CADASTER

CAse studies on the Development and Application of in-Silico Techniques for Environmental hazard and Risk assessment

Grant agreement no.: 212668

Collaborative Project

Sub-Priority ENV2007 3.3.1.1: In-silico techniques for hazard-, safety-, and environmental risk-assessment

CADASTER
D2.3 Overview of non-testing approaches available for implementation in REACH

Start date of project: 1 January 2009 Duration: 4 years

Lead Contractor: National Institute of Public Health and the Environment (RIVM), Laboratory for Ecological Risk Assessment

Corresponding authors of document: Emiel Rorije[1], Willie Peijnenburg[2]
1. RIVM, Expertise Centre for Substances, P.O. Box 1, 3720 BA Bilthoven, The Netherlands (emiel.rorije@rivm.nl)
2. RIVM, Laboratory for Ecological Risk Assessment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands (willie.peijnenburg@rivm.nl)

Deliverable no: 2.3 (Report: Overview of non-testing approaches available for implementation in REACH). Objective of this deliverable: To provide a brief overview of alternative testing that are potentially available for use within REACH in order to substitute for (experimental) generation of test data and to warrant optimal use of existing information on related chemicals.

Nature: Report publicly available

Project co-funded by the EU Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

WP 2: Database on experimental parameters and (Q)SARs for chemical and biological endpoints

Work Package Leader: Mojca Durjava (PHI, partner 2)

Task 2.3 - Overview of non-testing approaches available for implementation in REACH (Deliverable 2.3)

Summary

In this CADASTER report an overview is given of the non-testing options given under REACH to either replace experimental testing, or to strengthen confidence in experimental results. The latter is needed as the (in general scarcely available) experimental data for specific (SIDS) endpoints and for specific chemicals, might on their own not be sufficiently convincing as a proper reflection of the actual value of specific endpoints. The non-testing options available under REACH are: Quantitative Structure Activity Relationships (QSARs), read-across, category approaches, and exposure based waiving.

In other CADASTER reports attention has already been given in detail on the use of QSAR techniques to generate data for chemical risk assessment (Deliverable 2.2: Overview of (Q)SAR models and their specific features for assessing fate and effects – December 2009). Therefore, in this report the focus is on the possibilities to apply read-across and category approaches to the CADASTER selection of substances, and an overview of tools as well as guidance for the application of read across and category approaches is given.

The focus of CADASTER is on exemplifying the integration of information, models and strategies for carrying out safety-, hazard- and risk assessments for large numbers of substances. The integration will be performed on the basis of standard emission scenarios for specific compound classes. This implies that detailed quantitative information on the (multiple) uses of substances and the exposure scenario's that is required to conclude that exposure based waiving is possible, will in general be lacking for most (if not all) substances of interest. Discussion of possibilities of exposure based waiving within CADASTER are therefore restricted to considerations on the basis of substance properties, most notably limited to lack of bioavailability of the chemicals in specific environmental compartments.

Contents

1 - Introduction

CADASTER aims at providing the practical guidance to integrated risk assessment by carrying out a full hazard and risk assessment for chemicals belonging to four compound classes. QSAR models will be developed and validated, also externally, according to the OECD principles for the validation of QSAR. Prediction of data for chemicals of four selected classes will be performed for hazard and risk assessment, when experimental data are lacking. The main goal is to exemplify the integration of information, models and strategies for carrying out safety-, hazard- and risk assessments for large numbers of substances. An increase of the use of non-testing information for regulatory decisions, whilst meeting the main challenge of quantifying and reducing uncertainty, is stimulated by showing how this non-testing information is integrated for application within risk assessment.

In this report an overview is given of all options for using non-testing information under REACH. Especially the possibilities to apply read-across and category approaches for filling data gaps present in the selected substance classes in CADASTER are discussed in some more detail. Read-across and category approaches are mentioned in the REACH text as separate options from QSAR, and in the guidance the criteria for a “good” read-across or category approach are not as well defined as for QSARs. However, both read-across and category approaches can be thought of as simplified or limited (in numbers of substances) versions of a QSAR. Therefore, evaluation and validation of results generated by these methodologies can often be achieved using the same criteria and approaches as developed for QSARs.

1.1 Regulatory background - REACH

Around 100,000 different substances are registered as existing chemicals in the EU, of which an estimated 30,000 are manufactured or imported in quantities above 1 tonne. The previous regulatory system in 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 up to 2008. Although a programme of work has been drawn up, this EU legislation on chemicals had several drawbacks. Firstly, a substantial number of existing chemicals which are marketed in the EU 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 had 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 process of risk assessment and chemical management in general was 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 allocation of responsibilities is not appropriate: public authorities were 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 2006a, b] which covers both new and existing chemicals, and replaces approximately forty existing Community Directives and Regulations by one single regulation. This new legislation (REACH) came into force June 2008. 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 substances = substances classified as carcinogenic, mutagenic, reprotoxic: category 1 and 2 according to Directive 67/548/EEC; PBT substances = substances which are persistent, bioaccumulative, and toxic according to REACH criteria; or vPvB substances = very persistent, very bioaccumulative substances according to REACH criteria), 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 alternatives for testing under REACH

One of the consequences of REACH is that in a relative short time period the risks of large groups of chemicals need to be assessed. This 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 the OECD as well, there is an understanding that for reasons of animal welfare, costs and logistics, it is important to limit as much as possible the number of tests to be conducted. Annex XI of the REACH text states a number of options to replace or adapt the required testing that is set out in Annexes VII to X for the different tonnage levels. The generation of a comprehensive test dataset for every chemical will not be needed if these test data can be replaced by any of the following methods (for definitions of the non-testing methods: see sections 2.1 – 2.3):

- Non-testing methods:
  - The application of grouping (categories) and read-across
  - Computational methods (SARs, QSARs and biokinetic models)
  - Exposure based waiving
  - The use of existing experimental and historical data (including human data)
- Testing methods:
  - *In vitro* tests
  - Optimised *in vivo* tests
- Weight-of-evidence (WoE) using several independent sources of information, possibly combining results from both testing and non-testing methods.

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 alternative methods have up till now been used only to a limited 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 conclude on a classification or the need for more information in relation to hazard, risk and PBT assessment.

2 – Non-testing methods (QSARs, read-across, categories, Exposure Based Waiving) under REACH

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/or 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 (i.e. approaches in which chemicals are assigned to specific chemical classes on the basis of the presence of specific chemical moieties). A separate (non-testing) way to avoid testing of substances is exposure based waiving, i.e. if it can be argued that no risk will possible based on exposure scenario(s) developed in the Chemical Safety Report, no additional (toxicity) testing is required.

2.1 (Q)SARs

Within CADASTER much attention is given to application of existing QSARs and development of new QSAR models. We refer to CADASTER Deliverable 2.2 "Overview of (Q)SAR models and their specific features for assessing fate and effects" for a detailed description of the principles of QSAR models and the possibilities for application under REACH.

Validation of (Q)SAR models is essential for their regulatory use. The OECD Principles for validation of QSAR models (OECD, 2004) 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, 2004)

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

Although these principles were developed in order to guide the evaluation/validation of (Q)SAR models, they can be applied equally to read-across and chemical category approaches. Specifically for read-across, and to a lesser extent for chemical category approaches, it is not feasible or meaningful, to address principle 4 in the same way as normally done for QSAR models as the statistical measures used in QSAR are not meaningful for e.g. a read across case. However, the principle states that *appropriate* measures of model performance should be given, and this is doable for read across and category approaches as well. The same can be argued for the criteria developed for reporting QSAR results under REACH. In order to facilitate the reporting and the assessment of QSAR generated predictions, a QSAR Model Reporting Format (QMRF) and a QSAR Prediction Reporting Format (QPRF) have been developed [Rorije 2007, ECB 2007]. These reporting formats serve as check lists of information which is considered minimally necessary to be able to assess the reliability of a (QSAR) prediction. The Model Reporting Format should document the model characteristics, performance and mechanistical interpretation. This is very much in line with the OECD principles for the validation of QSARs. The Prediction Reporting Format should address the substance specific issues which might make an actual prediction more or less reliable, e.g. one of the questions to be answered is whether the substance for which a property is to predicted is part of the domain of applicability of the model.

2.2 Read-across and chemical category

Grouping approaches are strongly linked to SAR concepts. Annex XI of the REACH regulation [EC 2006a,b] 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 applies, substances shall be classified and labelled 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. In a recent RIVM project [Rila et al., 2006], guidance documents on read across and category approaches 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 provides only qualitative instead of quantitative guidance for deciding whether a category is robust. This means that the decision on categorisation remains to a large extent based on expert judgment, whereas it should be noted as well that quantitative data are usually needed for risk assessment purposes. In general the following issues need attention in categorisation:

(1) the group/category to which the chemicals are assigned should be indicated (please note that although this issue reads like a triviality, it is not always reported in practise),
(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 is to be noted that validation of grouping approaches is not explicitly mentioned as a requirement in REACH. As demonstrated above these approaches heavily rely on expert judgement. This implies that the process of expert judgement and the intermediate steps, 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 2004], part 3.2: Guidance on the Grouping of Chemicals.

2.3 Exposure based waiving (including Threshold of Toxicological Concern)

The basic principle behind any potential exposure based waiving is the recognition that there are situations in which human or environmental exposures are so low that acquisition of additional effects information for these exposure situations does not necessarily lead to an improvement in the ability to manage possible risks. In the Annexes VI-X of the REACH guidance, specific rules (column 2 of the tables) are presented regarding conditions in which information that is basic for risk assessment, may be omitted, triggered, replaced or adapted. Annexes IX and X include examples of waiving of certain tests based on exposure criteria. In addition, Annex XI presents the possibility of waiving of certain effects information in Annex VIII, IX and X based on exposure considerations. The approach is promising, especially when combined with (Q)SAR or read-across approaches, but it requires further investment in the development of exposure models and it also requires accurate information on the use pattern of the chemicals (e.g. downstream use information). The latter is usually one of the current bottlenecks regarding application of exposure based waiving.

An example in which information on effects, based on exposure considerations has been incorporated in the legislation, includes a Community procedure for flavouring substances used or intended for use in or on foodstuffs. In this case, the European Food Safety Authority (EFSA) has implemented the concepts 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 of the existence of 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 TTC for a specific compound 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 is close to the TTC, this could trigger the need to obtain further information on the toxicity of the chemical (including application of (Q)SARs). The TTC concept can however only be used to limit testing for those compounds for which adequate and detailed information is available on their use and subsequent exposure of man and the environment.

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

- The location where a substance is used; e.g. the case of restricted use within a well-characterised site with limited or no subsequent environmental exposure.
- 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.
- 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, or a solid produced/used in solution or dispersion only or a solid produced as non-abrasive large granules or flakes (e.g. marbles) that will reduce or even fully limit actual exposure of man and biota.

With respect to the environment, tests can be waived when information is available that one or more environmental compartments, or one or more specific groups 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 of soil or sediment 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].

If exposure-based arguments are used as a basis for a decision on the reduction of the required 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.

One of the main objectives of CADASTER is to exemplify the integration of information, models and strategies for carrying out safety-, hazard- and risk assessments for large numbers of substances. The integration will be performed on the basis of standard emission scenarios for specific compound classes. Because of the detailed information on use and emission scenarios that is required for exposure based waiving, waiving on the basis of detailed emission scenario's does not seem to be a general non-testing option which can play a role in the CADASTER project. This is especially the case for the group of substituted musks/fragrances, as musks and other fragrances by nature have diverse applications (consumer products, detergents, aerosols, perfumes), are marketed with the explicit intention of exposure (after all: a musk/fragrance is designed to be smelled, i.e. exposure is necessary), whereas the physico-chemical properties of musks/fragrances in general are such that they have the potential of exposure in all environmental compartments.

Apart from general exposure considerations, exposure based waiving for the four classes of "CADASTER chemicals" might be based upon consideration of physico-chemical properties only. Given the range of physico-chemical properties affecting the distribution of polyfluorinated compounds, substituted musks/fragrances, and (benzo)triazoles, it is to be concluded that possibilities for exposure based waiving for these compound classes for either the aquatic or the terrestrial compartment are limited on forehand. Polybrominated diphenylethers (PBDEs) on the other hand are characterised by their hydrophobicity and very low water solubility. Extreme low water solubility implies lack of exposure/bioavailability of PBDEs for pelagic organisms, and on forehand it is likely that aquatic testing of PBDEs is to be restricted to the sediment compartment. Thereupon, in common practise it is experienced that it is extremely difficult to maintain effective exposure concentrations of virtually insoluble chemicals at pre-set levels, this is very difficult to obtain even in a controlled laboratory setting. In itself this would already raise concern on the reliability of aquatic toxicity test data of PBDEs. After reviewing available experimental sediment effect data it is obvious that effects in sediment occur at relatively high concentrations (from 50 to 1500 mg/kg), and the equilibrium partitioning theory may be used to assess whether exposure based waiving is justified for PBDEs. Alternatively, bioconcentration data may be used to calculate toxicity endpoints, either using experimentally obtained critical body burdens (CBBs) for the various PBDE's, or by using QSAR approaches for predicting CBBs. The latter approach is advocated by for instance Hendriks et al. [2005].

3 – Read across and Category approaches for CADASTER

Following the general discussion on read across and category approaches, the possibilities for application of these non-testing approaches are discussed in this chapter in more detail for each of the four chemical classes that are the core of the CADASTER project.

3.1 Polybrominated Diphenylethers

The polybrominated diphenylethers form a very good example of a category of compounds for which application of read across and category approaches is potentially possible, to establish trends regarding both their fate and effect properties, on the basis of data-rich brominated diphenylethers in the CADASTER group. This will allow to fill essential data gaps and allow for prioritization of actual laboratory testing to obtain data essential for hazard and risk assessment. Possibly, if necessary due to lack of data, also chlorinated structural analogues can be used for reading across, or to show trends in impact of halogenation (in terms of number of halogen atoms and position of halogen atoms) on fate and toxicity characteristics. Read-across between structural analogues with different halogen substitution might be complicated for more complex toxicity endpoints such as endocrine disruption (see below in this paragraph).

The group of diphenylethers all share a common functional group – the phenyl-ether bridge. Those substances within this group that do not have this specific functionality (some halogenated phenols are also present in the group) are breakdown products (metabolites) resulting form cleavage of the ether-moiety. It is very probably that all substances in this category follow the same transformation routes in the environment, leading to a limited number of similar breakdown products. The importance of knowledge of metabolism of the PBDE is highlighted by reports in literature that the endocrine disrupting (ED) potential of PBDE's is strongly increased by hydroxylation of the aromatic ring(s) [Hamers 2006, Liu 2007]. This is an example of the conclusions in section 2.2, that knowledge of the metabolism is crucial for correct read-across and category formation (and QSAR), especially for the more complex toxicological endpoints such as ED, chronic toxicity, reproductive toxicity etc. Data from these metabolites (hydroxylated PBDEs), which are not part of the CADASTER selection, should possibly also be taken into account when attempting read-across or category formation of the more complex toxicity endpoints.

Physico-chemical data of non-brominated diphenylethers may well be used to establish trends and in the assessment of the chemical domain in case of QSAR development or application or QSARs.

3.2 Polyfluorinated chemicals

The category of polyfluorinated chemicals is, in a chemical sense, already more diverse than the compound class of diphenylethers, as it includes various structural (chemical) functionalities. Read across (and trend analysis within chemical categories) should therefore at best focus at subgroups of polyfluorinated chemicals which have identical chemical functionalities. Similar to the class of PBDE, information from polyhalogenated structurally related chemicals in general may be used as the basis for read across and categorization. It should be noted, however, that the molecular weight of the fluorine atom is the lowest of all halogen atoms. In all cases, additional attention is needed to warrant that fluorinated compounds are indeed part of the chemical domain spanned by non-fluorine based chemicals of similar basic chemical structure. The possibility to potentially use halogenated substances (other than fluorinated) for read-across or category approaches is again highly dependent on the endpoint for which this exercise is performed. This approach might be less suitable for chronic toxicity endpoints than e.g. physico-chemicals and fate parameters.

3.3 Substituted musks / fragrances

Opposed to the other three compounds classes, the group of substituted musks / fragrances is best typified by its diversity of chemical structures: this class of compounds does not share a specific chemical functionality, but instead shares its use pattern, i.e. all substances are used in fragrances. However, it seems that the class of substituted musks / fragrances is mainly made up of esters (majority), aldehydes, musks (nitroaromatics or esters), and some alcohols. Therefore it seems very well possible to establish categories of chemically very similar fragrances (for example the artificial musks), where a read across on the basis of high chemical similarity could be considered. The fact that these substances also share a similar use pattern might make it interesting to also consider the possibilities for specific Exposure Based Waiving.

3.4 Triazoles / Benzotriazoles

Within this group in CADASTER a number of pesticides are present, which might make category formation or read-across on the basis of shared chemical functionality potentially more difficult, as pesticides often have a very specific mechanism(s) of action. For each read-across or category within this group one has to address the possibility that the triazole derivative for which read across is applied, might possess a specific toxicity profile which was not present in the triazoles derivatives from which data is inter-/extrapolated (and *vice versa*). Especially for (aquatic) toxicity prediction, errors of several orders of magnitude might occur in this way.

4 – Available tools and guidance for Read Across and Category approaches

Both within the OECD as well as in the European Union (EChA) guidance is provided on the use of read-across and category. The efforts from EChA and the OECD are tuned, and a lot of the work on this guidance has been developed under OECD flag, but stimulated by EChA. This is for example also the case with the development of the OECD QSAR Application Toolbox.

- REACH Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals. EChA, 2008. http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?vers=20_08_08
- REACH Practical guide 6 : How to report read-across and categories. EChA 2010. http://echa.europa.eu/doc/publications/practical_guides/pg_report_readacross_categ.pdf
- OECD Guidance document no. 80. Guidance on grouping of chemicals. 2007 http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)28
- Various demonstration materials for the OECD QSAR Application Toolbox can be used as guidance on how to build valid chemical categories or do read-across in a meaningful way. Information on the Toolbox and the OECD QSAR project can be found under: http://www.oecd.org/document/23/0,3343,en_2649_34379_33957015_1_1_1_1,00.html

4.1 OECD HPV categories

In the OECD HPV program a number of substances have been discussed as a group. The assumption has been that the same conclusion for all endpoints (ecotoxic, human and environmental) can be applied to all substances in an OECD category. This assumption, which was used as the basis of proposing categories for discussion in the OECD in the past, is not optimal as it leads to categories that only contain actives or inactives, whereas trends in activity are not present within the category. This problem is highlighted again in the case study on monoethylene glycol ethers in Appendix I (see below in this paragraph). In current guidance (see above) it is recommended to incorporate both actives and inactives in one category.

In the REACH guidance these OECD categories are explicitly mentioned as possible categories. Where substances have been accepted as members of categories under other regulatory programs (for example the OECD HPV categories), the registrant should refer to them in the dossier. Nevertheless all available information (including information which became available after assessment in the other regulatory programme) should be included and, the validity of the category should be reassessed.

More information on the OECD HPV categories can be found online:
http://cs3-hq.oecd.org/scripts/hpv/
http://www.chem.unep.ch/irptc/sids/OECDSIDS/sidspub.html
http://webnet.oecd.org/hpv/ui/ChemGroup.aspx

A number of CADASTER selected substances have been discussed in the OECD HPV program as a category:

Polybrominated Diphenyl ethers:

No OECD HPV categories apply

Perfluoroalkylated substances:	
PFOS and its salts (5)	Perfluorooctane sulfonic acid (CAS 1763-23-1)
	Perfluorooctane sulfonic acid, ammonium salt (CAS 29081-56-9)
	Perfluorooctane sulfonic acid, diethanolamine salt (CAS 70225-14-8)
	Perfluorooctane sulfonic acid, lithium salt (CAS 29457-72-5)
	Perfluorooctane sulfonic acid, potassium salt (CAS 2795-39-3)
PFOA (2)	Ammonium Perfluorooctanoate (APFO) (CAS 3825-26-1)
	Perfluorooctanoic Acid (CAS 335-67-1)

Substituted musks and Fragrances :	
Menthols (4)	+)-Menthol (CAS 15356-60-2)
	D/L-Menthol (CAS 89-78-1)
	Menthol (CAS 2216-51-5)
	menthol (CAS 1490-04-6)

Triazoles and Benzotriazoles :

No OECD HPV categories apply

Recently, within the OECD, and also within the EU a discussion has been started on the (im)possibility to generate strict criteria (better defined than in the current guidance documents) on the type and amount of information necessary to build a category. Within the OECD this discussion was held during a joint meeting of the SIDS (Screening Information Data Set) Initial Assessment Meeting (SIAM 30, 19-22 April 2010, Paris, France) together with the OECD Working Group on QSARs. For this discussion a case study was worked out on the OECD category of monoethylene glycol ethers, targeted at the issue of developmental toxicity (or lack thereof) of the category members. This is an existing OECD category

and the case study tried to analyze which data would be required to extend this existing category with one or more substances which meet the initial definition of the category in terms of chemical structure. This same case study was brought to discussion in a Workshop on REACH Testing Proposals organized by the European Chemicals Agency (EChA) in Helsinki, Finland, April 26-27, 2010. In both international fora the consensus among the experts/decision makers present was that it will be impossible to set specific criteria on the type and amount of information needed to construct an acceptable, valid read-across or category approach. The conclusion was that this type of reasoning is very much case-by-case, as all kinds of information interplay. As an example, the actual use of a chemical was given as it can influence the level of detail which is required. Read across of a property like (absence of) carcinogenicity will be more easily accepted for a substance which is used as flame retardant in a matrix in which it will be trapped for the lifetime of the product (e.g. a plastic), as compared to a substance that might end up in consumer products and will have intended exposure. This is for instance the case for fragrances. Therefore the foundation of read across cases or categories for the CADASTER group of fragrances might require more "evidence" than for the group of (trapped) brominated flame retardants.

To exemplify the provision on guidance on grouping of chemicals, the case study on developmental toxicity of monoethylene glycol ethers is included in Annex I to this report as it is not publicly available and serves as a nice example on how to apply a category approach for a specific endpoint. Similar studies are foreseen for the four CADASTER classes of compounds.

4.2 OECD (Q)SAR Application Toolbox

To increase the regulatory acceptance of (Q)SAR methods, the OECD has started the development of a (Q)SAR Application Toolbox to make (Q)SAR technology readily accessible, transparent, and less demanding in terms of infrastructure costs. The Toolbox is a software application intended to be used by governments, chemical industry and other stakeholders in filling gaps in (eco)toxicity data needed for assessing the hazards of chemicals. The Toolbox incorporates information and tools from various sources into a logical workflow. Crucial to this workflow is grouping chemicals into chemical categories. More information on the toolbox, and the possibility to download the software can be found under: www.oecd.org/env/existingchemicals/qsar

The OECD QSAR Toolbox gives the user the possibility to characterize a substance using "Profiles". If a certain "Profile" applies to the substance of interest (e.g. substance belongs to HPV category PFOA, or substance contains a chemicals reactive site which is related to DNA binding), then the workflow within the Toolbox allows a user to search all (relevant) databases for substances to which the same Profile

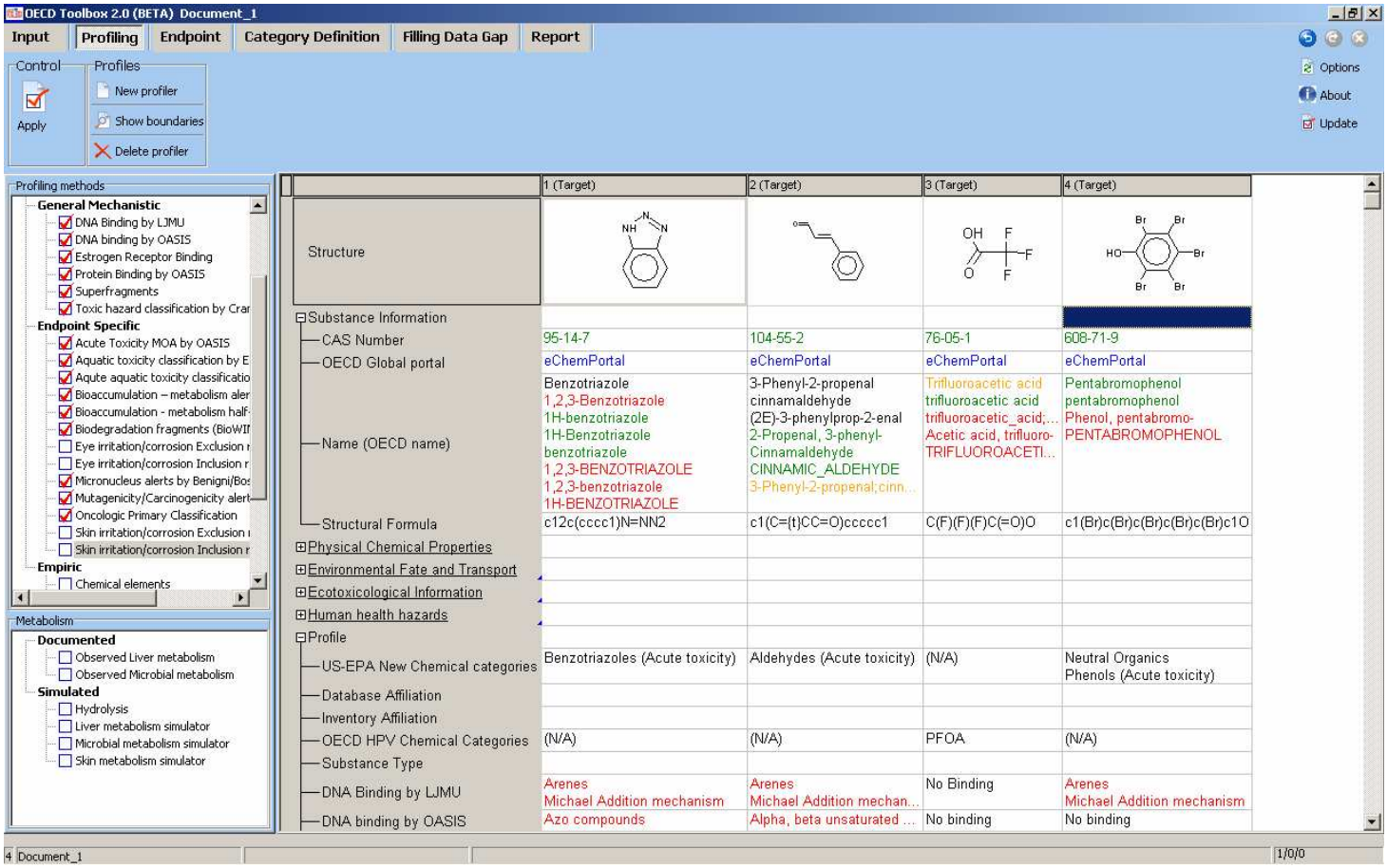
applies, and that have experimental data reported for the endpoint of interest. The basic idea is that a profile (or combinations of profiles) delivers the (mechanistical) basis for the application of read-across or category approaches. If a profile is mechanistically directly linked to the endpoint of interest this might be obvious (e.g. the link between DNA binding profiler and potential mutagenicity of a substance), but often profiles developed for one endpoint (e.g. skin sensitization / protein binding profiler) can be very relevant for unrelated endpoints (e.g. fish toxicity as substances that bind to proteins will also show specific (acute) fish toxicity).

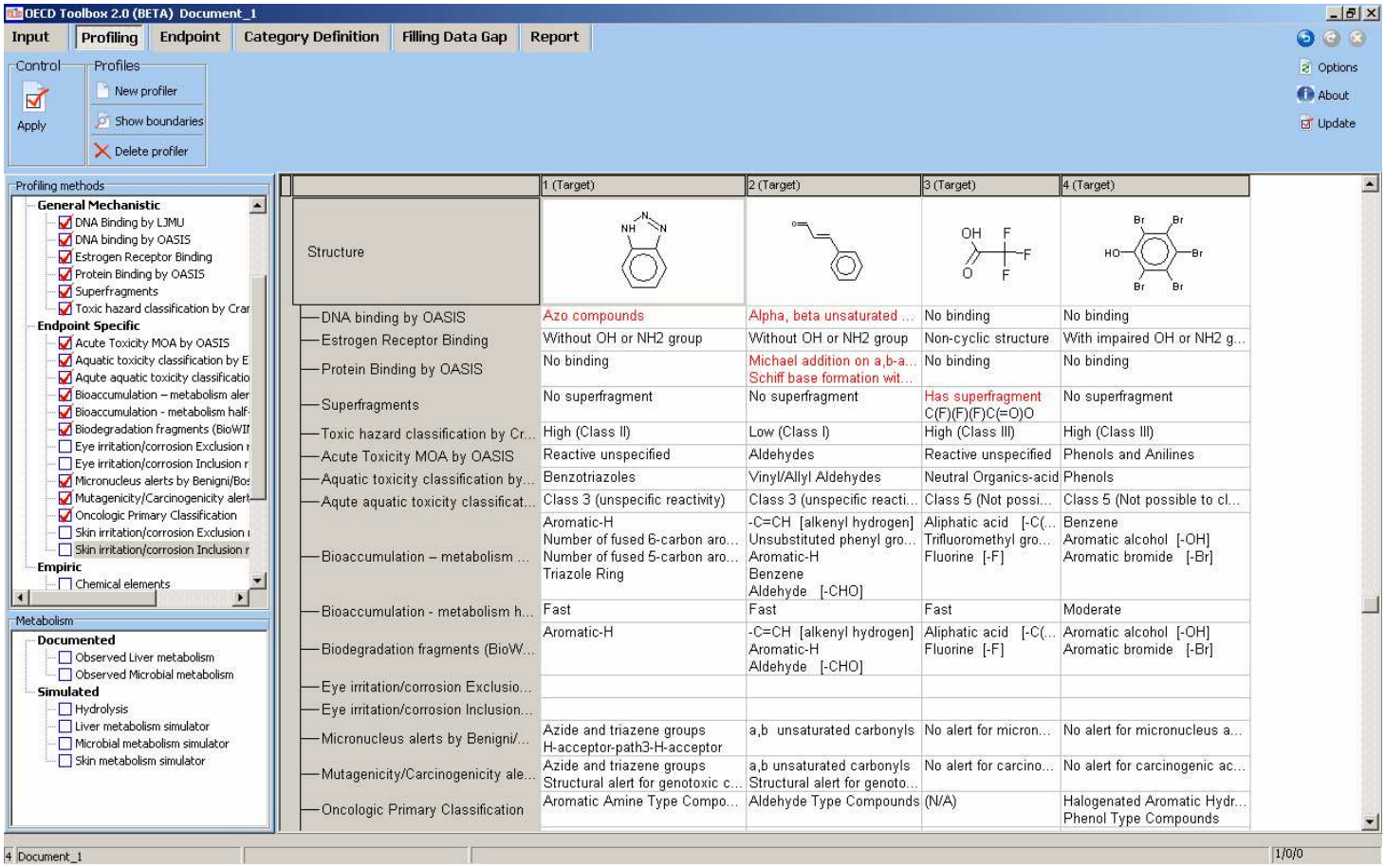
It seems worthwhile to use the OECD QSAR Application Toolbox for all substances selected in the four CADASTER groups to generate their profiles, and subsequently use these profiles to define subsets of substances which are used for read-across and category approach purposes. Optimally, such subsets are based on mechanistical profiles (e.g. protein binding) which are related to the endpoint of interest, but selection can also be performed on more empirical profiles, using e.g. chemical functionalities, and combinations e.g. with bioavailability profiles such a Lipinski's rule of five might we considered, all depending on the relevance for the endpoint of interest.

Because of the large number of possible profile / toxicological endpoint combinations it is suggested that specific data gaps (substance(s) + endpoint) are determined for which read-across / category approaches might be useful to fill this gap.

As an example of the profiles present in the current version of the OECD QSAR Application Toolbox (v2.0 Beta testing version) one substance at random from each CADASTER group is shown with its OECD QSAR Application Toolbox profile on the next two pages.

Figure 1. OECD QSAR Application toolbox screenshots showing the assigned profile classes for four substances from the different CADASTER selected groups. Benzotriazole (CADASTER group triazoles / benzotriazoles); Cinnamaldehyde (CADASTER group Fragrances); Trifluoroacetic acid (CADASTER group perfluoroalkylated substances) and pentabromophenol (CADASTER group of brominated diphenyl ethers).



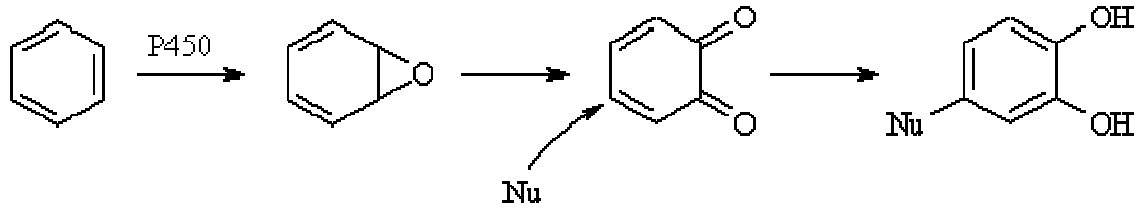


The "US EPA New Chemical categories" profile gives the (obvious) information that these 4 substances from the 4 different CADASTER groups all belong to different categories, according to this scheme; benzotriazoles (CADASTER group benzotriazoles), aldehydes (CADASTER group of fragrances), N/A (CADASTER group of perfluoroalkylated substances) and phenols (CADASTER brominated diphenyl ethers). Much more interestingly three of the four substances are identified as potential DNA binders (LJMU DNA binder profile), through the same reactivity mechanism (Arenes, Michael addition mechanism of the metabolites). In this case the profiler indicates the potential DNA binding of aromatic ring substances after P450 oxidation (see next page). The P450-generated metabolites are hypothesized to be able to bind to DNA via a Michael type addition, see the category documentation from the Toolbox below.

OECD QSAR Toolbox, LJMU DNA Binding Category: Arenes

Mechanism

A P450 mediated epoxidation followed by conversion to a reactive quinone has been postulated as the primary cause of benzene derivatives ability to bind to biological nucleophiles (via a Michael addition mechanism) (Saghir et al 2009, Ishihama et al 2008).



Nu = biological nucleophile

Mitigating factors

- No mitigating factors have been reported

Before actually deciding to build a category on this common fragment (the arene ring), first the possibility of P450 oxidation of the three substances should be evaluated, and the relevance of this type of reactivity for the endpoint of interest should be determined. One could in this case possible conclude that the P450 oxidation an subsequent Michael addition reactivity is not very likely for pentabromophenol due to the fact that the arene ring is fully substituted. Subsequently this substance would then be excluded from a read-across/category approach. However, this random and very hypothetical example immediately shows that it might be worthwhile to apply read-across and/or category approaches across the four CADASTER groups, and not limit a read-across or category to substances within the CADASTER chosen groups.

Another feature of the QSAR toolbox is the possibility to profile substances based on their (reported or simulated) metabolism. Knowledge (or at least some hypothesis) of the metabolism of a substance is crucial information to build a proper category or to perform read-across. Reading across a property of chemically very similar substances, which differ in the (bio)transformation, can potentially lead to very wrong conclusions. Even read-across of physico-chemical properties can result in errors if one of the substances transforms quickly (i.e. by hydrolysis). Especially for a meaningful risk assessment of the substance in the environment information on stability of the substance, and on possible metabolites is necessary.

If a category is formed by substances that show similar transformation patterns, or which have similar metabolites, then this fact strengthens the plausibility of the category interpolations or a read-across of properties.

4.3 TOXWIZ database

The ToxWiz software from CambridgeCellNetwork (http://www.camcellnet.com/index.php) could be a means to establish categories based on mechanism of action instead of assessment of categories based on chemical similarity. This would theoretically lead to more robust categories, as one assumption in the chemical category approach – similar structure leads to similar effect - is omitted. The database contains reported substance – protein interactions, which might lead to a specific toxic effect.

ToxWiz is a software solution for predicting toxic endpoints and for elucidating mechanisms of toxicity. It allows the user to understand on- and off- target mechanisms of action of the compounds of interest, thus minimizing the number of animal testing wherever possible, and it uses novel algorithms to predict toxic end-points. This approach to predictive toxicology offers a new perspective in this field, being highly complementary to well-established QSAR and other approaches. The ToxWiz software contains:

- The world largest data collection of chemical structures linked to protein targets - over 20 000 bioactive chemicals and growing;
- Across-species translation of effects for over 15 different organisms;
- Collection of over 900 organ and tissue specificity of toxic endpoints;
- Means to search by exact chemical structure, substructure, free text, gene and protein sequence;
- A wide spectrum of chemicals such drugs, metabolites, agrochemicals, food additives and industrial & environmental chemicals;
- A large database of manually curated data from textual databases (PubMed abstracts);
- A focus on pathways known to be involved in toxic endpoints or pathologies;
- The ability to compare results between ten different tox-relevant species;
- Thousands of known chemical/protein interactions relevant to toxicology, including known ligands, substrates, products, inducers & suppressors of all major drug metabolising enzymes, and nuclear hormone receptors;
- The ability to handle large sets of genes, proteins or chemicals and use these to predict toxic-endpoints or affected pathways.

RIVM currently acquired an evaluation licence (1-year) for this software and experience in using the software still needs to be built up. It is recommended nevertheless to investigate the CADASTER selection of substances on possible similarities in enzyme/protein interactions, as this might be linked to similar (toxic) effects. Grouping substances based on the fact that they all interact with the same enzymes should theoretically lead to categories that have the same mechanism of action for a specific toxicity endpoint.

The focus of the software is on (human) toxicology, therefore it will not be of direct use for ecotoxicological and environmental fate endpoints. However, possibilities for read-across are to be investigated aiming at extrapolation of human toxicological endpoints towards ecotoxicity endpoints. This might include read-across of data obtained for vertebrates like rats and mice towards aquatic toxicity for specific fish species. A (hypothetical) example of such a read-across (mentioned earlier in section 4.2) would be the group of aldehydes (part of the CADASTER class of substituted musks/fragrance). Aldehydes show specific skin sensitizing properties, but also increased toxicity to aquatic species (compared to base-line or narcosis type toxicity). If a trend is visible in skin sensitization data that some aldehydes are skin sensitizers and others are not, this trend might also be used to predict higher aquatic toxicity for those compounds, or vice versa. The read-across is then based on the mechanistical assumption that the reactivity towards proteins that is causing the skin sensitizing properties will also lead to increased toxicity to aquatic organisms. The reactivity towards proteins can be used as a category argument e.g. using the Toolbox, and selecting those substances to which the Protein Binding profile apply, or based data in the ToxWiz database which highlights interactions with specific enzymes or proteins known to be related to skin sensitization.

5 – References

EC 2006a. European Communities. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Official Journal of the European Union* **L396/1** of 30.12.2006. Accessible from http://publications.europa.eu.

EC 2006b. European Communities. Directive 2006/121/EC of the European Parliament and of the Council of 18 December 2006 amending Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances in order to adapt it to 60 Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) and establishing a European Chemicals Agency. *Official Journal of the European Union* **L 396/850** of 30.12.2006. Accessible from http://publications.europa.eu/

ECB 2007. QSAR Reporting Formats and JRC QSAR Model Database. Website information. http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=QRF

Hamers 2006. Hamers, T., Kamstra, J.H., Sonneveld, E., Murk, A.J., Kester, M.H.A., Andersson, P.L., Legler, J., Brouwer, A. 2006. *In vitro* profiling of the endocrine-disrupting potency of brominated flame retardants. Tox.Sci. 92, 157-173.

Hendriks 2005. Hendriks, A.J., Traas, T.P., Huijbregts, H.A.J. 2005. Critical Body Residues Linked to Octanol-Water Partitioning, Organism Composition, and LC50 QSARs: Meta-analysis and Model. Environmental Science and Technology 39, 3226-3236.

Ishihama 2008. Ishihama, M., Toyooka, T., Ibuki, Y. 2008. Generation of phosphorylated histone H2AX by benzene metabolites. Toxicology in Vitro 22, 1861-1868 .

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

Liu 2007. Liu, H., Papa, E., Walker, J.D., Gramatica, P. 2007. In silico screening of estrogen-like chemicals based on different nonlinear classification models. J. Mol. Graph. Model. 26, 135-144.

OECD 2004 Manual for investigation of HPV chemicals. Paris, France, Organization for Economic Co-operation and Development http://www.oecd.org/document/7/0,3343,en_2649_34379_1947463_1_1_1_1,00.html

Pedersen 2003. 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.

Rila 2006. 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. www.rivm.nl/bibliotheek/rapporten/601200009.pdf

Rorije 2007. E Rorije, Hulzebos E, Hakkert B (2007) The EU (Q)SAR Experience Project: reporting formats. Templates for documenting (Q)SAR results under REACH. RIVM rapport 601779001, www.rivm.nl/bibliotheek/rapporten/601779001.pdf

Saghir 2009. Saghir, S.A., Rick, D.L., McClymont, E.L., Zhang, F., Bartels, M.J., Bus, J.S. 2009. Mechanism of Ethylbenzene-Induced Mouse-Specific Lung Tumor: Metabolism of ethylbenzene by rat, mouse, and human liver and lung microsomes. Toxicological Sciences 107, 352-366.

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 F1SCISPBE.

Van der Jagt 2004. 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.

Vermeire 2007. T.Vermeire *et al*. (2007) Selected Integrated Testing Strategies (ITS) for the risk assessment of chemicals. RIVM report 601050001/2007. www.rivm.nl/bibliotheek/rapporten/601050001.pdf

Annex I

Guidance on Grouping of Chemicals: Expanding existing categories.

Developmental toxicity of monoethylene glycol ethers

Emiel Rorije, Sjofn Gunnarsdottir, Betty Hakkert, Aldert Piersma
RIVM, Bilthoven, The Netherlands
version 4 (April 9, 2010)

Abstract

In order to stimulate the discussion on data requirements for hazard assessment of chemicals within the OECD HPV program, a case study is presented where the extension of existing OECD category of monoethylene glycol ethers with two additional glycol ethers is discussed. Information at various levels is presented; this information may be used to support inclusion of these additional glycol ethers. The levels are ordered from simple to complex: 1) chemical similarity; 2) physico-chemical data; 3) QSAR predictions; 4) i*n vitro* experimental data; 5) *in vivo* experimental data. The question is asked what (level of) information would be sufficient to extend the category using interpolation and using extrapolation.

No definitive recommendations are derived on what information is required, as this case study is not supposed to actually build the case for inclusion of these substances in the existing OECD category. The goal is to discuss what information should be required, and what information would be needed for the OECD SIAM to accept extensions of an existing category.

Introduction

OECD Guidance document no. 80 (Guidance on grouping of chemicals) defines a chemical category as a "*group of chemicals whose physico-chemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristic)*".

In chemical assessment programs, data availability or (commercial) interests of the sponsor are factors that can play a role in the selection of members to be included in the category when a category is being formed. Such factors can lead to the formation of a category that does not contain all chemicals that fit the category definition. In the same way, the boundary of a category can also be based on factors

other than structural similarity or similarity in properties. It is often possible to expand the boundaries of such categories without changing the properties or primary characteristics of the category. It should be noted that a lot of the existing OECD categories were created for practical reasons of discussing (similar) substances in SIAM at once, and were not (necessarily) created following OECD Guidance document 80 on Grouping of Chemicals, as a large number of categories were proposed before the guidance document was drafted.

Recently, the OECD Task Force on Hazard Assessment encouraged the investigations into expanding existing chemical categories, for example by expanding the definition of a previously assessed category beyond previously defined boundaries or by applying conclusions from existing assessments for individual chemicals to other chemicals (ENV/JM/HA(2009)13).

In the case study presented in this paper, we explore the expansion of the category of monoethylene glycol ethers, targeted specifically for the endpoint developmental toxicity. The proposed category extension is not necessarily valid for other (toxic) endpoints. More specifically, we will examine which information is necessary to

a) add a category member using category interpolation; and
b) extend the category by extrapolation outside the (presumed) category boundaries

The monoethylene glycol ether category and the endpoint developmental toxicity were chosen for several reasons. Firstly, testing of the developmental toxicity potential of a chemical traditionally requires a (relatively) large number of animals, a lot of time and is considered expensive. Secondly, the amount of chemicals that needs to be assessed in the near future because of new legislations such as REACH[1], in combination with animal welfare concerns, underscore the need for faster, cheaper and more animal friendly ways to assess the developmental toxicity of (large numbers of) chemicals. Although non-*in vivo* methods for assessing developmental toxicity are available, e.g. several ECVAM validated *in vitro* methods, these methods are currently not used in regulatory settings, and are considered not (yet) adequate as stand alone methods to assess the developmental toxicity of a chemical, and are therefore currently not used in regulatory settings as such. The same can be said for the few *in silico* models available to assess developmental toxicity. Data for a chemical is often available from a wide variety of sources; these data can give direct or indirect information on the hazard of a chemical. Further discussion is necessary on which information is acceptable, adequate and sufficient to fulfill the information requirement on developmental toxicity for regulatory purposes.

In this paper, we will not be building or defending a case for the expansion of the monoethylene glycol ether category. Rather, we will discuss the variety of relevant information that may be available and can be used to come to a conclusion on the acceptability of adding a chemical to an existing category. For this exercise, we have organized the information into 5 different levels that follow a

[1] For a list of abbreviations, see section 0.

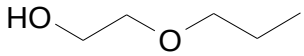
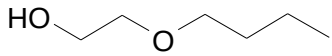
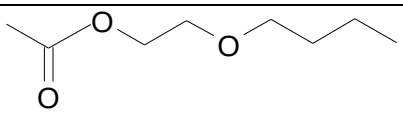
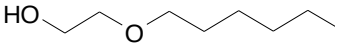
certain order, going from the simple to the more complex. The levels suggested in this exercise do not comprise an exhaustive list of information levels. Neither is all available information for all chemicals discussed; other types of relevant information are almost certainly available. In addition, the order that the information is examined may differ depending on data availability and the characteristics of the chemical being assessed, but assuming that no information at all is available at the start of the exercise makes it logical to go from simple (inexpensive) to complex (expensive).

Monoethylene glycol ethers

The monoethylene glycol ether category

Monoethylene glycol ethers are a group of chemicals that have the general structure HO-CH_2-CH_2-O-R where R can in theory be any functional group. During the 19th meeting of OECD's High Production Volume Chemicals Program in October 2004 (SIAM 19), a category of monoethylene glycol ethers containing the members shown in Table 1 was discussed.

Table 1: The members of the monoethylene glycol ether category discussed at SIAM 19.

Category member	Abbreviation	CAS no.	Structure
Ethylene glycol propyl ether	EGPE	2807-30-9	
Ethylene glycol butyl ether *	EGBE	111-76-2	
Ethylene glycol butyl ether acetate	EGBEA	112-07-2	
Ethylene glycol hexyl ether	EGHE	112-25-4	

* Ethylene glycol butyl ether is included in the category only to fill data gaps for mammalian toxicity. The data set was discussed and agreed during SIAM 6.

Boundaries of this monoethylene glycol category were not explicitly defined in the SIAP. In the category justification, it is stated that "the four substances of this category all have similar molecular structures, functionality and metabolic pathways. The category members demonstrate similar physicochemical properties and mammalian toxicity". Interpolation or extrapolation of the category is not foreseen, boundaries were therefore not defined. The OECD category just consists of its members.

Based on the structure of the current category members and the justification given, the category can (retrospectively) be defined as consisting of monoethylene glycol ethers with a linear alkyl C3 to C6 ether side chain and their corresponding acetates. Using this category definition – which does not specify anything about (toxicological) endpoints for which it is supposed to be valid – it can be seen that the category does not contain all possible members: ethylene glycol pentyl ether (EGPeE) and the acetates of the propyl, pentyl and hexyl ethers are currently not included.

Developmental toxicity of monoethylene glycol ethers

The developmental toxicity of ethylene glycol ethers is generally believed to be due to their alkoxyacetic acid metabolites [Louisse, 2010]. These are formed by the oxidation of the ethylene glycol by alcohol and aldehyde dehydrogenases [ECETOC, 2005; Louisse, 2010]. The adverse effects observed after exposure to monoethylene glycol ethers are structure dependent. Developmental toxicity, testicular atrophy, bone marrow depression and immunotoxicity have been observed after exposure to monoethylene glycol methyl and ethyl ethers but not after exposure to the longer chain ethers (C3 and higher). In contrast, haemolysis (anemia) has been observed in experimental animals administered the longer chain ethers; this anemia is considered not relevant for humans. The difference in systemic toxicity observed with chain length is considered to be due to different kinetics of the alkoxyacetic acid metabolites. MAA and EAA show relatively slow excretion rates, especially in larger animals. The half-life of MAA and EAA is 14-18.6 h and 7.6-10.1 h, respectively. The excretion rate of the longer-chain alkoxyacetic acid metabolites is faster. For example, the half-life of BAA is reported to be 1.5-3.2 h. The longer excretion half-lives result in higher plasma levels which may contribute to the higher toxicity observed with the smaller metabolites [ECETOC, 2005; De Jong, 2009].

Of mechanistic interest is that the adverse effects observed after exposure to the monoethylene glycol ethers are not observed after exposure to propylene glycol ethers which have a secondary alcohol group (α-isomer). These substances cannot be oxidized to the corresponding alkoxypropionic acid. In contrast, propylene glycol methyl ether in which the alcohol is pimary (β-isomer), can thus be oxidized to the alkoxypropionic acid metabolite and has been reported to cause developmental effects. This observation is a further argument for the view that the alkoxyacetetic acid metabolites are the mediators of the adverse effects observed after exposure to monoethylene glycol ethers [ECETOC, 2005; Louisse, 2010].

SIAM19 concluded that the category of monoethylene glycol ethers are not primary developmental toxicants but rather that developmental toxicity is a secondary effect due to maternal toxicity. This conclusion was based on results of developmental toxicity studies via the inhalation route during the gestation period for the category members EGPE, EGBE and EGHE in rats and rabbits (SIAP,

SIAM19). The two monoethylene glycol ethers that are considered developmental toxic according to SIAM evaluation, the methyl- and ethyl ethers EGME and EGEE, are not part of the category.

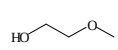
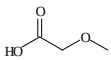
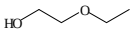
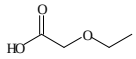
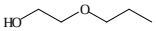
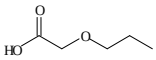
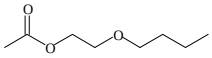
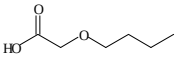
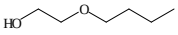
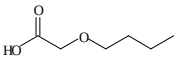
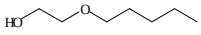
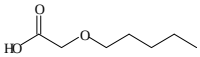
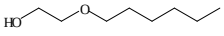
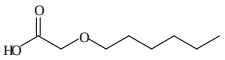
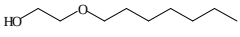
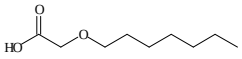
Expanding OECD category monoethylene glycol ethers category for the developmental toxicity endpoint

As discussed in section 0, the OECD category of monoethylene glycol ethers contains a gap: EGPeE is currently not included in the category although it would fit a category definition based on structure. If a sponsor would like to add EGPeE to the category, which information would be available and considered sufficient to justify inclusion of the EGPeE? And following up on this, there appears to be no chemical reason for limiting the category to ethers with a maximum alkyl chain length of six. In the following sections, several levels of information that may be available to justify the inclusion of EGPeE and monoethylene glycol heptyl ether (EGHepE) into the category of monoethylene glycol ethers will be discussed.

Level 1: Chemical Structure

Based on structural evidence, EGPeE could be included in the monoethylene glycol ether category. EGBE and EGHE, which are already members of the category, are structural analogs of EGPeE. These three chemicals are all monoethylene glycol ethers with straight aliphatic ether chains. The only difference is the length of the ether chain. Whereas EGPeE has a 5 carbon chain, EGBE and EGHE have a 4 and a 6 carbon chain, respectively (Table 2). Both EGBE and EGHE would be considered to be good structural analogs of EGPeE. EGHepE has the same chemical functionalities as the rest of the chemicals in the category. The length of the ether alkyl chain is seven carbons, one more than for EGHE which is not a developmental toxicant.

Table 2. Chemical structures of monoethylene glycol ethers and their metabolites

Substance		Ether alkyl chain length	Metabolite		OECD category monoethylene glycol ethers	SIAM conclusion on Developmental Toxicity
EGME		1	MAA		outside	Positive
EGEE		2	EAA		outside	Positive
EGPE		3	PAA		member	Negative
EGBEA		4	BAA		member	(negative)
EGBE		4	BAA		member	Negative
EGPeE		5	PeAA		interpolation	??
EGHE		6	HAA		member	Negative
EGHepE		7	HepAA		extrapolation	??

Furthermore, substances with similar structures are generally considered to have a similar mechanism of action. Based upon the structural similarity of EGBE, EGPeE and EGHE, a conclusion might be drawn that all three chemicals have the same mechanism of action and therefore cause the same toxicity.

> **Is this structural information sufficient to add EGPeE to the OECD category of monoethylene glycol ethers?**

> **Is it sufficient to add EGHepE to this category?**

Also, based on the similar structure, similar metabolism (conversion to alkoxy acetic acid metabolites) can be assumed for all substances in the series, although the rate of metabolism might differ in the series. The same arguments are valid for EGHepE. The (hypothesized) metabolites of the monoethylene glycol ethers are also given in table 2.

> **Is chemical structure combined with (hypothesized) metabolism information sufficient to add EGPeE to the OECD category of monoethylene glycol ethers?**

> **Is it sufficient to add EGHepE to this category?**

Although it is generally assumed that structurally similar compounds cause the same effects, there are several examples where this assumption does not hold true. An example is hexane which is structurally comparable to pentane and heptane in the same manner as EGPeE is structurally similar to EGBE and EGHE. However, hexane is a much more potent neurotoxicant than pentane and heptane. Another example can be seen in the series of phthalate esters with regards to reproductive toxicity, where the dibutyl- and 2-ethylhexyl esters of phthalate causes reproductive effects whereas the shorter (methyl, ethyl, propyl) and longer (octyl, nonyl) chain alkyl esters of phthalate do not show this effect. For both these examples, an interpolation as proposed for the monoethylene glycol ethers might have resulted in the wrong conclusions; i.e. that a substance would have been erroneously classified as non-hazardous. Extrapolation of a trend, to include EGHepE in the OECD category can give similar mistakes in reasoning.

Level 2: Physico-chemical paramaters and (toxico-)kinetics

Physico-chemical parameters such as log K_{ow} do affect the absorption, metabolism, distribution and excretion of a chemical. For example, absorption via the oral route is thought to become limited when substances become very hydrophobic (>log Kow 5, Lipinski's rules). For the monoethylene glycol ethers, the size of the ether chain will determine their log Kow value. The log Kow of EGPeE is estimated to be 1.06 (KowWin estimate), which comes as expected between the estimated log Kow values of EGBE (0.57) and EGHE (1.55). Based on the log Kow value, EGPeE appears to be a suitable member of the category. The log Kow of EGHepE is estimated to be 2.07 which suggests that it might be slightly less absorbed through the oral route. The increasing log Kow does continue the trend seen for the smaller carbon chain glycol ethers.

Information on the hypothesized embryotoxic mechanism of action of the glycol ethers - decrease of intracellular pH [Louisse, 2010] - makes it relevant to look at the (estimated) pKa values of the (metabolites of the) glycol ethers. Log Kow (experimental and estimated) and pKa (estimated using the ACE acidity calculator, available online; http://aceorganic.pearsoncmg.com/epoch-plugin/public/pKa.jsp) are summarized in table 3 for the series of monoethyleneglycol ethers.

> **Is this physico-chemical information, related to kinetics, sufficient information to add EGPeE to the OECD category of monoethylene glycol ethers?**

> **Is it sufficient to add EGHepE to this category?**

Table 3: log Kow and pKa estimates of monoethylene glycol ethers

Substance	exper. log Kow	KowWin estimate	Metabolite	pKa estimate	OECD category monoethylene glycol ethers	SIAM conclusion on Developmental Toxicity
EGME	-0.77	-0.91	MAA	3.8	outside	Positive
EGEE	-0.32	-0.42	EAA	4	outside	Positive
EGPE	-	0.08	PAA	4.1	member	Negative
EGBEA	-	1.57	BAA	4.3	member	(negative)
EGBE	0.83	0.57	BAA	4.3	member	Negative
EGPeE	-	1.06	PeAA	4.3	interpolation	??
EGHE	1.86	1.55	HAA	4.4	member	Negative
EGHepE	-	2.04	HepAA	4.2	extrapolation	??

The (seemingly) decreasing pKa of the EGHepE might raise doubt on the possibility to extrapolate the category to include this substance in the category. However, pKa estimates for even longer chain lengths (octyl, nonyl, decyl) give the same estimated value (pKa 4.2), showing that further increasing the chain length does not lower pKa more. Other pKa estimation methods (VCClab; http://www.vcclab.org/lab/alogps/start.html, or SPARC; http://sparc.chem.uga.edu/sparc/) produce identical values for the pKa of the whole series of ethylene glycol ether from methyl up to decyl (pKa of 3.8 and 3.75 for the two QSAR methods respectively).

In general, other physical chemical properties of EGPeE (such as molecular weight, vapour pressure, water solubility) will lie very close to, and probably in between EGBE and EGHE, and EGHepE will be similarly close to EGHE. Their pharmacokinetics are therefore also expected to be similar. Also, their metabolism is expected to be similar, although the rate of transformation is likely to follow a trend (increase or decrease) with increasing size (chain length) and/or log Kow. It should be kept in mind, however, that different metabolic pathways and/or the flux between different metabolic pathways can differ for these compounds. Such differences might be more pronounced or more likely for EGHepE than EGPeE.

It should be noted that kinetics are only one part of the mechanism of action. Furthermore, the hypothesis of similar kinetic behaviour for substances with similar physico-chemical properties also assumes for example the same (passive) absorption for all substances. If humans have a specific active uptake mechanism for EGPeE which does not function for EGBE (or EGHE), the hypothesis of similar behaviour of course fails. Information on a trend in excretion rates is still lacking, although this is indicated in the literature [ECETOC, 2005; de Jong, 2009] to be a determining factor differentiating between the developmental toxicity of the methyl and ethyl ethers and the non-developmental toxicity of the longer chain ethylene glycol ethers. Excretion rates are further discussed in section 3.5.

Level 3: QSAR predictions

QSAR models for developmental toxicity are scarce. The commercial TopKat package is one of the few that is intended to produce a prediction that can be used (interpreted) to predict the absence of a potential effect. Structural alert models, as e.g. implemented in DEREK for a number of human toxicological endpoints (but not developmental toxicity) are not meant to be conclusive if no alert is found, and subsequently have a high(er) rate of false positives associated with them. The developmental toxicity model in TopKat is extensively described in the manual of the program, and the relevant parts have been copied in Annex I to serve as (information required by) a QSAR Model Reporting Format (QMRFs, [Rorije, 2007]). Several examples of detailed model predictions from TopKat, which can be regarded as QSAR Prediction Reporting Formats (QPRFs, [Rorije 2007]) are also given in Appendix I to this case study of monoethylene glycol ethers. As TopKat predicts the animal test outcome of a parent substance, metabolism is implicitly incorporated in the model. This is apparent in the comparable probabilities computed for the ethylene glycol ethers and their acetic acid metabolites.

Table 4. TopKat QSAR predictions for monoethylene glycol ethers and their metabolites.

Substance	TopKat DevTox probability*	Metabolite	TopKat DevTox probability*	OECD category monoethylene glycol ethers	SIAM conclusion on Developmental Toxicity
EGME	**0.997**	MAA		outside	Positive
EGEE	**0.957**	EAA	**0.987**	outside	Positive
EGPE	**0.966**	PAA	**0.974**	member	Negative
EGBEA		BAA	**0.040**	member	(negative)
EGBE	**0.017**	BAA	**0.040**	member	Negative
EGPeE	**0.011**	PeAA		interpolation	??
EGHE	**0.001**	HAA		member	Negative
EGHepE	**0.001**	HepAA		extrapolation	??

* Probability values from 0.0 to 0.30 are considered low probabilities, and are likely to produce a negative response in an experimental assay; whereas probability values greater than 0.70 are considered high, and are likely to produce a positive response in an experimental assay. Probabilities greater than 0.30 but less than 0.70 are considered indeterminate.

> Is this QSAR information sufficient information to add EGPeE to the existing OECD category of monoethylene glycol ethers?

> Is it sufficient to add EGHepE to this category?

From the data in the detailed model prediction reports (Annex I), it seems that the data basis of positive developmental toxicants that are given as toxicological/structural analogues is identical to what

is available to us in the OECD category. In other words, the methyl, ethyl and propyl ethers are considered developmental toxicants and the butyl and hexyl ether are considered non-developmental toxicants. The TopKat prediction indicating propyl ether as a developmental toxicant is therefore a more conservative, precautionary interpretation then the OECD category conclusions). The QSAR model therefore offers us only a different type of descriptors that are related to developmental toxicants, but is not offering more in terms of training data. The question therefore is reduced to: do we gain an increase in confidence using a TopKat prediction which applies structure descriptors statistically shown to have a relationship with developmental toxicity, compare to predictions based on the simple trends signaled earlier in levels 2 and 3?

Level 4a: in vitro results (Embryonic Stem Cell Test)

In case structural similarity and comparable physical-chemical parameters are not considered sufficient to draw a conclusion on the addition of EGPeE or EGHepE to the category, *in vitro* data might be considered as the next level of information.

In vitro assays for developmental toxicity of chemicals have been developed. One of them is the Embryonic Stem Cell Test (EST). The EST has been scientifically validated by the European Centre for the Validation of Alternative Methods (ECVAM, http://ecvam.jrc.it/), however, due to several limitation this *in vitro* test can not be used as a stand alone method in REACH [Marx-Stoelting, 2009].

The INVITTOX protocol reports the ID50 values obtained for 16 chemicals in the EST. These 16 chemicals were grouped into three teratogenic groups: non-teratogens, weak/moderate teratogens and strong teratogens. Table 5 shows the range of ID50 values (in mM) obtained for these three groups. Further information on the chemicals comprising this data set along with additional data can be found in Appendix II to this case study of monoethylene glycol ethers and in the INVITTOX protocol (http://ecvam.jrc.it/).

Table 5: The range of ID50 values (in mM) for 16 chemicals whose classification was correctly predicted in the EST.

Embryotoxic potential	**ID50 (range in mM)**
Non	2.3 - 10.9 mM
Weak/moderate	0.05 - 1.38 mM*
Strong	$3x10^{-7}$ - 0.02 mM

* In addition, one moderate teratogen had an ID50 value of $4x10^{-5}$

De Jong and colleagues (De Jong et al., 2009) have tested EGME, EGEE and the acetic acid metabolites MAA, EAA and BAA in the EST in two independent laboratories. Table 6 shows the BMC_v50 (the concentration corresponding to a 50% decline in cell viability) and the BMC_d50 (the concentration corresponding to a 50% decline in the fraction of beating embryo bodies in comparison with solvent control) obtained in this study.

Table 6: Results of EST testing for monoethylene glycol ethers and their metabolites.

	BMC_v50 (mM)	**BMC_d50 (mM)**		**Cell viability at BMC_d50 (%)**
Substance	**Laboratory 1**	**Laboratory 1**	**Laboratory 2**	**Laboratory 1**
EGME	ND	ND	ND	ND
EGEE	ND	ND	ND	ND
MAA	6.7	2.3	2.5	105
EAA	11.9	2.9	3.9	103
BAA	ND	4.5	5.9	102

ND: No reduction in viability or differentiation.

These results demonstrate that the ethylene glycol alkoxyacetic acid metabolites inhibit the differentiation of ES cells at concentrations that do not affect cell viability. Furthermore, the inhibitory response is consistent between two independent laboratories. These results suggest that the EST assay is suitable to assess the differentiation inhibition potential of alkoxyacetic acid metabolites of monoethylene glycol esters. The finding that the parent monoethylene glycol esters did not show any activity is consistent with the fact that the developmental toxicity of the monoethylene glycol ethers is caused by the alkoxyacetic acid metabolite and not by the parent compound.

> Would an EST result for EGPeE and EGHepE, together with the historical data ranges from the validation set of the EST assay be sufficient to come to a conclusion whether these substances can safely be added to the existing OECD category?

When the *in vitro* results in table 6 are compared with the ranges of historical EST data reported in the INVITTOX protocol (Table 5), it can be seen that the BMC_d50 for the alkoxyacetic acid metabolites MAA, EAA, BAA and PAA fall in the range of ID50 values that are observed for non-teratogens: the BMC_d50 for the alkoxyacetic acid metabolites ranges between 2.3 and 7.8 mM whereas the range for the non-teratogens reported in INVITTOX no. 113 is 2.3 to 10.9 mM. Based on these observations (see also figures A1 and A2 in the Annex II), the conclusion might be drawn that *none of the alkoxyacetic acid metabolites are developmental toxicants*. However, this conclusion is not correct since EGME and EGEE, the parent compounds giving rise to the metabolites MAA and EAA, are well known

developmental toxicants. This shows that a reported *in vitro* test result cannot be used as an absolute indication of toxicological (developmental) effects, even if there may be historical data available to which this value can be compared. In the historical data set available, no glycol ethers or short chain carboxylic acids are present. It can (retrospectively) be argued that the test results for glycol ethers/alkoxy acetic acids can not be compared to the results for the substances that are used in the validation of the EST assay. The EST assay as described in the INVITTOX protocol is still subject to intense discussion [Marx-Stoelting, 2009]. The additional value of including the cytotoxicity of 3T3 cells is heavily debated, the extent of the validation is being discussed, and there are uncertainties about how to perform the assay, and about its predictivity and applicability domain.

Level 4b: In vitro EST results in relation to available in vivo data for other category members

Although it is difficult to use *in vitro* data as stand-alone to draw a conclusion on the developmental toxicity of a chemical, it might be possible to use results from the EST assay in combination with other information to give a comprehensive picture or a trend which can be used to support a decision.

Table 7 combines information obtained *in vivo* for the straight alkyl side chain monoethylene glycol ethers EGME, EGEE and EGBE with *in vitro* results obtained for the metabolites MAA, EAA and BAA. These data were extracted from de Jong et al. (2009). The SIAM conclusions on developmental toxicity of these substances have also been added to the table.

Table 7: Summary of *in vitro* and *in vivo* information

Substance	**BMC_d50 (mM) Lab 1**	**BMC_d50 (mM) Lab 2**	**BMD_{10}* fetal weight (mmol/kg bw/day)**	**BMD_{10}* malformations (mmol/kg bw/day)**	**Excretion half-life in vivo (h)**	**SIAM conclusion on developmental toxicity**
EGME/MAA	2.3	2.5	1.3	0.8	14-18.6	Positive
EGEE/EAA	2.9	3.9	4.9	13.9	7.6-10.1	Positive
EGBE/BAA	4.5	5.9	12.0	14.8	1.5-3.2	Negative

* BMD_{10}: The benchmark dose causing 10% of the effect of interest.

Table 7 shows that there is a positive correlation between the length of the side chain and the outcome of the *in vitro* tests. Similarly, there is a positive trend between the length of the side chain and the BMD_{10} for fetal weight and malformation. In contrast, there is a negative correlation between the length of the side chain and the excretion half-life of the monoethylene glycol ethers. These trends are illustrated in Figure 1.

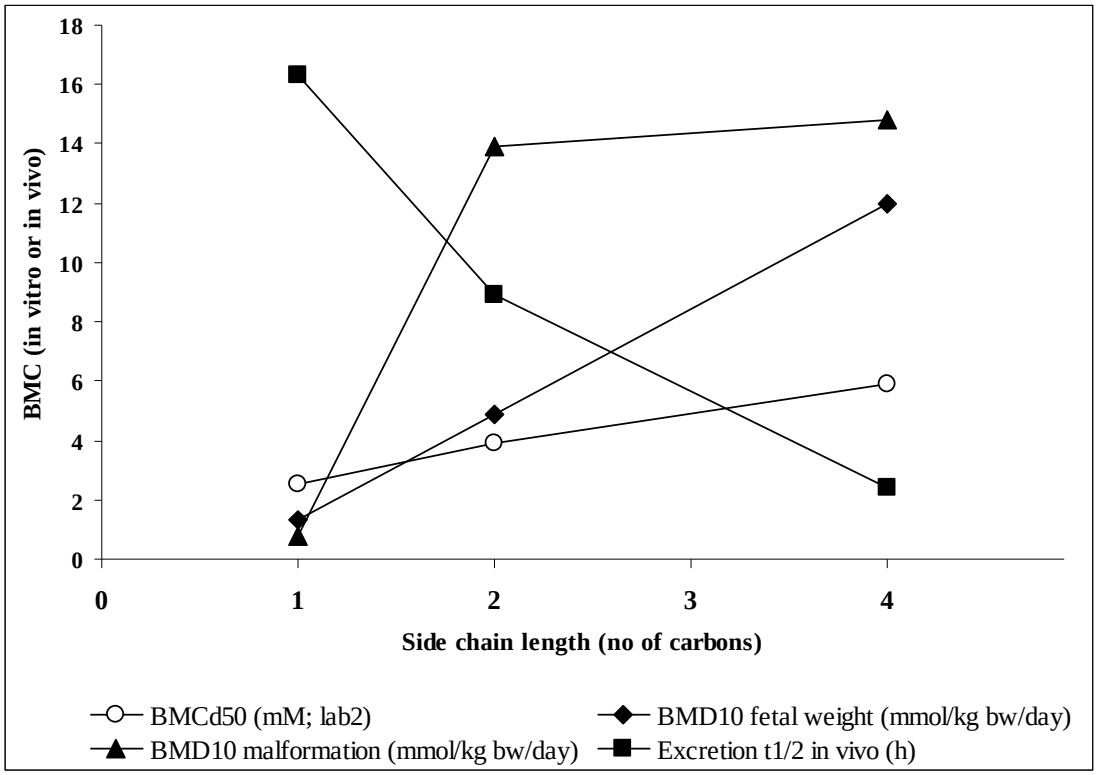


Figure 1: Correlation between *in vitro* results, *in vivo* results and side chain lengths

Although based on a limited number of data points, the results from the EST match well with the current knowledge about the mechanism of developmental toxicity of the monoethylene glycol ethers. In other words, the positive correlation between the results from the *in vitro* assay and the BMD_{10} for fetal weight and malformation shows that the monoethylene glycol ethers/alkoxyacetic acid metabolites that are developmental toxic *in vivo* are also more active in the *in vitro* assay.

Inclusion of EGPeE as well ass EGHepE into the category could be considered based on results obtained in the EST. In order for EGPeE and EGHepE to be included, the EST results would need to fit the general trend seen in Figure 1. In other words, the ID50 or BMD_d50 obtained for EGPeE and EGHepE in the EST assay would have to be similar (plateau) or higher than that of BAA. If this general trend does not hold true, inclusion of EGPeE and EGHepE would not be recommended based on EST results.

> Would *in vitro* EST results at or above effect levels seen for EGBE be sufficient to add EGPeE to the OECD category of monoethylene glycol ethers?

> Would the same also be sufficient for EGHepE?

Even if the ID50 or BMD_d50 for a chemical fits the trend discussed above which may suggest that there is no cause for concern for that particular chemical, it is possible that the chemical does act through a mechanism different from that measured in the assay. The EST assay, just like most other *in vitro* assays, only covers one mechanism of action. A negative response in an assay is therefore difficult to interpret. It can mean that the chemical does not cause the effect being measured (here, inhibition of differentiation of cardiomyocytes); the chemical is a true negative. Alternatively, it can mean that the chemical acts through a mechanism that is different from that being measured in the assay; a negative response would then be a false negative. Furthermore additional information, most notably on kinetics and metabolism (level 1 and 2 information) is considered essential for reaching a conclusion based on *in vitro* test results.

It is also of importance to note that although a case can be made for expanding category boundaries to include higher chain monoethylene glycol ethers based on EST results, a similar expansion to the lower chain ethers is not recommended. In fact, if ethylene glycol propyl ether (EGPE) had not already been included in the OECD category of monoethylene glycol ethers, it is uncertain whether a recommendation could have been made for including EGPE based on EST results. This is because EGPE is a borderline case; it is known that the one carbon smaller EGEE is teratogenic whereas the one carbon larger EGBE is not teratogenic. However, whether EGPE would give rise to the same *in vivo* effects as EGEE or EGBE would be difficult to predict based on results from the EST alone.

Level 4c: Additional in vitro assays

An *in vitro* assay is often limited to measuring one aspect of the toxicity or one mechanism of action. Therefore, multiple *in vitro* assays can be used to assess whether a chemical acts through another or more than one mechanisms of action. Assays that could additionally be used are for example the micromass test (INVITTOX protocol no. 122), embryotoxicity testing in post-transplantation embryo cultures (INVITTOX protocol no. 123), or the zebrafish embryo toxicity test.

In order to be of value in drawing a conclusion on the hazard of interest, it is of crucial importance to identify the mechanism that the assay is measuring and the limitations of the assay. It can be debated how many *in vitro* assays are necessary to cover all the complexity of real life embryo development and all the processes that are involved.

> **Would results from additional *in vitro* assays be sufficient to include EGPeE in the OECD category of monoethylene glycol ethers?**
> **Would the same also be sufficient for EGHepE?**

Level 5: In vivo testing

If the confidence in the information discussed in the previous levels is not considered sufficient, a traditional *in vivo* study in a rodent or a rabbit might be considered. These assays are currently accepted as producing information that can with sufficient confidence be used to conclude on the developmental toxicity of a chemical. *In vivo* testing using accepted guidelines will give results that will be accepted in regulatory frameworks. However, one might consider adapted or limited testing as an option that would yield sufficient reliability in combination with the information already obtained in the previous steps.

It should always be kept in mind that the traditional *in vivo* assays are also models for developmental toxicity in humans. Interspecies extrapolations always introduce a level of uncertainty to the results. Furthermore, because of species differences, not all test animals are equally suitable as a model for humans. Is it then necessary to test all chemicals in more than one species in order to have sufficient confidence that we have beyond doubt identified all true developmental toxicants?

Conclusions

In this case study, the different levels of information that are available for a chemical have been discussed to see which level is useful and deemed sufficient to justify the extension of (existing) categories. More specifically, we have explored interpolation and extrapolation for the monoethylene glycol ether category by considering the addition of EGPeE and EGHepE on the basis of structure, physico-chemical information, non-testing data such as QSARs, in vitro data and in vivo results.

A good case can be built for the inclusion of EGPeE (interpolation) based on structural information alone. Although this information requirement may seem minimal, this decision is not made in 'isolation'. Rather, in order to have confidence in the decision that information on structure only is sufficient, it is necessary to consider all available information on the other category members. In this case, a fair amount of information is available, including a generally accepted hypothesis that the alkoxy acetic acid metabolites of the monoethylene glycol ethers cause developmental toxicity. Having this information available allows us to consider the inclusion of EGPeE based on structure alone. Additional confidence in the decision would be gained by information about the kinetics and metabolism of EGPeE. For EGPeE, information obtained at the other levels, such as with QSARs and *in vitro* results, is unlikely to change the conclusions that were based on structural information alone.

In contrast, for the addition of EGHepE (extrapolation), structural information alone is not sufficient, even when all available information on the category members is taken into consideration. Additional information levels will need to be considered:

- QSAR predictions using TopKat suggest that EGHepE is not a developmental toxicant. However, we know that TopKat does not have any long-chain monoethylene glycol ethers in its training set.
- Information on the metabolism and kinetics of EGHepE would be necessary in order to assess whether it displays similar *in vivo* behavior as EGHE. It is possible that such a study reveals the formation of other (major) metabolites. This might trigger additional testing as the mechanism of action of EGHepE and/or its adverse effects might be different than for EGHE.
- In vitro data might be considered. However, as discussed above, many in vitro assays only assess one mechanism of action or do not take metabolism into consideration. Therefore, it is likely that a battery of *in vitro* tests might be required to assess EGHepE.
- Or an *in vivo* study might be carried because none of the additional levels will give us necessary confidence in the results to draw a conclusion.

Taken together, for the inclusion of EGHepE in the monoethylene glycol ether category, information on structure, metabolism and kinetics in addition to QSAR predictions, in combination with what is already known about the category could be considered sufficient.

An additional question that needs to be considered is how far a category can and should be extended? If we accept including EGHepE on the basis of e.g. QSAR predictions, is it then possible to include the octyl, decyl, hexadecyl etc. ethers using the same arguments? Which scientific information is necessary in order to make that decision? This issue needs further discussion.

We are not able, at this moment in time, to create generic rules or criteria on what information is sufficient or necessary in order to add a chemical to a category or extend the boundaries of an already existing category. None of the information at each individual level may be sufficient to make a conclusion; however, considering all the pieces together in a weight of evidence approach may give sufficient confidence for a decision to be made. An additional complicating factor is that what information is necessary for a chemical to be added also depends on what information is available for the category as such: adding chemicals to a data-rich category or categories that are well understood may require only a limited amount of targeted information whereas adding chemicals to data-poor categories may require more data.

Which level of information is sufficient to draw a conclusion on the hazard of a chemical or the expansion of current categories has been and will continue to be a subject to discussion. This continued discussion is necessary because new information is constantly being generated, new techniques are being developed and new regulatory or societal requirements are being implemented. However, what can be concluded is that any decision made should be transparently justified and well documented. Such justification and documentation will increase consistency of the decision making process and stimulate the necessary information exchange.

List of abbreviations

BAA	Butoxyacetic acid (metabolite of EGBE)
BMC_d50	The concentration corresponding to a 50% decline in the fraction of beating embryo bodies in comparison with solvent control
BMC_v50	The concentration corresponding to a 50% decline in cell viability
BMD_{10}	The benchmark dose causing 10% of a specified effect
EAA	Ethoxyacetic acid (metabolite of EGEE)
EGBEA	Ethylene glycol butyl ether acetate
EGBE	Ethylene glycol butyl ether
EGEE	Ethylene glycol ethyl ether
EGME	Ethylene glycol methyl ether
EGHE	Ethylene glycol hexyl ether
EGHeE	Ethylene glycol heptyl ether
EGPE	Ethylene glycol propyl ether
EGPeE	Ethylene glycol pentyl ether
ES cells	Embryonic stem cells
EST	Embryonic stem cell test
IC50	Concentration causing 50% inhibition of growth or cytotoxicity
ID50	Concentration causing 50% inhibition of differentiation
MAA	Methoxyacetic acid (metabolite of EGME)
REACH	Regulation (EC) No 1907/2006 on the Registration, evaluation, authorisation and restriction of chemicals.

Disclaimer

The views presented in this document do not necessarily represent the official Dutch opinion.

References

De Jong, E et al. (2009) Relative developmental toxicity of glycol ether alkoxy acid metabolites in the embryonic stem cell test as compared with the *in vivo* potency of their parent compounds. Tox Sci 110:117-124.

Marx-Stoelting P et al. (2009) A review of the implementation of the embryonic stem cell test (EST). ATLA 37, 313–328.

Louisse J, Y Bai, M Verweij, JJM van de Sandt, BJ Blaauboer, IMCM Rietjens (2010) Decrease of intracellular pH as possible mechanism of embryotoxicity of glycol ether alkoxyacetic acid metabolites. To appear in: Toxicology and Applied Pharmacology.

ECETOC (2005) Technical report no. 95. The Toxicology of Glycol Ethers and its Relevance to Man (Fourth Edition) Volume II - Substance Profile.

ECVAM. Protocol INVITTOX no. 113. http://ecvam.jrc.it/

Rorije E, Hulzebos E, Hakkert B (2007) The EU (Q)SAR Experience Project: reporting formats. Templates for documenting (Q)SAR results under REACH. RIVM rapport 601779001, www.rivm.nl/bibliotheek/rapporten/601779001.pdf

Appendix I: TopKat Documentation general & Developmental Toxicity Prediction

ADMET >

Theory - Toxicity Prediction (TOPKAT)

Overview

TOPKAT accurately and rapidly assesses the toxicity of chemicals solely from their 2D molecular structure. TOPKAT uses a range of robust, cross-validated Quantitative Structure-Toxicity Relationship (QSTR) models for assessing specific toxicological endpoints.

Uses of TOPKAT

The computational toxicology in TOPKAT is useful in:

- Rapidly assessing a broad range of toxicity of an organic compound solely from its 2D molecular structure
- Examining structure-toxicity relationships as a function of substructure and potential changes in structure and relating these to mechanism of action
- Ranking compounds for experimental testing or further development
- Designing new molecules

TOPKAT assessments

TOPKAT applies statistically robust, cross-validated and rigorously developed QSTR models to predict specific toxicological effects solely from chemical structure. A range of QSTR models are available.

TOPKAT is characterized by verified databases, information-rich descriptors, highly predictive QSAR-based models and prediction-validation techniques which permit users to determine the applicability of the model to the compounds being assessed.

TOPKAT also features a patented algorithm (US Patent 6, 036,349, issued March 14, 2000), which determines whether a query structure lies within the Optimum Prediction Space (OPS) of a respective model, and a set of hypothesis testing tools for determining the acceptability of an assessment.

TOPKAT models

TOPKAT generates assessments of various toxic effects of chemicals based on their molecular structure.

TOPKAT utilizes quantitative structure toxicity relationship (QSTR) models. TOPKAT currently supports assessment of:

- Developmental Toxicity Potential (DTP)
- Mutagenicity (Ames test)
- Rodent Carcinogenicity -- can be assessed as a sex and species specific endpoint based on either the NTP dataset, the FDA dataset, or as an overall Weight of Evidence score based on a combination of both data sets.
- Rat Chronic Oral LOAEL
- Skin Sensitization (GPMT)
- Skin Irritancy
- Rat Oral LD50
- Maximum Tolerated Dosage
- Fathead Minnow LC50
- Daphnia magna EC50
- VlogP
- Ocular Irritation
- Inhalational LC50
- Aerobic Biodegradability

QSTR models

The concept of quantitative structure toxicity relationship (QSTR) is based on the axiom that molecular properties (e.g., toxicity) can be explained by the information contained in the molecular structure. To develop robust, predictive QSTR models, scientists carefully examined the experimental data gathered from the open literature and various other sources and extracted only those data whose toxicity values have been generated under uniform conditions. This time-consuming process helps ensure that the resulting toxicity assessments are meaningful. Structural information in TOPKAT models is quantified by using descriptors which account for the complexities of the biochemical interactions in a non-mechanistic manner.

If we consider the biological aspects of the toxic response, we can rationalize that the response is a function of mainly two terms:

1. The ability of the molecule to reach a site.
2. The ability of the molecule to chemically interact with the biological system of the site to produce a toxic response.

TOPKAT uses descriptors that quantify the properties related to the transport of a chemical (e.g., molecular bulk, shape, symmetry), as well as descriptors that quantify the chemistry. For the latter, we use the information-rich electrotopological state (E-State) values developed by Kier and Hall to quantify the electronic attributes of molecular structure to account for interaction at the

site.

Applicability of the assessment

A model, whether a set of rules, or an equation, is a representation of a closed system; it cannot be relied upon to describe a point which is outside its prediction space. No model is universally applicable. This means that TOPKAT can generate a toxicity value for any structure for which you provide; but the computed value may not be meaningful unless the model is applicable to the query structure. TOPKAT automatically conducts a unique 3-stage analysis of the query structure to help you determine if the model applies to the structure.

Coverage of the training set

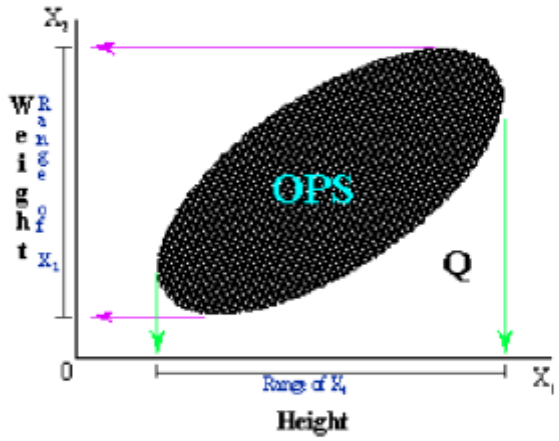
The first stage of this analysis warns you if a query molecule contains a substructure which was not considered during the model development process. TOPKAT performs this by comparing all 1- and 2-atom fragments in the query structure with the list of fragments from the training set of the model. Should the query structure contain an uncovered fragment (i.e., a fragment that is in the query structure but not in the training set), it will caution you as to the acceptability of the assessment.

OPS analysis

Because the model descriptor space is multivariate, a simple univariate examination of a query structure is not sufficient to determine the acceptability of the assessment. The third stage in the query analysis process determines if the query structure is within the Optimum Prediction Space (OPS) of a model. The OPS is unique multivariate descriptor space in which the model is applicable. In TOPKAT assessment of a chemical structure inside a models OPS may be accepted with confidence, subject to the results obtained from hypothesis testing.

For example below is depicted a number of observations for a 2-variable equation.

Figure 1.OPS region for weight vs height



These observations represent height and weight of a certain population of people, and we want to predict an anthropometric index from the data. Person Q's height is well within the range of the height represented by the data, and Q's weight is also within the range of weight represented by the data. But the JOINT combination of height and weight is not within the data space of the model. Subsequently, the model is not applicable to this query (i.e., the query is outside the OPS).

What do we mean by not applicable? Certainly any model may be applied to any query structure of interest, and TOPKAT will provide a numerical answer. When a query is within the OPS for a given model, that means that the probability of the assessment value being correct or accurate is as good as, but not better than, the cross-validated statistical performance of the model. Thus, if the statistics for the Female Rat submodel of the NTP Carcinogenicity model are 93 percent specificity and 91 percent sensitivity (which correspond to getting positives correct and negatives correct, respectively), then a compound assessed as a carcinogen that is within OPS has a 93 percent chance of being correctly assessed.

Such a statement is not possible about a compound outside of OPS. Simply because a query structure is outside of OPS does not mean the TOPKAT assessed value is incorrect; in fact, the value may be extremely accurate or correct. However, *there is no way of knowing how correct or how large the error bars are on an assessment outside of OPS.* Thus, when we say that a model is not applicable to a compound outside of OPS, we mean that the quality of the results are unknowable.

As an example, we trained the VlogP model on 6675 compounds to a coefficient of determination, R^2 = 0.986 and a standard error of estimate of 0.20. When applied to a test set (of compounds not used in training) of 113 compounds, 84 were inside of OPS while 29 were not. For those 84 compounds inside of OPS, the average deviation was 0.272, the maximum deviation was 1.34, and 95 percent of the chemicals were predicted within 0.71 of their experimental values. However, for those 29 compounds outside of

OPS, the average deviation was 1.38, the maximum deviation was 6.77, and 95 percent of the chemicals were predicted within 3.65 log units of their experimental values (Gombar and Enslein, 1999).

Hypothesis testing

A toxicity assessment can be thought of as a hypothesis which states that the model descriptors present in the query structure are the determinants of its toxicity. In this case, it is necessary to confirm the validity of the hypothesis.

The analysis of the applicability of the assessment, discussed in the previous section and which TOPKAT performs automatically, is necessary but not sufficient. This is because the performance of any given QSTR model is only as good as its cross-validated statistical accuracy (typically 85-95% for TOPKAT). In other words, no matter how good the model, sometimes it does not predict the right answer.

TOPKAT provides a means of validating the toxicity assessment through similarity searching of the models database. The underlying assumption is that if the model performs an accurate assessment on a compound from the data base, and if that compound is similar to the query structure, then the model should also perform an accurate and valid assessment of the query structure because the model is accurately predictive in that region of space within OPS.

Similarity is an interesting concept. Most of us look at two chemical structures as two-dimensional graphs, which may contain some stereochemical and three-dimensional information, and recognize patterns and substructures that are alike. In this case, we might consider the structures "similar". But what constitutes similarity?

Let's consider a non-chemical example of a red apple, a baseball, a banana, and a red truck. In terms of shape and size and weight, the apple is most similar to the baseball. However, if our purpose was to choose the most edible of the objects based on similarity, we would hopefully not choose the baseball but rather the banana, which is made up of similar chemical constituents as the apple despite its dissimilarity in color and shape. On the other hand, if we were trying to tell the automobile salesman what color of truck we want to purchase, the apple is similar to the red truck and would certainly help validate what we meant when we said red.

Thus, it is clear that similarity is meaningful only in a given context. For computational toxicology, that context is the specific toxic endpoint or model that is under consideration; e.g., Ames mutagenicity, and not across various toxicities. An example is the comparison of benzene and chlorobenzene shown in the table below.

	Benzene	Chlorobenzene
Ames Mutagenicity	Negative (Ashby and Tennant, 1988)	Negative (Ashby and Tennant, 1988)
Carcinogenicity in Female Mouse	Positive (NTP, 1986)	Negative (NTP, 1985)

Furthermore, within the TOPKAT approach to QSTR, structural similarity based on common functional groups, substructures, or patterns of atoms or bonds is not appropriate, since none of these are used in determining the QSTR. Remember, the structure-toxicity relationship is developed from a set of 1- and 2-atom E-State descriptors, shape and symmetry indices, and transport-related descriptors such as molecular weight and VlogP. Thus, similarity between two compounds must be similarity of descriptors and their values (i.e., similarity in descriptor space, not chemical structure space).

Applications

TOPKAT provides researchers with the capability to rapidly and confidently evaluate the toxic effects of chemicals directly from their molecular structure. Scientists can readily analyze the effects of modifying substituents on a parent compound, assess toxicity of the metabolytes, and compounds as yet unsynthesized, and generate toxicity profiles on large sets of compounds.

Utilizing TOPKAT to conduct computational toxicology experiments can play a major role in decreasing time to market, reducing animal experiments, assessing human health risks, and strategic planning of pharmaceutical and chemical development processes.

TOPKAT technology is currently used to:

- Optimize therapeutic ratios of lead compounds
- Prioritize promising compounds for further development/investment
- Evaluate intermediates, metabolites and pollutants
- Screen compounds generated via HTS systems
- Assess pharmaceutical, commercial, industrial and agricultural chemical products for potential safety problems
- Set dose-ranges for animal assays

Evaluating an assessment

If you consider a TOPKAT assessment of a query structure as a hypothesis that states that the model parameters present in the query structure are the determinants of its toxicity, then this hypothesis can be tested against similar compounds in the model's database. The Similarity Search function in TOPKAT will automatically rank all the compounds in the respective model database based on their QSTR similarity to the query structure.

The following information is available for each compound:

- The actual experimental result
- The TOPKAT predicted result
- Whether the compound was used in the training set
- The similarity distance from the query on a scale of 0.0 - 1.0

Note: the smaller the distance, the greater the similarity.

With this information you can determine:

- Whether the query structure lies in an information-rich region of the model data space
- If similar compounds are well predicted by the model

If you find similar compounds whose experimental and computed values are in concordance with the computed value of the query structure, the assessment is acceptable. Conversely, if you find evidence in the model database that the query structure lies in a region of OPS where model performance is poor, then that assessment is unacceptable.

How toxicity is computed by TOPKAT

TOPKAT computes a probable value of toxicity for a submitted chemical structure from a Quantitative Structure-Toxicity Relationship (QSTR) equation.

The equation is linear in the structure descriptors. The coefficients are optimized during the development of the equation.

The product of a structure descriptors value and its corresponding coefficient is the descriptors contribution to the probable toxicity. Contributions from the products may be either positive or negative; a positive contribution will increase the probability of the chosen property, whereas a negative contribution will decrease it.

Toxicity values are computed by summing the individual contributions. For assessing toxicity values such as LD50 or LC50, this sum is transformed into a weight/weight unit (mg/kg) or a weight/volume unit (mg/l); for 2-group classifications, such as carcinogens/non-carcinogens, this sum is transformed into a probability value between 0.0 and 1.0.

Probability values

Probability values from 0.0 to 0.30 are considered low probabilities, and chemicals with TOPKAT-computed probability values in this range are not likely to produce a positive response in an experimental assay; whereas probability values greater than 0.70 are considered high, and are likely to produce a positive response in an experimental assay. Probabilities greater than 0.30 but less than 0.70 are considered indeterminate (i.e., too near chance (0.50) for an assessment to be meaningful).

Query Structure Examination

TOPKAT always outputs a value of toxicity; however, whether the assessment is meaningful or not can only be answered by:

- A univariate analysis or Coverage Examination, that is, whether all of the structural fragments of the query structure are well represented in the database compounds which were used to develop the model (training set).
- A multivariate analysis, or OPS Examination, that is, whether the submitted structure fits within, or near the periphery of,

the Optimum Prediction Space (OPS) of the equation. These 2 steps are accomplished automatically in TOPKAT and results are output in terms of a confidence percentage.

Coverage Examination

Every QSTR model is associated with a certain training set of compounds, and these compounds contain a limited set of structural attributes.

A QSTR model, when extrapolated to chemical structures containing structural attributes which are not represented in the training set, may produce unreliable toxicity assessments. Therefore, it is important to determine whether the structural attributes of the query molecule are represented in the compounds used for the development of a QSTR. TOPKAT automatically determines whether the input structure contains molecular substructures which are foreign to the training set (a univariate analysis).

Additionally during this process, TOPKAT compares the values of the model descriptors for the query structure to the range of the values of the respective descriptors in the training set compounds.

Optimum Prediction Space

As well as determining its coverage, TOPKAT checks whether a query structure is located inside or outside the Optimum Prediction Space (OPS) of a QSTR (multivariate analysis).

The OPS of a QSTR is a multi-dimensional space, the number of dimensions being one more than the number of model parameters of the QSTR. An important characteristic of the OPS is that within and near its periphery the QSTR may be applied with confidence.

The OPS confidence contains information about both the Optimum Prediction Space, and the fragment coverage.

When a query structure is determined to be inside all dimensions of a models OPS, the computed value of toxicity can be considered acceptable (unless evidence exists to refute the assessment).

However, if a query structure is found outside one or more dimensions, the computed toxicity may or may not be acceptable depending on the querys distance from OPS.

The distance of a query structure from the OPS is a complex function of the querys location in each dimension. Every TOPKAT QSTR model has a permissible limit of distance from the OPS. If the query structures distance from the OPS is greater than this permissible limit, the TOPKAT-assigned toxicity value is considered unacceptable. The permissible limits of distance from the OPS for all QSTR models have been precalculated and stored in TOPKAT. For every query structure outside the OPS, TOPKAT reports the location of a query structure with respect to the permissible limit of distance from the OPS.

Developmental Toxicity Potential (DTP)

The Developmental Toxicity Potential (DTP) Module of the TOPKAT package comprises three statistically significant and cross-validated quantitative structure-toxicity relationship (QSTR) models, and the data from which these models are derived. Each model applies to a specific class of chemicals. Molecular structure is the only input required to conduct a Developmental Toxicity Potential assessment. These discriminant models, derived from uniform experimental studies selected after critical review of approximately 3,000 open literature citations, compute the probability of a submitted chemical structure being a developmental toxicant in the rat; a probability below 0.3 indicates no potential for developmental toxicity (NEG), and probability above 0.7 signifies developmental toxicity potential (POS). The probability range between 0.3 and 0.7 refers to the "indeterminate" zone (IND).

This Module is designed for operation with the TOPKAT interface, which (i) automatically determines whether the submitted structure belongs to the Optimum Prediction Space (OPS) of the model, and (ii) computes QSTR similarity distance from chemicals with experimental DTP data in order to evaluate the reliability of the QSTR-based assessment.

Following is the cross-validated (leave-one-out) accuracy of the three models:

Submodel Class	Number of Cmpds	Specificity (%)	Sensitivity (%)	Indeterminate (%)
Aliphatics	87	88.6	88.6	2.5
Carboaromatics	95	97.4	87.0	2.2
Heteroaromatics	91	86.0	86.1	5.5

The TOPKAT DTP model was developed from 374 open-literature references. The model includes only rat oral data. Two types of studies were removed from the database prior to further evaluation: single-dose studies in which developmental as well as maternal toxicity were observed at that dose, and studies in which neither developmental nor maternal toxicity was observed at the highest dose.

For the remaining studies, the following scoring scheme was adopted:

- Score 1: No developmental toxicity (DT) even at maternotoxic (MT) doses
- Score 2: Strict concordance between DT and MT, that is, no DT nor MT at one dose, and both DT and MT at a higher dose
- Score 3: DT at the dose preceding MT

- Score 4: DT at least 2 doses below that which produced MT

Scores 2, 3, and 4 were combined into one group, so that the resultant model distinguishes between no evidence of DTP vs. any evidence of DTP. Thus a probability of 0.8, for example, indicates that there is good likelihood that were a DTP assay performed with the chemical, it would be found to be a developmental toxicant. But no severity of DTP can be assigned to that probability.

A number of signs and symptoms were taken as evidence of MT. These included weight loss, increased mortality, decreased feed intake, and various forms of distress, irritation, etc. DT was evidenced by reduced fetal growth, fetal death, resorption, and teratology both external and visceral.

The probability value is rounded to three-significant figures.

TOPKAT Detailed DevTox Prediction Report for Ethylene Glycol Pentyl Ether (EGPeE)

1 Summary

Molecular Formula: $C_7H_{16}O_2$
Molecular Weight: 132.20074
ALogP: 1.295
Rotatable Bonds: 6
Acceptors: 2
Donors: 1

Prediction

Model: Developmental Toxicity Potential (DTP) (v3.1)

Computed Probability of DTP = 0.011

1 Developmental Toxicity Potential (DTP) (v3.1)

Molecular Formula: $C_7H_{16}O_2$
Molecular Weight: 132.20074
ALogP: 1.295
Rotatable Bonds: 6
Acceptors: 2
Donors: 1

Prediction

SubModel: Developmental Toxicity Potential Aliphatic Model

Computed Probability of DTP = 0.011

Discriminant Score = -4.510

Probability values from 0.0 to 0.30 are considered low probabilities, and are likely to produce a negative response in an experimental assay; whereas probability values greater than 0.70 are considered high, and are likely to produce a positive response in an experimental assay. Probabilities greater than 0.30 but less than 0.70 are considered indeterminate.

Similar Compounds

	Compound 1	Compound 2	Compound 3	Compound 4	Compound 5
Molecule					
Actual Endpoint	NEG	NEG	POS	NEG	NEG
Predicted Endpt	NEG	NEG	NEG	NEG	NEG
Distance	0.027	0.109	0.115	0.145	0.177

OPS Summary

Within OPS: True
Within OPS Limits: True
All Fragments Covered: True
Compound in Database: False

Descriptor Contribution

Descriptor	Value
Symmetry Index #4	-8.293
Shape Index # 7	-6.481
[Aliphatic O]	6.268
Shape Index # 5	-4.955
CONSTANT TERM	4.922

TOPKAT Detailed DevTox Prediction Report for Ethylene Glycol Ethyl Ether (EGEE)

1 Summary

Molecular Formula: $C_4H_{10}O_2$
Molecular Weight: 90.121
ALogP: -0.141
Rotatable Bonds: 3
Acceptors: 2
Donors: 1

Prediction

Model: Developmental Toxicity Potential (DTP) (v3.1)

Computed Probability of DTP = 0.957

1 Developmental Toxicity Potential (DTP) (v3.1)

Molecular Formula: $C_4H_{10}O_2$
Molecular Weight: 90.121
ALogP: -0.141
Rotatable Bonds: 3
Acceptors: 2
Donors: 1

Prediction

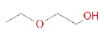
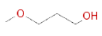
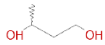
SubModel: Developmental Toxicity Potential Aliphatic Model

Computed Probability of DTP = 0.957

Discriminant Score = 3.101

Probability values from 0.0 to 0.30 are considered low probabilities, and are likely to produce a negative response in an experimental assay; whereas probability values greater than 0.70 are considered high, and are likely to produce a positive response in an experimental assay. Probabilities greater than 0.30 but less than 0.70 are considered indeterminate.

Similar Compounds

	Compound 1	Compound 2	Compound 3	Compound 4	Compound 5
Molecule					
Actual Endpoint	POS	NEG	POS	POS	POS
Predicted Endpt	POS	POS	POS	POS	POS
Distance	0.000	0.007	0.192	0.236	0.262

OPS Summary

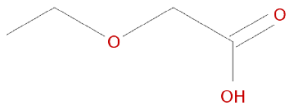
Within OPS: True
Within OPS Limits: True
All Fragments Covered: True
Compound in Database: True
Modeled Endpoint: POS

Descriptor Contribution

Descriptor	Value
Symmetry Index #4	-8.384
[Aliphatic O]	6.020
CONSTANT TERM	4.922
[*CH2*] * [Aliphatic O]	3.846
Shape Index # 5	-3.303

TOPKAT Detailed DevTox Prediction Report for Ethoxy Acetic Acid (EAA)

1 Summary



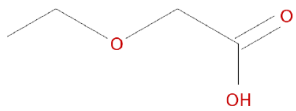
Molecular Formula: $C_4H_8O_3$
Molecular Weight: 104.10452
ALogP: -0.12
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Prediction

Model: Developmental Toxicity Potential (DTP) (v3.1)

Computed Probability of DTP = 0.987

1 Developmental Toxicity Potential (DTP) (v3.1)



Molecular Formula: $C_4H_8O_3$
Molecular Weight: 104.10452
ALogP: -0.12
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Prediction

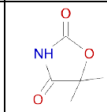
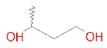
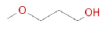
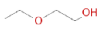
SubModel: Developmental Toxicity Potential Aliphatic Model

Computed Probability of DTP = 0.987

Discriminant Score = 4.344

Probability values from 0.0 to 0.30 are considered low probabilities, and are likely to produce a negative response in an experimental assay; whereas probability values greater than 0.70 are considered high, and are likely to produce a positive response in an experimental assay. Probabilities greater than 0.30 but less than 0.70 are considered indeterminate.

Similar Compounds

	Compound 1	Compound 2	Compound 3	Compound 4	Compound 5
Molecule					
Actual Endpoint	POS	POS	POS	NEG	POS
Predicted Endpt	POS	POS	POS	POS	POS
Distance	0.202	0.221	0.224	0.258	0.260

OPS Summary

Within OPS: True
Within OPS Limits: True
All Fragments Covered: True
Compound in Database: False

Descriptor Contribution

Descriptor	Value
[Aliphatic O]	10.365
Symmetry Index #4	-8.943
CONSTANT TERM	4.922
Shape Index # 5	-3.303
[*CH2*] * [Aliphatic O]	1.303

Appendix II: The Embryonic Stem Cell Test (EST)

The protocol for the test that is published on the ECVAM website (INVITTOX protocol no. 113) describes the EST as being based on the potential of embryonic stem (ES) cells from stable (permanent) mouse ES cell lines to remain in an undifferentiated stage in culture in the presence of a certain cytokine. When the cytokine is removed, the cells will differentiate into the major embryonic tissues under appropriate conditions. Furthermore, cytotoxicity data show that ES cells are more sensitive to toxic agents than adult cells. Therefore, in the EST the inhibition of differentiation is combined with the study of differences in sensitivity to cytotoxic damage between embryonic tissue (ES cells and adult tissues (mouse 3T3 fibroblasts). The three endpoints, inhibition of differentiation (ID50), and cytotoxicity (IC50) in ES cells and 3T3 cells are combined for the predicting the embryotoxic potential of chemicals.

Furthermore, INVITTOX protocol no. 113 states that the EST is applicable for differentiating embryotoxic chemicals into three groups: non-embryotoxic, weak/moderate embryotoxic and strong embryotoxic. The predictivity and precision for these three groups as obtained in the assay during the validation is given in Table 2. The overall accuracy was 78% for the 20 chemicals used in the validation.

Table A1: Predictivity and precision obtained for the EST assay

Embryotoxic potential	Predictivity (%)	Precision (%)
Non	72	70
Weak/moderate	70	83
Strong	100	81

Historical data taken from INVITTOX protocol no. 113 (http://ecvam.jrc.it/)

Table A2: The IC50 and ID50 for 16 chemicals in three teratogenic groups. The IC50/ID50 ES values that would lead to misclassification of compounds are given in red.

Test chemical	CAS no	MW	IC50 3T3 (mM)	IC50 ES (mM)	ID50 ES (mM)
Group 1: Non-teratogens					
Saccharin	82385-42-0	183.18	16.38	19.10	10.92
Penicillin G	69-57-8	334.40	4.74	8.82	10.32
Isoniazid	54-85-3	137.14	2.60	5.47	2.63
Ascorbic acid	134-03-2	176.12	0.14	0.78	2.32

Group 2: Weak	**teratogens**				
Aspirin	50-78-2	180.16	1.28	1.22	1.38
Caffein	58-08-2	194.19	0.80	0.85	0.95
Diphenhydramine	147-24-0	255.36	0.12	0.12	0.03
Diphenylhydantoin	630-93-3	252.69	0.14	0.11	0.08
Indomethacin	53-86-1	357.79	0.08	0.08	0.18
Dexamethasone	50-02-2	392.46	0.07	0.06	0.05
Methotrexate	59-05-2	454.44	0.00003	0.0002	0.00004
Group 3: Strong	**teratogens**				
Hydroxyurea	127-07-1	76.05	0.095	0.026	0.022
Busulphan	55-98-1	246.30	0.019	0.009	0.019
5-Fluorouracil	51-21-8	130.08	0.0013	0.0008	0.00022
Cytosine arabinoside	69-74-9	243.22	0.00014	0.00010	0.00012
Retinoic acid	302-79-4	300.44	0.00333	0.00002	0.0000003

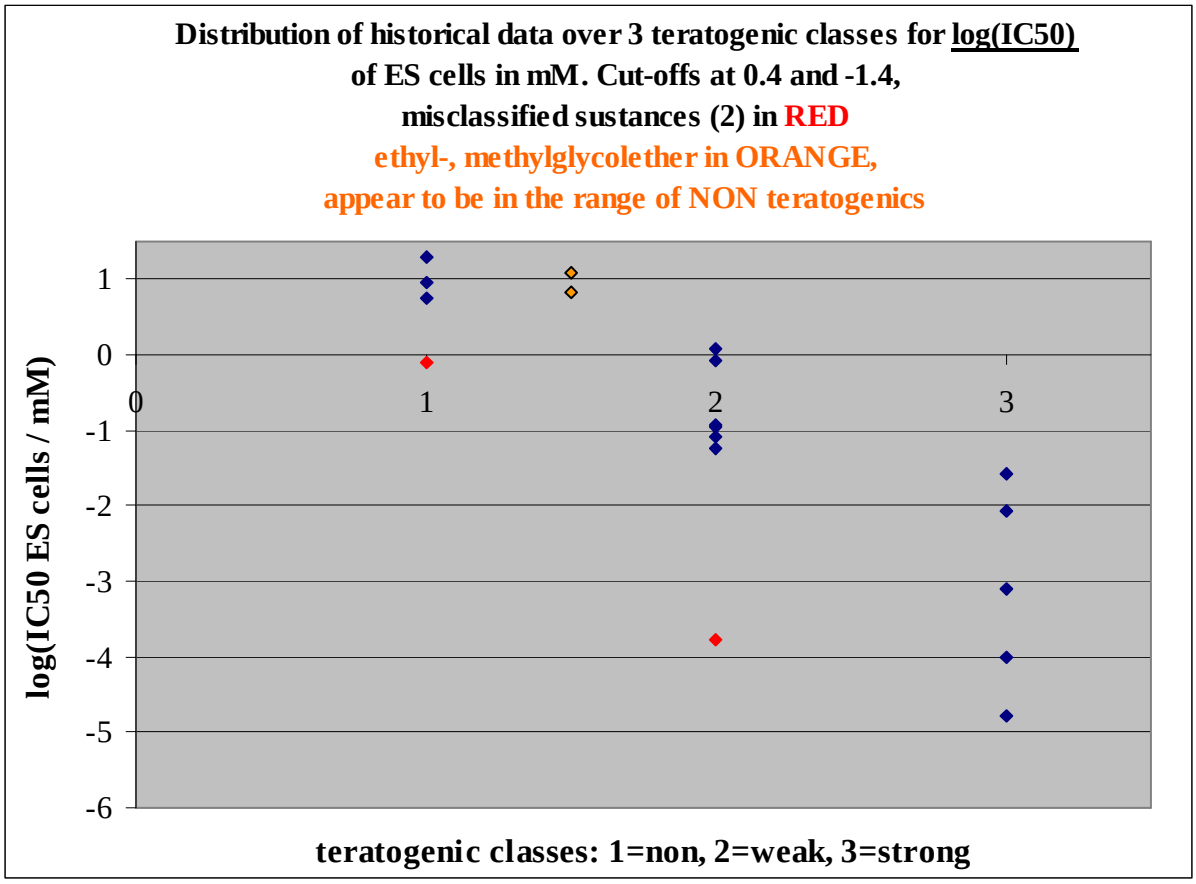


Figure A1: Distribution of historical IC50 value in ES cells, with the IC50 values of ethylglycol methyl ether and ethylglycol.ethyl ether given in orange. Their IC50 values seem to indicate that they belong to the class of non-teratogenics.

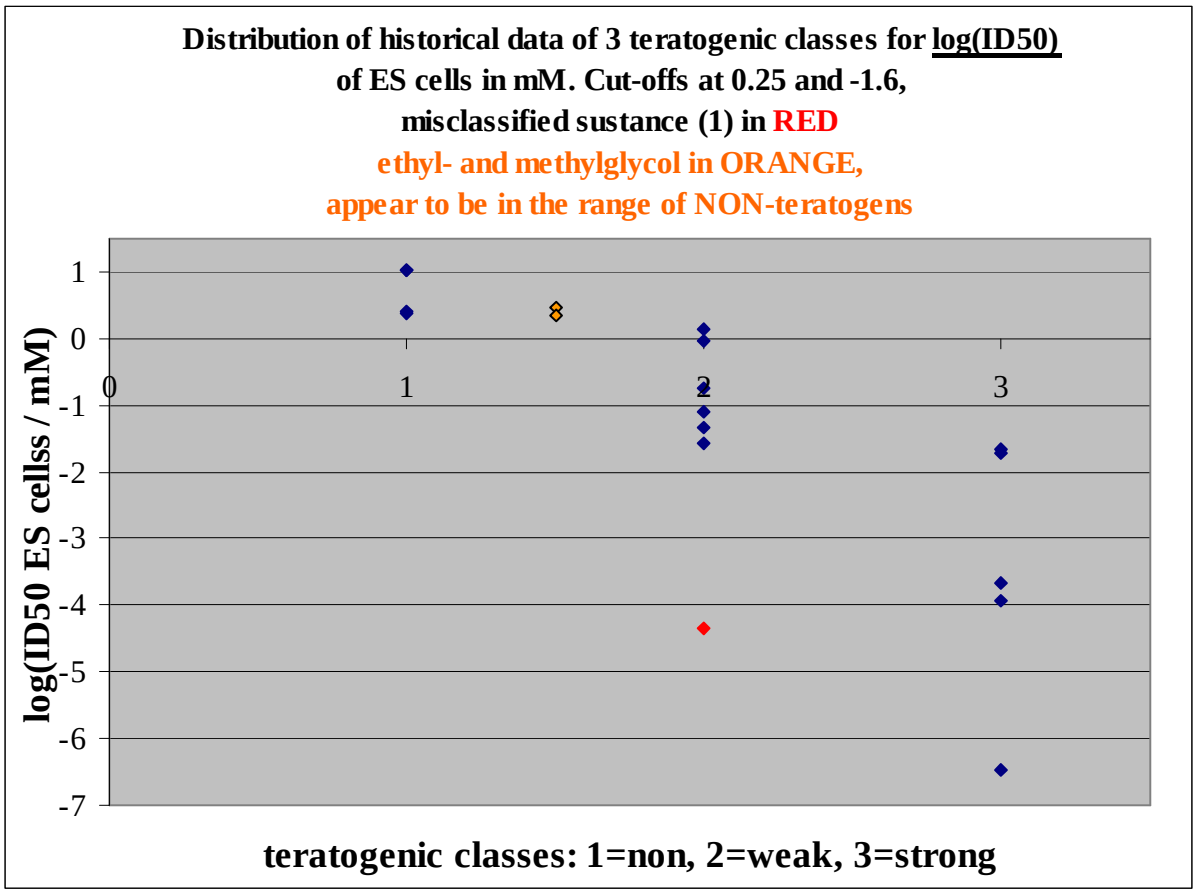


Figure A2: Distribution of historical IC50 value in ES cells, with the IC50 values of the metabolites of ethyleneglycol methyl ether and ethyleneglycol.ethyl ether given in orange. Their IC50 values seem to indicate that they belong to the class of non-teratogenics.