143	0.107	0
46	005259-88-1	oxycarboxin_(ISO)	C(=O)(C1=C(C)OCCS1(=O)=O)Nc1ccccc1	0.263	0.461	0.0668
46.1			C(=O)(O)C1=C(C)OCCS1(=O)=O	0.071	0.1737	0.4872
46.2			C1(C)=CS(=O)(=O)CCO1	0.105	0.2275	0.0207
46.4			C1(C)(O)C(O)S(=O)(=O)CCO1	0.118	0.107	0
48	000101-90-6	resorcinol_diglycidyl_ether	c1(OCC2CO2)cc(OCC2CO2)ccc1	0.292	0	0.1381
48.6			C(=O)(O)C(O)COc1cc(OCC(O)C(=O)O)ccc1	0.174	0.3224	1
48.8			C(=O)(O)C(O)COc1cc(OCC(=O)O)ccc1	0.225	0.2184	0.0244
48.1			c1(OCC(=O)O)cc(OCC(=O)O)ccc1	0.301	0.148	0.0006
48.11			c1(OC)cc(OCC(=O)O)ccc1	0.444	0.1003	0
49	000122-60-1	phenyl_glycidyl_ether	c1(OCC2CO2)ccccc1	0.409	0	0.0906
49.3			C(=O)(O)C(O)COc1ccccc1	0.353	0.3224	1
49.5			c1(OCC(=O)O)ccccc1	0.497	0.2184	0.0244
54	001918-00-9	dicamba_(ISO)	C(=O)(O)c1c(OC)c(Cl)ccc1Cl	0.203	0.1638	0.0918
54.1			C(=O)(O)c1c(O)c(Cl)ccc1Cl	0.041	0.5817	0.375
54.2			c1(Cl)c(O)c(O)c(O)cc1	0	0.2545	0.075
56	000094-75-7	2,4-D_(ISO)	c1(OCC(=O)O)c(Cl)cc(Cl)cc1	0.122	0.3224	0.0068
56.1			c1(OC)c(Cl)cc(Cl)cc1	0.18	0.111	0.0424
56.2			c1(O)c(Cl)cc(Cl)cc1	0.019	0.3942	0.0159
56.3			c1(Cl)c(O)c(O)cc(Cl)c1	0	0.1725	0.0083
57	000122-20-3	1,1',1''-nitrilotripropan-2-ol	C(C)(O)CN(CC(C)O)CC(C)O	0.079	0.2333	0

Id#	CAS #	Chem. Name	Smiles	BOD	Quantity	Rel.
57.1			C(C)(=O)CN(CC(C)O)CC(C)O	0.064	0.1789	0
57.2			C(C)(=O)CN(CC(C)=O)CC(C)O	0.042	0.1372	0
57.3			C(C)(=O)CN(CC(C)=O)CC(C)=O	0.009	0.4454	0
59	000096-13-9	2,3-dibromopropan-1-ol	C(Br)(CO)CBr	0.468	0.3034	0.4853
59.2			C(O)(CO)CBr	0.607	0.2113	1
59.8			C(=O)(O)C(O)C(=O)O	0.678	0.1564	0.6667
67	003813-05-6	3(2H)-BENZOTHIAZOLEACETIC_ACID,_4-CHLORO-2-OXO-	c1(Cl)c2c(ccc1)SC(=O)N2CC(=O)O	0.066	0	0.153
67.1			c1(Cl)c(N(C(=O)O)CC(=O)O)c(S)ccc1	0.017	0.461	1
67.2			c1(Cl)c(NCC(=O)O)c(S)ccc1	0.032	0.3749	0.7
67.3			c1(S)c(NCC(=O)O)c(Cl)c(O)c(O)c1	0	0.164	0.4667
68	007027-11-4	Cyclohexanecarbonitrile,_1,3,3-trimethyl-5-oxo-	C(#N)C1(C)CC(=O)CC(C)(C)C1	0.163	0	0.0542
68.1			C(#N)C1(C)CC(C)(C)COC(=O)C1	0.126	0.1005	1
68.4			C(=O)(O)C(C)(C)CC(C)(C#N)CC(=O)O	0.058	0.1489	0.383
68.5			C(=O)(O)C(C)(C)CC(C)(C(N)=O)CC(=O)O	0.024	0.242	0.2553
68.6			C(=O)(O)C(C)(C)CC(C)(C)C(N)=O	0.035	0.2345	0.0108
68.7			C(=O)(O)C(C)(C)CC(C)(C)C(=O)O	0.015	0.2129	0.0053
69	101657-77-6	Cyanic_acid,_methylenebis(2,6-dimethyl-4,1-phenylene)_ester	c1(C)c(N=C=O)c(C)cc(Cc2cc(C)c(N=C=O)c(C)c2)c1	0.043	0	0.1113
69.4			c1(C)c(N)c(C)cc(Cc2cc(C)c(N)c(C)c2)c1	0	1	1
70	123312-89-0	Pymetrozine	c1(C=NN2C(=O)NN=C(C)C2)cccnc1	0.462	0	0.1762
70.3			c1(C(O)NN2C(=O)N=NC(C)(O)C2)cccnc1	0.398	0.2359	1
70.4			C1(C)(O)CN(N)C(=O)N=N1	0.269	0.3523	0.65
70.7			C(C)(=O)CN(N)C(=O)O	0.539	0.1456	0
70.21			C(=O)(O)CN	0.764	0.1292	0
72	001823-59-2	Bis-(3-phthalyl_anhydride)_ether	C1(=O)c2c(C(=O)O1)c(Oc1c3C(=O)OC(=O)c3ccc1)ccc2	0.172	0	0.2944
72.2			C(=O)(O)c1c(C(=O)O)c(Oc2c(C(=O)O)c(C(=O)O)ccc2)ccc1	0.117	0.1118	1
72.4			C(=O)(O)c1c(Oc2c(C(=O)O)cc(O)c(O)c2)cc(O)c(O)c1	0.007	0.7834	0.25
74	001862-07-3	1-HEXANOL,_6-(DIMETHYLAMINO)-	C(O)CCCCCN(C)C	0.566	0	0.3594
74.6			C(=O)(O)CCCN(C)C	0.364	0.3224	1
74.7			C(CC)N(C)C	0.537	0.1598	0.0424
74.8			C(CC)NC	0.629	0.1221	0.0424
75	019060-15-2	1-Butanamine,_4,4-dimethoxy-	C(C)(CCN)(OC)OC	0.135	0.2359	0.0423
75.2			C(=O)(O)CC(C)(OC)OC	0.059	0.2463	0.2692
75.3			C(C)(C)(OC)OC	0.087	0.4021	0.0114
76	019247-05-3	Acetic_acid,_2,2'-hydrazonobis-	C(=O)(O)CNNCC(=O)O	0.143	0	0.3974
76.1			C(=O)(O)CN=NCC(=O)O	0	1	1
77	002122-19-2	PROPYLENE_THIOUREA	C1(=S)NC(C)CN1	0.026	0.7766	0
77.1			C1(=S)NC(CO)CN1	0.007	0.2198	0
78	002855-13-2	Cyclohexanemethanamine,_5-amino-1,3,3-trimethyl-	C1(C)(C)CC(C)(CN)CC(N)C1	0.141	0.2359	0.0343
78.5			C(=O)(O)C(C)(CC(C)(C)CC(=O)O)CN	0.056	0.2216	0.1473
78.6			C(=O)(O)C(C)(CC(C)(C)C)CN	0.083	0.1099	0.0062
78.8			C(=O)(O)C(C)(C(=O)O)CC(C)(C)C	0.024	0.2764	0.0017
79	035541-81-2	1,4-Cyclohexanedimethanol,_dibenzoate	C(=O)(c1ccccc1)OCC1CCC(COC(=O)c2ccccc2)CC1	0.469	0.1005	0.0357
79.6			C(=O)(O)C1CCC(C(=O)O)CC1	0.051	0.7031	0.1467
79.7			C(=O)(O)C1=CCC(C(=O)O)CC1	0.392	0.0119	0.1467
79.9			C(=O)(O)C1C(=O)CC(C(=O)O)CC1	0.409	0.0105	0.0314
79.22			C(=O)(O)c1ccccc1	0.775	0.2631	0.383
79.3			C(=O)(O)C(C(=O)O)C(=O)O	0.678	0.4629	0.0021
80	000357-57-3	Brucine	C123c4c(cc(OC)c(OC)c4)N4C(=O)CC5C(C6C(=	0.122	0.1638	0.0307

			CCO5)CN(C1C6)CC2)C34			
Id#	CAS #	Chem. Name	Smiles	BOD	Quantity	Rel.
80.1			C123c4c(cc(OC)c(O)c4)N4C(=O)CC5C(C6C(=CCO5)CN(C1C6)CC2)C34	0.079	0.137	0.375
80.2			C123c4c(cc(O)c(O)c4)N4C(=O)CC5C(C6C(=CCO5)CN(C1C6)CC2)C34	0.017	0.3223	0.1406
80.3			C123c4c(cc(O)c(O)c4)NC1C1C4C(=CCOC1CC(=O)O)CN(C2C4)CC3	0.011	0.1215	0.1406
80.4			C123c4c(cc(O)c(O)c4)NC1C1C4C(=CCOC1C)CN(C2C4)CC3	0.017	0.159	0.006
81	040649-36-3	4-Propylcyclohexanone	C1(=O)CCC(CCC)CC1	0.374	0	0.1902
81.1			C1(=O)CCC(CCC)CCO1	0.348	0.1005	1
81.8			C(=O)(O)C(CC(=O)O)CCC	0.08	0.7817	0.383
82	000463-56-9	Thiocyanic_acid	C(#N)S	0.424	0.1656	0.5334
82.2			C(=O)=S	0.016	0.8211	0.6667
83	057369-32-1	Pyroquilon	c12c3c(ccc1)CCC(=O)N3CC2	0.034	0.461	0.0256
83.1			c12c(c(CCC(=O)O)ccc1)NCC2	0.029	0.4684	1
83.12			C(=O)=O	0	0.1468	0
83.13			O	0	0.1448	0
84	000584-84-9	Benzene,_2,4-diisocyanato-1-methyl-	c1(C)c(N=C=O)cc(N=C=O)cc1	0.104	0	0.1818
84.4			c1(C)c(N)cc(N)cc1	0.004	0.967	1
87	006914-71-2	1,1-Cyclopropanedicarboxylic_acid_dimethyl_ester	C(=O)(C1(C(=O)OC)CC1)OC	0.43	0.1005	0.1109
87.2			C(=O)(O)C1(C(=O)O)CC1	0.066	0.2608	0.1467
87.3			C(=O)(O)C1CC1	0.098	0.4765	0.0062
88	069377-81-7	Fluroxypyr	c1(F)c(Cl)c(N)c(Cl)c(OCC(=O)O)n1	0.015	0.3224	0.0048
88.1			c1(F)c(Cl)c(N)c(Cl)c(OC)n1	0.022	0.6081	0.0424
89	081334-34-1	Imazapyr	C(=O)(O)c1c(C2=NC(C)(C(C)C)C(=O)N2)nccc1	0.374	0.1118	0.0119
89.1			C1(c2c(O)cccn2)=NC(C)(C(C)C)C(=O)N1	0.402	0.1332	0.25
89.4			C(=O)(O)C(C)(C(C)C)NC(O)c1c(O)cccn1	0.436	0.1781	0
89.5			C(=O)(O)C(C)(N)C(C)C	0.404	0.1361	0
89.6			C(=O)(O)C(=O)C(C)C	0.678	0.1421	0
89.2			C(=O)(O)CN	0.764	0.1086	0
90	082558-50-7	N-[3-(1-Ethyl-1-methylpropyl)-5-isoxazolyl]-2,6-dimethoxybenzamide	C(=O)(c1c(OC)cccc1OC)NC1=CC(C(C)(CC)CC)=NO1	0.266	0	0.1184
90.4			C(=O)(c1c(OC)cccc1OC)NC(=O)CC(=O)C(C)(CC)CC	0.25	0.1638	1
90.5			C(=O)(c1c(O)cccc1OC)NC(=O)CC(=O)C(C)(CC)CC	0.23	0.137	0.375
90.6			C(=O)(c1c(O)cccc1O)NC(=O)CC(=O)C(C)(CC)CC	0.197	0.3223	0.1406
90.8			C(=O)(O)C(C)(CC)CC	0.018	0.2927	0.0685
91	083016-70-0	Ethanol,_2-[[2-[2-(dimethylamino)ethoxy]ethyl]methylamino]-	C(O)CN(C)CCOCCN(C)C	0.393	0.2359	0.2317
91.1			C(O)CN(C)CCOCCNC	0.437	0.1802	1
91.2			C(O)CN(C)CCOCCN	0.492	0.1377	0.65
92	083164-33-4	3-Pyridinecarboxamide,_N-(2,4-difluorophenyl)-2-[3-(trifluoromethyl)phenoxy]-	C(=O)(c1c(Oc2cc(C(F)(F)F)ccc2)nccc1)Nc1c(F)cc(F)cc1	0.024	0.461	0.032
92.2			C(F)(F)(F)c1cc(Oc2c(O)cccn2)ccc1	0.003	0.4753	0.1218
92.33			c1(N)c(F)cc(F)cc1	0.004	0.5212	0.4872
93	000091-08-7	Benzene,_1,3-diisocyanato-2-methyl-	c1(N=C=O)c(C)c(N=C=O)ccc1	0.094	0	0.1857
93.4			c1(N)c(C)c(N)ccc1	0.004	0.967	0.5

Appendix IV Comparison between CATABOL predicted metabolites and metabolites found in soil for a group of carbamates and O-P and S-P esters (Est. = Predicted by CATABOL, exp. = observed metabolite in soil).

Carbamates	Transformation	Remarks
1-Naphthalenol, Methylcarbamate	1-Naphthalenol, methylcarbamate	In domain
Metabolite 1 - est.	1-Naphthalenol	This is no carbamate: cleavage of the carbamate functionality
Metabolite 1 - exp.	1-Naphthalenol	Prediction correct
2-(1-Methylethoxy) phenol, Methyl carbamate	Propoxur	In domain
Metabolite 1 - est.	Propoxur, hydroxylated at the 2- en 3-position next to the carbamate moiety	Prediction incorrect, this metabolite cannot be formed
Metabolite 1 - exp.	2-isopropoxyphenol - formed by cleavage of the carbamate moiety	Out of the CATABOL-predicted metabolites
2-Methyl-2-(methylthio)propionaldehyde, O-(Methyl-carbamoyl)oxime	Aldicarb	Out of structural domain
Metabolite 1 - est.	Hydrolysis of the C=N bond: C=N becomes OH-C-N-H	Prediction incorrect.
Metabolite 2 - est.	Cleavage of the carbamate functionality of metabolite 1	Prediction incorrect.
Metabolite 1 - exp.	aldicarb sulfoxide	This is the metabolite in which the S-atom is oxidized once: (S=O) bond
Metabolite 2 - exp.	aldoxycarb	This is the metabolite in which the S-atom is oxidized twice: (O=S=O) bond
1,2-Ethanediylbis-carbamothioic acid, disodium salt	Nabam	Out of structural domain
Metabolite 1 - est.	H-S-C=S replaced by H-S-C=O	
Metabolite 2 - est.	H-S-C=O replaced by H-O-C=O	
Metabolite 3 - est.	Cleavage of the carboxylic group	
Metabolite 1 - exp.	No information available	
2,3-Dihydro-2,2-dimethyl-7-benzofuranol, Methylcarbamate	Carbofuran	Out of structural domain
Metabolite 1 - est.	Cleavage of the carbamate functionality of carbofuran; yields the alcohollevert alcohol	Prediction incorrect.
Metabolite 1 - exp.	3-hydroxycarbofuran	
Metabolite 2 - exp.	3-ketocarbofuran	
Carbamates	**Transformation**	**Remarks**

3,5-Dimethyl-4-(methylthio)phenol, Methylcarbamate Metabolite 1 - est. Metabolite 1 - exp. Metabolite 2 - exp.	Methiocarb Cleavage of the carbamate functionality of methiocarb; yields the alcohol 3,5-dimethyl-4-(methylsulfinyl)phenol - cleavage carbamate functionality and oxidation C-S bond p-(methylsulfonyl)phenol - oxidation C-S bond and cleavage of 2 CH3-groups	Out of structural domain Prediction incorrect.
Methylcarbamate 2-(1-methylethyl) phenol Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Phenol,2-(1-methylethyl)-,methylcarbamate Hydroxylation of the aromatic ring: yields the diol Further hydroxylation of the aromatic ring: yields the triol No info in database	In domain
Trimethacarb Metabolite 1 - est. Metabolite 1 - exp.	N-Me-3,4,5-triMePhenyl carbamate Cleavage of the carbamate functionality Cleavage of the carbamate functionality	Out of structural domain Prediction correct.
2-(1-Methylpropyl) phenol, Methyl-carbamate Metabolite 1 - est. Metabolite 1 - exp.	Phenol, 2-(1-methylpropyl)-, methylcarbamate Cleavage of the carbamate functionality No info in database	In domain
N-[[(Methylamino) carbonyl]oxy]ethanimidothioic acid methyl ester Metabolite 1 - est. Metabolite 1 - exp.	Methomyl Cleavage of the carbamate functionality No info in database	Out of structural domain
2,2-Dimethyl-1,3-benzodioxal-4-ol methylcarbamate Metabolite 1 - est. Metabolite 1 - exp.	1,3-Benzodioxol-4-ol,2,2-dimethyl-,methylcarbamate (bendiocarb of isoprocarb) Hydroxylation of the aromatic ring: yields the diol 2,2-dimethyl-1,3-benodioxol-4-ol: hydrolysis carbamate moiety	Out of structural domain Prediction incorrect.
Dimethylcarbamic acid, 2-(Dimethylamino)-5,6-dimethyl-4-pyrimidinyl ester Metabolite 1 - est. Metabolite 1 - exp.	Pirimicarb Hydrolysis of the carbamate moiety Hydrolysis of the carbamate moiety as primary transformation	Out of structural domain Prediction correct.
2-(Dimethylamino)-N-[[methylamino-carboxy]oxy]-2-oxo, Methyl ester ethan-imidothioic acid Metabolite 1 - est.	Oxamyl Hydrolysis of the carbamate moiety	Out of structural domain
Carbamates	**Transformation**	**Remarks**

Metabolite 1 - exp.	No info in database	
Butoxycarboxim Metabolite 1 - est. Metabolite 1 - exp.	Butoxycarboxim Hydrolysis of the C=N bond: C=N becomes OH-C-N-H No info in database	Out of structural domain
N,N'-[Thiobis [(methylimino)carbonyloxy]]bisethanimido thioic acid, Dipentyl ester Metabolite 1 - est. Metabolite 1 - exp.	Thiodicarb - symmetrical carbamate Hydrolysis of one of the carbamate groups - cleavage of O-N-binding Methomyl - formed by cleavage of N-S-bond in middle of the molecule	Out of structural domain Prediction incorrect.
2-Methyl-4-(1-methylethyl)-7-oxo-8-oxa-3-thia-2,4-diazadecanoic acid, 2,3-Dihydro-2,2-dimethyl-7-benzofuranyl ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Benfuracarb Hydrolysis of the ester functionality Additional hydrolysis of the carbonyl group thus formed: cleavage acetic acid Carbofuran	Out of structural domain Prediction incorrect – site of carbamate hydrolysis wrongly predicted

O-P and S-P esters	Transformation	Remarks
(2,2,2-Trichloro-1-hydroxyethyl)phosphonic acid, Dimethyl ester	Dipterex	Out of structural domain
Metabolite 1 - est.	Hydroxylation of the CCl3 group: CCl3 transformed into the carboxylic acid	
Metabolite 2 - est.	Hydrolysis of the carboxylic acid, followed by hydrolysis of the P-O-CH3 moiety	
Metabolite 1 - exp.	Dichloroethanol and acetic acid - dichloroethanol is not predicted.	Prediction incorrect.
O,O-Dimethyl O-[3-Methyl-4-(methyl thio)phenyl]ester phosphorothioic acid	Fenthion	Out of structural domain
Metabolite 1 - est.	Transformation of SP-ester into the OP-ester	
Metabolite 2 - est.	Hydroxylering van de aromatische CH3-groep!	
Metabolite 1 - exp.	Fenthion sulfoxide en fenthion sulfone	Prediction incorrect.
Phosphorothioic acid, O,O-Diethyl O-(4-nitrophenyl) ester	Parathion	Out of structural domain
Metabolite 1 - est.	Transformation of SP-ester into the OP-ester	
Metabolite 2 - est.	4-nitrofenol en dimethylfosfaat-ester: splitsen van de groep aan de fosfaat-functionaliteit	
Metabolite 1 - exp.	No info in database	
Phosphorodithioic acid, O,O-Dimethyl S-[2-methylamino)-2-oxoethyl]ester	Dimethoate	In domain
Metabolite 1 - est.	Transformation of SP-ester into the OP-ester	
Metabolite 2 - est.	Afsplitsen van de groep aan de fosfaat-functionaliteit	
Metabolite 1 - exp.	No info in database	
Phosphoric acid, 2,2-Dichloroethenyl dimethyl ester (Dichlorvos)	Phosphoric acid, 2,2-dichloroethenyl, dimethyl ester	Out of structural domain
Metabolite 1 - est.	Demethyldichlorvos (een van de O-CH3 methyl-groepen is hierbij afgesplitst)	
Metabolite 2 - est.	Ook de 2e methylgroep wordt afgesplitst	
Metabolite 1 - exp.	Demethyldichlorvos (een van de O-CH3 methyl-groepen is hierbij afgesplitst)	Prediction correct.
O,O,-Dimethyl S-[(4-oxo-1,2,3-benzotriazin-3(4H)-yl)methyl] ester, Phosphorodithioic acid	Methylazinphos	Out of structural domain
Metabolite 1 - est.	Transformation of SP-ester into the OP-ester	
Metabolite 2 - est.	Ringopening van N-C=O-binding: substituent aan het fosforatoom	
Metabolite 1 - exp.	No info in database	

O-P and S-P esters	Transformation	Remarks
((Dimethoxyphosphinothioyl)thio)butanedioic acid, Diethyl ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Malathion Transformation of SP-ester into the OP-ester Cleavage substituent Malathion monocarboxylic acid and malathion dicarboxylic acid.	Out of structural domain Prediction incorrect.
O,O-Dimethyl O-(3-methyl-4-nitrophenyl ester phosphorothioic acid Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Fenitrothion Transformation of SP-ester into the OP-ester Cleavage substituent 3-methyl-4-nitrophenol, cleavage substituent	In domain Prediction correct.
Phosphorothioic acid, O,O-Dimethyl-O-(p-nitrophenyl)ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Parathion, methyl Transformation of SP-ester into the OP-ester Cleavage substituent No info in database	In domain
Phosphorothioic acid, O,O-Diethyl O-(2-isopropyl-6-methyl-4-pyrimidinyl) ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Diazinon Transformation of SP-ester into OP-ester Hydrolysis CH3-functionality of the aromatic ring of the substituent to yield the carboxylic acid 2-isopropyl-4-methyl-6-hydroxypyrimidine (cleavage substituent)	In domain Prediction incorrect.
Phosphoric acid, 2-Chloro-1-(2,4-dichlorophenyl)ethenyl diethyl ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Chlorfenvinphos Cleavage 1 Me group of P: P-O-C becomes P-O-H Cleavage 2nd Me group of P: P-O-C becomes P-O-H No info in database	Out of structural domain
S-[(1,3-Dihydro-1,3-dioxo-2H-isoindol-2-yl)methyl]O,O-dimethyl ester, phosphorodithioic acid Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Phosmet Transformation of SP-ester into the OP-ester Ring opening of the substituent No info in database	Out of structural domain
Ethylphosphonodithioic acid, O-Ethyl S-phenyl ester Metabolite 1 - est. Metabolite 2 - est.	Fonophos Transformation of SP-ester into the OP-ester Cleavage substituent	Out of structural domain

O-P and S-P esters	Transformation	Remarks
Metabolite 1 - exp.	No info in database	
Phosphorothioic acid, O,O-Diethyl O-(2-(ethylthio) ethyl)ester, Mixt. with O,O-diethyl S-(2-(ethylthio) ethyl)ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Demeton Transformation of SP-ester into the OP-ester Oxidation S atom of the substituent. This yields the sulfoxide anfd the sulfon No info in database	Out of structural domain
Carbamates	**Transformation**	**Remarks**
Phosphoramidothioic acid, O,S-Dimethyl ester Metabolite 1 - est. Metabolite 1 - exp.	Methamidphos Hydrolysis of the S-Me group: O-S-Me becomes OH No info in database	Out of structural domain
Phosphoric acid, 2-Chloro-3-(diethyl amino)-1-methyl-3-oxo-1-propenyl dimethyl ester Metabolite 1 - est. Metabolite 1 - exp.	Phosphamidon Cleavage of a part of the substituent, yielding the carboxylic acid No info in database	Out of structural domain
Phosphorodithioic acid, O-Ethyl S,S-dipropyl ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Ethoprophos Cleavage of one of the S-CH2-CH2-CH3 groups Cleavage of the 2nd S-CH2-CH2-CH3 group No info in database	Out of structural domain
Phosphorothioic acid, O-(4-Bromo-2-chlorophenyl)-O-ethyl-S-propyl ester Metabolite 1 - est. Metabolite 2 - est. Metabolite 1 - exp.	Profenofos Cleavage of the S-CH2-CH2-CH3-group Cleavage of the aromatic ring No info in database	Out of structural domain