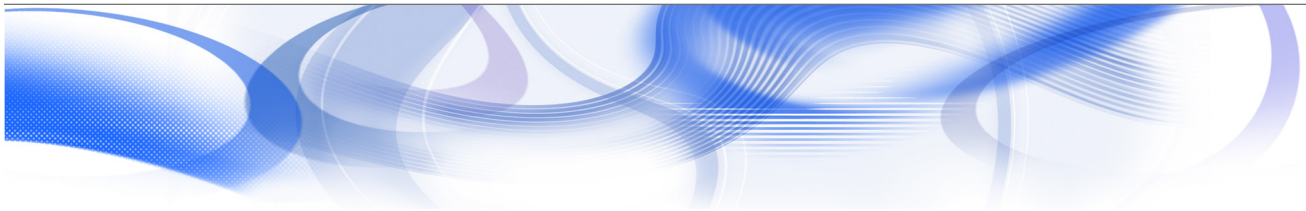
JRC Scientific and Technical Reports



The Use of Computational Methods in the Grouping and Assessment of Chemicals - Preliminary Investigations

Andrew Worth, Arianna Bassan, Evelin Fabjan, Ana Gallegos Saliner, Tatiana Netzeva, Grace Patlewicz, Manuela Pavan and Ivanka Tsakovska

EUR 22941 EN - 2007

JRC EUROPEAN COMMISSION

ihcp Institute for Health and Consumer Protection

The mission of the IHCP is to provide scientific support to the development and implementation of EU policies related to health and consumer protection.

The IHCP carries out research to improve the understanding of potential health risks posed by chemical, physical and biological agents from various sources to which consumers are exposed.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information
Address: European Chemicals Bureau TP 581
E-mail: andrew.worth@ec.europa.eu
Tel.: +39 0332 789566
Fax: +39 0332 786717

http:// http://ecb.jrc.it/QSAR
http://www.jrc.cec.eu.int

JRC 40287

EUR 22941 EN
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Quality control insert

	Name	*Signature*	*Date*
Report Prepared by:	Andrew Worth		17/09/07
Reviewed by: (Scientific level)	Grace Patlewicz		17/09/07
Approved by: (Head of Unit)	Steven Eisenreich		18/09/07
Final approval: (IHCP Director)	Elke Anklam		19/09/07

ABSTRACT

This document presents a perspective of how computational approaches could potentially be used in the grouping and assessment of chemicals, and especially in the application of read-across and the development of chemical categories. The perspective is based on experience gained by the authors during 2006 and 2007, when the Joint Research Centre's European Chemicals Bureau was directly involved in the drafting of technical guidance on the applicability of computational methods under REACH. Some of the experience gained and ideas developed resulted from a number of research-based case studies conducted in-house during 2006 and the first half of 2007. The case studies were performed to explore the possible applications of computational methods in the assessment of chemicals and to contribute to the development of technical guidance. Not all of the methods explored and ideas developed are explicitly included in the final guidance documentation for REACH. Many of the methods are novel, and are still being refined and assessed by the scientific community. At present, many of the methods have not been tried and tested in the regulatory context. The authors therefore hope that the perspective and case studies compiled in this document, whilst not intended to serve as guidance, will nevertheless provide an input to further research efforts aimed at developing computational methods, and at exploring their potential applicability in regulatory assessment of chemicals.

LIST OF ABBREVIATIONS

AAR	Activity-Activity Relationship
BfR	German Federal Institute for Risk Assessment
BMD	Benchmark Dose
BMD LCL	95% lower confidence limit on the Benchmark Dose
CAS	Chemical Abstracts Service
CEFIC (LRI)	European Chemical Industry Council (Long Range Initiative)
DNEL	Derived No Effect Level
EC	European Commission
ECB	European Chemicals Bureau
ECx%	Effective Concentration x%
EPA	Environmental Protection Agency
ESIS	European chemical Substances Information System (ECB)
ESR	Existing Substances Regulation
EINECS	European Inventory of New and Existing Chemical Substances
EU	European Union
HPV	High Production Volume
ITS	Integrated (Intelligent) Testing Strategy
JRC	Joint Research Centre
LCx%	Lethal concentration %
LDx%	Lethal dose %
LOAEL	Lowest-Observed-Adverse-Effect-Level
NOAEL	No-Observed-Adverse-Effect-Level
OECD	Organisation for Economic Cooperation and Development
PBT	Persistent Bioaccumulative and Toxic
vPvB	very Persistent and very Bioaccumulative
PCA	Principal Components Analysis
POR	Partial Order Ranking
QSAAR	Quantitative Structure-Activity-Activity Relationship
QMRF	(Q)SAR Model Reporting Format
(Q)SAR	(Quantitative) Structure Activity Relationship
REACH	Registration, Evaluation, and Authorisation of Chemicals
RIP	REACH Implementation Project
RTECS	Register of Toxicology Effects of Chemical Substances
SIAM	OECD Screening Information Assessment Meeting
SMILES	Simplified Molecular Input Line Entry System
TTC	Threshold of Toxicological Concern
TOR	Total Order Ranking
WoE	Weight of Evidence

CONTENTS

APPENDICES

1. Introduction

This document presents a perspective of how computational approaches could potentially be used in the development of chemical categories and in the application of read-across. It also contains a compilation of case studies that were developed by the ECB during 2006 and the first half of 2007. The case studies were performed to explore the possible applications of computational methods in the assessment of chemicals and to contribute to the development of technical guidance on the regulatory use of such methods, especially in the framework of REACH, which entered into force on 1 June 2007 (EC, 2006).

1.1 Basic concepts of non-testing methods

This section explains the various types of non-testing, and in particular *in silico*, methods that can be used to provide information about the basic physicochemical and fate properties of chemicals, as well as their ecological and human health effects. Computer-aided toxicity prediction makes use of the relationship between chemical structure and biological activity to compute such properties, thus generating non-testing data on the effects of the chemicals on humans and the environment. The different types of non-testing methods include qualitative and quantitative Structure Activity Relationship models (i.e. SARs and QSARs) (Cronin, 2004); Activity-Activity Relationships (AARs) and Quantitative Structure-Activity-Activity Relationships (QSAARs) (Lessigiarska *et al*, 2006); and expert systems (Dearden *et al*, 1997). Non-testing data can also be generated by less formalised chemical grouping approaches, referred to as the analogue and chemical category approaches. All of these non-testing methods are based on the premise that the properties (including biological activities) of the chemical depend on its intrinsic nature and can be directly predicted from its molecular structure and inferred from the properties of similar compounds whose activities are known.

1.1.1 (Q)SARs

A (Q)SAR is an umbrella term referring to both SARs and QSARs, and is often used to refer to any theoretical model that can be used to predict the physicochemical, biological (e.g., toxicological) and fate properties of molecules from knowledge of chemical structure.

More specifically, a SAR is a qualitative relationship (i.e. association) between a molecular (sub)structure and the presence or absence of a given biological activity, or the capacity to modulate

a biological activity imparted by another substructure. The term substructure refers to an atom, or group of adjacently connected atoms, in a molecule. A substructure associated with the presence of a biological activity is also called a structural alert. A SAR can also be based on the ensemble of steric and electronic features considered necessary to ensure the intermolecular interaction with a specific biological target molecule, which results in the manifestation of a specific biological effect. In this case, the SAR is sometimes called a 3D SAR or pharmacophore.

A Quantitative Structure-Activity Relationship (QSAR) is a quantitative relationship between a biological activity (e.g., toxicity) and one or more molecular descriptors that are used to predict the activity. A molecular descriptor is a structural or physicochemical property of a molecule, or part of a molecule, which specifies a particular characteristic of the molecule and is used as an independent variable in a QSAR.

Similar to a QSAR, a Quantitative Activity-Activity Relationship (QAAR) is a mathematical relationship between two biological endpoints, which can be in the same or different species. QAARs are based on the assumption that knowledge about the mechanism or mode of action, obtained for one endpoint, is applicable to the same endpoint in a different species, or to a similar endpoint in the same species, since the main underlying processes are the same (e.g. partitioning, reactivity, enzyme inhibition).

1.1.2 Expert systems

An expert system refers broadly to any formalised system, generally computer-based, which enables a user to obtain rational predictions about the properties or biological activity of chemicals. Expert systems may be classified as knowledge-based (when the rules are based on expert knowledge), induction rule-based (when statistical methods are used to automatically derive the rules) or hybrid (when both approaches are present). One or more databases may additionally be integrated in the system.

1.1.3 Read-across

Read-across is a non-formalised approach in which endpoint information for one chemical (called a "source chemical") is used to make a prediction of the endpoint for another chemical (called a "target chemical"), which is considered to be similar in some way (usually on the basis of structural similarity). In principle, read-across can be applied to characterise physicochemical properties, fate, human health effects and ecotoxicity, and it may be performed in a qualitative or quantitative

manner. Read-across can either be qualitative or quantitative, depending on the whether the data being used is categorical or numerical in nature.

To estimate the properties of a given substance, read-across can be performed in a one-to-one manner (one analogue used to make an estimation) or in a many-to-one manner (two or more analogues used). Within the context of a chemical category, the read-across can also be performed in a one-to-many manner or in a many-to many manner.

1.1.4 Chemical category

Chemical category formation is another non-formalised approach. The assessment of a group of chemicals as a category represents a departure from the traditional approach to property/hazard assessment in which chemicals are assessed on a substance-by-substance basis to an approach in which the category is assessed as a whole.

By definition, a chemical category is a group of chemicals whose physicochemical and human health and/or environmental toxicological 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).

Accordingly, a chemical category is selected based on the hypothesis that the properties of a series of chemicals with common (structural) features will show coherent trends in their physicochemical properties, and more importantly, in their toxicological (human health/ecotoxicity) effects or environmental fate properties. The presence of common behaviour or coherent trends is generally associated with a common underlying mechanism of action.

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

The trend analysis may involve the development of an "internal" model, i.e. a computational model such as a QSAR that is based entirely on the data in the category. This term is used in distinction to "external" model, which refers to a computational model developed using a different or more extensive dataset. In principle, a (Quantitative) Activity-Activity Relationship ([Q]AAR) could also provide a means of performing trend analysis and filling data gaps, although at present experience in the regulatory use of these models is limited.

Within a chemical category, data gaps can therefore be filled by using several approaches, namely: a) read-across; b) trend analysis and use of computational methods based on internal models; and c) use of computational methods based on external models. In this context, the term "model" refers to any formalised method for estimating the properties of chemicals, such as a (Q)SAR, a (Q)AAR or an expert system.

1.1.5 Analogue and category approaches

Ideally, there should be sufficient members in a chemical category to enable the detection of trends across endpoints. As the number of chemicals being grouped into a category increases, the potential for developing hypotheses and making generalisations about the trends will also increase. This increases the robustness of the evaluation. However, in the case of a limited number of analogues being identified, there might be insufficient data to establish a trend. In the REACH guidance documentation (ECB, 2007), grouping approaches reflecting these two extremes referred to as the "category approach" and the "analogue approach", respectively.

Thus, the analogue approach is used when the grouping is based on a very limited number of chemicals, such that trends in properties are not apparent. In such cases, data gaps can be filled by read-across. The category approach is used when there is a more extensive dataset that allows the detection of one or more trends. In such cases, data gaps can be filled by read-across but also by trend analysis and the development of an internal model. An external model could also be used to fill data gaps in a category, but this possibility remains irrespective of whether an analogue or category approach is being followed.

In this document, the term grouping is used in the broadest sense to refer to the formation of any group of (structurally) related chemicals, irrespective of size. The focus here is on the usefulness of computational methods in developing such groups. The key point is that result of this computer-based grouping can often provide support for applying the analogue approach (i.e. performing and

documenting a read-across assessment) or for applying the category approach (i.e. assessing the properties of a group of related chemicals in a chemical category).

1.2 The application of grouping approaches under REACH

Under REACH (EC, 2006), testing requirements for individual substances are largely tonnage-dependent and based on the specific information requirements shown in Annexes VI to X. As an alternative approach, Annex XI opens the possibility of evaluating chemicals not on a one-by-one basis, but in groups.

Annex XI contains the following wording for the use of grouping methods (read-across and chemical categories):

> "Substances whose physico-chemical, 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 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 Agency, after consulting with relevant stakeholders and other interested parties, shall issue guidance on technically and scientifically justified methodology for the grouping of substances sufficiently in advance of the first registration deadline for phase-in substances.
>
> The similarities may be based on:
>
> (1) a common functional group,
> (2) the common precursors and/or the likelihood of common breakdown products via physical and biological processes, which result in structurally similar chemicals, or
> (3) a constant pattern in the changing of the potency of the properties across the category.
>
> If the group concept is applied, 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 referred to in Article 13(3)
> - cover an exposure duration comparable to or longer than the corresponding test method referred to in Article 13(3) if exposure duration is a relevant parameter, and
> - adequate and reliable documentation of the applied method shall be provided."

1.3 Technical guidance on the analogue and category approaches

Guidance on the application of the analogue and category approaches for the assessment of chemicals under REACH has been developed in the context of one of the REACH Implementation Projects (RIPs), namely RIP 3.3 (ECB, 2007). With the exception of the REACH-specific considerations, essentially the same guidance has also been adopted by the Organisation for Economic Cooperation and Development (OECD) so there is harmonisation between the guidance at the EU and international levels (OECD, 2007b). In addition, a compilation of case studies that are referenced in the REACH (and OECD) guidance is available from the ECB website (Worth & Patlewicz, 2007). This includes a summary of how grouping approaches have been used in the EU in support of classification and labelling (Gallegos Saliner *et al*, 2007a) and risk assessment (Tsakovska *et al*, 2007).

The REACH (and OECD) guidance provide stepwise schemes for applying the analogue and category approaches, considerations for assessing their adequacy, and reporting formats for documenting their results. The guidance also gives examples of read-across assessments and chemical categories that have been accepted for different regulatory purposes under various regulatory programmes, such as the risk assessment and classification and labelling in the EU, the assessment of high production existing chemicals in the OECD High Production Volume (HPV) Chemicals Programme, as well as the assessment of chemicals by the US Environmental Protection Agency (EPA).

The ECB was actively involved in the developing the REACH (and OECD) guidance. The case studies published here were part of the work carried out in-house to explore the possibilities offered by computational methods in the grouping and assessment of chemicals. Not all of the methods and ideas developed on the basis of these case studies are explicitly included in the REACH/OECD guidance. Many of the methods are novel, and are still being explored in the scientific community. At present, many of these methods have not been tried and tested in the regulatory context. The authors therefore hope that this compilation will provide an input to further efforts aimed at developing computational methods, and at exploring their potential applicability in regulatory assessment of chemicals.

2. The use of computational methods in the assessment of chemicals

For the purposes of this document, the term "computational method" is a general term to cover the following types of models: a) SARs and QSARs; b) AARs and QAARs; and c) expert systems. Numerous models based on these approaches have been published in the scientific literature or are currently the subject of research investigations, but few are accessible for end-users in the form of commercially or publicly available software tools. Therefore, some of the main software tools that may be useful to non-specialist end-users who need to develop or assess chemical categories are briefly outlined, including both existing tools and tools under development.

2.1 Regulatory applications of computational methods

In principle, computational methods can be used in the following ways for hazard and risk assessment:

a) to identify groups of similar chemicals and express this similarity in qualitative and/or quantitative terms, thereby supporting the formation of chemical categories and the application of read-across
b) to provide mechanistic information, thereby supporting the interpretation of experimental data
c) to fill data gaps, thereby replacing the need for (animal) testing
d) to supplement available test data, thereby supporting a weight-of-evidence (WoE) assessment
e) to identify chemicals of potential concern, in order to guide or prioritise testing

Some of these applications are illustrated in the case studies appended to this report. For example, the case studies in Appendices 1 and 4 illustrate how a range of different QSAR methods can be used to support the grouping the chemicals into categories (application a). The case study in Appendix 2 explores approaches to data gap filling by focusing on some human health endpoints of the existing category of ethylene glycols (application c). The case study in Appendix 3 explores the use of ranking approaches in the grouping (application a) and sorting of chemicals (application e).

2.2 Use of computational methods for grouping and expressing chemical similarity

Chemical similarity is a widely used concept in toxicology, and is based on the hypothesis that similar compounds should have similar biological activities. This forms the underlying basis for performing read-across, developing chemical categories and (Q)SARs.

It is beyond the scope of this document to provide a detailed review of different methods that can be used for the grouping of chemicals and/or for providing measures of chemical similarity. Therefore, a few basic concepts are explained, and the reader is referred to other literature.

Chemical similarity is often thought of as structural similarity. However chemical similarity can also be assessed by comparing numerical chemical information generated from a structure. The usual starting point in any computational approach to grouping and the assessment of similarity is to obtain a quantitative description of molecular structure. In general, the representation of a chemical can be considered in terms of its constitution, configuration, and conformation. Descriptors have been developed to capture all levels of molecular description, and the development of new descriptors is an ongoing field of research. Constitution refers to information about the types and numbers of atoms present (zero-dimensional descriptors), the types and numbers of substructural fragments (one-dimensional descriptors), the sequence of their bonding and the two-dimensional structure of the molecule (two-dimensional descriptors). Configuration is defined by the three-dimensional spatial arrangement of atoms (three-dimensional descriptors), whereas conformations represent thermodynamically stable spatial arrangements of the atoms in three dimensions. It has been estimated that more than 3000 descriptors have been proposed (Karelson, 2000). Most of these can be generated by software packages, such as Molconn-Z (eduSoft, LC, USA), DRAGON (TALETE srl, Italy), TSAR (Accelrys, USA), Cerius2 (Accelrys, USA), MDL® QSAR (MDL, USA), Adriana.Code (Molecular Networks, Germany), ADAPT (PennState University, USA), OASIS (LMC, Bulgaria), and CODESSA (MolCode Ltd, Estonia).

Having obtained a quantitative description of molecular structure for the chemical of interest, a comparison of one or more of the attributes with those of other chemicals can be performed either to identify a set of analogues or to determine the similarity between the identified analogues. To enable such comparisons, a variety of similarity indices have been developed, as described elsewhere (Gallegos *et al*, 2005; Gallegos-Saliner, 2006). These indices may be based on both physicochemical properties and/or fragment information. The development of new descriptors and similarity metrics is a field of ongoing research. Examples that have been investigated for their

similarity searching capabilities include Euclidean distance measures, correlation type indices (e.g. Hodgkin – Richards (Hodgkin and Richards, 1987), Tanimoto (Tou and González, 1974), Carbó (Carbó *et al*, 1996), Maximum Common Substructures (Raymond & Willett, 2002) and atom-environment descriptors (Bender *et al*, 2004).

More detailed accounts of methods for chemical similarity assessment can be found in the following reviews (Rouvray, 1995; Gillet *et al*, 1998; Martin *et al*, 2002; Nikolova & Jaworska, 2003; Bender & Glen, 2004; Sheridan *et al*, 2004; Jaworska *et al*, 2005; Gallegos Saliner, 2006; Jaworska & Nikolova-Jeliazkova, 2007).

An important notion that is expressed in many of these papers is that similarity is context-dependent. Thus, there is no absolute measure of similarity - it is only meaningful to say that chemical *X* is similar to chemical *Y* with respect to activity *Z*. This implies the need to explore the use of different measures of chemical similarity in the context of defined endpoints and/or modes/mechanisms of action.

Standard exploratory data analysis methods available in statistical software packages provide a convenient means of visualising relationships between chemicals to explore their similarities as well as to identify possible outliers. Some of the commonly used approaches include principal components analysis (PCA), cluster analysis and k-Nearest Neighbours.

2.3 Use of computational methods for providing mechanistic information

In combination with experimental data, mechanistically-based QSAR models can be used to provide information on the mechanism of action. For example, when risk assessments have been carried out in the EU under the Existing Substances Regulation (ESR), QSAR estimates for aquatic toxicity have sometimes been compared with experimental data to conclude whether a substance acts via non-polar narcosis. An example has been 2-methoxy-2-methylbutane (TAME). Other examples are given in Tsakovska *et al* (2007).

2.4 Use of computational methods for filling data gaps

If a computational method is used to fill a data gap, i.e. to directly replace a test result, the validity of the model used should firstly be established. The justification for selecting and applying a

particular model needs to be clearly reported. The OECD Guidance Document on the Validation of (Quantitative) Structure Activity Relationship Models (OECD, 2007a) provides guidance on how to evaluate specific models with respect to the OECD principles for the validation, for regulatory purposes, of (Q)SAR models. The justification should be documented by using an appropriate QSAR Model Reporting Format (QMRF). The OECD guidance on (Q)SAR validation is based almost entirely on guidance developed earlier by the ECB (Worth *et al,* 2005).

AARs and QAARs are not strictly covered by the OECD validation principles. However, similar considerations could be applied, especially in relation to the characterisation of such models in terms of their applicability domain, statistical properties, and mechanistic plausibility.

It is not generally recommended to use computational methods for predicting physicochemical properties, since these are key properties used within the risk assessment process and reliable experimental data are normally available (or easily obtainable). Some physicochemical properties, such as the octanol-water partition coefficient, can be predicted with some confidence for a wide range of chemicals, but other physicochemical properties (for example, the boiling point or explosive properties) often cannot.

Within the ESR, QSARs have been used routinely for key environmental fate parameters of organic substances. For example, QSARs have been used to estimate the adsorption of chemicals to soil (Koc), abiotic degradation by hydrolysis and photooxidation, partitioning between air and water (Henry constant), and the partitioning between octanol and water (Kow). Examples are given in Tsakovska *et al* (2007).

2.5 Use of computational methods for supplementing experimental data

Computational methods can be used to supplement the available experimental data for a particular substance, or to supplement the available data for a series of related substances.

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

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

For a series of related substances, experimental data may be available for some but not all members in the series. In such a case, a QSAR model developed specifically for the series of substances could be used to scale the results of a quantitative read-across. In other words, the coefficients of the QSAR model can be used as proportionality factors in the interpolation between two experimental values. In the REACH guidance documentation, this type of QSAR is called an "internal" QSAR, in contrast to an "external" QSAR which is developed for a different or wider set of substances.

2.6 Use of computational methods to guide or prioritise testing

Since computational methods can be used to predict trends across a series, they could be used to identify which chemicals are expected to show the highest and lowest toxicities, which can then be tested on the grounds that the endpoint values for the remaining members of the series can be interpolated.

Computational methods may additionally indicate possible outliers to a predicted trend (e.g. due to metabolism), in which case testing may be performed to check whether or not the potential outlier is actually an outlier. Similarly, such methods may help to rationalise actual outliers to an observed trend.

Ranking methods can also be used to characterise trends. Within the EU, the EURAM method has been used to set priorities for the assessment of HPV chemicals in the context of the Existing Substances Regulation (Hansen *et al*, 1999; van Haelst & Hansen, 2000). For the purposes of REACH, guidance on the application of such methods in the Evaluation and Authorisation procedures is being developed within RIPs 4.3 and 4.5. The development of ranking methods is an active area of research. Since a review of such methods is outside the scope of this document, the reader is referred to the scientific literature (e.g. Pavan, 2003). An illustration of the application of ranking methods to a data set of phthalate esters is given in Appendix 3.

3. Possible applications of computational methods in the grouping of chemicals

Traditionally, the development of chemical categories has made little or no use of computational methods. The key message of this document is that computational methods have the potential to facilitate the grouping of chemicals into categories by complementing more traditional approaches. This chapter explains how computational methods can be used to propose and support the grouping of certain types of chemicals (particularly organic chemicals). In the REACH guidance documentation, a stepwise approach to category formation is proposed (ECB, 2007). Computational methods could be useful at several of these steps, as proposed in the following sections.

3.1 Assessing membership of existing categories

In the case of existing categories, computer based approaches could be used to identify which category (or categories) a chemical (not already included in the category definition) belongs to (if any). In particular, chemical similarity tools could be used to help determine whether or not a new chemical is sufficiently similar to the members of an existing category. Such tools may provide qualitative or quantitative measures of similarity, and may also predict reactivity and modes (or mechanisms) of chemical action. Tools that can be helpful in this regard include those described in Chapter 4.

It is stressed that the application of such tools would need to be supplemented with expert judgement. For example, the user would need to make a number of choices, including: a) the properties according to which the chemicals should be compared; b) the choice of similarity measure(s); and c) criteria (cut-off points) for similarity. Furthermore, expert knowledge of mechanisms of action should also be taken into account where possible.

Since QSARs (and certain expert systems) provide quantitative estimates of endpoints, they can be used to investigate trends in one or more endpoints across category members, thereby identifying possible outliers (i.e. members that might need to be excluded), as well as possible breakpoints and boundaries in the trends.

3.2 Developing a category hypothesis and category definition

In the case of new category proposals, computational methods can help to develop the category hypothesis (rationale) and to define the category in terms of its endpoints and members. The choice of computational method(s) is likely to depend on the starting point of the investigation. For example, the user may start from a single chemical or a small group of chemicals, with the intention of building up a category by drawing on data from multiple sources (bottom-up approach). Alternatively, the user may start from a predefined group of chemicals (e.g. an inventory or subset of an inventory whose members have been decided on a particular basis), with the intention of grouping some or all of the members into one or more categories (top-down approach). The identification of analogues by the bottom-up approach is illustrated in Appendix 1 (in this case study, an initial dataset of seven phthalate esters is used to identify a total of 341 analogues). The application of the top-down approach is also illustrated in Appendix 3 (application of ranking methods to a data set of phthalate esters) and Appendix 4 (subgrouping of the EINECS inventory into seven mechanistic domains considered to underlie skin sensitisation potential). These two approaches reflect different starting points in the development of a category. Of course, a combination of bottom-up and top-down approaches could also be used; for example, the bottom-up approach could be used to expand a dataset of analogues, after which the top-down approach could be used to identify subgroups. Alternatively, the top-down approach could be used to identify subgroups within a dataset/inventory after which the bottom-up approach could be used to identify additional analogues and refine the subgroups.

Irrespective of the approach followed, the computational methods used in developing the category hypothesis and definition are likely to fall into one of the following classes: knowledge-based, analogue-based, unsupervised, and supervised.

Knowledge-based approaches typically encompass “human experience about mechanism” encoded as structural alerts. Some of these structural alerts are available as SARs, QSARs or have been built into expert systems. An example of a knowledge based approach is the Cramer Threshold of Toxicological Concern (TTC) approach which has been encoded into Toxtree (see Chapter 4). These mechanistic insights can be used as the basis for developing a category hypothesis. For example, a SAR or expert rule could provide the grouping rationale in terms of a common substructure or mechanism of action, and the applicability domain of the SAR or rule (if known) could help to define the applicability domain (potential membership) of the category. In addition, such structural rules may also be used as seeds to form larger groupings from a starting inventory

(or part inventory) of chemicals. Thus knowledge-based approaches can be useful in both bottom-up and top-down category development.

Analogue-based methods are useful in the case of a bottom-up approach, to identify candidate analogues and to provide quantitative measures of chemical similarity that could be used to identify the "closest" analogues. Many of the tools described in Chapter 4 have analogue searching capabilities.

A top-down approach comprises either unsupervised or supervised methods. Unsupervised approaches involve the use of statistical techniques to split a dataset/inventory of chemicals into smaller groupings. The approach relies on a starting dataset/inventory of chemicals and computing different numerical parameters for those chemicals or characterising them through the use of fingerprints/structural features. No assumptions are made about which parameters are relevant so a grouping can be performed on the basis of as much information as possible or only parameters that are thought to be influential for a given endpoint. The sorts of statistical approaches that can be used to split the dataset vary. Common techniques include principal components analysis (PCA), clustering methods and self organising maps (SOMs).

Supervised learning approaches are similar to unsupervised ones except that information about the activity/toxicity of chemicals is taken into account in addition to the structural/descriptor information. For example clustering techniques may still be employed but the criterion is that the clusters are extracted to discriminate for the toxicity present. Other techniques might include recursive partitioning where the aim is to find active or statistically correlated subsets based on the presence or absence of a particular combination of substructural features/fingerprints.

Ranking methods are also useful for the identification of subgroups incorporating activity profiles (see Appendix 3). Ranking methods provide a powerful means of sorting and grouping chemicals on the basis of multiple properties (e.g persistence, bioaccumulation and toxicity). As illustrated in Appendix 3, ranking methods are useful not only for sorting chemicals according to their relative level of "concern" (i.e. for identifying trends and defining subgroups based on different levels of concern), but also for identifying different profiles of toxicological behaviour (which can also be regarded as subgroups). Thus, ranking methods provide a means of comparing chemicals in terms of both the quantitative and qualitative differences in their toxicity profiles, and consequently provide a means of performing trend analysis and subgrouping.

In the case of a top-down approach, computer-based approaches could be used to explore different substructures within the dataset, some of which may reflect different mechanisms or modes of action.

Having identified the category rationale, QSARs, ranking methods (and some expert systems) could be used to highlight trends (increasing, decreasing or constant), to identify the "safe" boundaries of an endpoint, as well as possible trend breakers and subcategories.

3.3 Gathering data for the analogues

The data gathered for analogues may include estimated data as well as experimental data. Computational methods may be used to obtain estimates in cases where experimental data are missing (direct replacement of test data) or to supplement available experimental data.

To facilitate the data gathering process, computational methods could be used to identify analogues (and corresponding data) that are included in or more databases. In addition, combinatorial methods exist for identifying, *a priori*, the possible permutations of the substituents on a given substructure. Examples of tools capable of this include TSAR or Cerius2. Having identified a range of possible chemicals, one or more databases could then be searched to identify those chemicals for which data are available.

3.4 Evaluation of data adequacy

In general, experimental data are used in preference to estimated data. The adequacy of experimental data can be evaluated with reference to Klimisch codes (Klimisch *et al*, 1997) or the OECD Guidance for Determining the Quality of Data, which is given in section 3.1 of the OECD Manual for Investigation of HPV Chemicals (OECD, 2005).

In cases where computer-generated estimates are available in addition to experimental data, they can add to the weight of evidence in making a decision based on the experimental data point. For example, the estimated data may be used to increase the level of confidence in the experimental data. Alternatively, a reliable QSAR estimate could be used to select an experimental value when a range of test data are available and of uncertain quality.

Reliable (Q)SAR predictions can contribute to the evaluation of data adequacy by providing supplementary information. For example, the deviation of a chemical from an experimentally observed trend could be attributable to a change in the mechanism of toxic action, as reflected by the presence of a structural alert. Alternatively, the chemical could have a different toxicokinetic behaviour, which might be predicted by a QSAR method. In other cases, the deviation of a chemical from an experimentally observed trend that cannot be rationalised in terms of mechanism or experimental artefact could simply be a reflection of an unreliable measurement.

3.5 Assessing the adequacy of the read-across or category

The results generated by one or more computational methods could be used to support a read-across argument. An important proviso is that the same methods should not have been used to develop the hypothesis or identify the analogue(s). In such a case, the argument would be a self-fulfilling circular argument.

Similarly, computational methods could be used to support the robustness of a category provided that the same methods have not already been used to develop the hypothesis or define the category in terms of its endpoints and members.

3.6 Guiding further testing

Computational methods can be used to guide strategic testing. For example, following a preliminary assessment of a category, if a need for additional experimental data is determined, QSARs or ranking methods can help to identify which chemicals should be tested. Ranking methods are especially useful when it is important to assess multiple properties in combination (e.g. Persistence, Bioaccumulation and Toxicity; Appendix 3).

4. Useful software tools and databases

4.1 Grouping of chemicals

This section provides a brief review of various software tools that are either publicly or commercially available. Guidance on the use and interpretation of these tools is outside the scope of this document. Some of the tools have been developed by the ECB in order to make computational tools for the regulatory assessment of chemicals freely available (Worth *et al*, 2007).

4.1.1 Analog Identification Methodology (AIM)

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

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

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

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

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

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

4.1.2 Leadscope

Leadscope is a software tool developed and commercialised by Leadscope Inc. (http://www.leadscope.com). It possesses a unique chemical hierarchy containing over 27,000 chemical fingerprints which represent functional groups, chemical groupings and pharmacophores. The software can be purchased with a toxicity database and/or known drugs database. The toxicity database contains integrated information on over 160,000 chemical structures from multiple sources. The database covers a range of endpoints including acute and multiple dose studies, such as subchronic liver, carcinogenicity, genetic toxicity, reproductive and irritation. The database can be searched by structure (such as substructure or similarity), type of study, toxic effect, species, sex, dosage, duration and route of exposure. Results can be viewed and exported in convenient formats, such as Excel files.

4.1.3 Ambit

Ambit is freely available software for data management and QSAR applications, including databases and tools for searching and applicability domain assessment. It was developed by Ideaconsult Ltd (Sofia, Bulgaria) with funding from the CEFIC LRI project, and is available from (http://ambit.acad.bg). Search options include searching by name, CAS number, SMILES, substructures and structure-based similarity, and by descriptor ranges. It can also apply grouping approaches based on mechanistic understanding, such as the Verhaar classification scheme. The suite of software tools includes a module for QSAR applicability domain assessment, Ambit Discovery.

4.1.4 Toxtree

Toxtree, developed by Ideaconsult Ltd for the ECB, is a freely available application downloadable from the ECB website (http://ecb.jrc.it/QSAR) which is able to estimate different types of toxic hazard by applying structural rules. The development of Toxtree was a follow-up to an ECB workshop on chemical similarity and TTC approaches (Patlewicz *et al*, 2007). At the time of

writing (September 2007), Toxtree includes options for applying the Cramer decision tree, the Verhaar scheme as well as the BfR and SICRET rules for skin irritation/corrosion. Additional rulebases can be added in a flexible manner.

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

- Class 1 contains substances of simple chemical structure with known metabolic pathways and innocuous end products which suggest a low order of oral toxicity;
- Class 2 contains substances that are intermediate. They possess structures that are less innocuous than those in Class 1 but they do not contain structural features that are suggestive of toxicity like those in Class 3;
- Class 3 contains substances with structures that permit no strong initial impression of safety and may even suggest a significant toxicity.

The Verhaar scheme is a widely used scheme for determining the mode of action of chemicals that display aquatic toxicity. It divides chemicals into four groups: non-polar narcotics, polar narcotics, reactive chemicals and specifically-acting chemicals (Verhaar *et al*, 1992, 1995).

The BfR and SICRET rules predict skin irritation and corrosion on the basis of physicochemical exclusion rules and structural alert inclusion rules (Walker *et al*, 2005; Gallegos Saliner *et al*, 2007b).

4.1.5 Danish QSAR database

The Danish Environmental Protection Agency (DK EPA) constructed a database of (Q)SAR predictions made by some 70 models for about 166,000 organic chemicals for a wide range of different endpoints. An internet-accessible version of this database is available from the ECB website (http://ecb.jrc.it/QSAR). Different types of searching are possible including structure (substructure/exact match) searching, ID (CAS number, name) searching and parameter (endpoint) searching. The (Q)SAR models encompass endpoints for physicochemical properties, fate, eco-toxicity, absorption, metabolism and toxicity.

4.1.6 OECD QSAR Application Toolbox

The OECD QSAR Application Toolbox, for which a pilot version is under development during 2006-2007, is an application linking a number of existing tools as well as a library of existing (Q)SAR models which will allow a user to:

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

4.1.7 ECB QSAR Inventory

The ECB QSAR Inventory, which is currently under development, will be a searchable tool for linking chemicals of interest to a collection of robust summaries of (Q)SAR models. The summaries are being compiled by using a standard (Q)SAR Model Reporting Format (QMRF). A database with a web-based interface will be implemented to allow on-line access to the inventory via the ECB website. The inventory will be integrated with ECB's European chemical Substances Information System (ESIS). Different search options will be possible, such as by chemical (CAS or EC number, structure), endpoint, descriptors, and model author.

4.1.8 Toxmatch

Toxmatch, developed by Ideaconsult Ltd for the ECB, is a freely available application downloadable from the ECB website (http://ecb.jrc.it/QSAR) which can be used to facilitate the development of generic and endpoint-specific categories. The tool includes a functionality to facilitate read-across, as well as to compare chemicals of interest with existing categories. The first public release (version 1.05) uses several endpoint datasets as examples including those for skin sensitisation, skin irritation, aquatic toxicity and bioaccumulation.

4.1.9 TOXNET

TOXNET (http://toxnet.nlm.nih.gov) which is managed by the Toxicology and Environmental Health Information Program (TEHIP) in the Division of Specialized Information Services (SIS) of the National Library of Medicine (NLM). This is a free web-based system of integrated databases on toxicology, hazardous chemicals, environmental health and related areas. Several databases can be queried through TOXNET: HSDB, IRIS, ITER, GENE-TOX, CCRIS, HazMap, Household Product Database, TOXMAP, TOXLINE, DART, TRI, ChemIDPlus.

ChemIDPlus (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CHEM) has the facility to perform structural, substructural and similarity based searches to retrieve analogues and their associated data. The database contains over 368,000 chemical records, of which 200,000 include chemical structures.

4.1.10 Chemfinder

ChemFinder (http://www.chemfinder.com) is a free chemical searching tool that has been on-line since 1995. The index provides chemical structures, physical properties, and hyperlinks to other data sources such as RTECS® and TOXNET. Searching can be done by (sub)structure using a free plug in as well as on the basis of Tanimoto similarity.

4.1.11 Pipeline Pilot

Scitegic's Pipeline Pilot is a sophisticated data mining application that comprises a set of tools to analyse and visualise data, build workflows to manipulate structures and data, as well as a host of other QSAR related algorithms. The software's key feature is in automating data manipulation. Pipeline Pilot can retrieve or join data from independent databases, files, or other applications. It can be used directly to read chemistry, sequence, text, and numeric data from all popular formats and analyse data from multiple sources in real-time, without the need to first create a centralised database. Data can be easily manipulated through building up workflows from available components. These provide functionality for chemical processing, statistics, modelling, clustering, and reporting. The components allow structures to be modified, molecular properties and fingerprints to be calculated, QSAR models to be derived, compounds to be clustered and maximal common substructures to be extracted. Appendix 4 highlights some aspects of this functionality in more detail.

4.1.12 ECOSAR

ECOSAR (http://www.epa.gov/oppt/exposure/docs/episuitedl.htm) uses a number of class-specific log K_{ow}-based QSARs in order to predict the toxicity of chemicals to aquatic organisms (fish, daphnids, and green algae). The QSARs are based on measured test data that have been submitted by industry to the US EPA. The ECOSAR Class Program separates inorganics, organometallics, polymers, dyes, and surfactants as special classes (user defined). The "all others" class contains 55 chemical classes than can be identified from SMILES.

4.1.13 TOPKAT

TOPKAT (http://www.accelrys.com/products/topkat) is commercial product of Accelrys Inc. that assesses the toxicity of chemicals from 2D molecular structure. (Q)SAR models (so called submodels) are available for different chemical classes and the program automatically selects the equation from the structural input. TOPKAT also makes visible experimental test data for similar analogues if available (presumably taken from the (Q)SAR training set). For each model, a model-specific similarity distance between a query structure and a database compound can be calculated.

4.2 Assessment of metabolism

Metabolic transformations may form the basis for a category definition, if a series of structurally related chemicals are involved. However, it is more often the case that metabolism accounts for chemicals being outliers to an expected trend. A variety of databases and software tools have been developed to help in the assessment of metabolism. Some of these are highlighted in the following paragraphs. For more detailed information, literature reviews are available (Payne, 2004). Guidance on the use and interpretation of these tools is outside the scope of this document.

4.2.1 COMPACT

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

4.2.2 META

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

4.2.3 MetabolExpert

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

4.2.3 METEOR

Meteor is a commercially available tool that uses a knowledge-base of structure-metabolism rules to predict the metabolic fate of a query chemical structure. The system is developed and marketed by Lhasa Ltd (Leeds, UK) and evolved from the Derek system for toxicity prediction (Greene *et al*, 1999). Meteor’s biotransformation rules are generic reaction descriptors rather than simple entries in a reaction database. To limit over prediction, Meteor has an integrated reasoning engine based on a system of non-numerical argumentation, which uses a repository of higher level reasoning rules. The reasoning model allows the system to evaluate the likelihood of biotransformation taking place and to make comparisons between potentially competing biotransformations. The user can choose to analyse queries at a number of available search levels. At the “high likelihood” level, only the more likely biotransformations are requested for display. The system is also supplied with a knowledge base editor so that users can add their own (proprietary) rules. The metabolic tree can be searched and metabolites of specific molecular mass and or molecular formula highlighted. The generated tree is also structure-searchable. Individual biotransformations can be viewed with generalised graphical descriptions of their scope. It is possible to generate sequences automatically and to generate metabolites from an individually chosen biotransformation. It is possible to search

for either phase I or phase II biotransformations only. Additionally, Meteor is provided with a link to ClogP to identify biotransformations that are not likely to occur, due to very low lipophilicity.

4.2.4 TIMES

The Tissue MEtabolism Simulator (TIMES) is a commercially available system that aims to produce plausible biotransformation pathways from a query molecule by using rules developed from a comprehensive library of biotransformations (Mekenyan *et al*, 2004). The system is developed by the Laboratory of Mathematical Chemistry (LMC; Bourgas, Bulgaria). The generation of metabolites by TIMES can be limited to the most likely ones or can be extended to include less likely ones. The developers have also integrated reactivity models for various macromolecular interactions, for example for mutagenicity and sensitisation, to simulate the generation of reactive metabolites by specific metabolising systems, such as S9.

4.2.5 MDL Metabolite

MDL Metabolite (http://www.mdli.com) is a commercial database containing a browsing interface. The database is the uses information from multiple studies to assemble structural metabolic database entries for particular parent compounds. The focus is on xenobiotic compounds and biotransformations of medicinal drugs. Experimental data is abstracted from *in vitro* and *in vivo* studies.. In addition to structural information, the database contains enzyme information, species information, physiological activity, parent compound toxicity, bioavailability, analytical methodology, route of administration, excretion routes, quantitative and qualitative yield, CAS number of parent compound and references to the original literature.

4.2.6 The Accelrys Biotransformation database

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

4.2.7 KEGG

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) is a freely available bioinformatics resource being developed by Kyoto University and the University of Tokyo

(http://www.genome.jp/kegg). The KEGG project was initiated in May 1995, with a view to providing a tool that helps to understand the basic principles and practical applications of the relationships between genomic information and higher order functional information.

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

4.2.8 SciFinder

SciFinder is a commercially available research tool providing access the world's largest collection of biochemical, chemical, chemical engineering, medical, and other related information (http://www.cas.org/SCIFINDER). It provides a means of using a single source to obtain scientific information in journals and patent literature from around the world. It is possible to explore the database by chemical name, structure, substructure, biological sequence and reaction, as well as by research topic, author, and company.

4.2.9 University of Minnesota Biocatalysis/Biodegradation Database

The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://umbbd.ahc.umn.edu/) contains compound, enzyme, reaction and pathway information for microbial catabolism. It is a growing database comprising >160 pathways and >1100 reactions of >1000 compounds catalyzed by >700 enzymes. Along with pathway data, Biochemical Periodic Tables (http://umbbd.ahc.umn.edu/periodic) and a Biodegradation Pathway Prediction System (PPS) (http://umbbd.ahc.umn.edu/predict) are also available.

5. References

Bender A & Glen RC (2004). Molecular similarity: a key technique in molecular informatics. *Organic and Biomolecular Chemistry* **2**, 3204–3218.

Bender A, Mussa HY & Glen RC (2004). Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *Journal of Chemical Information & Computer Science* **44**, 1708-1718.

Carbó R, Besalú E, Amat L, Fradera X (1996). On quantum molecular similarity measures (QMSM) and indices (QMSI). *Journal of Mathematical Chemistry* **19**, 47-56.

Cramer GM, Ford RA & Hall RL (1978). Estimation of Toxic Hazard - A Decision Tree Approach. *Journal of. Cosmetics. Toxicology.***16**, 255-276.

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

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

Cronin MTD (2004). Predicting chemical toxicity and fate in human and the environment – an introduction. In *Predicting Chemical Toxicity and Fate*. Cronin MTD & Livingstone DJ, eds. Boca Raton, FL, USA: CRC Press. pp. 3-13.

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

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

EC (2006). 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**, 374-375.

ECB (2007). Final report of REACH Implementation Project (RIP) 3.3-2. Technical Guidance Document to Industry on the Information Requirements for REACH. European Commission-Joint Research Centre, Ispra, Italy. Accessible from: http://ecb.jrc.it/REACH/RIP_FINAL_REPORTS

Gallegos A, Patlewicz G & Worth AP (2005). *A Similarity Based Approach for Chemical Category Classification. European Commission report*. EUR 21867 EN. European Commission-Joint Research Centre, Ispra, Italy. Available from ECB website: http//ecb.jrc.it/QSAR.

Gallegos Saliner A (2006). Mini-review on chemical similarity and prediction of toxicity. Current *Computer-Aided Drug Design* **2**, 105-122.

Gallegos Saliner A, Tsakovska I, Pavan M, Patlewicz G & Worth AP (2007). Evaluation of SARs for the prediction of skin irritation/corrosion potential. Structural inclusion rules in the BfR decision support system. *SAR & QSAR in Environmental Research* **18**, 331-342.

Gallegos Saliner A, Langezaal I & Worth AP (2007a). Review of examples of read-across/category approaches discussed by the TC C&L. In: *A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read-Across*. Worth A & Patlewicz G (Eds). European Commission report EUR 22481 EN. European Commission-Joint Research Centre, Ispra, Italy. Available from ECB website: http//ecb.jrc.it/QSAR.

Gallegos Saliner A, Tsakovska I, Pavan M, Patlewicz G & Worth AP (2007b). Evaluation of SARs for the prediction of skin irritation/corrosion potential. Structural inclusion rules in the BfR decision support system. *SAR & QSAR in Environmental Research* **18**, 331-342.

Gillet VJ, Wild DJ, Willett P & Bradshaw J (1998). Similarity and Dissimilarity Methods for Processing Chemical Structure Databases. *The Computer Journal* **41**, 547-558.

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

Hansen BG, van Haelst AG, van Leeuwen K & van der Zandt P (1999). Priority setting for existing chemicals: European Union risk ranking method. *Environmental Toxicology & Chemistry* **18**, 772-779.

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

Hodgkin EE & Richards WG (1987) Molecular similarity based on electrostatic potential and electric field. *International Journal of Quantum Chemistry* **32**, 105-110.

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

Jaworska J & Nikolova-Jeliazkova N (2007). How can structiral similarity analysis help in category formation QSAR applicability domain estimation by the training set. *SAR & QSAR in Environmental Research* **18**, 195-207.

Karelson M (2000). Molecular Descriptors in QSAR/QSPR. Wiley-InterScience: New York.

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

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

Lessigiarska IV, Worth AP, Netzeva TI, Dearden JC & Cronin MTD (2006). Quantitative structure-activity-activity and quantitative structure-activity investigations of human and rodent toxicity. *Chemosphere* **65**, 1878-1887.

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

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

Martin YC, Kofron JL & Traphagen LM (2002). Do structurally similar biological activity? *Journal of Medicinal Chemistry* **45**, 4350-4358.

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

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

OECD (2005). Manual for Investigation of HPV Chemicals. Chapter 3 – Data Evaluation. Organization for Economic Cooperation and Development.

OECD (2007b). Guidance Document on the Grouping of Chemicals. Organisation of Economic Cooperation and Development. Paris, France. Available from: http://www.oecd.org/document

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

Patlewicz G, Gallegos Saliner A, Pavan M, Worth A, Benigni R, Aptula A, Bassan A, Bossa C, Falk-Filipsson A, Gillet V, Jeliazkova N, McDougal A, Mestres J, Munro I, Netzeva T, Safford B, Simon-Hettich B, Tsakovska I, Wallén M & Yang C (2007). *Chemical Similarity and Threshold of Toxicological Concern (TTC) Approaches. Report of an ECB Workshop held in Ispra, November 2005*. European Commission report EUR 22657 EN. European Commission-Joint Research Centre, Ispra, Italy. Available from ECB website: http//ecb.jrc.it/QSAR.

Pavan M (2003). Total and Partial Ranking Methods in Chemical Sciences. PhD Thesis. University of Milano-Bicocca: Milan, Italy.

Pavan M, Netzeva TI & Worth AP (2006). Validation of a QSAR Model for Acute Toxicity. *SAR & QSAR in Environmental Research* **17**, 147-171.

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

Raymond JW & Willett P (2002). Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *Journal of Computer-Aided Molecular Design* **16**, 521-533.

Rouvray DH (1995) Similarity in chemistry: past, present and future. In *Topics in Current Chemistry*, Vol 173. Sen K, ed. Springer-Verlag: Berlin, pp. 1-30.

Sheridan R, Feuston RP, Maiorov VN & Kearsley S (2004). Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *Journal of Chemical Information and Computer Science* **44**, 1912–1928.

Tou JT & González RC (1974). *Pattern recognition principles.* Addison-Wesley: Reading, MA.

Tsakovska I, Fabjan E & Worth AP (2007). Summary of Discussions on the Use of QSARs, Read-Across and Grouping in the Technical Committee on New and Existing Substances (TC NES) In: *A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read-Across*. Worth A & Patlewicz G (Eds). European Commission report EUR 22481 EN. European Commission-Joint Research Centre, Ispra, Italy. Available from ECB website: http//ecb.jrc.it/QSAR.

van Haelst AG & Hansen BG (2000). Priority setting for existing chemicals: automated data selection routine. *Environmental Toxicology & Chemistry* **19**, 2372-2377.

Verhaar HJM, van Leeuwen, CJ & Hermens JLM (1992). Classifying environmental pollutants. 1. Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere* 25, 471-491.

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

Walker JD, Gerner I, Hulzebos E & Schlegel K (2005). The Skin Irritation Corrosion Rules Estimation Tool (SICRET). *QSAR & Combinatorial Science* **24**, 378-384.

Worth AP, Bassan A, de Bruijn J, Gallegos Saliner A, Netzeva TI, Patlewicz G, Pavan M, Tsakovska I & Eisenreich S (2007). The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods. *SAR and QSAR in Environmental Research* **18**, 111-125.

Worth AP & Patlewicz G, Eds (2007). *A Compendium of Case Studies that helped to shape the REACH Guidance on Chemical Categories and Read Across.* European Commission report EUR 22481 EN. European Commission-Joint Research Centre, Ispra, Italy. Available from ECB website: http://ecb.jrc.it/QSAR.

Worth AP, Bassan A, Gallegos A, Netzeva TI, Patlewicz G, Pavan M, Tsakovska I & Vracko M (2005). *The Characterisation of (Quantitative) Structure-Activity Relationships: Preliminary Guidance*. European Commission report EUR 21866 EN. European Commission-Joint Research Centre, Ispra, Italy. Available from ECB website: http//ecb.jrc.it/QSAR.

Appendix 1
Possible application of QSAR methods to organic chemicals

Andrew Worth, Arianna Bassan, Ana Gallegos, Tatiana Netzeva, Grace Patlewicz, Manuela Pavan & Ivanka Tsakovska

European Chemicals Bureau

1 Possible application of QSAR methods to organic chemicals

1.1 Summary

This appendix aims to summarise and illustrate some of the different ways in which QSAR methods can be used in developing chemical categories. In the context of this document, the term QSAR method is interpreted in the broadest sense to incorporate analogue-searching methods, statistical methods for exploratory data analysis and model development, as well as existing computational prediction methods based on SARs and/or QSARs.

A more detailed description of one aspect of this ECB investigation (including methods, materials and detailed results) has been published separately (1).

There are a number of steps in the development and assessment of a chemical category as described below. In principle, QSAR methods can be used in any one or more of these stages.

1) Assessing membership of existing categories
2) Developing a category hypothesis and category definition
3) Gathering data for the analogues
4) Evaluation of data adequacy
5) Assessing the adequacy of the read-across or category
6) Guiding further testing

The findings from this limited investigation demonstrate that QSAR methods can provide a useful supplement to non-formalised approaches, especially when developing the category hypothesis and defining the category in terms of its members and endpoints (stage 2).

1.2 Background information on regulatory assessments

We emphasise that the main aim of this investigation was to explore and illustrate how different QSAR methods could be used in the formation of chemical categories, using a dataset of phthalate esters as an example of a category of organic chemicals. The purpose was *not* to re-evaluate any substance-specific data nor the conclusions made in the above-mentioned regulatory assessments of specific phthalate esters.

Solely for completeness and as background information, it is noted that various regulatory assessments have been conducted on phthalate esters:

a) an OECD SIAM category on a set of seven high-molecular weight phthalate ester (HMWPE) mixtures has been developed (2)
b) EU risk assessments have been completed for two higher molecular weight esters (3,4)
c) EU harmonised classifications have been agreed for seven phthalate esters (5)
d) A total of 14 phthalate esters were considered during an initial screening exercise by the EU PBT Working Group (6).

1.3 Assessing membership of existing categories

A question that is likely to arise under REACH is whether a new substance is sufficiently similar to the existing members of a category to be reasonably regarded as a member of that category and consequently to be evaluated in the same or similar way. This question can be addressed, to some extent, by using computational approaches. However, in order to apply QSAR modelling approaches, it is first necessary to characterise the existing members of the category in terms of their chemical identities.

In this investigation, the SIAM category of HMWPEs (2) was examined to determine membership and scope (Table 1). The seven members of this category are all multi-component substances, which makes it more complicated to characterise them for modelling purposes. In the case of complex multi-component substances, the computational toxicologist has several options for the representation of each substance:

a) to consider all possible components in the mixture (if they are known or suspected, and if the composition indicates approximately equal proportions of the components);
b) to consider the dominant component in the mixture, i.e. to select a representative structure on a basis of the relative (molecular) weight;
c) to select a representative structure that reflects the chain length and the branching of the components.

Furthermore, since the substances belong to a category, and the question being asked is whether a chemical is a reasonable member of that category, an additional consideration might be to include in the selection structures that represent the boundaries of the category (and possibly any subcategories). The inclusion of such structures needs to be judged on the basis of the available information on the applicability domain (and subdomains) of the category. This is not necessarily straightforward because, for administrative reasons, chemical categories tend to be defined in terms of the actual members (of commercial interest and subject to the requirements of a particular regulatory programme), rather than all possible members that can be conceived by applying the considerations of combinatorial chemistry, irrespective of whether such members are actually produced or marketed. For example, in addition to listing its seven members, the HMWPE category is defined as “esters with an alkyl carbon backbone with 7 carbon atoms or greater”. This provides no indication of the upper limit of the carbon atoms in the side chains nor for the degree of branching that might result in similar physicochemical and toxicological profiles. This adds to the arbitrary nature of selecting representative structures for the category.

For the purposes of this study, and in the absence of composition information, the selection of a single representative structure for each mixture was carried out for simplicity. CAS numbers and other information on chemical identity provided in the OECD report (2) were therefore used to draw a representative 2D chemical structure for each ester (Table 2). These 2D structures were then used as seeds to illustrate hypothesis development and analogue identification (section 1.4). The QSAR predictions subsequently made for these chemicals apply to individual chemicals and do not aim to re-evaluate the SIAM HMWPE category, which is based on mixtures of isomers.[1]

[1] If it were the intention to evaluate and further develop the existing SIAM category, it would be desirable to obtain data on the composition of the HMWPE mixtures, both in terms of the discrete chemical structures and their mass distribution.

1.4 Developing a category hypothesis and category definition

Our starting hypothesis was that phthalate esters as individual chemicals show common properties and trends on account of the phthalate moiety. Clearly, on the basis of existing knowledge, we know this generalisation to be flawed, but for the purposes of exploring how different computational (and other) methods can be used to define suitable applicability domains for different endpoints, it was a convenient starting point. It also should be emphasised that this investigation aimed to simulate a situation where no category already existed within any regulatory programme.

1.4.1 Strategy for analogue searching

An early step in performing read-across, in developing a new category, or extending an existing one, is to search for possible analogues for which experimental data might be available. In this investigation, we used seven analogues (Table 2) which were considered representative structures of the seven mixtures in the SIAM category[2]. These analogues were used to help search for both higher and lower molecular weight phthalate esters.

The following publicly available and commercial tools were used to identify analogues:

a) US EPA Analog Identification Method (AIM): http://esc.syrres.com/analog
b) AMBIT (IDEA Ltd): http://ambit.acad.bg
c) Danish (Q)SAR Database: http://ecbqsar.jrc.it
d) Chemfinder: http://www.chemfinder.com
e) ChemID plus: http://chem.sis.nlm.nih.gov/chemidplus
f) Leadscope: http://www.leadscope.com

The main searching methods included substructural searches and the use of fingerprints or fragments in combination with the Tanimoto or modified Tanimoto coefficient as the similarity metric (Figure 1). A total of 558 analogues were retrieved, including replicates of the same chemicals. Removal of these replicates gave rise to a unique set of 341 chemicals.

The set of 341 chemicals had the following characteristics:

a) The total number of carbons varied between 8 and 58 (the simplest molecular formula was $C_8H_6O_6$ and the most complex was$C_{58}H_{84}O_{12}S_3Sn$);
b) Molecular weights varied between 164.2 and 1188.2;
c) Log Kow values varied between -1.38 and 18.36.

Figure 2 gives the distributions of molecular weight and hydrophobicity within the compiled set of phthalate esters and within the set of representative HMWPE structures. The figure illustrates the potential to extend the original HMWPE category with additional chemicals, both in terms of MW and hydrophobicity. Chemicals with molecular weight higher than approximately 600 tend to break the normal distribution of this property, and are statistical outliers in the distribution. The inclusion of such chemicals in a new or extended category should be considered with caution, since they are

[2] The appropriate treatment of mixtures, for example by choosing representative structures, depends on the purpose of the investigation. Representative could mean most prevalent in the mixture or most typical in terms of physicochemical, environmental and/or toxicological properties. In this study, there was no attempt to analyse the structural variation within each mixture, because it was not the purpose to produce assessments for the SIAM chemicals.

atypical of the majority of chemicals in the distribution. The distribution of hydrophobicity (logKow) is further examined in Figure 3, which is suggestive of two overlapping distributions, with median values around logKow values of 3.5 and 9.0.

Figure 2 also shows that the representative HMWPE structures occupy a physicochemical domain above the mean MW of the larger, compiled distribution (which is consistent with the name of the SIAM category). A large number of phthalate structures with lower molecular weights were also identified.

As a result of the analogue searching, it was found that:

a) analogue searching can be performed in multiple ways, depending on the specific search engine used. The databases underpinning these search engines differ in size and scope. In addition the search options offered can vary e.g. use of more or less detailed substructures and different similarity measures;
b) due to the lack of a unique correspondence between chemical name and CAS number, searching by chemical name can result in multiple CAS numbers being retrieved. Similarly, searching by CAS number can give rise to multiple names (and different structures). Thus, it is necessary to check the results obtained.

1.4.2 Exploratory analysis: identifying similarities, differences and outliers

In the absence of any experimental data, various exploratory data analysis methods were applied to the data set of 341 analogues. Such methods provide a convenient means of visualising relationships between chemicals to explore their similarities as well as to identify possible outliers.

In addition to visualising possible outliers in a single descriptor space (as shown in Figures 2 and 3), a common statistical method for exploratory data analysis is principal components analysis (PCA). This can be used to visualise the chemical space of a dataset. PCA is a method that manipulates a multi-dimensional dataset to provide two-dimensional or three-dimensional cross-sections that capture as much of the variance in the data set as possible.

Before applying PCA, a large number of descriptors were generated by using the following commercial programs: Accord for Excel (Accelrys Inc), TSAR (Version 3.3, Accelrys Inc) and DRAGON (Talete srl). Descriptors were computed for 323 of the 341 phthalates. For the remaining 18 chemicals either the SMILES code (which is the input to the software) was unavailable or could not be processed by one of above-mentioned software programs (e.g. in the case of salts or polymers). The following types of descriptors were calculated:

a) constitutional descriptors (40 descriptors), which are zero-dimensional counts of atomic features within a molecule
b) topological descriptors (89 descriptors), which capture 2D information including structural features such as shape, symmetry, branching and cyclicity
c) molecular connectivity indices and information indices (80 indices), which account for bond accessibility within intermolecular interactions (connectivity indices) and total information content (descriptors based on the application of information theory to chemical graphs)
d) geometrical descriptors, which capture 3D information (295 descriptors)
e) physicochemical properties (12 properties), i.e. descriptors that estimate experimental properties such as logKow and boiling point.

PCA plots were generated for different combinations of these descriptors, as well as for all descriptors (to capture as much chemical information as possible). As an example, Figure 4 shows a PCA plot based on the combined use of connectivity and information indices. The plot captures 71.6% of the total variance in the data and reveals a number of statistical outliers (data points outside the ellipse). It can also be seen that the seven representative members of the SIAM category (red points) are within the ellipse and surrounded by many other analogues. In other words, the representative HMWPE chemicals are similar to the many of the analogues on the basis of the chosen descriptors.

The PCA plots are useful in two ways:

a) they help to identify statistical outliers, i.e. chemicals that are likely to behave differently, and hence are candidates for exclusion from a category.
b) they illustrate how the analogues are clustered together, hence which analogues are similar with respect to the chemical information computed (i.e. the connectivity and information indices).

Figure 4 indicates that there are four significant statistical outliers in the lower left corner, which are illustrated in Figure 5. These are chemicals that responded to some of the analogue search criteria but are in fact different from the others for some other reason. Therefore, if experimental data were available for these outliers, it would not be advisable to read-across these data to the other candidates and *vice versa*. The analysis of these outliers emphasises the point that when an automatic search of analogues is performed, the identified candidates might not all be suitable for the selection purpose (i.e. they might not be suitable candidates for the category) and should therefore be analysed critically. Figure 4 illustrates the other useful feature of the PCA plots, namely their utility in identifying chemicals that are grouped closely together with very similar structures. Read-across between such chemicals is more likely to be adequate (but this should still be analysed critically).

1.4.3 Exploratory analysis: building a matrix of estimated endpoints

In the absence (or paucity) of experimental data (as was the case with the data set here), existing QSAR models can be used to: a) identify chemicals with common hazards; b) predict trends in the potencies of such hazards; and c) identify possible breakpoints in trends, therefore possible subcategories.[3]

To illustrate this approach, the data set of analogues was extended to include predictions for a number of endpoints. The result was a large data matrix in which the phthalates were represented by different rows and the following human health and environmental endpoints were represented by different columns:

1) skin sensitisation, calculated with TOPKAT
2) skin irritation, calculated with TOPKAT
3) acute oral mammalian toxicity, calculated with TOPKAT
4) developmental toxicity, calculated with TOPKAT
5) persistence, calculated with BIOWIN3 (ultimate biodegradation model)

[3] If sufficient reliable experimental data are available, it should also be possible to develop new QSARs that reflect the observed trends. In such a case, new QSAR model development is an integral part of the formation of the category.

6) bioaccumulation, calculated with BCFWIN
7) acute fish toxicity (96h fathead minnow), calculated with ECOSAR and TOPKAT

These endpoints were chosen to cover a variety of endpoints that might normally be of interest in the development of a category. The choice was not intended to be comprehensive, nor were the endpoints chosen considered to be of any particular importance for the assessment of phthalate esters, since the exercise was based on the assumption of little or empirical knowledge.[4]

The different software packages used were chosen simply for convenience (available in-house and capable of running in batch mode). BIOWIN3, BCFWIN and ECOSAR are available as free-ware from the US EPA (http://www.epa.gov/oppt/exposure/docs/episuitedl.htm). TOPKAT (Version 6.2) is developed and commercialised by Accelrys Inc (San Diego, CA, USA).

The data matrix generated is the usual starting point for any (Q)SAR modelling or chemometric ranking exercises. In fact, the data for persistence (P), bioaccumulation (B), and fish toxicity (T) were used to explore the applicability of ranking methods as a means of evaluating chemicals according to their PBT properties (see Appendix 3).

A major effort in constructing the data matrix was in obtaining the identities of the chemicals and, as far as possible, the structural representations, (e.g. SMILES codes). This process can involve some arbitrary decisions. For example, as mentioned above, in the case of the seven SIAM phthalates, each member is defined by a CAS number but this actually refers to a mixture of isomers, with varying lengths of the side chains and branching patterns. In such a case, it is either necessary to identify a "representative" structure for each isomeric mixture, or to analyse the structural variation within the mixture, depending on the purpose of the investigation.

On the basis of such a matrix of estimated endpoints, it should be possible to gain an impression of whether the group of analogues is likely to form a robust category, or whether it needs to be redefined by removal and/or addition of analogues. It should also be possible to identify subgroups containing different trends that might form the basis for subcategories.

In the matrix of 323 phthalate esters generated in this study, different trends were observed for different endpoints (which is not surprising given the size of the matrix). These trends are not discussed at length here. However, two examples are provided in an attempt to illustrate how phthalate esters could be grouped (in an endpoint-specific manner) for aquatic toxicity and skin sensitisation.

1.4.3.1. Exploratory analysis: identifying possible subcategories for aquatic toxicity

The availability of QSAR estimates for the members of a potential category enables trends to be identified in the absence of experimental data, as well as possible breakpoints in those trends. This allows a preliminary identification not only of possible category boundaries but also possible divisions into subcategories, some of which may be endpoint-specific.

[4] In reality, evidence from regulatory assessments points to concerns for developmental toxicity and aquatic toxicity, at least for certain phthalate esters. However, this was not directly relevant to this exercise exploring the possible uses of QSAR methods.

Under the current EU classification and labelling system, the following risk phrases are used for environmental hazard classification on the basis of aquatic toxicity:[5]

- R50 (Very toxic to aquatic organisms)
  96h LC_{50} (fish) $\leq$ 1 mg/L
- R51 (Toxic to aquatic organisms)
  96h LC_{50} (fish) > 1 and $\leq$ 10 mg/L
- R52 (Harmful to aquatic organisms)
  96h LC_{50} (fish) > 10 and $\leq$ 100 mg/L

The same cut-off values also apply if 48h EC50 to Daphnia or other crustaceans, and 72 or 96h EC50 for algae or other aquatic plants, are used. If the LC50 (EC50) value is above 100 mg/L, it is considered that there is no concern (NC) for acute toxicity to aquatic organisms.

For the purposes of the illustration in this section, it is assumed that a category is a chemical group in which a trend (e.g. increasing toxicity with increasing hydrophobicity) can be observed while a subcategory is a smaller group, in which the toxicity varies within a narrower range due to subtle structure modifications (but where the mechanism of action does not change). The assumption of a common mechanism of action is the "unifying feature" between the category members but the category as a whole does not provide a suitable basis for read-across unless subcategories are also defined. Thus, a read-across between members of the same category should be allowed because the result will not change the classification result (classification categories and cut-offs being preferably defined in formal way). However, read-across between members of different subcategories might produce predictions that are as unreliable as from members of different categories.

To explore the trends in acute toxicity to fish, ECOSAR was used to predict the LC50 values for toxicity to fathead minnow[6]. The predicted LC50 values generated by ECOSAR were compared with those generated by TOPKAT (only predictions in the TOPKAT optimum prediction space were considered). Figure 6 shows how the two models correlate with each other (r^2 = 0.8). Two additional observations can be made from Figure 6:

a) TOPKAT makes more conservative predictions than ECOSAR (i.e. predicts higher toxicity – lower concentrations, for the same chemicals,) for a large number of chemicals. This is particularly evident for the higher predicted values. Systematic differences in the results obtained by different QSAR models should be taken into account if the model estimates on their own are to be relied upon for classification and labelling and/or risk assessment purposes.
b) A large number of phthalates are predicted as very toxic by both programs. The reasons for this observation was analysed further and is further discussed below.

ECOSAR has multiple models, each of which is applicable to a specific chemical class. The model for esters (Equation 1) was used, unless one of the other models gave a more conservative value (lower LC50).

$$\text{Log LC50} = -0.535\ \text{logKow} + 0.25 \qquad [1]$$

[5] In addition, R53 may be assigned if there is potential for long-term adverse effects in the environment. The prediction of R53 was not considered in this investigation.

[6] The ECOSAR and EPIWIN programs were able to process 324 smiles codes and therefore make predictions for 324 phthalate esters.

ECOSAR produces warnings relating to the applicability of the estimates (e.g. when the water solubility is very low, or when the prediction is outside the range of logKow values in the model training set).

If used alone, Equation 1 assumes that toxicity increases in a linear fashion as hydrophobicity increases (i.e. a monotonic increasing linear trend), yet numerous studies (e.g. Hermens *et al*, 7) have shown that a parabolic (i.e. non-monotonic) relationship between logLC50 and logKow is more realistic and better predicts toxicity at higher values of logKow (Figure 7). Therefore, to take this into account, a simple bilinear relationship was derived (Figure 8).

This relationship was used to develop a decision rule approach, i.e. rules-of-thumb for classifying chemicals for acute fish toxicity on the basis of estimated logKow values (Table 3). The rules do not take into account possible metabolism of the esters the possible effects of metabolism on the toxicity.

In developing the decision rules, the relationship between the solubility limit and LC50 was also taken into account. If the solubility is less than the predicted LC50, then the solubility is limiting. On this basis, it was assumed that if the water solubility of a phthalate ester, calculated with WSKOWWIN, was less than 0.002 mg/L, then the chemical was not sufficiently soluble to be classified as toxic. This solubility value corresponds to a logKow of approximately 8 and places a large number of the category candidates in the NC group (Figure 8). However, this does not exclude the possibility of long-term chronic toxicity effects.

At the border line of the solubility limit, more data should be compiled to define it better, especially for the toxicity of di-C_6-C_7 (and even C_8) analogues, for which little or no data were found in this study. It is possible that around the limit of aqueous solubility, the branching of the side chains plays a significant role, but this hypothesis needs to be studied further.

Application of these rules to the seven SIAM phthalate representative structures led to predictions of NC, which corresponds with available experimental data (Table 4).

We emphasise that these decision rules are based on a preliminary investigation, and are presented to illustrate the application of a QSAR approach for defining possible subcategories based on structural and/or physicochemical rules. A larger set of experimental data would be needed to validate (and most likely refine) these decision rules. Validation of the decision rules would be especially important if they were used as the basis for (or to support) classification and labelling. Furthermore, it is anticipated that in regulatory practise, such rules would be used as “rules of thumb” rather than as prescriptive rules. In particular, expert judgement would be especially important for those chemical lying on or close to classification boundaries.

1.4.3.2. Exploratory analysis: identifying possible subcategories for skin sensitisation

A qualitative read-across can be regarded as equivalent to the use of a SAR based on the common substructure. In developing a SAR, the greater the number of analogues identified, the greater the possibility of understanding the consequences of structural variation, i.e. the greater the possibility of defining an applicability domain (AD) for the SAR. Ideally, the AD should be defined by using experimental data to identify the known consequences of structural variation. However, such data is not always available, especially in the case of large homogeneous datasets (e.g. the dataset of 341 phthalate esters). However, in such cases, an exploratory analysis can be performed on the basis of predicted data, to identify the possible consequences of structural variation.

As an example, a substructure-based clustering method within Leadscope (Leadscope Inc, Columbus, OH, US) was applied to 79 predicted non-sensitising phthalates and 29 predicted sensitising phthalates. This identified two main "substructural signatures" that were representative of non-sensitisers and sensitisers, respectively (Figure 9). The signatures themselves do not represent any single chemical in the group; instead they sum up the overall features of that group. This analysis indicated that phthalate esters containing a single ester group, or containing an additional functionality on the benzene ring (e.g. a nitro or chloro group), or containing branching on the acyl chain, were more likely to form a different subcategory to the typical SIAM diesters (i.e. these structural variants were more likely to be sensitising rather than non-sensitising).

1.5 Gathering data – filling data gaps with estimated values

Under REACH, in the absence of reliable experimental data, it is possible to fill data gaps by using valid QSAR models. Models used should be validated in accordance with the OECD principles for QSAR validation. As far as possible, the predictions and trends established by QSAR methods should be verified by comparison with experimental data. This should be feasible for category approaches, since experimental data should be available for at least some of the category members.

It may be the case that a QSAR predicts the trend correctly, but the numerical values of the predictions are consistently high or low. In such cases, the QSAR could be used to support the trend, but the experimental data should be relied upon when performing an interpolation (or even extrapolation).

In this study, the predictions made by the above-mentioned models for the seven representative structures were compared with experimental data for the seven SIAM phthalates. Table 4 summarises the experimental data for the seven endpoints, to the extent it was provided in the SIDS Initial Assessment Report (SIAR; 3), and Table 5 provides the corresponding QSAR predictions.

Irrespective of the endpoint predicted, TOPKAT predictions are associated with an indication of their reliability, according to whether the chemical of interest lies within the AD of the model. This is an assessment of whether the chemical lies within the structural and physicochemical descriptor space of the training set. However it is up to the end-user to decide how to interpret this information. For the purposes of this study, a strict criterion was adopted – a prediction was only considered to be reliable if the chemical was found to be within the structural and descriptor space.

For developmental toxicity and acute fish toxicity, none of the TOPKAT predictions was reliable since all HMWPE were outside the optimum prediction space of TOPKAT models. For skin sensitisation, four phthalates were correctly predicted to be non-sensitisers, while the other three were outside the AD. In the case of skin irritation, six predictions of non-irritancy were made, of which four were correct and two could not be evaluated, due to lack of experimental data in the SIDS Initial Assessment Report (SIAR). For eye irritation, TOPKAT made six predictions, including three correct predictions of known non-irritants, a prediction of non-irritancy for a phthalate with no experimental data, a prediction of mild irritancy for a phthalate with no experimental data, and a prediction of mild irritancy for a non-irritant.

For bioconcentration, there was full concordance between the seven BCFWIN predictions of no bioaccumulation potential and the experimental data reported in the SIAR.

For biodegradation, there was a good concordance between the BIOWIN predictions by experimental data reported in the SIAR (where data were available). Experimentally it was observed that lower molecular weight ranges are expected to biodegrade to a high extent (greater

than 60% after 28 days), while higher molecular weight members are expected to biodegrade to an extent less than 60%. BIOWIN predictions are in agreement with these results, with Di-phC10 PE (53306-54-0) and Di-C11 PE (3648-20-2) being predicted to biodegrade within weeks, and Di-C13 PE (68515-47-9) and Di-C13 PE (119-06-2) in months.

The TOPKAT model for acute oral toxicity (rat oral LD50) predicted all seven phthalates to be non-toxic, but five of these predictions were out of the domain.

1.6 Assessing the adequacy of experimental data

In the case of test data for aquatic toxicity, comparison of predicted and experimental results can explain "unreliable" test results. For example, a measured LC50 for acute fish toxicity that lies above the theoretical minimum LC50 predicted by a baseline log Kow QSAR model could indicate that adsorption or volatilisation has occurred in the test system. Another pitfall may be caused by low water solubility of the chemicals, which is relevant to the HMWPE category.

1.7 Assessing the adequacy of read-across (impact of structural variation)

One of the difficulties in applying read-across is the uncertainty of whether a small structural difference between the source and target chemicals could invalidate the read-across.

QSAR analysis can help to establish the adequacy of the read-across in cases where the (Q)SAR captures the factor(s) responsible for driving the (eco)toxicological effect. For example acute aquatic toxicity is typically modelled with logKow. For illustrative purposes, it can be assumed that data for this endpoint is missing for diethyl phthalate (Figure 10), but there are several choices of analogue for the read-across. Assuming the same mechanism of action holds, read-across from 1,2-benzenedicarboxylic acid, bis(2-hydroxyethyl) ester is likely to lead to an underestimation of the acute aquatic hazard, whereas read-across from diethyl 3,4,5,6-tetrachlorophthalate is likely to lead to overestimation. In other words, neither read-across would be adequate, even though the three example chemicals have two carbon atoms in the ester side chains. This is because introduction of hydrophilic groups (in the 1,2-benzenedicarboxylic acid, bis(2-hydroxyethyl) ester) reduces the logKow (hydrophobicity) to 0.12, whereas introduction of halogen atoms in the benzene ring increases the logKow to 5.22 (the latter chemical, however, might show a deviation from the predicted toxicity due to the possible effect of the halogen atoms on the hydrolysis rate). Conversely, another analogue (1,2-Benzenedicarboxylic acid, bis(2-ethoxyethyl) ester) has a similar logKow of 2.10. Thus, read-across from this analogue is therefore likely lead to the correct hazard classification (harmful), although looking at the side chain, it looks more dissimilar to diethyl phthalate than the other two analogues. Thus, the use of a valid QSAR can help to choose an appropriate analogue and support the adequacy of the read-across.

The general assumption that all chemicals sharing a common substructural fragment show similar (eco)toxicological profiles fails when the structural analogues are able to act via different (or multiple) mechanisms of action (see Figure 10). For example, allyl 2,3-epoxypropyl phthalate can act as an alkylating agent towards proteins and DNA, dimethyl 3,6-dihydroxyphthalate can easily transform to a strong electrophile (quinine form), and 1,2-benzenedicarboxylic acid, 2-hydroxyethyl 2-[(1-oxo-2-propenyl)oxy]ethyl ester can be a strong electrophile itself due to the presence of an acrylate moiety in the molecule. Such chemicals should not normally be predicted by the same QSAR (unless the QSAR is developed specifically to capture more than one mechanism). Furthermore, such chemicals should either be excluded from a category based on a narcotic mode of action, or they should be included as a subcategory.

In other words, a category should be developed with mechanistic considerations in mind, to enable adequate read-across to be carried out within the same category or subcategory. The scientific challenge is to define appropriate structural rules and/or physicochemical cut-off values for defining these subcategories.

1.8 Conclusions

The investigation described here, which used the existing SIAM category of phthalate esters as a starting point, aimed to explore and illustrate some of the ways in which QSAR methods can be applied in the development of chemical categories. It was not the aim of this investigation to further justify the SIAM category itself, to develop an alternative and extended category of phthalate esters, or to make proposals for the classification and labelling of specific chemicals. It was rather intended to simulate some of the questions that arise in *de novo* category development, and explore how computational toxicology can help.

In particular, the investigation resulted in the following learnings:

a) There are a number of search engines available for the identification of analogues for read-across. These provide different analogues on account of the database covered and the similarity measure used. The different approaches vary in terms of their ease-of-use and their capacity for data-mining. Errors were discovered which propagated between different databases, which highlighted the importance of checking any results obtained.
b) Data exploration tools, such as PCA and clustering, are useful to enable visualisation (in 2D or 3D) of the chemical domain of a set of compounds to look for obvious groups of "like" compounds. These approaches rely on a starting dataset of chemicals and computing different numerical parameters (such as geometrical, topological, structural, physicochemical, electronic descriptors) for those chemicals or characterising them through the use of structural fingerprints.
c) The use of predictions from existing QSARs can be useful to explore trends within groups of chemicals or help in the assessment of data adequacy. Structural fingerprints and cut-off values along descriptors (e.g. physicochemical properties) can be useful to gain insights about the scope and boundaries of a category (and subcategories).
d) QSARs that encode the descriptor(s) driving an endpoint can be helpful in assessing the adequacy of a read-across (i.e. assessing how similar an analogue is to the chemical of interest with respect to a given endpoint).

Finally, QSAR methods should be seen as supplementary tools that are useful in the development of categories of certain kinds of chemicals, rather than as an automated substitute for more conventional approaches to category formation.

1.9 References

1. Netzeva TI & Worth AP (2007). Classification of phthalates according to their (Q)SAR predicted acute toxicity to fish. A Case study. JRC report EUR 22623 EN. European Chemicals Bureau, Joint Research Centre, European Commission, Ispra, Italy. http://ecb.jrc.it.

2. OECD (2004). SIDS Initial Assessment Report for SIAM 19. Berlin, Germany, October 2004.

3. EC (2003). EU RAR on 1,2-Benzenedicarboxylic acid, di-C8-10-branched alkyl esters, C9-rich (DINP), CAS. No. 68515-48-0 Volume 35. Available from the ECB website: http://ecb.jrc.it/esis/esis.php?PGM=ora

4. EC (2003). EU RAR on 1,2-Benzenedicarboxylic acid, di-C9-11-branched alkyl esters, C10-rich (DIDP), CAS. No. 68515-49-1. Volume 36. Available from the ECB website: http://ecb.jrc.it/esis/esis.php?PGM=ora

5. The ECB CLASSLAB database (http://ecb.jrc.it/classification-labelling) identifies the following entries:

   i) Annex I entry 607-317-00-9: bis(2-ethylhexyl) phthalate; di-(2-ethylhexyl) phthalate; DEHP (CAS No. 117-81-7; EC No. 204-211-0): Repr. Cat. 2; R60; Repr. Cat. 2; R61.

   ii) Annex I entry 607-318-00-4: dibutyl phthalate; DBP (CAS No. 84-74-2; EC No. 201-557-4): Repr. Cat. 2; R61; Repr. Cat. 3; R62; N; R50.

   iii) Annex I entry 607-426-00-1: 1,2-benzenedicarboxylic acid, dipentylester, branched and linear (CAS No. 84777-06-0; EC No. 284-032-2): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

   iv) Annex I entry 607-426-00-1: n-pentyl-isopentylphthalate (CAS No. -; EC No. -): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

   v) Annex I entry 607-426-00-1: di-n-pentyl phthalate (CAS No. 131-18-0; EC No. 205-017-9): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

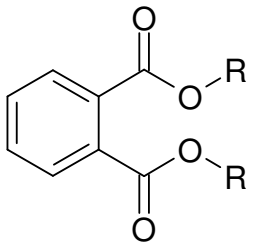
   vi) Annex I entry 607-426-00-1: diisopentylphthalate (CAS No. 605-50-5; EC No. 210-088-4): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

   vii) Annex I entry 607-480-00-6: 1,2-benzenedicarboxylic acid di-C7-11-branched and linear alkylesters (CAS No. 68515-42-4; EC No. 271-084-6): Repr. Cat. 2; R61; Repr. Cat. 3; R62.

6. ECB (2002). Identification of potential PBTs or vPvBs among the IUCLID High Production Volume Chemicals. Doc ECB 4/14/02(pbt strategy – report)_rev1. Date 21 November 2002.

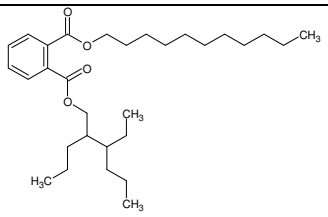
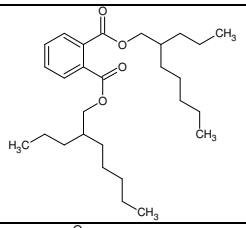
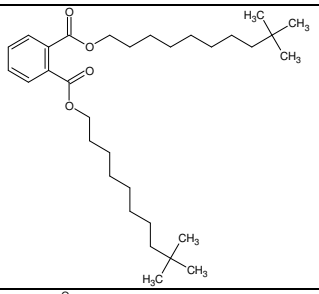
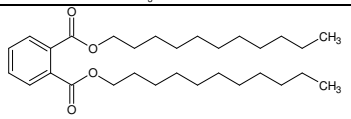
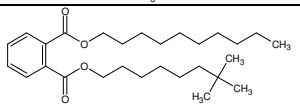
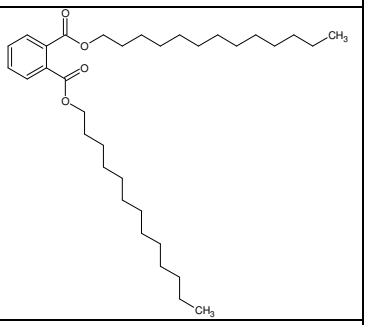
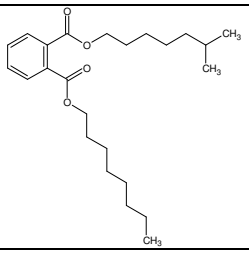
7. Hermens J., Leeuwangh P. & Musch A. (1984). Quantitative structure-activity relationships and mixture toxicity studies of chloro- and alkylanilines at an acute lethal toxicity level to the guppy (Poecilia reticulate). Ecotox Environ Safety 8: 338-394.

Table 1. Membership of the SIAM category of High Molecular Weight Phthalate Esters



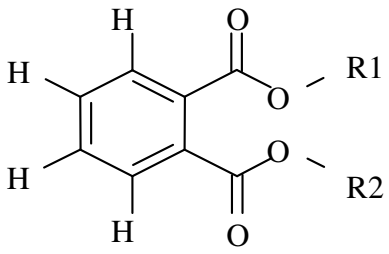
CAS	Name	Formula
53306-54-0	1,2-benzenedicarboxilic acid, di-2-propylheptyl ester	R = $C_{10}H_{21}$ (propyl branched) [100% branched]
68515-41-3	1,2-benzenedicarboxilic acid, di-C7-9-branched and linear alkyl esters	R = C_7H_{15} to C_9H_{19} (branched and linear) [>80% linear]
85507-79-5	1,2-benzenedicarboxilic acid, di-C11-branched and linear alkyl esters	R = $C_{11}H_{23}$ (branched, essentially methyl, and linear)
68515-43-5	1,2-benzenedicarboxilic acid, di-C9-11-branched and linear alkyl esters	R = C_9H_{19} to $C_{11}H_{23}$ (branched and linear) [>80% linear]
3648-20-20	1,2-benzenedicarboxilic acid, di-C11-alkyl ester	R = $C_{11}H_{23}$ (branched)
685151-47-9	1,2-benzenedicarboxilic acid, di-C11-14-branched alkyl esters, C13 rich	R = $C_{11}H_{23}$ (branched, essentially methyl)
119-06-2	1,2-benzenedicarboxilic acid, di-C13-alkyl ester	R = $C_{11}H_{23}$ (branched)

Table 2. Representative structures for the seven SIAM phthalate esters

CAS number	Structure	CAS number	Structure
1) 85507-79-5		5) 53306-54-0	
2) 68515-47-9		6) 3648-20-2	
3) 68515-43-5		7) 119-06-2	
4) 68515-41-3			

[1] 1,2-Benzenedicarboxylic acid, di-C11-branched and linear alkyl esters (Di-C11 PE)
[2] 1,2-Benzenedicarboxylic acid, di-C11-14-branched alkyl esters, C13-rich (Di-C13 PE)
[3] 1,2-Benzenedicarboxylic acid, di-C9-11-branched and linear alkyl esters (Di-C9-11 PE)
[4] 1,2-Benzenedicarboxylic acid, di-C7-9-branched and linear alkyl esters (Di-C7-9 PE)
[5] 1,2-Benzenedicarboxylic acid, di-2-propylheptyl ester (Di-phC10 PE)
[6] 1,2-Benzenedicarboxylic acid, di-C11-alkyl ester (Di-C11 PE)
[7] 1,2-Benzenedicarboxylic acid, di-C13-alkyl ester (Di-C13 PE)

Table 3. Rules-of-thumb for the classification of phthalates on the basis of estimated logKow



R1, R2	Log Kow range	Toxicity group	Example Chemical	Acute fish toxicity value from IUCLID[a] (mg/L)
-	log Kow < 1.5	No concern		
R1 = R2 = CH3 R1 = R2 = C2H5	1.5 < log Kow < 3.2	Harmful	Diethyl phthalate (CAS 84-66-2)	29.60; 17.00; 17.00; 16.80; 31.80
R1 = R2 = C3H7 R1 = R2 = C4H9	3.2 < log Kow < 5.0	Toxic	Dibutyl phthalate (CAS 84-74-2)	2.60; 0.71; 1.00; 1.30
R1 = R2 = C5H11 R1 = R2 = C6H13	5.0 < log Kow < 7.0	Very toxic	Dipentyl phthalate (CAS 131180) Dihexyl phthalate (CAS 68515-50-4)	N[b]; R50[c] 0.82[d]
R1 = R2 = C7H15	7.0 < log Kow < 8.0	Toxic		Data not found
R1 = R2 = C8H17 R1 = R2 => C8H17	log Kow > 8.0	No concern	Di-sec-octyl phthalate (CAS 117-81-7)	NTBLAS[d,e]

[a] Due to a large number of values available and for simplicity, only LC50 to fathead minnow is shown
[b] N: Classified in Annex I of Directive 67/548/EEC as "Dangerous for the environment"
[c] Classified in Annex I of Directive 67/548/EEC as "Very toxic to aquatic organisms"
[d] Collected from http://www.epa.gov/opptintr/chemtest/pubs/alkpht.pdf
[e] Not toxic below the limit of water solubility

Table 4. Experimental data for the seven SIAM phthalates

CAS number	Abbreviation	Developmental Toxicity	Acute fish toxicity	Skin Irritation	Eye Irritation	BCF	Biodegradation Non-acclimated Inoculum	Biodegradation Acclimated Inoculum	Rat Oral LD50, g/kg	Skin Sensitisation
85507-79-5	Di-C11 PE	No[a]	NT[b]	No data	No data	Not B	No data	No data	>60 (NT)	NEG
68515-47-9	Di-C13 PE	No[a]	NT[b]	NI	NI / mild	Not B	12.80%	No data	>10 (NT)	NEG
68515-43-5	Di-C9-11 PE	No[a]	NT[b]	NI	NI	Not B	No data	No data	>19.7 (NT)	NEG
68515-41-3	Di-C7-9 PE	No[a]	NT[b]	NI	NI	Not B	No data	No data	>19.3 (NT)	NEG
53306-54-0	Di-phC10 PE	No[a]	NT[b]	NI	NI	Not B	67 to 75%	No data	>5 (NT)	No data
3648-20-2	Di-C11 PE	No[a]	NT[b]	NI	NI	Not B	57.40%	76.00%	>60 (NT)	NEG
119-06-2	Di-C13 PE	No[a]	NT[b]	No data	No data	Not B	42.00%	37.00%	>2 (NT)	NEG

[a] No developmental toxicity signs are observed below maternotoxic doses

[b] Data were provided in the SIAR for several different species of fish (but not fathead minnow), invertebrates, and algae. All valid studies show that members of the HMWPE Category do not produce acute toxicity to fish and invertebrates or toxicity to algae at their maximum solubility in the various media

Table 5. Predicted endpoints for the seven (representative structures of) SIAM phthalates

CAS number	Abbreviation	TOPKAT Developmental Toxicity	TOPKAT Acute fish toxicity (96h LC_{50} to *P. promelas*)	Acute fish toxicity (96h LC_{50} to *P. promelas*)[a]	TOPKAT Skin Irritation	TOPKAT Eye Irritation	BCFWIN BCF	BIOWIN 3 Ultimate degradation	TOPKAT Rat Oral LD50, g/kg	TOPKAT Skin Sensitisation
85507-79-5	Di-C11 PE	NA	NA	NTBLAS	NEG	NEG	3.16	2.7288	≥10, NT (NA)	NA
68515-47-9	Di-C13 PE	NA	NA	NTBLAS	NA	NA	3.16	1.9443	≥10, NT (NA)	NA
68515-43-5	Di-C9-11 PE	NA	NA	NTBLAS	NEG	NEG	20.07	2.5787	≥10, NT (NA)	NEG
68515-41-3	Di-C7-9 PE	NA	NA	NTBLAS	NEG	NEG	3.16	2.9148	6.8, NT	NEG
53306-54-0	Di-phC10 PE	NA	NA	NTBLAS	NEG	NEG	3.16	3.0892	≥10, NT	NA
3648-20-2	Di-C11 PE	NA	NA	NTBLAS	NEG	MILD	3.16	3.0272	≥10, NT (NA)	NEG
119-06-2	Di-C13 PE	NA	NA	NTBLAS	NEG	MILD	3.16	2.9032	≥10, NT (NA)	NEG

[a]The acute fish toxicity (96-h LC_{50} to *P. promelas*) in this column was judged on a basis of octanol-water partition coefficient (log Kow) and water solubility (WS, mg/L), calculated by WSKOWIN v. 1.41. All representative chemicals had calculated log Kow ≥ 8.5 and calculated water solubility below 0.001 mg/L. Therefore, they were considered not (acutely) toxic because they are below the limit of aqueous solubility (NTBLAS). This does not exclude possible concern for prolonged/chronic toxicity.

NA = not applicable (out of domain)
NT= non-toxic

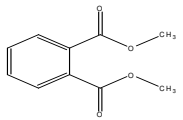
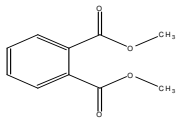
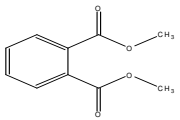
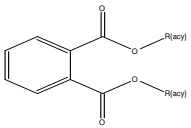
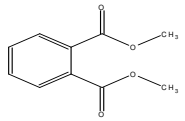
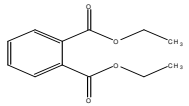
	AIM	AMBIT	ChemID Plus	ChemFinder	DK DB	Leadscope
Search Criteria	Analogues of the 7 phthalate esters belonging to the original SIAM category	Fingerprints/ Tanimoto distance AND Substructure search: 	Substructure search 	Substructure search 	Substructure search 	Modified Tanimoto + data contained within Leadscope  AND 
Hits	23	113	176	41	182	53

Figure 1. Multiple ways of analogue searching for seven phthalate esters

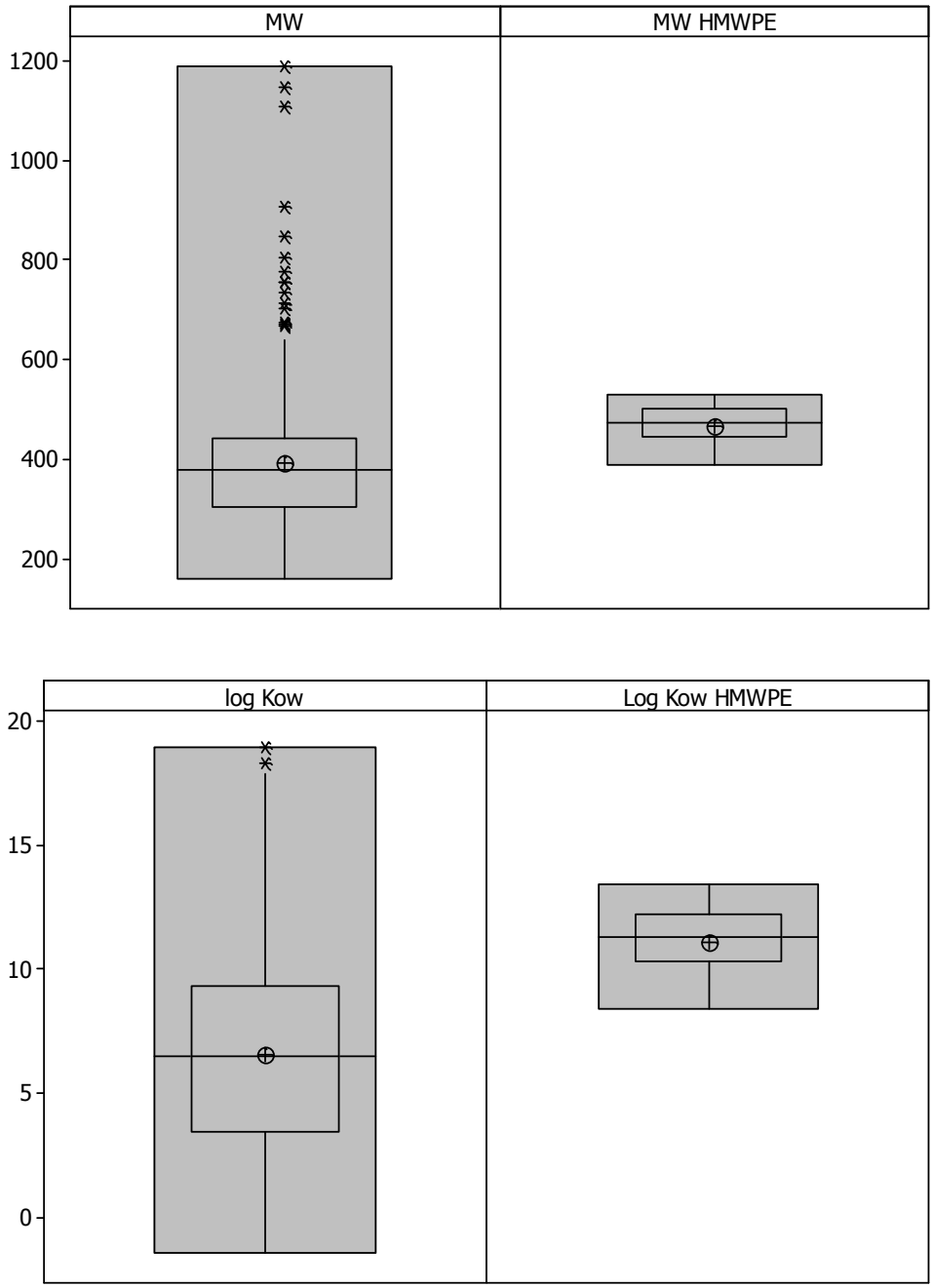


Figure 2. Plots of molecular weight (top) and hydrophobicity (bottom) for the compilation of phthalates (first column) and representative structures for the HMWPE category (second column). Mean symbol and median line are shown along with the min/max box and statistical outliers indicated with an asterisk.

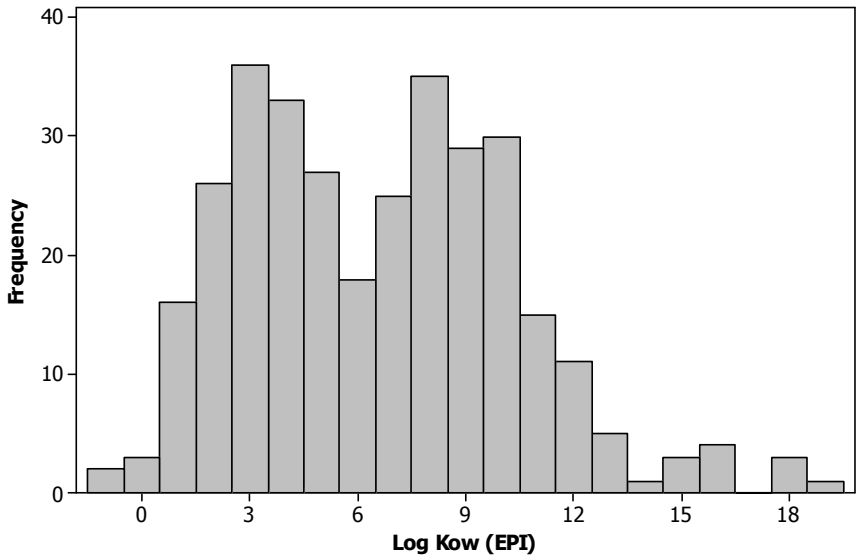


Figure 3. Distribution of logKow values for 324 phthalate esters

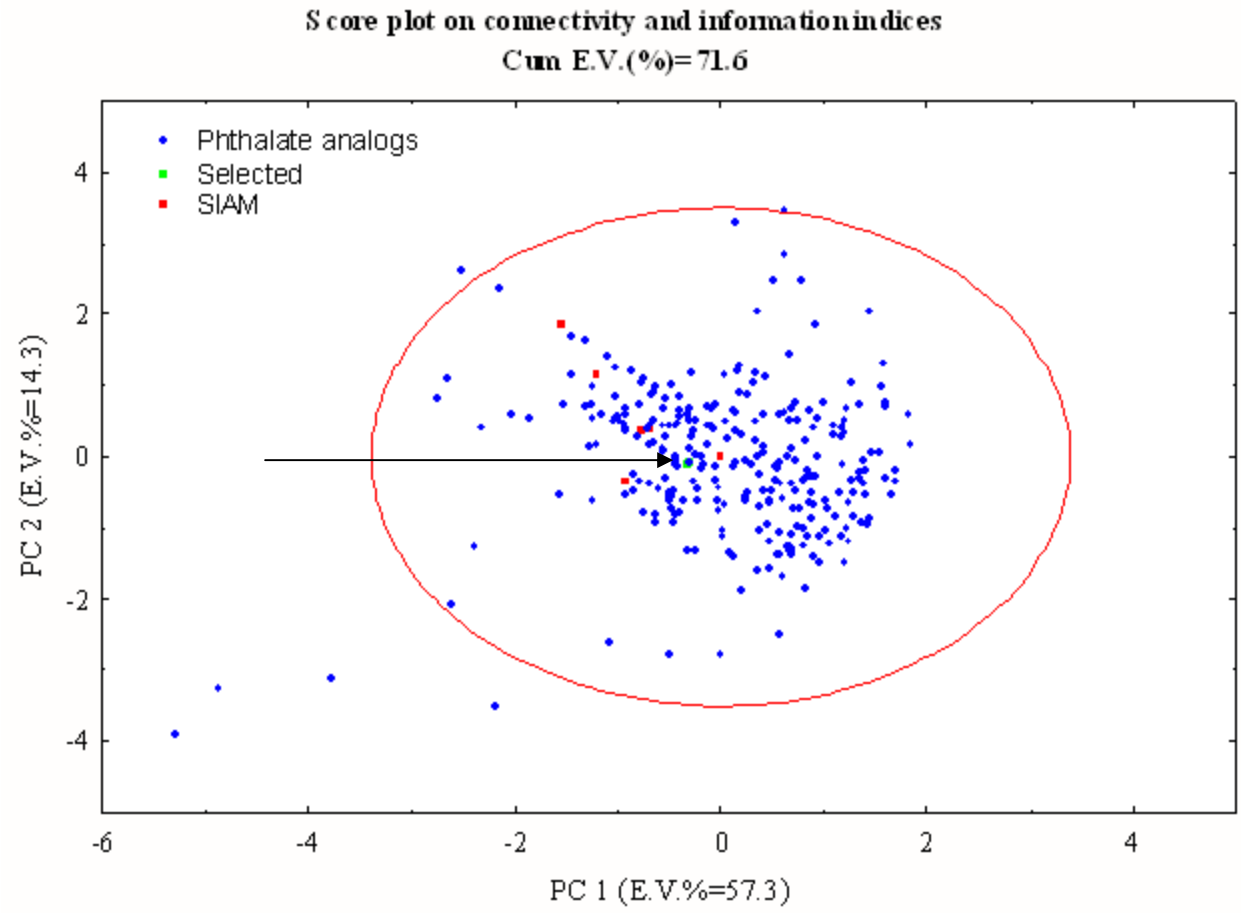
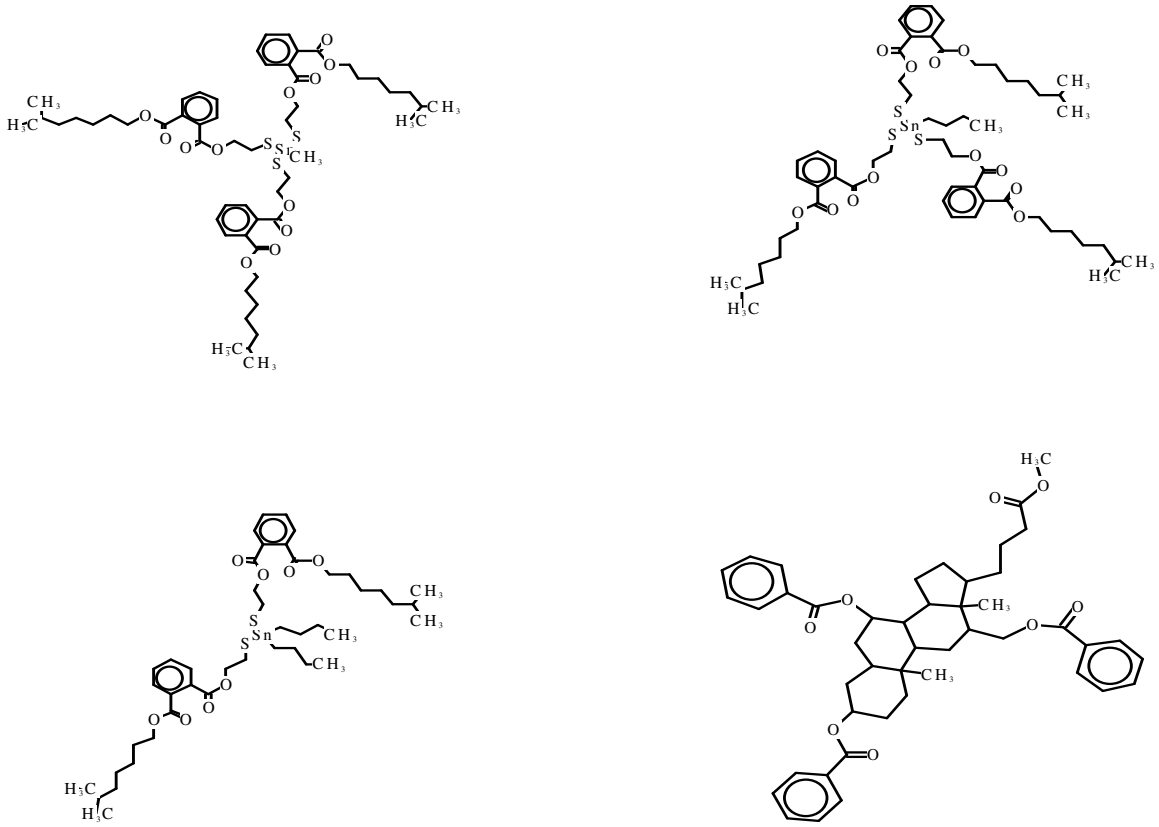


Figure 4. Plot of the first two principal components calculated from connectivity and information indices.

Examples of outliers in the PCA plot (chemicals outside the ellipse in the lower left corner)



Examples of overlapping chemicals in the PCA plot (indicated with an arrow)

Figure 5. Examples of structures that have substantially different chemical features compared to the rest of selected chemicals (outliers), and of very similar structures that are clustered together (very close analogues).

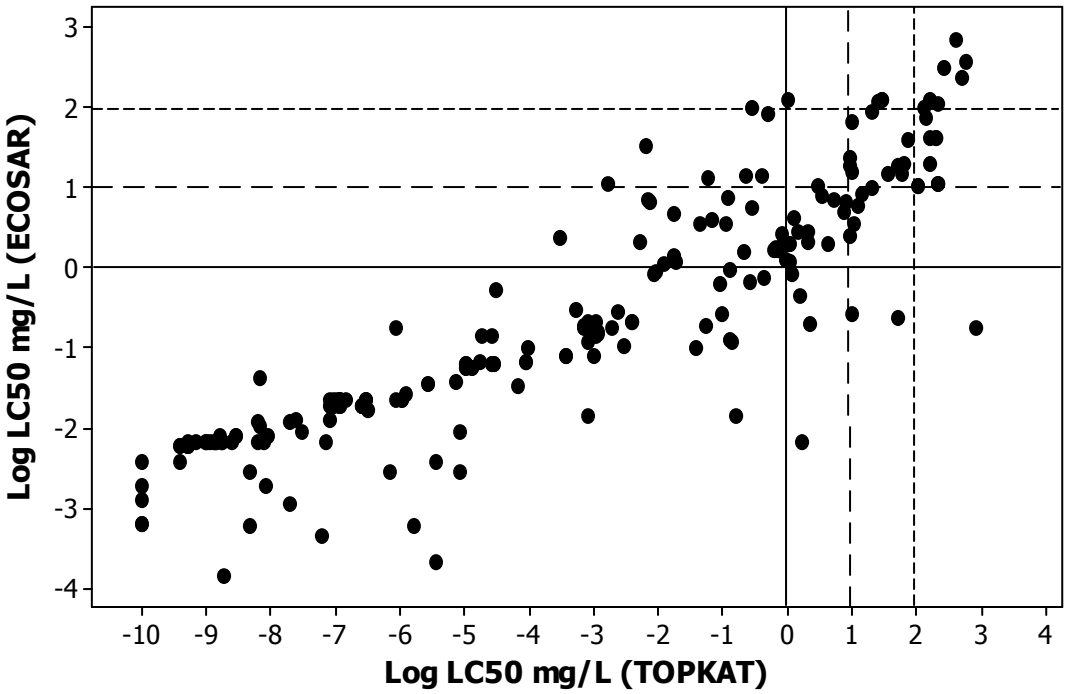


Figure 6. Correlation between log toxicity values to fathead minnow predicted by ECOSAR and TOPKAT. The lines are drawn according to regulatory cut-offs: 1 mg/L (solid line), 10 mg/L (dashed line), and 100 mg/L (dotted line). The chemicals in the lower left corner are predicted to be very toxic by both programs and the chemicals in the upper right corner are predicted to of no concern. Only chemicals predicted within the TOPKAT optimum prediction space were used.

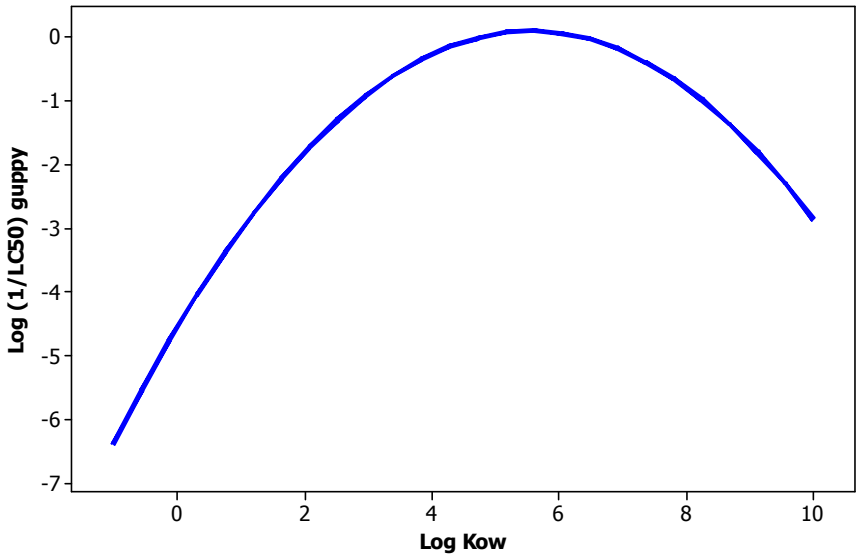


Figure 7. Parabolic relationship between logKow and toxicity. The function, observed by Hermens *et al*. (8), was determined until log Kow of approximately 6. It was prolonged in this plot until 10 as a maximum upper limit for measuring of log Kow.

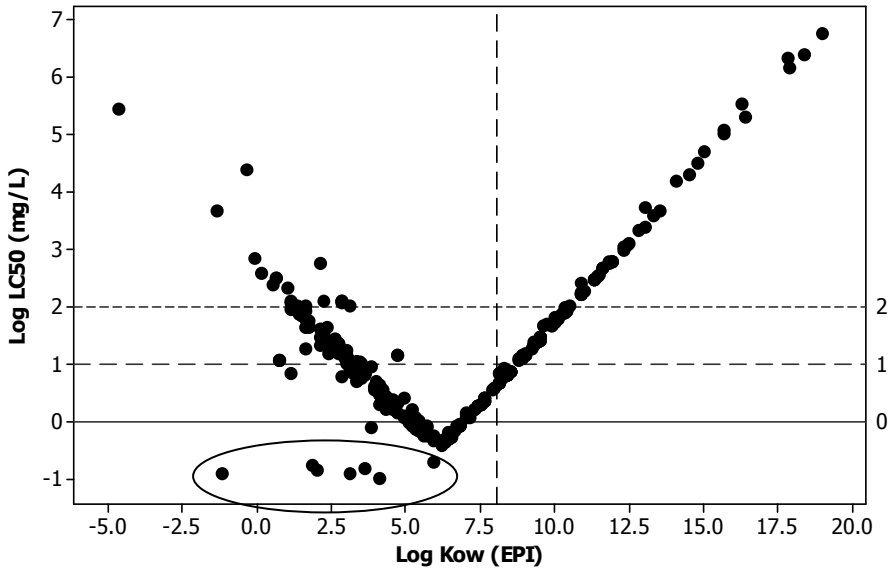


Figure 8. Bilinear relationship between logKow and toxicity. The three reference lines indicate the regions of the four toxicity categories: very toxic (VT) below 0; toxic (T) between 0 and 1; harmful (H) between 1 and 2, and no concern (NC) above 2. The dashed line represents the logKow value, which corresponds to the solubility limit of the phthalate esters. The circled chemicals are possible outliers due to the possibility of acting by a more reactive mechanism.

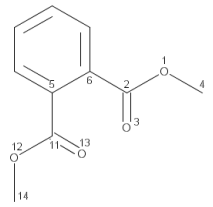
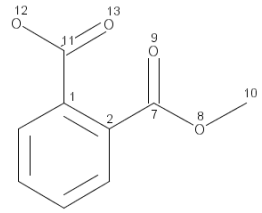
Cluster	Signature
1) 79 predicted non-sensitisers	 Cluster 1
2) 27 predicted sensitisers	 Cluster 2

Figure 9. Substructural fingerprints for discriminating between sensitising and non-sensitising phthalate esters

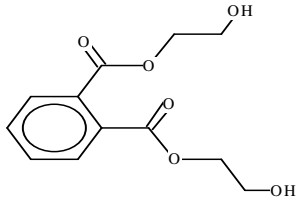
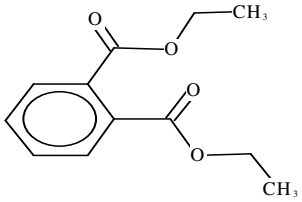
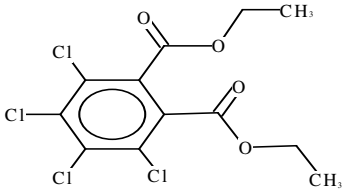
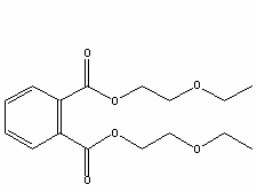
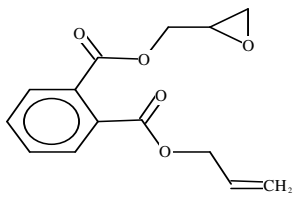
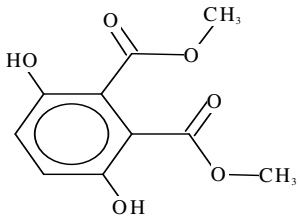
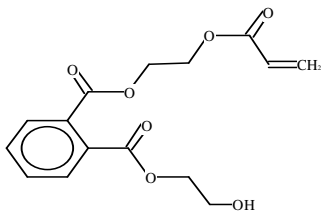
 Log Kow (estimated): 0.12	 Log Kow (estimated): 2.65	 Log Kow (estimated): 5.22
No concern	Harmful	Very toxic
logKow = 0.12	logKow = 2.65	logKow = 5.22
1,2-benzenedicarboxylic acid, bis(2-hydroxyethyl) ester	diethyl phthalate	diethyl 3,4,5,6-tetrachlorophthalate
		
	Harmful	
	logKow = 2.10	
	1,2-Benzenedicarboxylic acid, bis(2-ethoxyethyl) ester	
Examples of possible trend breakers (should not be predicted by the ester model)		
		
logKow = 2.05	logKow = 1.96	logKow = 1.08
Allyl 2,3-epoxypropyl phthalate	Dimethyl 3,6-dihydroxyphthalate	1,2-Benzenedicarboxylic acid, 2-hydroxyethyl 2-[(1-oxo-2-propenyl)oxy]ethyl ester

Figure 10. Modifications in chemical structure can change the toxicity classification

Appendix 2
Data gap filling in chemical categories: Case study using human health endpoints of the ethylene glycols category

Evelin Fabjan, Ivanka Tsakovska and Andrew Worth

European Chemicals Bureau

2 Data gap filling in chemical categories: Case study using human health endpoints of the ethylene glycols category

2.1 Summary

In the absence of sufficient relevant and reliable experimental data for a chemical, one or more data gaps can be filled by non-testing methods to finalise the hazard and/or risk assessment. The following approaches for filling data gaps in chemical categories can be applied:

a) Read-across
b) Trend analysis and use of computational methods based on internal models (internal QSARs)
c) Use of computational methods based on external models (external QSARs)

The data obtained by applying these methods has to be suitable for classification and labelling and/or risk assessment. Therefore it has to be suitable also for derivation of reference doses, such as the Derived No Effect Level (DNEL), in cases where the respective testing protocol allows the derivation of a dose descriptor value (NOAEL/LOAEL/BMDL).

When applying these methods for filling data gaps there is always some uncertainty on the accuracy of the predicted value, as there is when applying any model. The level of uncertainty depends in this case on the robustness of the category, which depends on various factors. In risk assessment, when, a dose descriptor such as the Lowest-Observed-Adverse-Effect Level (LOAEL), the No-Observed-Adverse-Effect Level (NOAEL) or Benchmark Dose (BMD) is derived by application of read-across or computational methods is used as the quantitative basis for deriving DNELs, the question arises as to whether an additional assessment factor is necessary to account for the uncertainty in the application of these methods.

The aim of this case study was to illustrate the application of the three approaches for filling data gaps in the category approach, and to test the use of category reporting format included in the REACH guidance documentation (1). In addition, the treatment of uncertainty in read-across estimates was considered. The case study is largely based on a category that was established and assessed in the OECD HPV Chemicals Programme at SIAM 18 in April 2004. The information used to build the case study was provided from the SIDS Initial Assessment Report for this category (2). It is emphasised that the study was not intended to be an evaluation of the OECD category, nor was it intended to derive "no-effect levels" for risk assessment.

2.2 Methods

2.2.1 Generation of QSAR predictions using TOPKAT

TOPKAT is a statistical system developed by Accelrys, Inc (3) consisting of a suite of QSAR models for a range of different endpoints. There are currently 16 modules for the following endpoints: aerobic biodegradability, Ames mutagenicity, *Daphnia magna* EC50, developmental toxicity, fathead minnow LC50, FDA rodent carcinogenicity, NTP rodent carcinogenicity ocular irritancy, logKow, rabbit skin irritancy, rat chronic LOAEL, rat inhalation toxicity LC50, rat Maximum Tolerated Dose (MTD), rat oral LD50, skin sensitisation, and weight-of-evidence rodent carcinogenicity (3).

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

The rat oral LD50 module of the TOPKAT (v3.1) includes 19 QSAR regression models. The models are based on a number of structural, topological and electrotopological indices, and make predictions of the oral acute median lethal dose in the rat (LD50). The models report results in units of chemical weight/body weight. The TOPKAT rat oral LD50 models are based on experimental values of 4000 chemicals from the RTECS.

The rat inhalation LC50 module contains five submodels related to different chemical classes. For the model development only exposure times in the range of 0.5 to 14 hours were accepted. In order to normalise the data from different durations of exposure, it was assumed that, within the range of adjustment, toxicity was proportional to duration. Thus, the units that were modeled were mg/m3/hour, and the predicted values are in the same units.

The chronic rat LOAEL module comprises five QSAR models. For the model development 393 uniform experimental LOAEL values were used. Each model predicts chronic LOAEL values in the rat in weight/body weight units, along with 95% confidence limits. Data for this TOPKAT module was derived from three sources: US EPA documents, National Cancer Institute/National Toxicology Program (NCI/NTP) Technical Reports, and the open scientific literature. All data were for oral rat chronic studies of at least 1year's duration.

The developmental toxicity potential module comprises three statistically significant and cross-validated QSAR models. Each model applies to a specific class of chemicals. The discriminant models 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, and the probability above 0.7 signifies developmental toxicity potential.

2.2.2 Generation of QSAR predictions using the BfR rulebase

The BfR rulebase (4-10) is based on the combined use of two predictive approaches: a) physicochemical exclusion rules to identify chemicals with no skin irritation/corrosion or eye irritation/corrosion potential; and b) structural inclusion rules to identify chemicals with skin irritation/corrosion or eye irritation/corrosion potential. The current (2005) version of the DSS is based on a training set of 1358 chemicals with experimental data for skin and eye irritation and corrosion. Only pure substances were considered (95% purity). The training set was compiled from confidential data submitted under the EU New Chemicals Notification procedure, and contained within the New Chemicals Database (NCD).

2.3 Results: category reporting format

1.	**Category definition and its members**
1.1.	Category Definition
1.1.a.	Category Hypothesis The category includes ethylene glycol (EG) and higher glycols (di-, tri-, tetra-, and penta-), which are closely related in molecular structure and have similar physicochemical properties, that differ in a regular and expected way as a result of increasing molecular weight and consistent functionality of a relatively less stable hydroxyl moiety on each end of the molecule. The potential for toxicological effects is also expected to change consistently - as the molecular weight increases, the potential for systemic, reproductive and developmental effects/toxicity is expected to decrease.
1.1.b.	Applicability domain (AD) of the category The category applies to substances that can be represented by the following generic molecular structure: $HO(CH_2\ CH_2O)_n\ H$, where n= 1,2,3,4, or 5 for the category members. All category members are composed of two primary alcohol (hydroxy) groups and the members differ from each other only in the number of oxyethylene units. All category members except EG contain at least one ether linkage.
1.1.c.	List of endpoints covered The category approach was applied to mammalian toxicity endpoints (except for genotoxicity/mutagenicity and carcinogenicity which are not reported in this format)
1.2.	Category Members See Table 1.
1.3.	Purity / Impurities See Table 2.
2.	**Category justification** The category members are all liquid substances of low volatility and high water solubility. The physico-chemical properties such as melting point, boiling point and density increase as the number of oxyethylene units increases, while the vapour pressure, partition coefficient, and surface tension generally decrease as the number of oxyethylene units increases. Based on available data, it can be concluded that as the molecular weight increases (above DEG in a homologous series), the potential for systemic, reproductive, and developmental toxicity decreases. This pattern of toxicity is consistent with a likely decrease in absorption with increasing molecular weight, though available data to serve this basis for comparison are limited and inconclusive. All substances are expected to be well absorbed by the oral route. The normal synthesis of these compounds involves the hydrolysis of ethylene oxide (EO) to initially produce EG, which then reacts with subsequent molecules of EO to produce the higher glycols in increasing order. All category members can be represented by the following generic molecular structure: $HO(CH_2\ CH_2O)_n\ H$ Where n = 1, 2, 3, 4, or 5 for the category members. All category members therefore possess two primary alcohol (hydroxy) groups and the members differ from each other only in the number of oxyethylene units. All category members except EG contain at least one ether linkage, but the ether functionality is extremely stable relative to the hydroxy functionality. Because of this and other similar chemistries, it is appropriate to classify EG and the higher glycols as a single group. The limits of this category

	are fairly well defined since at n=6 to 8 the absorption from ingestion decreases significantly (He *et al*, 1998; 11) and the materials start to become solids. Above n=8, the physicochemical attributes change very little and are represented well by another category named poly(ethylene oxide).
3.	**Data matrix** A matrix on data availability was created for physico-chemical properties (Table 3) and mammalian toxicity endpoints (excluding genotoxicity/mutagenicity and carcinogenicity), Table 4. From the available data a clear trend in physicochemical properties and toxicity is observed. Toxicity seems to decrease with increasing chain length.
4.	**Conclusions per endpoint for C&L, PBT/vPvB and dose descriptor**

2.4 Results: data gap filling for individual endpoints and category members

Details of the data gap filling are given in Table 4. This is not intended to be a comprehensive evaluation of the category (e.g. evaluation of toxicity data) and derivation of the "critical" no-effect levels to be used in risk assessment, but more an illustration of possible methods/approaches that can be used for this purpose; this also in view of the very high doses used in most studies and very high LD(C)50s/NOAELs which are not of great relevance for risk management.

2.4.1 Acute toxicity

Experimental information from animal studies is available for all members of the category and all substances are of low acute oral toxicity. For this endpoint however, an example of a derivation of an internal QSAR that can be used for prediction of toxicity is given below.

All substances appear to be also of low inhalation and dermal toxicity. For EG information from animal experiments is missing for both acute inhalation and dermal toxicity.

These data gaps could be filled with:

a) read-across from the closest analogue DEG,
b) trend analysis (internal QSAR)
c) an external model (e.g. TOPKAT 6.1 prediction).

Rat oral LD50

Prediction for the category:

SMILES	Compound ID	Within OPS	Compound in TOPKAT Database	Reported Endpoint, mg/kg	Computed Rat Oral LD50, mg/kg
OCCO	EG	TRUE	TRUE	4700	6500
OCCOCCO	DEG	TRUE	TRUE	12565	9900
OCCOCCOCCO	TEG	TRUE	TRUE	17000	>10000
OCCOCCOCCOCCO	tetraEG	TRUE	TRUE	29000	>10000
OCCOCCOCCOCCOCCO	pentaEG	TRUE	FALSE		>10000

Example of a derivation of an internal model (internal QSAR):
Based on the experimental data for acute toxicity, internal QSAR modelling was applied. For this purpose the experimental acute toxicity values as reported in TOPKAT training set for four members of the category (EG, DEG, TEG and tetraEG) were used. A reasonable correlation was outlined between acute toxicity and log Kow. The following internal QSAR equation was derived:

$\log(1/LD50) = -2.48 + 0.93\ \log Kow \qquad (R^2 = 0.703)$.

A plot of the log (1/LD50) against the logKow is shown in Figure 1.

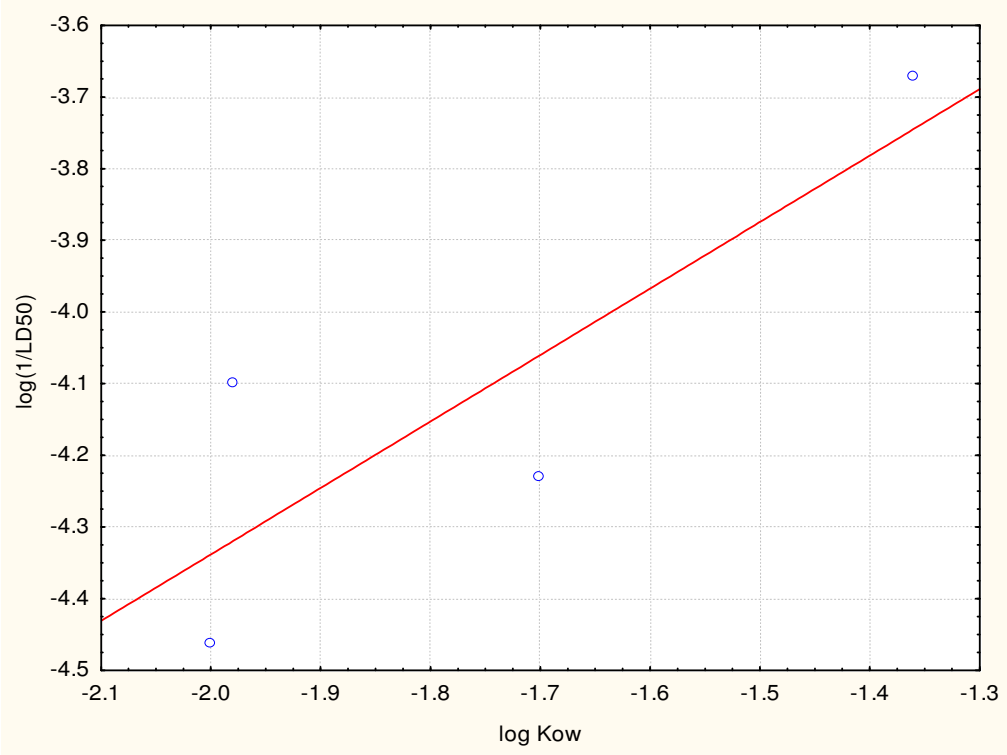


Figure 1. Internal QSAR for the rat oral LD50 for ethylene glycols

On this basis the oral LD_{50} value for Pentaethylene Glycol (pentaEG) was predicted to be 41591 mg/kg. However, a prediction based on a model derived by using such a low number of data is usually not considered to be very reliable.

Acute dermal toxicity
Given that EG seems to be the most toxic compound in this category and taking into account information on dermal absorption, performing a read-across from DEG could give an overestimation of the LD50 value for dermal toxicity (i.e. an underestimation of the dermal toxicity). Therefore, in this case and assessment factor might be considered, or the application of read-across might not be appropriate. No TOPKAT 6.1 prediction was possible and in view of the data available the application of trend analysis is also not possible for this endpoint.

Acute inhalation toxicity
Again, EG could be considered as being the most toxic within the category, so the application of read-across is questionable. An LC50 value was calculated by application of the TOPKAT 6.1 model pointing to low acute inhalation toxicity, however due to the type of data available (very low

toxicity), this value can not be quantitatively compared to the values for other members of the category.

Prediction for the category:

SMILES	Compound ID	Within OPS	Computed Rat Inhalational LC50, mg/m3/H
OCCO	EG	TRUE	6800
OCCOCCO	DEG	FALSE	254.8
OCCOCCOCCO	TEG	FALSE	268.6
OCCOCCOCCOCCO	tetraEG	FALSE	455
OCCOCCOCCOCCOCCO	pentaEG	FALSE	825.1

Only the prediction for EG is reliable.

In order to fill the data gaps similarity search was performed. In TOPKAT similarity is measured between two molecules with reference to a specific property. For every model TOPKAT computes a model-specific similarity distance between a query structure and a database compound with the smaller the distance, the greater the similarity. A TOPKAT assessment of a query structure is based on the hypothesis that the model parameters present in the query structure are the determinants of its toxicity. Therefore this hypothesis can be tested against similar compounds in the model's database. The similarity is scaled from 0– 1; the smaller the distance, the greater the similarity.

Similarity search outlined 2-(2-ethoxyethoxy)ethanol (Figure 2) as substance "similar" to the substances with ID=2÷5 (difference estimated by TOPKAT ,0.3) and presented in the TOPKAT training set. The rat inhalation LC50 for this substance is 5240 mg/m3/4H.

OCCOCCOCC

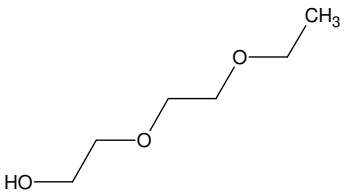


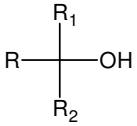
Figure2. Structure of 2-(2-ethoxyethoxy)ethanol

2.4.2 Skin/eye irritation

Some information on skin/eye irritation was available for all the members of the category and substances appear to be weakly irritating to skin and eye. These endpoints were also checked with the BfR rulebase and the findings direct to possible skin and eye irritation potential of these substances

Category predictions:

In the list of BfR structural alerts for presence of skin and eye irritation is the following alert:



R=aliphatic chain; $R_{1,2}$=H or aliphatic chain

The finding indicates possible skin and eye irritation potential of the substances.

2.4.3 Skin sensitisation

Based on available data from animal studies (GPMT test) EG, DEG, TEG and tetraEG are not skin sensitizers. This information is missing for pentaEG but by applying read-across from the other members of the category, pentaEG could also be considered as non-sensitising.

2.4.4 Repeated dose toxicity (oral)

There is a rather large variability in the number and type of studies performed (in terms of duration and animal species used) among the different members of the category. This is an example which illustrates how diverse the available data can be within a category. In this case comparable studies in terms of duration with the same species are available for some of the members (EG and DEG) but not for others. From the available data, a clear trend of decreasing in toxicity with increasing molecular weight is observed, in particular for substances having higher molecular weight than DEG.

For EG and DEG, information from animal studies is available for both sub-acute/sub-chronic and chronic toxicity, for TEG and tetraEG chronic toxicity data is missing, while for pentaEG no data on repeated dose toxicity is available.

To fill in the data gaps, the possibility of applying trend analysis, computational methods based on internal models, analogue read-across and external models was explored.

Trend analysis and computational methods based on internal models

Given the small size of the category and high variability in the data, the relevance of performing a trend analysis or development of an internal QSAR is questionable.

Concerning the NOAEL values, although the trend is seen, there is a big difference between the NOAEL values for DEG and TEG. For example, in case that the sub-chronic NOAEL would be estimated by interpolation from the available sub-chronic information from both EG and TEG (as for example the average of the two values) this would most probably be overestimated. Based on the available data, it might be appropriate in this case, to divide the category in two subcategories for repeated-dose toxicity endpoints and consider TEG as a breakpoint chemical. However, there are also some indications that kidney toxicity observed after exposure to EG might be a result of a different mode of action than the kidney toxicity caused by DEG. The renal toxicity of EG upon repeated dosing in rats appears to be due to crystal nephropathy caused by calcium oxalate crystals, indicating that kidney oxalic acid concentrations are key to EG renal effects. Oxalic acid has been shown to be a minor metabolite of DEG and TEG (low levels observed in urine) and this suggests that oxalic acid is probably not the cause of renal toxicity observed after repeated exposure to these compounds.

Read-across

Given the difficulty to perform trend analysis the data gaps for sub-acute/sub-chronic toxicity of pentaEG could be filled by:

- read-across from EG, which is the most toxic; this approach can be considered as very conservative since a clear trend of decreasing toxicity with increasing molecular weight is observed in the category;
- read-across from the closest analogue for which information is available, namely from tetraEG for sub-acute and from TEG for sub-chronic toxicity. Due to the trend of decreasing toxicity with increasing molecular weight, it is assumed that pentaEG would be less toxic

than the other two chemicals; therefore the values could also be considered conservative, resulting in no need for an additional assessment factor.

Regarding chronic toxicity, information from chronic animal studies is available only for EG and DEG. The lowest (sub)-chronic NOAEL has been obtained in a 225 days feeding study with DEG in rats. Since DEG is the closest analogue with chronic toxicity data to TEG, tetraEG and pentaEG, the data gap for chronic toxicity could be filled by read-across from DEG. Due to the trend of decreasing toxicity with increasing molecular weight the NOAEL value from DEG could be considered conservative and result in no need for an additional assessment factor.

Another approach could be to apply an assessment factor of e.g. 2 to the NOAEL from the sub-chronic study with TEG (sub-chronic to chronic) and an assessment factor of 6 to the NOAEL of the sub-acute study for tetraTEG (sub-acute to chronic), to account for the differences in exposure duration in studies and to derive the chronic DNEL. However, in this case the question arises as to whether a trend analysis on these derived values provide a reliable means of filling the data gaps for substances without data.

Use of computational methods based on external models

A TOPKAT 6.1 prediction for chronic rat LOAEL was performed, resulting in an overestimation of the LOAEL for DEG if compared with the lowest experimental LOAEL available (Table 4) and an estimated LOAEL for TEG, which was higher than the one for DEG. TetraEG and pentaEG were outside the applicability domain of the model.

Predictions for the category:

SMILES	Compound ID	Within OPS	Compound in TOPKAT training set	Reported Endpoint	Computed Chronic LOAEL, mg/kg
OCCO	EG	TRUE	TRUE	1000 mg/kg	828.6
OCCOCCO	DEG	TRUE	FALSE		1500
OCCOCCOCCO	TEG	TRUE	FALSE		3400
OCCOCCOCCOCCO	tetraEG	FALSE	FALSE		4900
OCCOCCOCCOCCOCCO	pentaEG	FALSE	FALSE		8600

Substance 1 is presented in the TOPKAT training set and the reported LOAEL is 1000 mg/kg. Substances 4 and 5 are outside the applicability domain of the model.

2.4.5 Reproductive toxicity

Data from a two-generation study are available for EG and data from studies conducted by the Reproductive Assessment by Continuous Breeding (RACB) protocol are available for DEG and TEG. No information on reproductive toxicity is available for tetraEG and pentaEG. A trend of decreasing toxicity with increasing molecular weight is observed. The data gaps for tetraEG and pentaEG could be filled by:

- read-across from EG which is the most toxic; this approach can be considered as very conservative as a clear trend of decreasing toxicity with increasing molecular weight is observed in the category;
- read-across from TEG, the closest analogue with available data. Due to the observed trend within the category, it can be assumed that tetra and pentaEG are less toxic than TEG and therefore the NOAEL from TEG can be considered conservative for tetraEG and pentaEG so it should not be necessary to add an additional assessment factor.

2.4.6 Developmental toxicity

Several developmental toxicity studies in mice and rat were available for EG, DEG and TEG while information on developmental toxicity was available for tetraEG and pentaEG. To fill in the data gaps, the possible application of trend analysis and computational methods based on internal models, analogue read-across and use of computational methods based on external models was explored.

Trend analysis and computational methods based on internal models:

Similarly as in case of repeated dose toxicity, given the small size of the category it appeared not appropriate to perform the trend analysis or to develop an internal QSAR. Taking into consideration only the lowest NOAELs observed in studies with mice, a clear trend of decreasing toxicity with increasing molecular weight cannot be observed. Based on NOAELs, EG appears to be the most toxic, followed by TEG and DEG. It appears that in this case the shape of the trend in the category is distorted due to the dose spacing used in different studies. However, when comparing the BMDs of these three substances, the trend of decreasing toxicity with increasing molecular weight is observed: EG shows the highest toxicity, followed by DEG and then TEG. The same trend is observed in the rat studies.

Read-across

Given the difficulty to perform trend analysis to fill the data gaps for developmental toxicity of tetraEG and pentaEG, the possibilities to fill these data gaps could be the following:

- read-across from the substance that has the lowest NOAEL value – EG, which would be a very conservative approach and would indicate that tetraEG and pentaEG are potential developmental toxicants;
- read-across from the substance that is considered to be the closest analogue for which information is available - TEG. In this case, if only information on NOAELs from mice were available (or considered), no apparent trend in toxicity within the category could be observed. The NOAEL value for TEG is considerably lower than the NOAEL value for DEG. In this case the NOAEL for tetraEG and pentaEG would be underestimated. To read across from TEG, BMD LCL values (95% lower confidence limit on the benchmark dose) would give a better prediction. From the rat studies a clear trend of decreasing toxicity with increasing molecular weight is observed. Due to the observed trend within the category, it can be assumed that tetra and pentaEG are less toxic than TEG. Therefore the NOAEL/BMDL from TEG can be considered conservative for tetraEG and pentaEG and application of an additional assessment factor should not be necessary.

Use of computational methods based on external models

A TOPKAT 6.1 prediction for developmental toxicity was performed, resulting in a prediction of possible developmental toxicity for EG, DEG and TEG. However, a dose descriptor value was not calculated. TetraEG and pentaEG were outside the domain of applicability of the model.

Predictions for the category:

SMILES	Compound ID	Within OPS	Compound in TOPKAT training set	Computed Probability of Developmental Toxicity
OCCO	EG	TRUE	FALSE	1
OCCOCCO	DEG	TRUE	FALSE	1
OCCOCCOCCO	TEG	TRUE	FALSE	0.997
OCCOCCOCCOCCO	tetraEG	FALSE	FALSE	0.965
OCCOCCOCCOCCOCCO	pentaEG	FALSE	FALSE	0.573

The results point to possible developmental toxicity potential for the first three substances. Substances 4 and 5 are outside the applicability domain of the model. None of the substances is present in the TOPKAT training set.

2.5 Discussion

2.5.1 Read-across

Within the context of a chemical category, read-across can be applied in the following ways:

- one-to-one (one source chemical, one target chemical)
- many-to-one (many source chemicals, one target chemical)
- one-to-many (one source chemical, many target chemicals)
- many-to-many (many source chemicals, many target chemicals)

When read-across is applied from one-to-one chemical or one–to–many chemicals, the uncertainty is usually considered greater than in cases where there are multiple source chemicals. To address this uncertainty, the application of an assessment factor might be considered. However, the question arises as to what assessment factor should be used.

The uncertainty in read-across is lower when a clear trend is observed and when a conservative value for the target chemical can be determined. This conservative value could for example be the lower 95^{th} confidence limit (i.e. the 5^{th} percentile) of the values of the analogues or, in a homologous series where a clear trend is observed, the value of the closest analogue that is considered to be of higher toxicity than the target chemical. In these cases this approach might be considered sufficiently conservative to account for the uncertainty without the need for an assessment factor.

In cases where there is no apparent trend in toxicity within the category and the dose descriptor values for the relevant endpoint vary considerably among the members, it might be more appropriate not to apply the read-across at all, rather than apply an assessment factor.

Cases in which the application of an assessment factor might also be considered are those where there is confidence that the two (or more) analogues are comparable, but there are reasons to suspect that the target chemical might be more potent than the source chemical(s). Again, in this case it might be more appropriate not to apply read-across, rather than to apply an assessment factor to account for the uncertainty.

In some cases it might not be necessary to derive a specific value (number), but it might be sufficient to define a dose range within which a NOAEL (DNEL) resides.

2.5.2 Trend analysis and use of computational methods based on internal models

The use of trend analysis and computational methods based on internal models will to a great extent depend on the quality, reliability, variability and the level of details available (e.g. availability of the confidence intervals of the NOAELs) of the experimental data. These methods have not yet been extensively used in human toxicology for regulatory purposes. No examples of application of these methods for evaluation of existing categories were found. This might be due to the fact that usually these approaches can be applied only to large categories, for which a trend is observed and for which suitable and comparable data are available for a sufficient number of category members.

One of the problems in applying these methods in human toxicology and in particular in case of existing substances is that there usually is a large variability in the experimental data available (e.g. with respect to the type of study, duration, dose spacing, species used). So, for example for the repeated dose toxicity, there might be studies of different duration, species and very different dose spacing available for different members of the category (see Table 4). Furthermore, the NOAEL depends critically on study design, choice of doses, dose spacing and group size. Two studies using the same chemical that are performed according to a completely identical study design (except for the dose spacing) can identify very different NOAELs because dose-spacing is a major determinant of this value. This might distort the shape of the trend in the category (see example on developmental toxicity in mice in Table 4) and make the modelling difficult (less reliable) or not applicable. In this respect it would be better to use the benchmark dose (BMD) approach (if possible), as this approach makes use of all the dose-response data and is less dependent on the study design, dose spacing or group size, compared to the NOAEL approach (see example on developmental toxicity in mice in Table 4).

The use of trend analysis and computational methods based on internal model is expected to work best (and the predictions are likely be most reliable) for homologous series of chemicals, for which the parameter(s) associated with the change (trend) in toxicity (e.g. physicochemical properties) can easily be identified. Furthermore, the proportionality factor that links the structural differences and the change in toxicity between the source and the target chemical should be more or less constant within the category, for the prediction to be accurate. This is particularly true in cases where there are many data gaps within the category and the spacing between the target chemical and the chemical from which the value is interpolated/extrapolated is large.

In general, if a trend analysis is performed, there should be sufficient confidence in the trend to make a correction (e.g. using the proportionality factor of an internal model) or a conservative estimations without the need to apply an additional assessment factor. In cases where the predictions are not considered accurate, it might be more appropriate not to apply trend analysis in the first place (and perhaps apply just the read-across), rather than apply an assessment factor to account for the uncertainty.

In some cases it might not be necessary to derive a specific value (number), but to define a dose range within which the dose descriptor (e.g. NOAEL), and consequently the DNEL, resides could be sufficient.

2.5.3 Use of computational methods based on external models

If relevant internal (Q)SAR models cannot be obtained, external models can be sought in the literature, or in external databases and tools. The predictions of the external models will be suitable only if the query compound falls within the applicability domain of the model, meaning that the model is likely to generate a reliable prediction for the compound. The applicability domain

assessment is necessary since (Q)SAR models are based on empirical knowledge about specific chemicals and therefore they are associated with limitations in terms of chemical structures, physicochemical properties and the mechanisms of action for which the models can reliably be used. A thorough analysis of ways to formulate applicability domains for (Q)SAR models is given in (12-15).

In cases where relevant and reliable external models are used, it should not be necessary to apply an assessment factor specifically to account for the uncertainty in the prediction. The (average) uncertainty (error) in the model predictions should be known from the statistical characteristics of the model. However, depending on the actual endpoint predicted and the endpoint of interest, some conversion may be necessary (e.g. from mol/L to g/L).

2.6 Conclusions

On the basis of the results of this study, it is concluded that:

1. Read-across is likely to be the method most often used in category approach, as it appears that a reliable trend analysis and development of internal QSARs will be possible only in the case of larger categories, having reliable and comparable data.

2. Wherever possible, the uncertainty in the read-across should be quantified (e.g. by providing confidence limits).

3. If there is a clear trend or a fairly constant value within a category for the endpoint of interest (or for a related endpoint), it should be possible to obtain a reliable dose descriptor (e.g. a DNEL) by read-across without the need to apply an additional assessment factor to account specifically for the uncertainty of using this method. For example, the nearest conservative value of a source substance within the category could be used for the target substance, or a proportionality factor could be applied to correct the estimate.

4. In cases where there is no trend in toxicity and there are differences in potency between members, or in cases in which the target substance is on the boundary of the category, it might be tempting to apply an assessment factor just to account for the uncertainty. In such a case, however, the question arises as to what assessment factor should be used, and how this should be decided. Case-by case decisions might lead to an inconsistent use of assessment factors. Thus, if there is insufficient data to assess the uncertainty in the read-across, or if the uncertainty is considered too high for the specific purpose of the read-across, it might be preferable not to apply the read-across.

2.7 References

1. ECB (2007). Final report of REACH Implementation Project (RIP) 3.3-2. Technical Guidance Document to Industry on the Information Requirements for REACH. European Commission-Joint Research Centre, Ispra, Italy. Accessible from: http://ecb.jrc.it/REACH/RIP_FINAL_REPORTS

2. OECD (2004). SIDS Initial Assessment Report for SIAM 18 (draft for submission, January 22, 2004).

3. TOPKAT, Accelrys, Inc. 10188 Telesis Court, Suite 100 San Diego, CA 92121 USA, http://www.accelrys.com.

4. Gerner I, Zinke S, Graetschel G & Schlede E (2000). Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Creation of fundamental rules for a decision support system *Alternatives to Laboratory Animals* **28**, 665-698.

5. Gerner I, Graetsche G, J. Kahl J & Schlede E (2000). Development of a Decision Support System for the Introduction of Alternative Methods into Local Irritation/Corrosivity Testing Strategies. Development of a Relational Database. *Alternatives to Laboratory Animals* **28**, 11-28.

6. Gerner I, Schlegel K, Walker JD & Hulzebos E (2004). Use of physicochemical property limits to develop rules for identifying chemical substances with no skin irritation or corrosion potential. *QSAR & Combinatorial Science* **23**, 726-773.

7. Hulzebos E, Walker JD, Gerner I & Schlegel K (2005). Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential. *QSAR & Combinatorial Science* **24**, 332-342.

8. Hulzebos EM, Maslankiewicz L & Walker JD (2003). Verification of literature derived SARs for skin irritation and corrosion. *QSAR & Combinatorial Science* **22**, 351-363.

9. Walker JD, Gerner I, Hulzebos E & Schlegel K (2004). (Q)SARs for Predicting Skin Irritation and Corrosion: Mechanisms, Transparency and Applicability of Predictions. *QSAR & Combinatorial Science* **23**, 721-725.

10. Gerner I, Liebsch M & Spielmann H (2005). Assessment of the Eye Irritating Properties of Chemicals by Applying Alternatives to the Draize Rabbit Eye Test: The Use of QSARs and *In vitro* Tests for the Classification of Eye Irritation *Alternatives to Laboratory Animals* **33**, 215-237.

11. He YL, Murby S, Warhurst G, Gifford L, Walker D, Ayrton J, Eastmond R & Rowland M (1998). Species differences in size discrimination in the paracellular pathway reflected by oral bioavailability of poly(ethylene glycol) and D-peptides *Journal of Pharmaceutical Sciences* **87**, 626-633.

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

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

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

15. Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J & Mekenyan O (2005). A stepwise approach for defining the applicability domain of SAR and QSAR models. *Journal of Chemical Information and Computer Sci*ence **45**, 839-49

Table 1. Members of the Ethylene Glycol Category

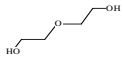
Chemical Name	IUPAC Name	CAS No.	Molecular Weight	Molecular Formula	Structural Formula (Smiles)
Ethylene Glycol (EG)	1,2-ethanediol	107-21-1	62.1	C2H6O2	HOCH2CH2OH (OCCO) 
Diethylene Glycol (DEG)	2,2'-oxybisethanol	111-46-6	106.1	C4H10O3	HO(CH2CH2O)2H (OCCOCCO) 
Triethylene Glycol (TEG)	2,2'-(1,2-ethanedylbis(oxy)bis ethanol)	112-27-6	150.2	C6H14O4	HO(CH2CH2O)3H (OCCOCCOCCO) 
Tetraethylene Glycol (tetraEG)	2,2-(oxybis(1,2-ethaneidyloxy)bis-)ethanol	112-60-7	194.2	C8H18O5	HO(CH2CH2O)4H (OCCOCCOCCOCCO) 
Pentaethylene Glycol (pentaEG)	--	4792-15-8	234.3	C10H22O6	HO(CH2CH2O)5H (OCCOCCOCCOCCOCCO) 

Table 2. Composition and impurity profile for the members of the Ethylene Glycol Category category

Chemical Name	CAS Number	Composition
Ethylene Glycol (EG)	107-21-1	≥ 99 % pure
Diethylene Glycol (DEG)	111-46-6	≥ 99 percent pure (technical grade)
		High purity grade
		DEG ≥99.5 %
		EG ≤ 0.04 %
		TEG ≤ 0.1 %
		Water ≤ 0.1%
Triethylene Glycol (TEG)	112-27-6	≥ 99 % (technical grade)
		≥ 99.5 % (high purity grade)
Tetraethylene Glycol (tetraEG)	112-60-7	TetraEG, ≥ 96.0 %
		DEG, ≤2 %
		TEG ≤3 %
		PentaEG ≤1 %
		Water ≤ 0.2 %
Pentaethylene Glycol (pentaEG)	4792-15-8	Crude pentaEG (30-85 % pentaEG, 10-70% tetraEG, 2-15 % hexaethylene glycol, ≤ 5% TEG, ≤1% DEG
		PentaEG, 98%

Table 3. Data matrix for physicochemical properties of the members of the category

Category Member	**EG**	**DEG**	**TEG**	**tetraEG**	**pentaEG**
CAS No.	**107-21-1**	**111-46-6**	**112-27-6**	**112-60-7**	**4792-15-8**
Physical state	Liquid	Liquid	Liquid	Liquid	Liquid
Melting point (oC)	-13	-8	-5	-6.2	<0
Boiling point (oC)	197.6	245	287	327	211-347
Density (g/cm3)	1.1088	1.118	1.1274	1.1285	1.1372
Vapour pressure (hPa)	0.104	0.0104	0.00133	6.2×10^{-5}	$1.44\text{-}3.99 \times 10^{-7}$
Partition coefficient (Log Kow)	-1.36 -1.20	-1.98 -1.47	-1.7	-2.0	-2.3
Water solubility (mg/l)	Miscible	Miscible	Miscible	Miscible	Miscible
Henry's Law Constant (atm-m3/mole)	1.31 × 10-7 (Epiwin, 2003)	2.03 × 10-9 (Epiwin, 2003)	3.16 × 10-11 (Epiwin, 2003)	4.91 × 10-13 (Epiwin, 2003)	7.62 × 10-15 (Epiwin, 2003)
Surface Tension (dynes/cm)	48.4	48.5	45.2	44	ND
Flash point (oC)	111-116 (closed cup)	138-143 (closed cup)	117 (open cup)	182 (open cup)	>110
Autoignition temperature (oC)	398	229	375	358	ND
Flammability	Lower limit 3.2%	1.6-10.8%	0.9-9.2%	ND	ND

Table 4. Data matrix for Mammalian toxicity endpoints (excluding mutagenicity and carcinogenicity) for the members of the category. Data reported in red were derived by applying read-across or QSAR estimation

Category Member	EG	DEG	TEG	tetraEG	pentaEG
CAS No.	**107-21-1**	**111-46-6**	**112-27-6**	**112-60-7**	**4792-15-8**
Acute oral toxicity (rat)	LD50: 4000 – 13000 mg/kg	LD50: 25300 mg/kg	LD50 : 17000-22000 mg/kg	LD50 : 34700 mg/kg	LD50 :: > 16000 mg/kg LD50 :: 41591 mg/kg (estimation by an internal QSAR model)
Acute inhalation toxicity (rat)	LC50: 6800 mg/m3/h (TOPKAT 6.1 prediction)	0/10 deaths at substantially saturated vapour	0/10 deaths at 50 mg/L (aerosol)	0/6 deaths substantially saturated vapour	0/12 deaths at 2516 mg/m3 aerosol
Acute dermal toxicity(rabbit) LD50	Acute toxicity expected to be higher than for DEG.	12500 mg/kg	> 18000 mg/kg	22600 mg/kg	> 18200 mg/kg
Skin irritation (human)	Some evidence of irritation (humans)	Minimal irritation (humans)	Minimal irritation (humans)	Minimal irritation (humans)	Minor irritation (rabbit)
Eye irritation	Minimal irritation	Minimal irritation	Minimal irritation	Minor transient irritation	Minor transient irritation
Skin sensitization	Non-sensitizing (GPMT)	Non sensitizing (GPMT)	Non sensitizing (GPMT)	Non-sensitizing (GPMT)	Non sensitizing (read-across)
Repeated dose toxicity (oral)	16 weeks study (rat)* NOAEL: 71 mg/kg/d LOAEL: 180 mg/kg/d (kidney effects) 16-weeks study (rat; feed): 0, 150, 500, 1000 mg/kg/d NOAEL: 150 mg/kg/d LOAEL: 500 mg/kg/d (kidney effects) 2 years study (rat; feed): 0, 40, 200, 1000 mg/kg/d NOAEL: 200 mg/kg/d LOAEL:1000 mg/kg/d (several effects on urine parameters, urine calcium oxalate crystals, increased absolute and relative kidney weight etc.) 2 years study (mice; feed): 0, 1500, 3000, 6000, 12000 mg/kg/d NOAEL:1500 mg/kg/d LOAEL: 3000 mg/kg/d	32 days study (rat; feed): 0, 11 46, 180, 850 mg/kg/d* NOAEL: 150 mg/kg/d LOAEL: 850 mg/kg/d (increased kidney weight) 225 days study (rat, feed): 0, 51, 105, 234 and 1194 mg/kg/d NOAEL: 105 mg/kg/d LOAEL: 234 mg/kg/d (oxalate crystalluria and mild defects of renal function) 2-years study (rat, feed): 0, 1200, 2300 mg/kg/d NOEL: 1200 mg/kg/d LOEL: 2300 mg/kg/d (few bladder stones one papilloma) Chronic LOAEL (rat): 1500 mg/kg/d (TOPKAT 6.1 pred.)	13 weeks study (rat; feed): 0, 748, 1522, 3849 mg/kg/d NOAEL: 1522 mg/kg/d LOAEL: 3849 mg/kg/d (decreased body weight gains, altered urine values, increases in kidney weights, hyperplasia, hypertrophy,..) Chronic DNEL derived by application of AF of 2 to the NOAEL of the sub-chronic study chronic NOAEL (rat): > 100 mg/kg/d (read-across from DEG) chronic LOAEL (rat): 3400 mg/kg/d (TOPKAT 6.1 prediction)	14 days study (rat; drinking water): 0, 92, 391, 2355, 6387 mg/kg/d NOAEL:> 6387 mg/kg/d 33 days study (rat; drinking water): 0, 220, 660, 2000 mg/kg/d NOAEL: > 2000 mg/kg/d Chronic DNEL derived by application of AF of 6 to the sub-chronic study chronic NOAEL (rat): > 100 mg/kg/d (read-across from DEG)	subacute NOAEL (rat): > 2000 mg/kg/d (read-across from tetraEG) sub-chronic NOAEL (rat): > 1500 mg/kg/d (read-across from TEG) chronic NOAEL (rat): > 100 mg/kg/d (read-across from DEG)

Category Member	EG	DEG	TEG	tetraEG	pentaEG
CAS No.	**107-21-1**	**111-46-6**	**112-27-6**	**112-60-7**	**4792-15-8**
Reproductive toxicity (oral)	Two generation study (mice; drinking water): 0, 410, 840, 1640 mg/kg/d P(NOEL): 1640 mg/kg/d F1(NOAEL): 840 mg/kg/d; F1(LOEL): 1640 mg/kg/d (lower number of live pups/litter, unusual facial features, skeletal defects);	RACB test (mice; drinking water): 0, 610, 3060, 6130 mg/kg/d P (NOEL): 3060 mg/kg/d F1 (NOEL): 3060 mg/kg/d F1 (LOAEL): 6130 me/kg/d (decreased number of litters per fertile pair)	RACB (mice; drinking water): 0, 590, 3300, 6780 mg/kg/d P (NOAEL): > 6780 mg/kg/d F1 (NOAEL): > 6780 mg/kg/d	NOAEL (mice): > 6000 mg/kg/d (read-across from TEG)	NOAEL (mice): >6000 mg/kg/d (read-across from TEG)
Developmental toxicity (oral)	Mice (gavage) GD 6-15: 0, 50, 150, 500, 1500 mg/kg/d NOAEL: 500 mg/kg/d LOAEL: 1500 mg/kg/d (no apparent maternal toxicity, decreased pups bw, fused ribs and arches, poor ossification) BMD LCL = 440 mg/kg/d Mice (gavage): GD 8-14 NOAEL: 700 mg/kg/d LOAEL: 2500 mg/kg/d (decrease in number of live implants, increase in number of dead implants;) NOAEL: (maternal): 2500 mg/kg/d. Mice (gavage): GD 6-15: LOAEL: 750 mg/kg/d (decreased bw, increased number of malformations) NOAEL (maternal) 750 mg/kg/d Rat (gavage): GD 6-15: NOAEL: 500 mg/kg/d LOAEL: 1000 mg/kg/d (decreased fetal bw) Possible developmental toxicity (TOPKAT 6.1 prediction)	Mice (gavage) GD 6-15: 0, 590, 2950, 11800 mg/kg/d NOAEL: 2950 mg/kg/d LOAEL: 11800 mg/kg/d (decreased fetal body weight) BMD LCL= 1652 mg/kg/d Rat (gavage) GD 6-15: 0, 1118, 4472, 8944 mg/kg/d NOAEL: 1118 mg/kg/d LOAEL: 4472 mg/kg/d (reduced fetal weight and delayed ossification) Possible developmental toxicity (TOPKAT 6.1 prediction)	Mice (gavage) GD 6-15: 0, 565, 5650, 11300 mg/kg/d NOAEL: 563 mg/kg/d LOAEL: 5630 mg/kg/d (decreased fetal bw) BMD LCL= 2373 mg/kg/d Rat (gavage) GD 6-15: NOAEL: 5630 mg/kg/d LOAEL: 11260 mg/kg/d (decreased fetal bw, increased bilobed thoracic centrum - skeletal variation) Rat (gavage) GD 1-21: NOAEL: > 4500 mg/kg/d Possible developmental toxicity (TOPKAT 6.1 prediction)	NOAEL (mice): because there is no apparent trend in mice NOAELs, the value of the closest analogue could be used – read across from TEG BMD LCL (mice): > 2000 mg/kg/d (extrapolation from TEG) NOAEL (rat): > 4500 mg/kg/d (read-across from TEG)	NOAEL (mice): because there is no apparent trend in mice NOAELs, the value of the closest analogue could be used – read-across from TEG BMD LCL (mice): > 2000 mg/kg/d (read-across from TEG) NOAEL (rat): > 4500 mg/kg/d (read-across from TEG)

*BMD LCL = 95 % lower confidence limit on the benchmark dose

Appendix 3
Possible application of ranking methods to organic chemicals

Manuela Pavan & Andrew Worth
European Chemicals Bureau

3 Possible application of ranking methods to organic chemicals

3.1 Summary

This appendix aims to summarise and illustrate different ways in which chemometric ranking methods could be used in the development of chemical categories. To illustrate possible applications of ranking methods, a data set of phthalate esters was investigated. Ranking methods were applied to estimated data generated by QSARs, which reflects the worse-case scenario that insufficient or no suitable experimental data are available. When sufficient experimental data are available, ranking methods can be applied directly to the experimental data. In general, however, when applying the chemical category approach, a combination of experimental and estimated data is likely to be available.

In chemical risk assessment, ranking methods are usually associated with the priority setting of large numbers of chemicals, since the trends established by these methods can be used to guide strategic testing. This application remains in the case of chemical categories, although large numbers of chemicals are rarely present, and simpler approaches could be used to guide the strategic testing.

As an additional application, it is proposed that ranking methods also provide a useful means of developing the initial category hypothesis. In particular, the ability of these methods to sort and group chemicals on the basis of *multiple* endpoints means that it is possible to compare chemicals in terms of both the quantitative and qualitative differences in their *toxicity profiles*. In other words, ranking methods provide a means of:

a) sorting chemicals according to their relative levels of concern, thereby providing the basis for analysing trends across multiple endpoints and defining subcategories in terms of different levels of concern.
b) identifying different profiles of toxicological behaviour, which might also be regarded as different subcategories.

3.2 Background information on regulatory assessments

It is emphasised that the general purpose of this investigation was to explore and illustrate how ranking methods could be used in the formation of chemical categories, using a dataset of phthalate esters as an example of a category of organic chemicals. It was *not* the purpose to re-evaluate any substance-specific data or conclusions made in existing regulatory assessments of specific phthalate esters.

For completeness and for background information, it is noted that various regulatory assessments have been conducted on phthalate esters:

a) an OECD SIAM category on a more restricted set of seven high-molecular weight phthalates has been developed (1)
b) EU risk assessments have been completed for two higher molecular weight esters (2,3)
c) EU harmonised classifications have been agreed for seven phthalate esters (4)
d) A total of 14 phthalate esters were considered during an initial screening exercise by the EU PBT Working Group. However, as a result of further evaluation, none of these was considered as potential PBTs (5).

3.3 Introduction to chemometric ranking methods

In the scientific literature, “ranking” and “priority setting” are sometimes used synonymously. However, in the regulatory assessment of chemicals, it is useful to make a distinction between ranking

methods, which are mathematically based and which can be automated in the form of computer-based algorithms, and priority setting procedures, which include additional considerations, such as expert judgement and concerns by regulatory authorities. In the context of REACH, priority setting procedures are foreseen for the use in the Evaluation and Authorisation procedures. Technical work is being carried out in the context of RIPs 4.3 and 4.5, to develop proposals for ranking methods, taking into account the challenges and data requirements of the new legislation.

From the scientific perspective, two main types of ranking methods are distinguished: total order and partial order methods (6). Methods for total and partial order ranking are described in detail elsewhere (7).

3.3.1 Total order ranking

Total order ranking (TOR) methods are scalar techniques that can be used to rank chemicals on the basis of more than one criterion. The different criteria values are combined into a global ranking index, and chemicals are ordered sequentially according to the numerical value of the ranking index. Since criteria are not always in agreement, i.e. can be conflicting, there is a need to find an overall optimum that can deviate from the optima of one or more of the single criteria. While a variety of TOR methods have been proposed in the literature, three commonly used methods are based on the desirability function, the utility function and the dominance function. Each of these methods was used in this investigation.

The desirability function transforms each criterion (variable) independently into a desirability d_i by using an arbitrary function that transforms the actual value of each chemical into a value between 0 and 1. The overall desirability D of each chemical is generated by combining all of the individual desirabilities (for that chemical) through a geometrical mean. The desirability function is very strict: if any desirability d_i is equal to 0, the overall desirability D_i will be zero, whereas the D_i will be equal to one if (and only if) all the individual desirabilities are equal to one. Once the overall desirability (D_i) for each chemical has been calculated, the full set of chemicals can be ranked according to their D values. This type of ranking is useful for highlighting trends based on highly conservative assumptions.

The utility function is similar to the desirability functions u_r in that it transforms the actual value of each chemical into a value between 0 and 1. The overall utility U of each chemical is defined as combining all the utilities (for that chemical) through an arithmetic mean. The overall utility is less severe: the overall utility of a chemical can be high even if a single utility function is zero.

The dominance function works differently: instead of transforming each criterion by a quantitative function, it is first established whether the best condition is satisfied by a minimum or maximum value of the selected criterion, and then all chemicals are compared with each other in a pairwise manner. For each pair of chemicals (i,j) the number of criteria where i dominates j, i.e. where i is better than j, is calculated. A C_{ij} value equal to 1 means equivalence of the two chemicals; $C_{ij} > 1$ means that the chemical i is, on the whole, superior to the chemical j, whereas $C_{ij} < 1$ means that the chemical i is, on the whole, inferior to the chemical j. The obtained C_{ij} values are normalised and a global score for each chemical is then calculated.

3.3.2 Partial order ranking

Partial order ranking (POR) methods are vectorial approaches that recognise that different criteria are not always in agreement, but can be conflicting, which means that not all chemicals can be directly

compared with others. The Hasse diagram is a means of illustrating partial order ranking. It was introduced by Halfon (8) and refined by Brüggemann (9). Each chemical is represented by a small circle. Comparable chemicals which belong to an ordered relation are linked, while incomparable chemicals are not connected. A typical Hasse diagram is shown in Figure 1.

3.4 Identifying trends and different levels of concern

In this ECB investigation, a chemically diverse set of 323 phthalate esters, including the seven members of the SIAM category on high molecular-weight phthalates esters, were investigated and ranked according to their predicted PBT behaviour. Total and partial ranking methods were applied to three main properties determining the PBT behaviour: persistence, the bioconcentration factor (BCF) and acute aquatic toxicity (96h fathead minnow), as calculated with BIOWIN, BCFWIN and ECOSAR, respectively. To simplify this illustration, additional types of toxic effect, such as chronic aquatic toxicity, chronic mammalian toxicity, carcinogenicity, mutagenicity and reproductive toxicity, were not taken into account.[1]

The predictions generated by each model were coded into a scale of 1 to 4, corresponding to low (score=1), low/moderate (score=2), moderate/high (score=3) and high concern (score=4), as shown in Table 1. In the case of acute aquatic toxicity, the lowest level of concern was based not only on the predicted LC50 values, but also on the predicted aqueous solubility. If the aqueous solubility of a substance was estimated by WSKOWWIN to be less than 0.001 mg/L, the substance was considered to be of no concern due to insufficient concentration in the aqueous phase.[2] The estimated value of 0.001 mg/L corresponds with an experimental solubility limit of 0.01 mg/L (it was found that for this data set, the WSSKOWIN predictions tend to be lower than the experimental values by a factor of 10).

3.4.1 Total order ranking of phthalates based on the desirability function

Since the "best" condition for each property (P, B and T) is related to the minimum score, each property was independently transformed into a desirability (and utility) by an inverse linear transformation (Figure 2). Thus, the best condition, corresponding to the chemicals predicted to be safest, has a desirability equal to 1, whereas the worst condition, corresponding to the chemicals predicted to be the most hazardous, has a desirability of 0.

The three properties were equally weighted in the ranking procedure and for each chemical the PBT hazard score was calculated as 1 - *Di* (*Ui*), where *Di* (*Ui*) is the overall desirability *Di* (or utility *Ui*) of the chemicals. Thus, the PBT hazard score ranges from 0, for chemicals with the least PBT concern, to a maximum of 1 for chemicals with the highest PBT concern (Figure 3).

The ranking based on the desirability function is severe: it gave a PBT hazard score of 1 if *any of the three* properties (P, B and T) had a score of 4, and only gave a PBT hazard score of 0 if *all of the three* properties had scores of 0. As shown in Figure 3, one of the SIAM members (CAS 68515-47-9) received the maximal score of 1, whereas four of seven SIAM phthalates had a lower PBT hazard score (score of 0.306), and two others had an even lower ranking (score of 0.126).

---

[1] In the EU PBT assessment strategy, evidence of such effects would also be considered when deciding whether a substance meets the "T" criteria.

[2] This is a simplification, because in principle, chronic toxicity could still arise even in the case of insoluble substances, and even acute toxicity could arise through the uptake of particles to which the insoluble chemical is adsorbed.

TOR based on the *desirability function* provides a means of sorting chemicals in a conservative manner. Thus, TOR based on the desirability function could be used to identify subcategories in cases where it is useful to define a subcategory reflecting a high level of concern for *any* endpoint or a low level of concern for *all* endpoints.

3.4.2 Total order ranking of phthalates based on the utility function

The application of the desirability function resulted in a large number of phthalate analogues appearing to be of high concern, which was considered unrealistic in view of the known properties of some of these chemicals. Therefore, the utility function was applied to rank the chemicals in a less severe manner.

The ranking based on the utility function allows better discrimination between chemicals based on their overall PBT profile (Figure 4). It can be seen that four the seven SIAM phthalates are considered to have the same PBT hazard score (score of 0.223), whereas one of the SIAM members has a higher ranking (score of 0.334), and two have a lower ranking (score of 0.112). Thus, the utility function produced the same relative order between the SIAM phthalates as the desirability function, but the absolute differences were less exaggerated.

The ranking based on the utility function gave a PBT hazard score of 1 if (and only if) *all three* properties (P, B and T) had a score of 4. This result was obtained for only two of the 323 chemicals: dipropyl 3,4,5,6-tetrachlorophthalate and tris(2-chloroethyl) 4,5,6-trichloro-1,2,3-benzenetricarboxylate. Because the utility function assigns the highest ranking only when all three hazard scores have maximal values, it could in principle be exploited in the identification of potential PBTs. According to the EU PBT criteria, a substance is identified as a PBT if it meets all three criteria for P, B and T. In the case of this particular dataset, the two chemicals with the highest PBT hazard ranking of 1 (mentioned above) failed to meet EU criteria for PBT assignment. In fact, the chemical with the lowest predicted LC50 value in the dataset was dipropyl 3,4,5,6-tetrachlorophthalate (LC50=0.45 mg/L), which is above the EU criterion for T assignment of 0.1 mg/L.

The utility function does not resolve whether the concern results from P, B or T. For example, if one of the three properties has a score of 4 (high concern for a single property), and the other two properties have scores of 1 (low concern), the PBT hazard score is the same, irrespective of whether the high concern results from P, B or T (Table 2).

Thus, TOR based on the utility function could be used to identify subcategories if it is sufficient to distinguish between chemicals based on their "average" behaviour across several properties.

3.5 Identifying and visualising different profiles of toxicological behaviour

3.5.1 Total order ranking based on the dominance function

To obtain a full discrimination between chemicals based on their individual P, B and T properties, i.e. to identify different profiles of PBT behaviour, TOR based on the dominance function can be used. For example, if a chemical has two properties with a score of 3, and one property with a score of 4, there are three possible combinations of the scores (Table 3). By applying the dominance function, each combination is distinguished by a different PBT hazard score (Table 3).

As illustrated in Figure 5, the use of the dominance function enables qualitative differences between the phthalates to be detected, resulting in the identification of 25 different PBT profiles. The different profiles could be regarded as different subcategories within the larger category of 323 phthalates.

Thus, TOR based on the dominance function could be used to identify subcategories based on different profiles of behaviour.

3.5.2 Partial order ranking

Partial order ranking overcomes the main limitation of total order ranking that information on conflicting properties is lost. Partial order ranking encodes both quantitative and qualitative information of the trends analysed. As an illustration, the application of partial order ranking to the set of 323 phthalates identified nine levels of PBT hazard concern (Figures 6-7). In level 8, all 19 chemicals have moderate/high concern for one of the three properties and high concern for the other two. However, the level contains two clusters, distinguishing between 17 chemicals with moderate/high concern for P and high concern for B and T, and two chemicals with high concern for P and B, and moderate/high concern for T (Figure 6). POR also provides an analysis of whether chemicals are comparable or incomparable

Thus, POR could be used to identify subcategories based on different profiles of behaviour and to identify which subcategories are comparable and which are incomparable.

3.5.3 Principal components analysis

Another way of visualising the toxicological profile of a set of chemicals is to apply principal component analysis (PCA) to the different levels of concern (Table 1). This method provides an additional means of visualising similarities and dissimilarities in the PBT profiles of the phthalate analogues.

PCA was applied to the predicted PBT data for the 323 phthalate analogues, to identify the orthogonal directions of maximum variance in the original data set and to project the data into a two-dimensional space formed by the two highest-variance components. Figure 8 shows the biplot of the first and second components. The cumulative explained variance of the first two principal components is 84.3%. The Hotelling T2 ellipse (in red) indicates the distance of each chemical from the model hyperplane. The ellipse was computed with a 95% confidence level.

It can be seen that the first principal component (PC1), explaining 49.7% of the total information, corresponds to a quantitative macrovariable, which can be interpreted as a PBT hazard score. High values of the first component are associated with compounds having a “safe” PBT profile, while low values of the first component are associated with compounds having a PBT profile of high concern. Thus, PC1 separates the safest compounds ones (right hand side of the plot) from the more hazardous ones (left hand side of the plot).

The second principal component (PC2), explaining 34.7% of the total information, discriminates between different profiles of PBT behaviour. In particular, PC2 separates persistence and bioaccumulation from toxicity. High values of PC2 are associated with high persistence and bioconcentration but low toxicity, whereas low values correspond with high toxicity but low persistence and bioconcentration. Thus, the upper left part of the plot contains chemicals characterised by high persistence and bioconcentration, but relatively low or moderate toxicity, whereas the lower

left part of the plot contains compounds with high toxicity, but relatively low or moderate persistence and bioconcentration.

3.6 Conclusions on the applicability of ranking methods

Ranking methods allow chemicals to be sorted and sub-grouped according to their relative levels of concern and different profiles of toxicological behaviour. Ranking methods provide a means of combining data from multiple endpoints, and thus provide additional ways of ordering and subgrouping chemicals, which might be useful when applying the top-down approach to category formation.

It should be noted that the numerical values of ranking scores have no absolute meaning, because if chemicals are added or deleted from the dataset, and the ranking algorithm is performed again, the scores will change. However, the ranking scores are meaningful with respect to each other, and can be used to sort the chemicals (according to their numerical values) and to define sub-groups of chemicals (having the same scores).

Rankings based entirely on QSAR data can be used to predict chemicals with the highest level of concern as well as the lowest level of concern. Chemicals at the extremes of the predicted trend could be selected for strategic testing to confirm the boundaries of the trend. In addition, selected chemicals in the middle of the predicted trend could also be selected for testing, to check whether there are any deviations.

The different levels of concern identified by ranking methods for subgroups could be used as the basis for identifying subcategories based on different levels of toxicity or concern. In particular, the ability of ranking methods to combine quantitative information from multiple properties could be exploited to define different subgroups based on multiple endpoints. For example, different levels of the PBT hazard ranking could be regarded as different subcategories. TOR based on the *desirability function* provides a means of sorting and sub-grouping chemicals in a conservative manner, reflecting a high level of concern for any single endpoint. In contrast, TOR based on the *utility function* provides a useful means of sorting and sub-grouping chemicals based on their "average" behaviour across multiple toxicological endpoints.

Ranking methods can also be used to identify subgroups based on different toxicological profiles (e.g. high P & B & T at one extreme *vs* low P & B & T at the other extreme). TOR based on the *dominance function* could be useful in this respect.

If it is desirable to compare chemicals both in terms of the quantitative differences in their hazard rankings and the qualitative differences in their hazard profiles, the method of choice is *partial order ranking*. The qualitative and quantitative differences can be visualised by using the Hasse diagram.

In this investigation, only estimated properties were used as the input to the ranking algorithms. This demonstrates how ranking methods could be used in combination with QSAR methods in cases where there are insufficient experimental data to develop the initial category hypothesis (or proposal). While the investigation focussed on environmental properties, the same general approach could also be applied to combinations of human health endpoints (e.g. carcinogenicity, mutagenicity and reproductive toxicity).

It is proposed that the trends, boundaries, and subcategories predicted by using QSARs and ranking methods could be used to help develop the initial category hypothesis, and to identify chemicals for strategic testing, in order to assess the robustness of the category.

3.7 References

1. OECD (2004). SIDS Initial Assessment Report for SIAM 19. Berlin, Germany, October 2004.

2. EC (2003). EU RAR on 1,2-Benzenedicarboxylic acid, di-C8-10-branched alkyl esters, C9-rich (DINP), CAS. No. 68515-48-0 Volume 35. Available from the ECB website: http://ecb.jrc.it/esis/esis.php?PGM=ora

3. EC (2003). EU RAR on 1,2-Benzenedicarboxylic acid, di-C9-11-branched alkyl esters, C10-rich (DIDP), CAS. No. 68515-49-1. Volume 36. Available from the ECB website: http://ecb.jrc.it/esis/esis.php?PGM=ora

4. The ECB CLASSLAB database (http://ecb.jrc.it/classification-labelling) identifies the following entries:

   i) Annex I entry 607-317-00-9: bis(2-ethylhexyl) phthalate; di-(2-ethylhexyl) phthalate; DEHP (CAS No. 117-81-7; EC No. 204-211-0): Repr. Cat. 2; R60; Repr. Cat. 2; R61.

   ii) Annex I entry 607-318-00-4: dibutyl phthalate; DBP (CAS No. 84-74-2; EC No. 201-557-4): Repr. Cat. 2; R61; Repr. Cat. 3; R62; N; R50.

   iii) Annex I entry 607-426-00-1: 1,2-benzenedicarboxylic acid, dipentylester, branched and linear (CAS No. 84777-06-0; EC No. 284-032-2): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

   iv) Annex I entry 607-426-00-1: n-pentyl-isopentylphthalate (CAS No. -; EC No. -): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

   v) Annex I entry 607-426-00-1: di-n-pentyl phthalate (CAS No. 131-18-0; EC No. 205-017-9): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

   vi) Annex I entry 607-426-00-1: diisopentylphthalate (CAS No. 605-50-5; EC No. 210-088-4): Repr. Cat. 2; R60; Repr. Cat. 2; R61; N; R50.

   vii) Annex I entry 607-480-00-6: 1,2-benzenedicarboxylic acid di-C7-11-branched and linear alkylesters (CAS No. 68515-42-4; EC No. 271-084-6): Repr. Cat. 2; R61; Repr. Cat. 3; R62.

5. ECB (2002). Identification of potential PBTs or vPvBs among the IUCLID High Production Volume Chemicals. Doc ECB 4/14/02(pbt strategy – report)_rev1. Date 21 November 2002.

6. Carlsen L (2004). Giving molecules an identity. On the interplay between QSARs and partial order ranking. *Molecules* **9**, 1010-1018.

7. Pavan M (2003). Total and Partial Ranking Methods in Chemical Sciences. PhD Thesis. University of Milano-Bicocca, Milan, Italy.

8. Halfon E (1989). Comparison of an Index Function and a Vectorial Approach Method for Ranking of Waste Disposal Sites. *Environmental Science & Technology* **23**, 600-609.

9. Brüggemann R & Bartel H-G (1999). A Theoretical Concept to Rank Environmentally Significant Chemicals. *Journal of Chemical Information and Computer Sciences* **39**, 211-217.

Figure 1. The Hasse diagram for partial order ranking

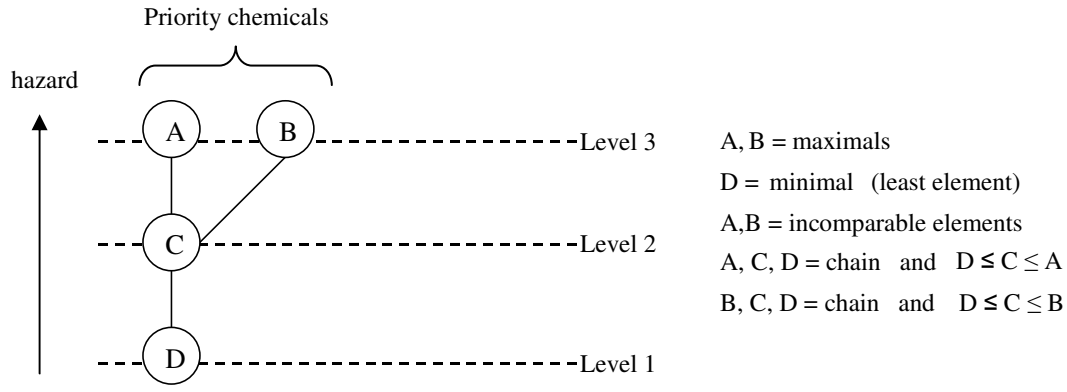


Figure 2. Inverse relationship between the ranking score for a property and its desirability or utility

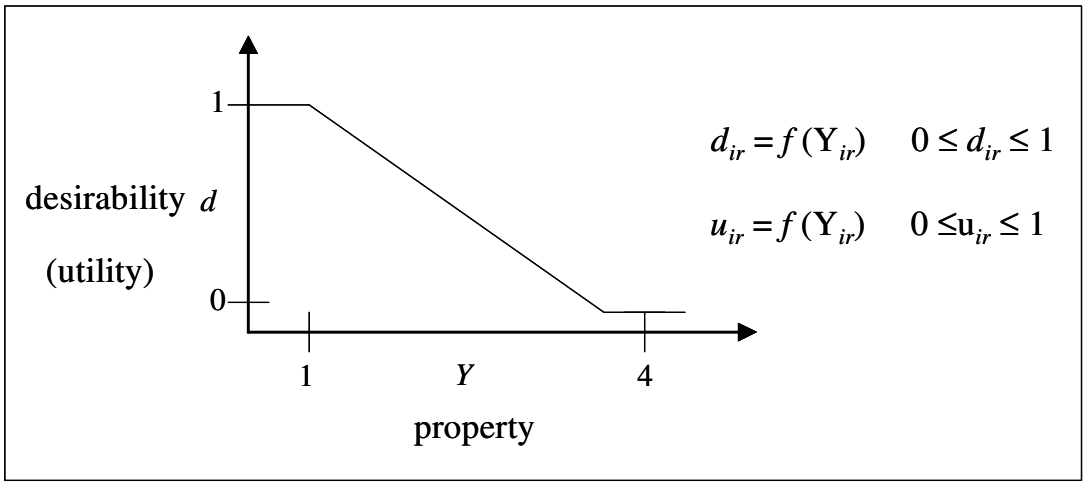


Figure 3. Total order ranking of phthalates based on the desirability function

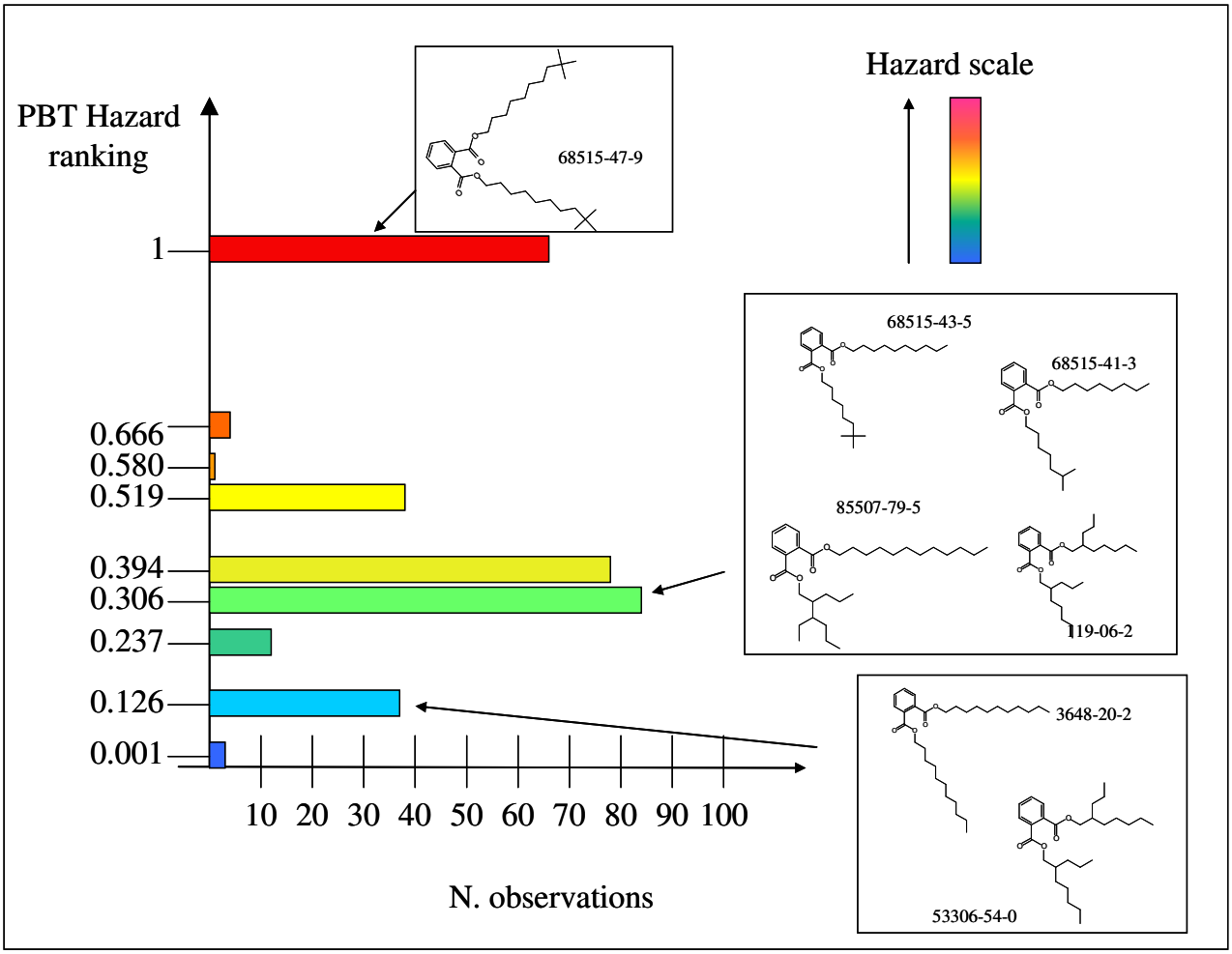


Figure 4. Total order ranking of phthalates based on the utility function

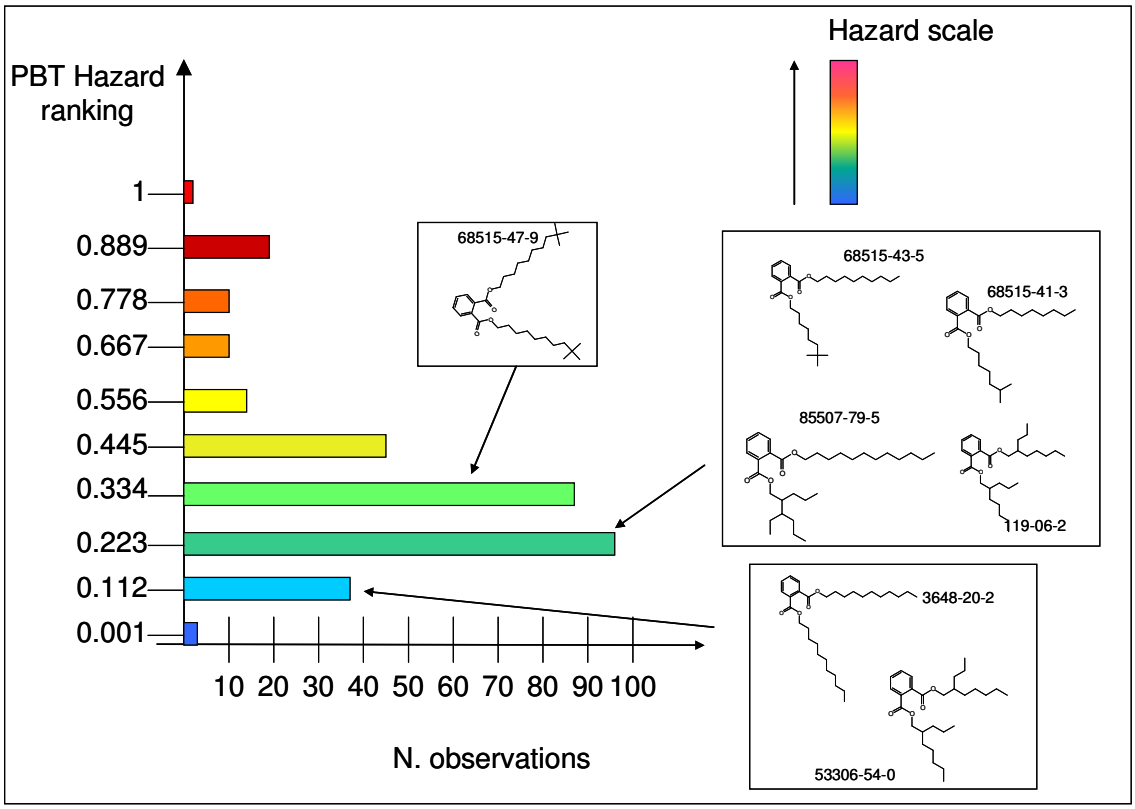


Figure 5. Total order ranking of phthalates based on the dominance function

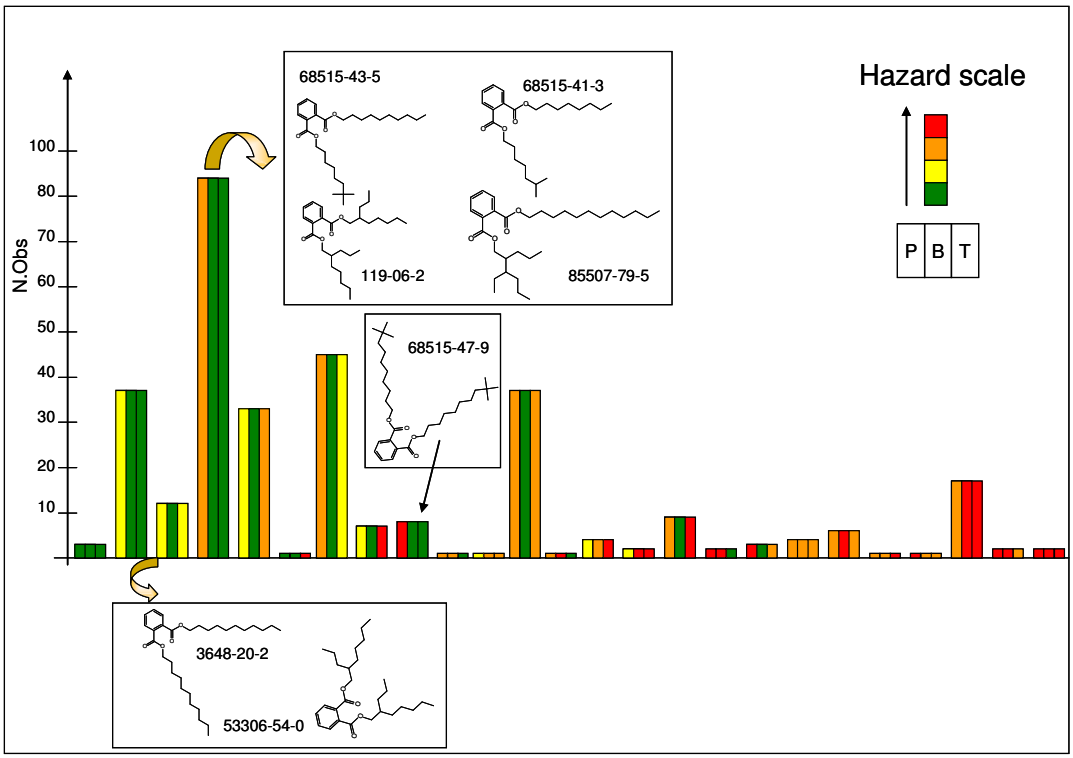


Figure 6. Partial order ranking of phthalates using the Hasse diagram

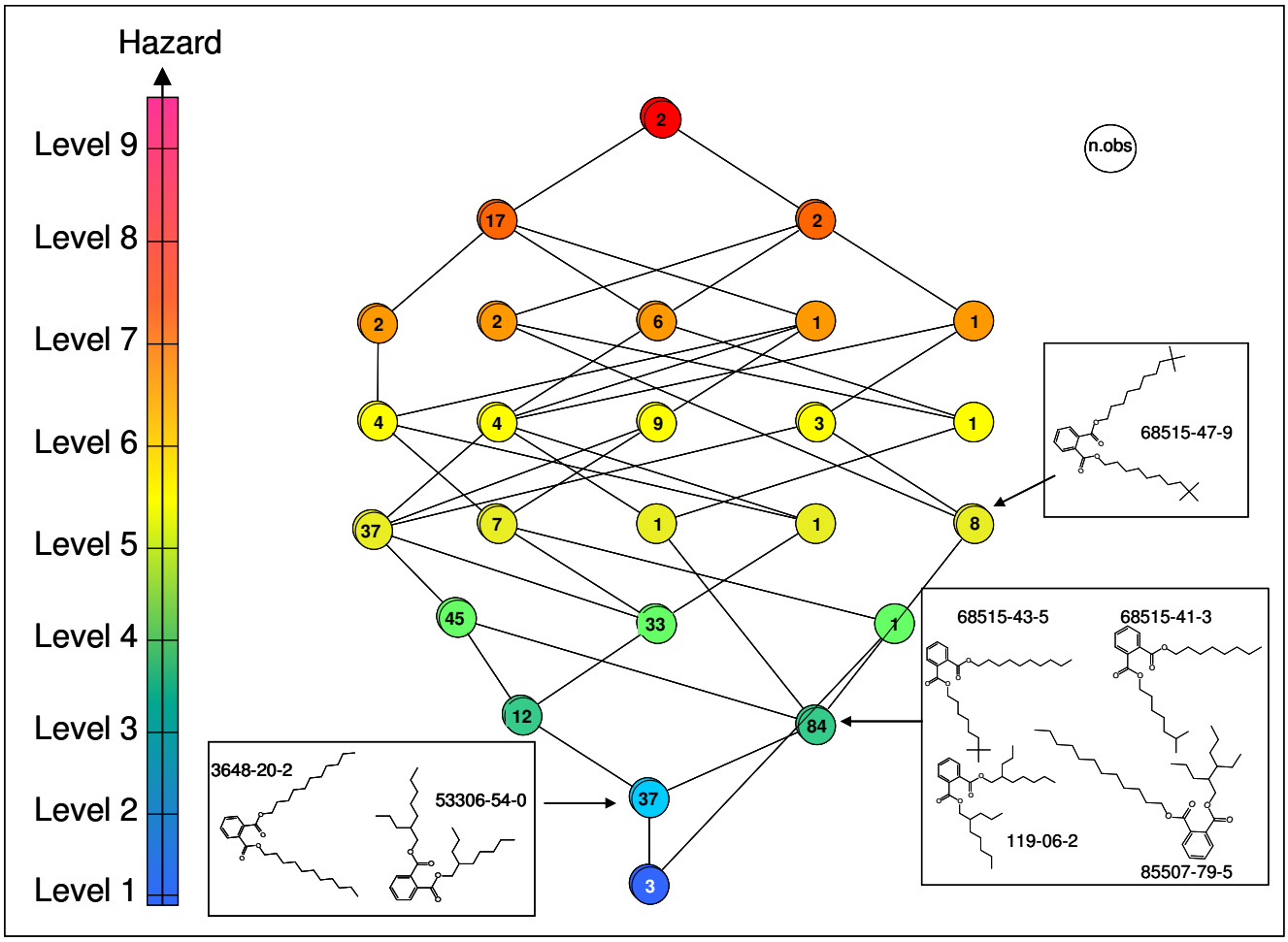


Figure 7. Distribution of phthalates across levels of concern defined by partial order ranking

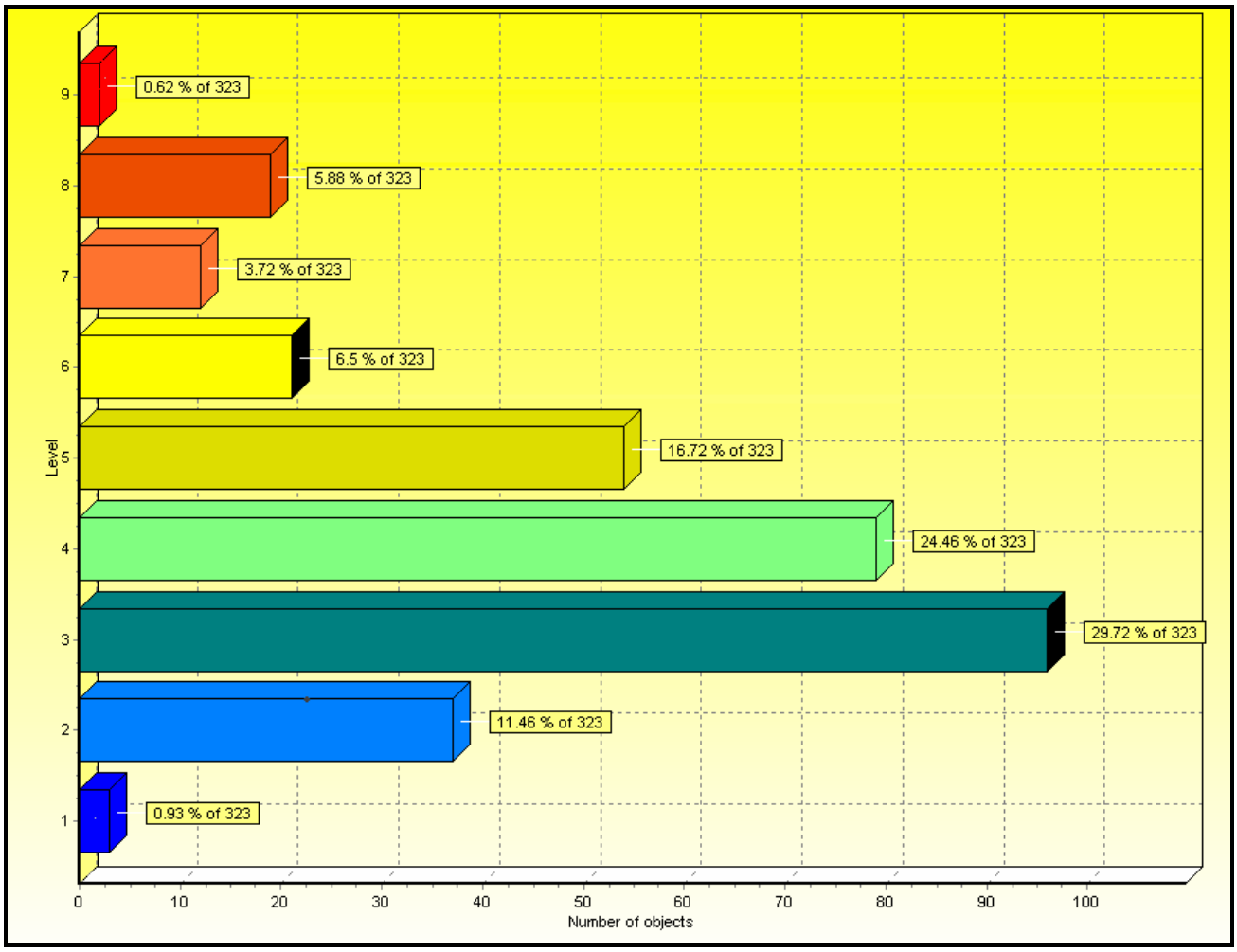


Figure 8. Visualisation of PBT profile by Principal Components Analysis

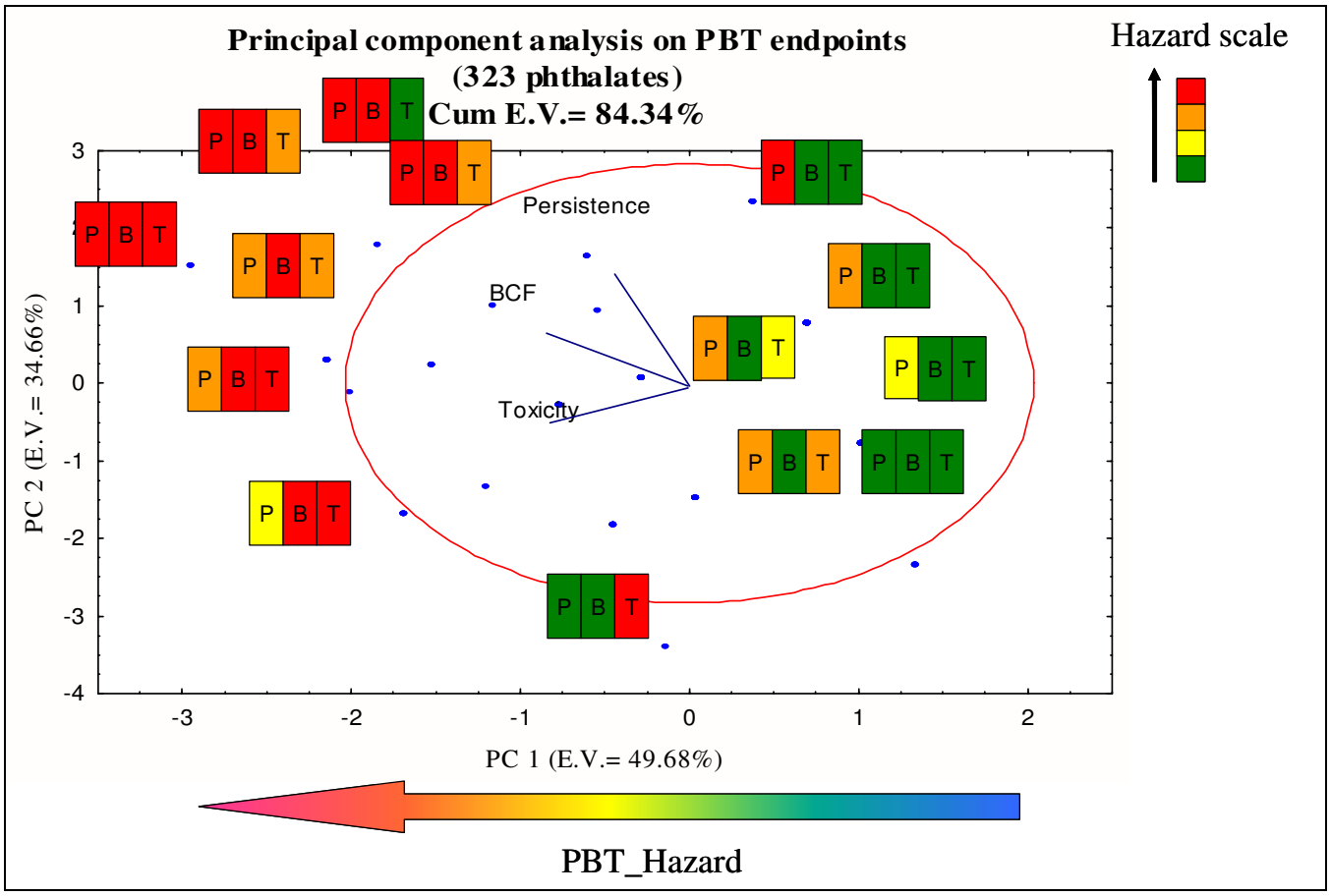


Table 1. Conversion of P, B and T predictions in different levels of concern

Ultimate persistence prediction[1]	BCF	Toxicity (LC50 (mg/L))[2]	Concern score
P ≤ 2	BCF > 2000	LC50 ≤ 1	4
2< P ≤ 3	1000 < BCF ≤ 2000	1 < LC50 ≤ 10	3
3< P ≤ 3.5	1000 < BCF ≤ 2000	10 < LC50 ≤ 100	2
P > 3.5	BCF ≤ 1000	LC50 > 100	1

[1] In the Biowin3 (ultimate biodegradation) model the ratings correspond to the following time units: 5 = hours; 4 = days; 3 = weeks; 2 = months; 1 = longer.
[2]The toxicity bands are equivalent to the EU R-phrases R50 (LC50≤1), R51 (1<LC50≤10), R52 (10<LC50≤100) and unclassified (LC50>100).

Table 2. Generation of a PBT hazard score by using the utility function

Ultimate persistence concern score	BCF concern score	Toxicity concern score	PBT Hazard score
4	1	1	0.334
1	1	4	0.334

Table 3. Generation of a PBT hazard score by using the dominance function

Ultimate persistence concern score	BCF concern score	Toxicity concern score	PBT Hazard score
3	4	3	0.870
3	3	4	0.897
4	3	3	0.917

Appendix 4
An investigation into the feasibility of developing categories using the top-down approach using automated workflows

Grace Patlewicz
European Chemicals Bureau

4 An investigation into the feasibility of developing categories using the top-down approach using automated workflows

4.1 Summary

Grouping similar chemicals into categories is a focus that has gained much interest particularly with the advent of the new REACH legislation. This short study investigated the use of SciTegic's Pipeline Pilot for grouping chemicals by applying the top-down approach to a large inventory of chemicals, the European Inventory of New and Existing Chemical Substances (EINECS). As a starting point, the study focused on a set of published skin sensitisation data in order to develop rules which characterised probable mechanistic pathways and models to predict the likely relative skin sensitisation potency. These models and rules were applied to EINECS as means of forming groups which were endpoint specific rather than merely structurally based. The findings, insights and some of the practical challenges encountered are described here.

4.2 Introduction

In the formation of chemical categories, the choice of computational method(s) is likely to depend on the starting point of the investigation. For example, it may start from a single chemical or a small group of chemicals, with the intention of building up a category by drawing on data from multiple sources (bottom-up approach). Alternatively, it may start from a predefined group of chemicals (e.g. an inventory or subset of an inventory whose members have been decided on a particular basis), with the intention of grouping some or all of the members into one or more categories (top-down approach).

In this feasibility study, we focused on exploring the formation of endpoint specific groups from a top-down approach, utilising the (European Inventory of New and Existing Chemical Substances) EINECS inventory together with available published skin sensitisation information. A secondary objective involved using the skin sensitisation information available to estimate the likely prevalence of skin sensitisers in the EINECS inventory.

4.3 Methods

Existing substances that comprise the EINECS inventory are limited in terms of their available toxicity data. If data exists, it has tended to have been generated for substances that have been manufactured or imported at high tonnage levels. In addition, the quality of any toxicity data can be quite variable, not always reported to a consistent standard and not systemically stored in a structured format. As a result, endpoint specific chemical groupings have not been readily developed as the toxicity data needed to derive them has been limited. The approach explored here investigated the feasibility of deriving endpoint specific chemical groups using predicted data. A collection of QSAR models were derived using a set of published skin sensitisation data (1). The data used, comprised a set of 210 chemicals that had been tested in the local lymph node assay (LLNA), the *in vivo* test of choice under REACH. The mechanisms underpinning skin sensitisation are sufficiently well understood and have resulted in the development of mechanistic QSAR models. Specifically, the rate determining step of skin sensitisation induction is believed to be the covalent binding that occurs between the chemical electrophiles and skin proteins (nucleophiles). These electrophilic-nucleophilic reactions can be conveniently described by standard organic reactions characterised by the following mechanistic domains; Michael addition, Schiff Base formation, S_NAr, S_N2, Acyl formation. These reaction mechanistic domains were first outlined in Aptula *et al*. (2) with respect to a dataset of 41 chemicals tested in the LLNA. The domains were subsequently described in more detail in Aptula and Roberts (3) culminating in a set of structural rules (somewhat akin to structural alerts). An approach for

estimating the skin sensitisation potential and potency of new chemicals has also been described (3). The key step is to identify the appropriate mechanistic domain and hence apply a QSAR model that has been derived for that mechanistic domain. The predictions from such a model should be more robust and interpretable then than would be from a general model derived from a set of sensitisation data. The structural rules describing the mechanistic domains have been recently applied to the dataset of Gerberick *et al*. (1). Roberts *et al*. (4) summarised the reaction chemistry for the dataset, assigning each chemical to its likely mechanistic domain. In this study, the dataset in (4) was extracted as a training dataset. The data included LLNA outcomes such as the test concentrations, simulation indices, EC3 values (the effective concentration that gives rise to a simulation index of 3) and potency category information. The potency category is a measure of the relative sensitising potency as defined by bands of EC3 values. The potency categories were first proposed by Kimber *et al*. (5) and take the form of "extreme", "strong", "moderate" sensitisers etc. e.g. where <0.1% would imply an extreme sensitiser.

The training dataset was exploited in two ways. Firstly a model to predict the most probable mechanistic domain for a given compound was derived. This was an attempt to see whether it was possible to translate the structural rules derived by human experts into structural features that could be routinely calculated by a QSAR program. A second set of models were then derived for each of the mechanistic domains to predict the likely sensitising potency of a chemical. It is important to note that robust mechanistic models for sensitisation that have been successfully developed recently rely on the hypothesis that sensitisation is a function of two parameters – reactivity and hydrophobicity (6). Understanding the likely mechanistic domain should identify the appropriate mechanistic model to apply, whereas the combination of hydrophobicity and reactivity information should discriminate the strength of sensitising potency. The models developed were then applied to the EINECS inventory to predict the likely mechanistic domains of the EINECS substances and their sensitising potency. Subsequent steps included breaking down the initial mechanistic domain chemical groups into smaller more manageable groups using clustering techniques. Maximal common substructures were also extracted for selected clusters to summarise the significant features characterising that cluster and to aide visualisation. All analysis steps were conducted using the datamining software tool PipelinePilot (SciTegic Inc, Accelrys Inc, San Diego, CA, USA).

4.3.1 Development of a model to predict mechanistic domain

The LLNA dataset with annotated mechanistic domains (Non-sensitisers, Michael acceptors, Schiff base formers, S_N2 reactors, Acyl formers, S_NAR reactors, and Special cases) was imported into PipelinePilot (SciTegic Inc). The component, "Learn Molecular Categories" was used to derive models that would predict the likelihood of a chemical to belong to a certain mechanistic domain. This Bayesian categorisation component uses probabilities to classify objects (in this case chemicals) into one of a set of categories (mechanistic domains). A set of models were built for each of the mechanistic domains using functional class fingerprints and other molecular properties such as Log P and molecular weight. A set of scores were then output, the highest of which determined which domain the chemical was most likely to be a member of.

The descriptors used were ALogP, Molecular Weight, Number of Hydrogen Donors, Number of Hydrogen Acceptors, Number of Rotatable Bonds, Molecular Fractional Polar Surface Area and ECFP_12. ECFP signifies Extended Connectivity Fingerprints, where 12 refers to the maximum diameter of the fingerprint. Extended Connectivity Fingerprints (ECFP) form one class of fingerprints that are proprietary to SciTegic. Each feature represents the presence of a structural unit, i.e. an exact structure with limited, specified attachment points. ECFPs represent a much larger set of features than what is common for other fingerprints. The virtual size of the fingerprint is four billion different features. For a given molecule, only a small subset of those features is present. This means the fingerprints are usually stored as a list of features that are present, rather than as a binary bit array. The

diameter represents the desired neighbourhood size, i.e. ECFP_12 generates features around each atom up in larger and larger structural neighbourhoods up to a diameter of 12. More information about the fingerprints can be found in the SciTegic Manual (7).

The developed model (named RXNDOMAIN) was applied to the EINECS inventory (compiled by ECB, version March 2007) to assign each of the 68,993 chemicals into their respective mechanistic domains. This formed a set of initial chemical groups which were similar with respect to their predicted skin sensitisation mechanism.

4.3.2 Development of a model to predict the prevalence of sensitisers

A set of models were then derived to predict potency. A separate model was developed for each mechanistic domain. PipelinePilot's workflow functionality was exploited to facilitate the automation of this process. The LLNA dataset was initially processed through the mechanistic class predictor model RXNDOMAIN to filter out all estimated non-sensitisers. The remaining chemicals were then processed through a set of filters to split the dataset into the training sets for the development of mechanistic domain specific models. Thus, first a filter identified predicted Michael acceptors so that a model to predict potency could be derived for these. A subsequent filter identified predicted Schiff Base formers to develop a potency model for these and so on. The resulting workflow assigned a chemical into one of the mechanistic domains using the RXNDOMAIN model and then predicted its likely sensitising potency using the mechanistic domain specific models. The resulting models were applied to the EINECS inventory to gain an overview of the distribution of predicted sensitisers and non-sensitisers.

4.3.3 Development of clusters for each mechanistic domain

The next stage was to take the EINECS chemicals that were assigned to a mechanistic domain and cluster each set on the basis of their fingerprint information and property information. A combination of the Euclidean distance and Tanimoto distance were used as the similarity index. The clustering method used was a relocation method based on maximal dissimilarity partitioning encoded within Pipeline Pilot. This begins by randomly choosing a data record as the first cluster centre. The record maximally distant from the first point is selected as the next cluster centre. The record maximally distant from both current points is selected after that. The process repeats itself until there are a sufficient number of cluster centres. The non-selected objects are then assigned to the nearest cluster centre to determine the cluster membership. The following parameters were used as quantitative representations of the chemicals in the clustering algorithm: FCFP_6, ALogP, Molecular Weight, Number of Hydrogen Bond Donors, Number of Hydrogen Bond Acceptors, Number of Rotatable Bonds, Number of Atoms, Number of Bonds, Number of Rings, and Number of Aromatic Rings. FCFP_6 represents the set of functional class fingerprints of maximum diameter 6. These are similar to the extended connected fingerprints except in their initial atom code assignment. The generation of an ECFP or FCFP fingerprint for a molecule begins with the assignment of an initial atom code for each heavy (non-hydrogen) atom in the molecule. For ECFPs, the initial atom code is derived from the number of connections to the atom, element type, charge and atom mass e.g. chlorine would be differentiated from bromine. For FCFPs, the initial atom code is based on the quick estimate of the functional role the atom plays thus chlorine and bromine are seen as equivalent instances of halogen atoms.

Each predicted reaction domain subset of EINECS was clustered in this way and a frequency table showing the number of clusters produced for each reaction domain was then produced.

4.3.4 Development of Maximal Common Substructures for selected mechanistic clusters

For a selection of clusters for at least two of the predicted reaction domains, the identification of maximal common substructures contained within a proportion of the input structures was performed using Pipeline Pilot. The Maximal Common Substructure Search (MCSS) is the process of finding the largest structure that is a substructure of all the molecules in a given set. This is a well-known method that is computationally intense. For a selection of clusters, the largest substructure common to at least 20 percent of the input molecules was performed for several of the smaller clusters whereas a set of substructures were identified for the clusters containing larger numbers of chemicals. This was a useful means of visualising the common structural motifs (or scaffolds) in the set of chemicals.

The MCSS within Pipeline Pilot accepts a number of input molecules, processes the data, and then outputs (generates) new molecules that represent the discovered maximal substructure or substructures. In the most common use, a single molecule is output (representing the largest substructure). Pipeline Pilot's method is based on an extension of extended-connectivity fingerprints (ECFPs). An ECFP "bit" is a 32-bit number; the fingerprint has about 4 billion possible bit values. (For a given molecule, only a very small subset of this large number of bits is "on" or present in the molecule.) Each "bit" represents the presence of a specific substructure, centred on some atom, comprised of all atoms and bonds within some radius. As it iterates, the radius is increased by one bond, and a new substructure (and a new "bit") is created. As large as the number of bits is, the space of bits is certainly smaller than the space of possible substructures, and so a single bit may be turned "on" by different substructural features. However, this collision rate is very low, and for most purposes, has an insignificant effect on most uses of the fingerprint. That the possibility of an occasional collision is not a problem can be seen by the use of explicit folding to reduce large fingerprints to a small, fixed-length space of, say, 2048 bits. These greatly shrunken fingerprints still performed admirably for the task of molecular comparisons (7). This suggests that the much smaller collision rate for ECFPs should not be important for most purposes.

4.4 Results and Discussion

4.4.1 Development of the RXNDOMAIN model

The LLNA training dataset of 210 chemicals was processed through the RXNDOMAIN model to determine its performance. The results were promising with 93.81% (195/210) of chemicals being assigned to the correct mechanistic domain i.e. correct predictions. Table 1 lists the 13 chemicals which were misassigned. No investigation was undertaken to rationalise these mispredictions since the main purpose of the model was to provide a screen to assign chemicals into their mechanistic groups. Figure 1 shows the distribution of chemicals into their predicted and assigned domains.

Applying the RXNDOMAIN model to the EINECS inventory provided a perspective of the domain distribution as shown in Figure 2. Obviously using a model developed on such a small dataset and applying it to a large inventory has limited value as no consideration of the applicability domain for each the different mechanistic domains was attempted. However it did provide a useful means of filtering a large inventory into smaller subsets that could be more readily evaluated whilst incorporating some aspect of endpoint similarity. Worth noting, is that the assignment of chemical to a specific domain does not equate to that chemical automatically being a sensitiser. A chemical could be assigned to the Michael acceptor domain but still fail to sensitise on account of its hydrophobicity and reactivity combination being insufficient to trigger induction. For this reason, the assignments predicted were further investigated to filter additional predicted non-sensitisers from each of the domain groups. The outcome would provide a better indicator of the prevalence of sensitisers within the EINECS inventory.

4.4.2 Development of a model to predict the prevalence of sensitisers

Potency models were developed for each of the domains using the LLNA dataset with exception to the non-sensitiser category. Twenty two of the 26 chemicals in the acyl formers training set were correctly predicted i.e. 84.6% correctly predicted. Table 2 lists the set of 26 chemicals together with their actual and predicted potencies. Four chemicals were underpredicted including short chain Azlactones (C4, C6 and C9) and phenyl benzoate.

For the S_NAR potency set, all potencies were correctly predicted though the reactivity domains were incorrect in two cases. Vinylidene dichloride (a potential Michael acceptor) and 4-Nitrobenzyl bromide (a potential S_N2 reactor) were incorrectly assigned.

For the non-sensitiser set, two chemicals Formaldehyde and Ethylenediamine were incorrectly assigned as non-sensitisers when their domains should have been Schiff Base formers.

83% of the predicted Michael Acceptors had their potencies correctly estimated (47/56 chemicals). Three chemicals were misassigned to this group namely (Propiolactone [SN2], Benzaldehyde [Schiff Base former], Vanillin [Non-sensitiser]). In two of these cases, this did not affect the outcome of the sensitising potency but in the case of propiolactone, the prediction was overcautious (predicted extreme vs. predicted strong). In the 8 other cases, there was a discrepancy between the actual and predicted potency as shown in Table 3.

All predicted Schiff Base formers were correctly assigned. 82% of Schiff Bases had their potencies corrected predicted (32/39). Seven compounds had their potencies under predicted. These are shown in Table 4.

The worse performing model was that for S_N2 reactors, here the correctly predicted potency was only 47.6% (20/42). In addition, three compounds were misassigned as non-sensitisers. Table 5 lists the set of compounds that were predicted to be assigned to this group together with their predicted and actual potency scores.

For the special case, only one chemical was misassigned (chlorobenzene [Non-sensitiser]). All potency calculations were however correctly predicted.

The EINECS inventory was processed through these models in the same way. This provided a primary assignment of mechanistic domain together with an estimate of predicted potency. This enabled an estimate of the likely prevalence of sensitisers within the EINECS inventory to be made. Figure 3 presents an overview of how the workflow was structured whereas Figure 4 presents a plot of the distribution of the different chemicals with respect to the overall sensitisation prediction and mechanistic domain.

Using the workflow in Figure 3, the percentage of predicted sensitisers within EINECS was estimated as 54.74%. Table 6 provides the breakdown of numbers used to generate this percentage. This percentage seemed inordinately high. Further work using another dataset, a subset of **E**uropean **LI**st of **N**otified **C**hemical **S**ubstances (ELINCS), was then conducted to investigate whether a more substantiated estimate could be derived. Structures for a subset of ELINCS (504) were generated and the models processed through this set to derive frequencies of estimated sensitisers. The results (not shown) revealed a 50:50 split between predicted sensitisers and non-sensitisers. For many of the substances, toxicity testing had been conducted to derive Risk labels, in this case R43 would classify a substance as a sensitiser. Reviewing the dataset on the basis of R43 labels found that 165 substances out of 504 were classed as R43 (i.e. a 32.7% prevalence of sensitisers). A prevalence of 32.7% is still quite high since those substances that did not have labels could either have not been tested due to their tonnage level or been tested for sensitisation and found to be negative. Further work to refine this

percentage would need to be carried out, specifically to consider the tonnage levels of each of the unlabelled substances to concretely identify whether they were in fact non-sensitisers or whether they had not been tested. In addition, the chemical space as characterised by this ELINCS data would need to be compared in more detail with that of the EINECS inventory and LLNA dataset to determine the extent to which they were chemically similar.

4.4.3 Development of clusters for each mechanistic domain

Figure 5 shows the workflow followed to assign EINECS chemicals within their respective domain using the RXNDOMAIN model and hence to perform the clustering. For convenience, plots depicting the frequency of chemicals within each set of clusters for each predicted mechanistic domain are shown to demonstrate the number, diversity and density of the clusters derived.

Figure 6 shows that there were 100 clusters identified for predicted acyl formers. Seventy five different clusters were derived from the 18641 chemicals assigned to this category as shown in Figure 7. 264 clusters were developed from the 13,191 predicted Michael acceptor category (Figure 8), 208 clusters were derived from 10,352 Schiff Base chemicals (Figure 9), 159 clusters from 7909 $S_N2$chemicals (Figure 10), 114 clusters were derived from 5654 S_NAR compounds (Figure 11) and 83 clusters from 8247 Special case compounds (Figure 12).

4.4.4 Development of Maximal Common Substructures (MCS) for selected mechanistic clusters

For two of the subsets; Acyl and Non-sensitisers, the feasibility of deriving MCS was then investigated for at least two clusters within these two sets.

Three clusters were selected from the clustering output of Non-sensitisers. One containing 33 compounds (Cluster 2), one containing 400 compounds (Cluster 14) and the final one containing 2532 compounds (Cluster 33). For the Cluster 2, the largest possible maximal subgraph was derived. For the Cluster 14 with 400 compounds, (up to 20) diverse maximal subgraphs were derived and for the Cluster 33 with 2532 compounds (up to 100) diverse maximal subgraphs were derived. 20 diverse maximal subgraphs were ultimately extracted for Cluster 14 and 7 MCS for Cluster 33 as query substructures. Figures 13-15 shows the subgraphs for each of these Clusters.

For the Acyl domain, two clusters were selected and maximal subgraphs derived. Cluster 93 with 20 compounds and Cluster 32 with 568 compounds. The largest maximal subgraph was derived for Cluster 93 and (up to 20) diverse maximal subgraphs for Cluster 32. In total, 11 MCS were derived for Cluster 32 as query substructures. Figures 16-17 shows the extracted subgraphs.

These query substructure MCS are difficult to interpret without reference to the parent structures, the original LLNA dataset or indeed the structural rules (3) that were first derived. To evaluate whether the MCS derived were meaningful from a sensitisation perspective, clusters were also derived for each of the mechanistic domains (Table 7) and MCS were derived for those clusters containing the largest number of compounds. 4 MCS were extracted for both Acyl forming compounds and non-sensitisers (Figures 18-19).

Comparison of the MCS for non-sensitisers and those extracted from the LLNA dataset showed how different they were. Inspection of the MCS from a chemistry perspective showed that in the majority of cases, the MCS extracted from the EINECS set were chemicals that contained an absence of electrophilic features. However there were some notable exceptions, including one of the MCS extracted for Cluster 14, an aliphatic aldehyde which could react by Schiff base formation. The MCS derived from Cluster 93 was of an azodye which is a highly reactive substance and capable of inducing

sensitisation. Presumably the absence of such analogues within the original LLNA data would explain why this particular MCS has been extracted incorrectly as a fragment indicative of non-sensitising behaviour.

Inspection of the 11 MCS derived from Cluster 32 and those from the original LLNA (1) showed them to be markedly different too. The MCS substructures were also thought to be too generic to encode specific mechanistic information for the acyl mechanistic domain. Inspection of the parent MCS structures was more helpful in providing the neighbourhood context of the query fragment but comparison of these parent MCS (Figure 20) with the extracted MCS from the LLNA (Figure 18) or the original acyl formation rule (Figure 21) suggested significant uncertainty in the predicted mechanism.

Whilst the query MCS derived from Cluster 93 was unlike any of the extracted acyl formers (Figure 18), the similarity in mechanism as outlined in the original rules (Figure 21) did appear to be feasible.

The MCS extraction is a potentially useful means of systemically characterising clusters of chemicals but these need to be interpreted with caution for the endpoint under consideration. The apparent inconsistencies observed here could be attributed in part to the fact that the chemistry domains had not been taken into account.

4.5 Conclusions

This short investigation highlights the means by which a large inventory such as EINECS can start to be broken down into more manageable groups for further evaluation, through using computational tools. Here, mechanistic information derived from a dataset of skin sensitisation was encoded into a predictive model and applied to the EINECS inventory to derive subsets of chemicals likely to favour different mechanistic pathways. The groups formed were then further split on the basis of predicted potency profile to gain a perspective of the likely potency profile of sensitisers within the EINECS list. Mechanistic domain subsets were also clustered to formulate smaller subsets of chemicals – each of which could then be potentially investigated further using experimental data to formulate trends and explore the robustness of the chemical groups.
To visualise the sort of substructures that characterised these clusters, a maximal common substructural analysis was performed to extract out the typical substructures that dominated a handful of selected clusters. Each of these scaffolds could potentially serve as a means of building up a category (seeds) by identification of other analogues with associated experimental toxicity data.

The investigative work highlights a few of the avenues available in exploiting computational approaches as encoded in the datamining tool Pipeline Pilot for the formation of chemical categories from top-down approaches.

It highlights the problems of forming endpoint chemical groups without sufficient experimental data and the problems of applicability when using predicted toxicity data. It also flags the need to interpret any results with caution, making reference to expert mechanistic insights where feasible and noting the scope of derived models with respect to the inventories that they are applied to. Clearly there is a pressing need to characterise inventories such as EINECS to appreciate the extent to which existing QSAR models can be robustly applied. Determining the prevalence of sensitisation within the EINECS inventory is equally difficult to determine in the absence of experimental data and faces many of the same issues when using predicted data.

Further work to explore how toxicological data can be better incorporated into refining the derived models such that they have greater predictive scope would help substantiate some of the proposed groupings. More efforts are also needed to explore the scope of existing QSAR models relative to

larger inventories in order to determine their applicability of use. This should be the subject of on-going work.

4.6 References

1. Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz GY & Basketter DA (2005). Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis* **16**, 157-202.

2. Aptula AO, Patlewicz G & Roberts DW (2005). Skin Sensitization: Reaction Mechanistic Applicability Domains for Structure-Activity Relationships. *Chemical Research in Toxicology* **18**, 1420-1426.

3. Aptula AO & Roberts DW (2006). Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chemical Research in Toxicology* **19**, 1097-1105.

4. Roberts DW, Patlewicz G, Kern PS, Gerberick F, Kimber I, Dearman RJ, Ryan CA, Basketter DA & Aptula AO (2007). Mechanistic Applicability Domain Classification of a Local Lymph Node Assay Dataset for Skin Sensitization. *Chemical Research in Toxicology* **20**, 1019-1030.

5. Kimber I, Basketter DA, Butler M, Gamer A, Garrigue JL, Gerberick GF, Newsome C, Steiling W & Vohr HW. (2003). Classification of contact allergens according to potency: proposals. *Food and Chemical Toxicology* **41**, 1799-1809.

6. Patlewicz G, Aptula AO, Roberts DW & Uriarte E. (2007). A mini review of available skin sensitization (Q)SARs/Expert systems. *QSAR & Combinatorial Science submitted*

7. SciTegic User Manuals. Basic Chemistry Collection and Data Modeling User Guides. (2006).

Table 1. List of 13 chemicals whose mechanistic domain was incorrectly predicted

Name	Mechanistic Domain*	Predicted Mechanistic Domain
trans-2-Decenal	MA	SB
Vinylidene dichloride	MA	S_NAR
4-Nitrobenzyl bromide	S_N2	S_NAR
Propiolactone	S_N2	MA
Benzyl benzoate	S_N2	SB
Formaldehyde	SB	Non
Ethylenediamine free base	SB	Non
Benzaldehyde	SB	MA
Sodium lauryl sulphate	Non	S_N2
Isopropyl myristate	Non	S_N2
Chlorobenzene	Non	Spec
Hexane	Non	S_N2
Vanillin	Non	MA

*as assigned in reference (4)

Table 2. List of assigned acyl formers

Name	Actual Potency Class	Predicted Potency Class
Oxazolone	extreme	extreme
Tetrachlorosalicylanilide	extreme	extreme
Fluorescein-5-isothiocyanate	strong	strong
2-Methyl-4H,3,1-benzoxazin-4-one (Product 2040)	strong	strong
C6-Azlactone*	moderate	weak
2-Mercaptobenzothiazole	moderate	moderate
Nonanoyl chloride	moderate	moderate
C4-Azlactone*	moderate	weak
Methyl 2-sulphophenyl octadecanoate	moderate	moderate
Isononanoyl chloride	moderate	moderate
3,5,5-Trimethylhexanoyl chloride	moderate	moderate
C9-Azlactone*	moderate	weak
3-Propylidenephthalide	moderate	moderate
3,4-Dihydrocoumarin	moderate	moderate
Sodium 3,5,5-trimethylhexanoyloxybenzenesulphonate	moderate	moderate
Palmitoyl chloride	moderate	moderate
1,2,4-Benzenetricarboxylic anhydride (Trimellitic anhydride)	moderate	moderate
Pationic 138C (Sodium Lauroyl Lactylate)	weak	weak
C11-Azlactone	weak	weak
C15 Azlactone	weak	weak
C17 Azlactone	weak	weak
Phenyl benzoate*	weak	non-sensitiser
Imidazolidinyl urea	weak	weak
C19-Azlactone	weak	weak
Penicillin G	weak	weak
Saccharin	non-sensitiser	non-sensitiser

Where * indicates incorrect prediction

Table 3. List of chemicals (predicted Michael Acceptors) whose potencies are incorrectly predicted

Name	Actual Potency Class	Mechanistic Domain	Predicted Mechanistic Domain	Predicted Potency Class
Isopropyl isoeugenol	strong	MA	MA	extreme
2-Hydroxyethyl acrylate	moderate	MA	MA	extreme
HC Red No3	moderate	MA	MA	strong
3-Aminophenol	moderate	MA	MA	strong
a-Methyl cinnamic aldehyde	moderate	MA	MA	weak
2-Methoxy-4-methyl-phenol	moderate	MA	MA	non-sensitiser
Dihydroeugenol	moderate	MA	MA	weak
Ethyl acrylate	weak	MA	MA	extreme
Propiolactone	strong	SN2	MA	extreme

Table 4. Seven Schiff base formers whose potencies were underpredicted

Name	Actual Potency Class	Predicted Potency Class
2-Methylundecanal	weak	strong
2,3-Butanedione	weak	strong
1-Phenyloctane-1,3-dione	weak	non-sensitiser
1-(2Æ,5Æ-Dimethylphenyl)butane-1,3-dione	weak	non-sensitiser
cis-6-Nonenal	weak	strong
2,2,6,6-Tetramethyl-heptane-3,5-dione	weak	strong
3-Ethoxy-1-(2Æ,3Æ,4Æ,5Æ-tetramethylphenyl)propane-1,3-dione	weak	non-sensitiser

Table 5. List of S_N2 reactors

Name	Actual Potency Class	Mechanistic Domain	Predicted Potency Class
1-Chloromethylpyrene	extreme	S_N2	extreme
Dimethyl sulfate	strong	S_N2	strong
Benzyl bromide	strong	S_N2	strong
Methyl dodecane sulphonate	strong	S_N2	strong
Methyl hexadecene sulphonate	strong	S_N2	strong
Bisphenol A-diglycidyl ether	moderate	S_N2	moderate
1-Bromohexadecane	moderate	S_N2	non-sensitiser
Diethyl sulfate	moderate	S_N2	moderate
2-Bromotetradecanoic acid	moderate	S_N2	moderate
1-Bromoheptadecane	moderate	S_N2	non-sensitiser
1-Bromopentadecane	moderate	S_N2	non-sensitiser
1-Bromoeicosane	moderate	S_N2	non-sensitiser
12-Bromo-1-dodecanol	moderate	S_N2	moderate
Methyl methanesulphonate	moderate	S_N2	strong
1-Bromodocosane	moderate	S_N2	non-sensitiser
Dodecyl methane sulphonate	moderate	S_N2	moderate (false positive)
1-Chlorohexadecane	moderate	S_N2	non-sensitiser
1-Bromotetradecane	moderate	S_N2	non-sensitiser
1-Bromohexane	weak	S_N2	non-sensitiser
1-Bromotridecane	weak	S_N2	non-sensitiser
1-Iodododecane	weak	S_N2	non-sensitiser
1-Iodotetradecane	weak	S_N2	non-sensitiser
1-Bromooctadecane	weak	S_N2	non-sensitiser
1-Chlorooctadecane	weak	S_N2	non-sensitiser
1-Bromododecane	weak	S_N2	non-sensitiser
12-Bromododecanoic acid	weak	S_N2	weak
1-Iodohexadecane	weak	S_N2	non-sensitiser
1-Bromoundecane	weak	S_N2	non-sensitiser
1-Chlorotetradecane	weak	S_N2	non-sensitiser
7-Bromotetradecane	weak	S_N2	weak
1-Iodononane	weak	S_N2	non-sensitiser
Oleyl methane sulphonate	weak	S_N2	moderate (false positive)
Butyl glycidyl ether	weak	S_N2	weak
1-Bromobutane	non-sensitiser	S_N2	non-sensitiser
1-Bromononane	non-sensitiser	S_N2	non-sensitiser
1-Chlorononane	non-sensitiser	S_N2	non-sensitiser
1-Iodohexane	non-sensitiser	S_N2	non-sensitiser
1-Iodooctadecane	non-sensitiser	S_N2	non-sensitiser
Methyl hexadecyl sulphonate	non-sensitiser	S_N2	strong
Sodium lauryl sulphate	moderate (false positive)	Non	moderate (false positive)
Isopropyl myristate	weak	Non	weak
Hexane	non-sensitiser	Non	non-sensitiser

Table 6. Breakdown of EINECS chemicals by predicted sensitisation potential and mechanistic class

Predicted Class and Potency	# of Chemicals
MA_weak	929
MA_extreme	2790
*MA_non	2900
MA_strong	4303
MA_mod	2269
SB_weak	2059
SB_mod	4026
SB_strong	3212
*SB_non	1055
S_N2_strong	1042
S_N2_weak	748
S_N2_extreme	71
S_N2_mod	738
S_N2_mod(fp)	3679
*S_N2_non	1631
Acyl_mod	1867
Acyl_weak	1251
*Acyl_non	1146
Acyl_extreme	328
Acyl_strong	407
*S_NAR_non	1186
S_NAR_extreme	3842
S_NAR_weak	626
SPEC_mod	1581
*SPEC_non	4664
SPEC_extreme	2002
*Assigned Non-Sensitiser based on RXNDOMAIN model	18641
Total	68,993

Table 7: Distribution of clusters for LLNA dataset

Mechanistic domain	Number of clusters
Michael Addition	12
Acyl formation	6
Schiff Base	8
S_N2	9
S_NAr	1
Special	3
Non-sensitisers	6

Figure 1. Distribution of the LLNA dataset within the actual and predicted reaction domains

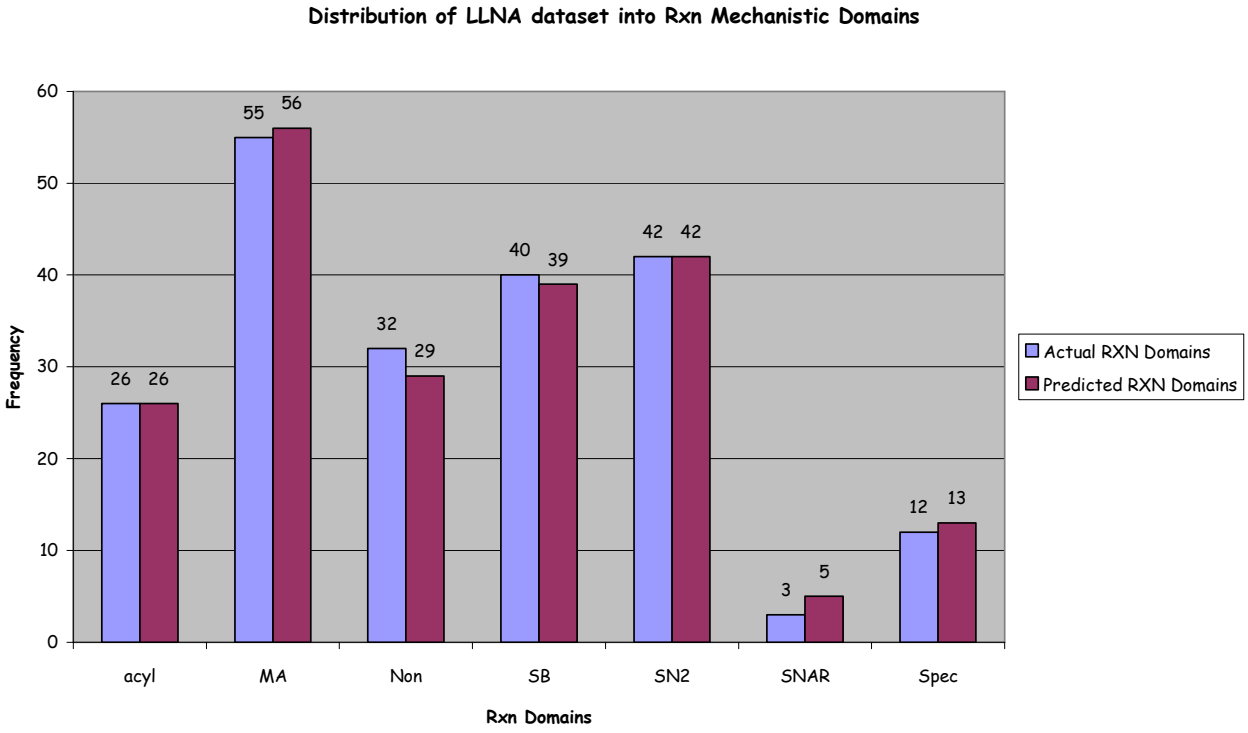


Figure 2. Predicted distribution of EINECS chemicals within the reaction domains

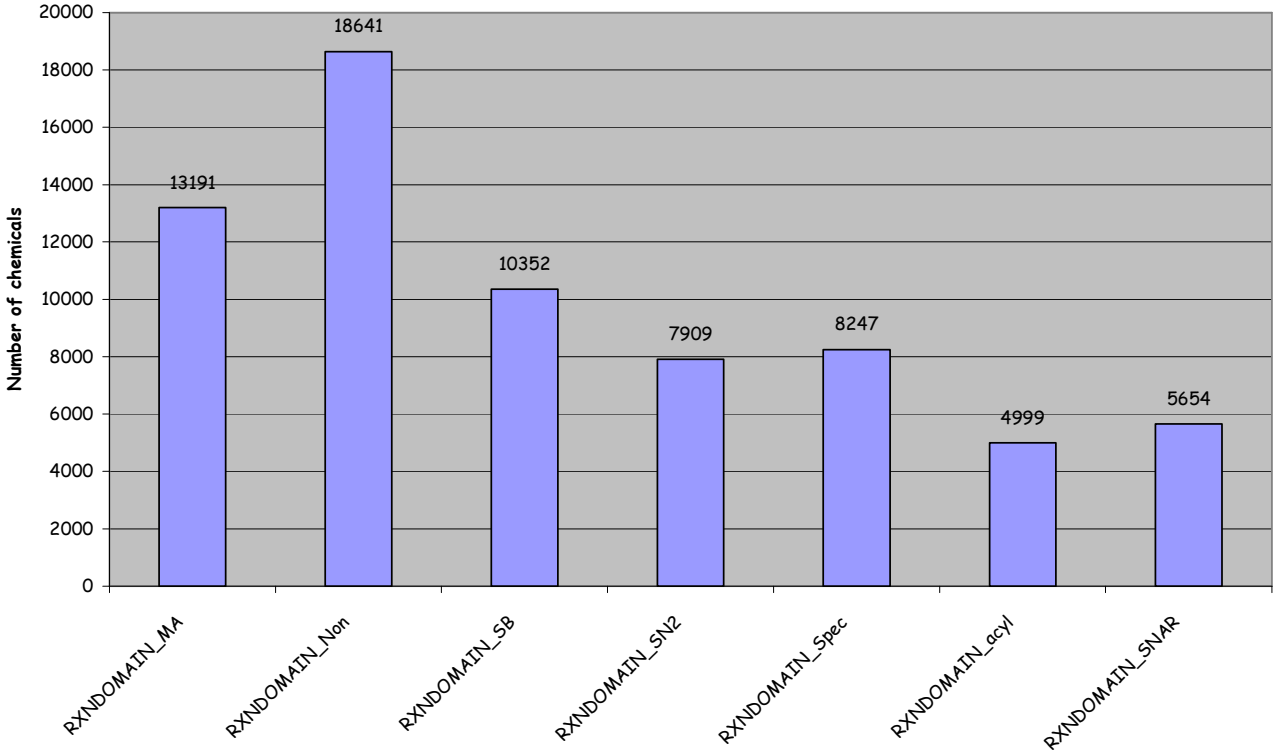


Figure 3. Pipeline Pilot workflow for assigning mechanistic domains and predicting potency

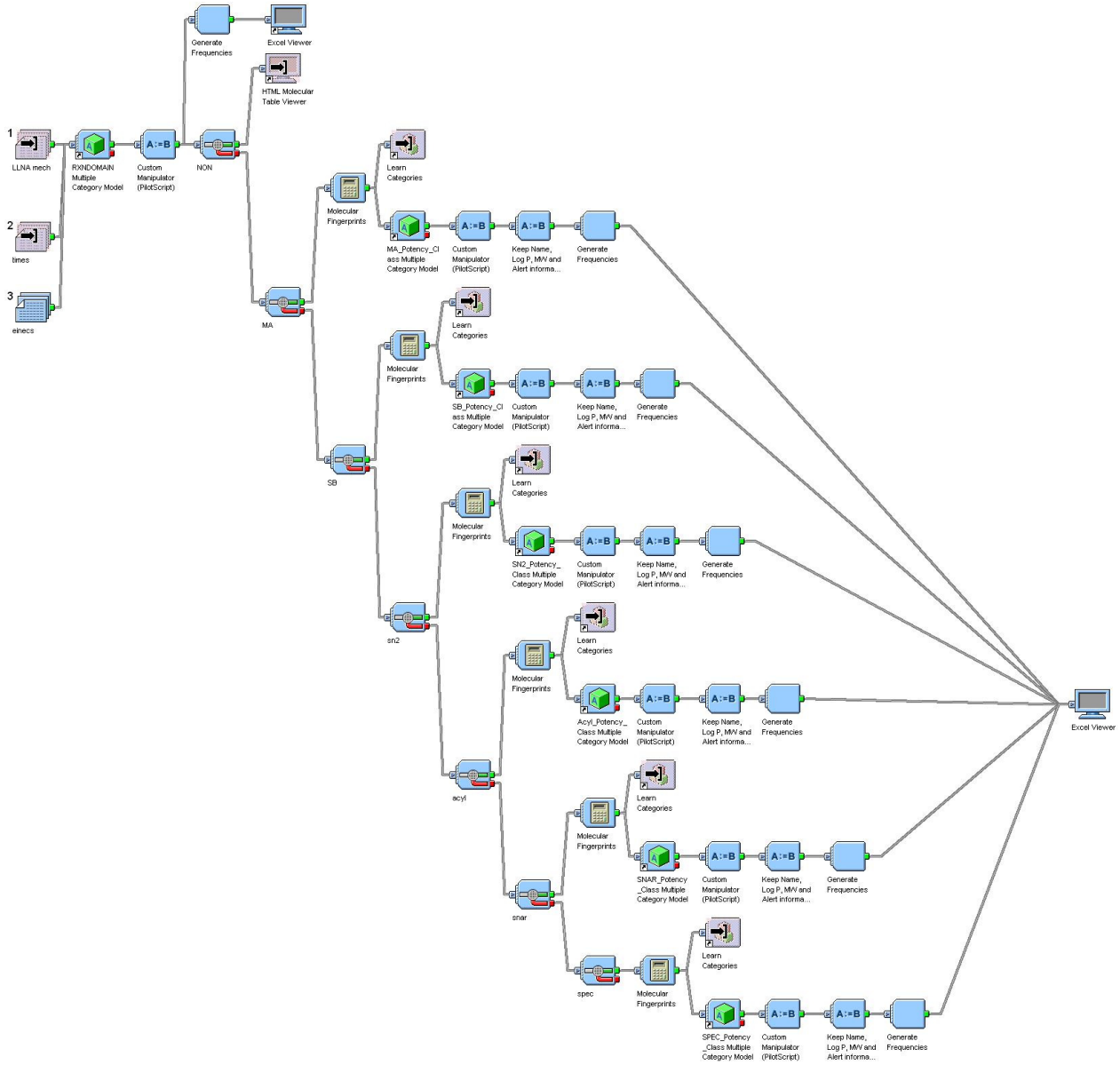


Figure 4. Distribution of EINECS chemicals by predicted sensitisation potential and mechanistic domain

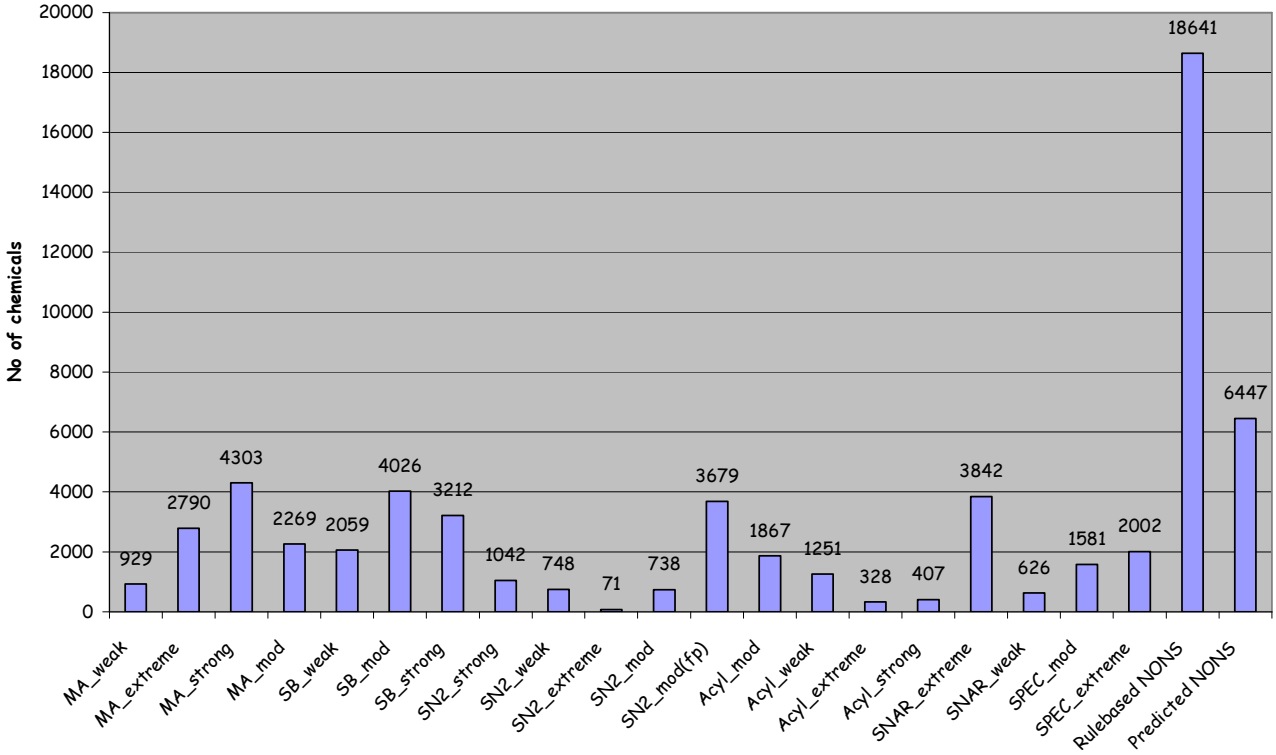


Figure 5. Pipeline Pilot workflow for assigning EINECS chemicals into mechanistic domains

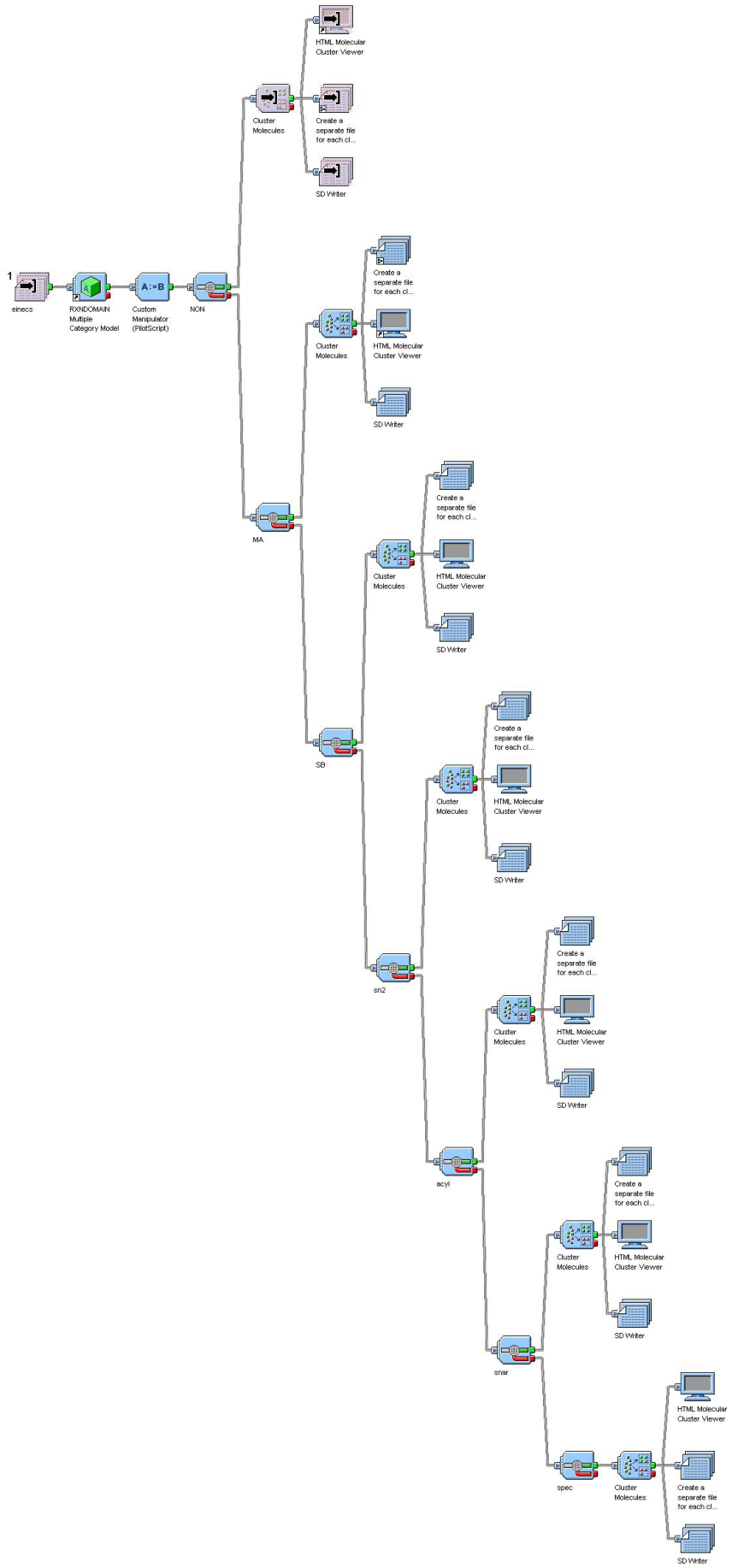


Figure 6. Cluster distribution for predicted acyl formers

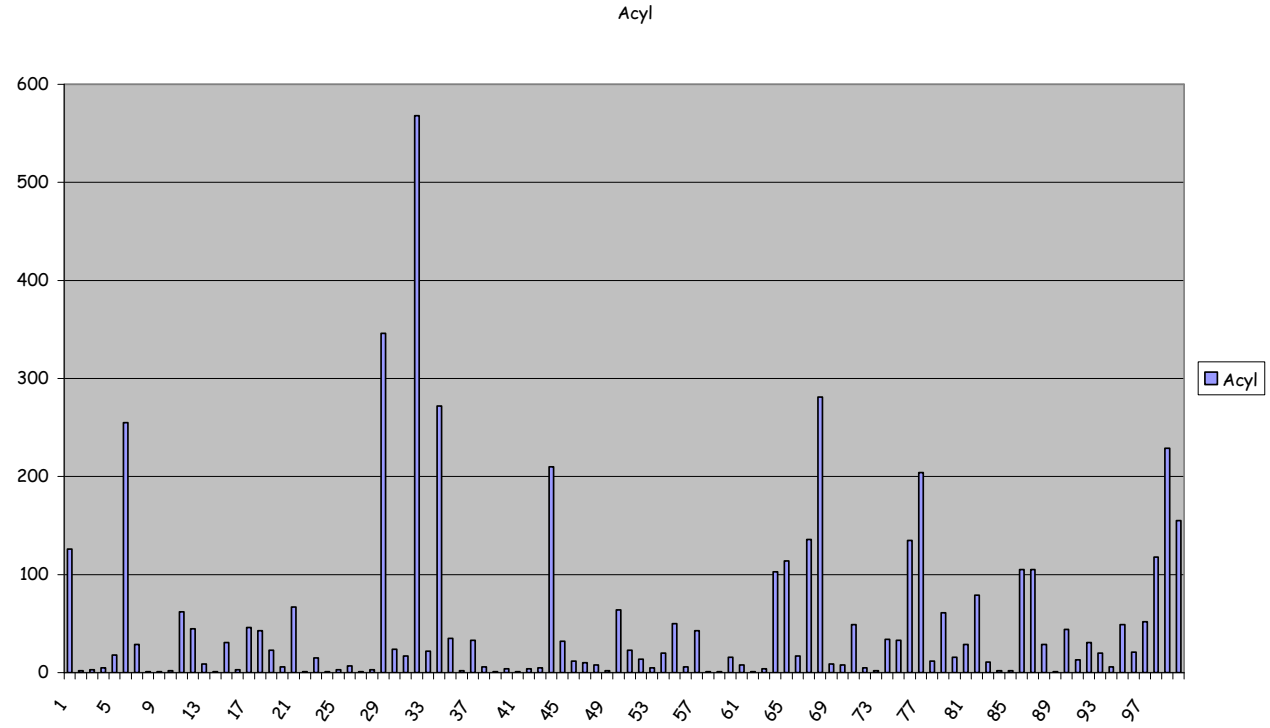


Figure 7. Distribution of clusters for predicted non-sensitisers

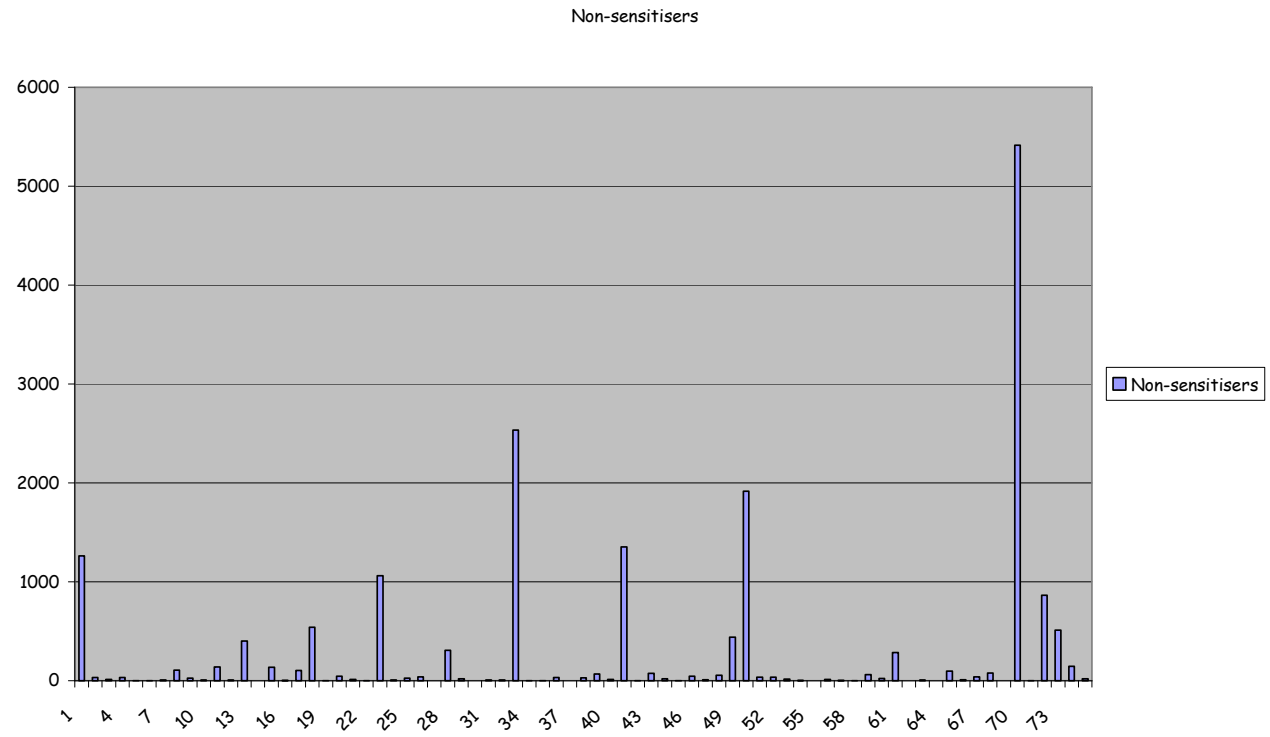


Figure 8. Distribution of clusters for predicted Michael acceptors

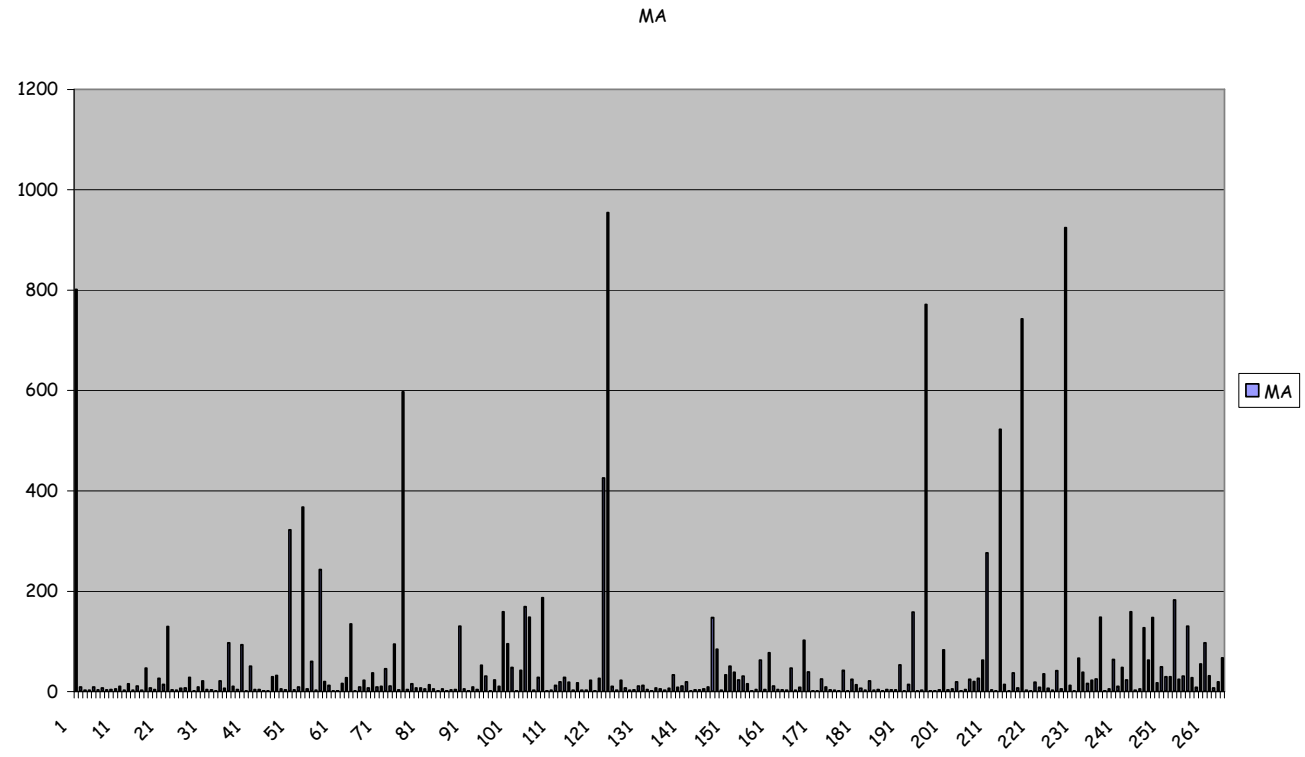


Figure 9. Distribution of clusters for the predicted Schiff base formers

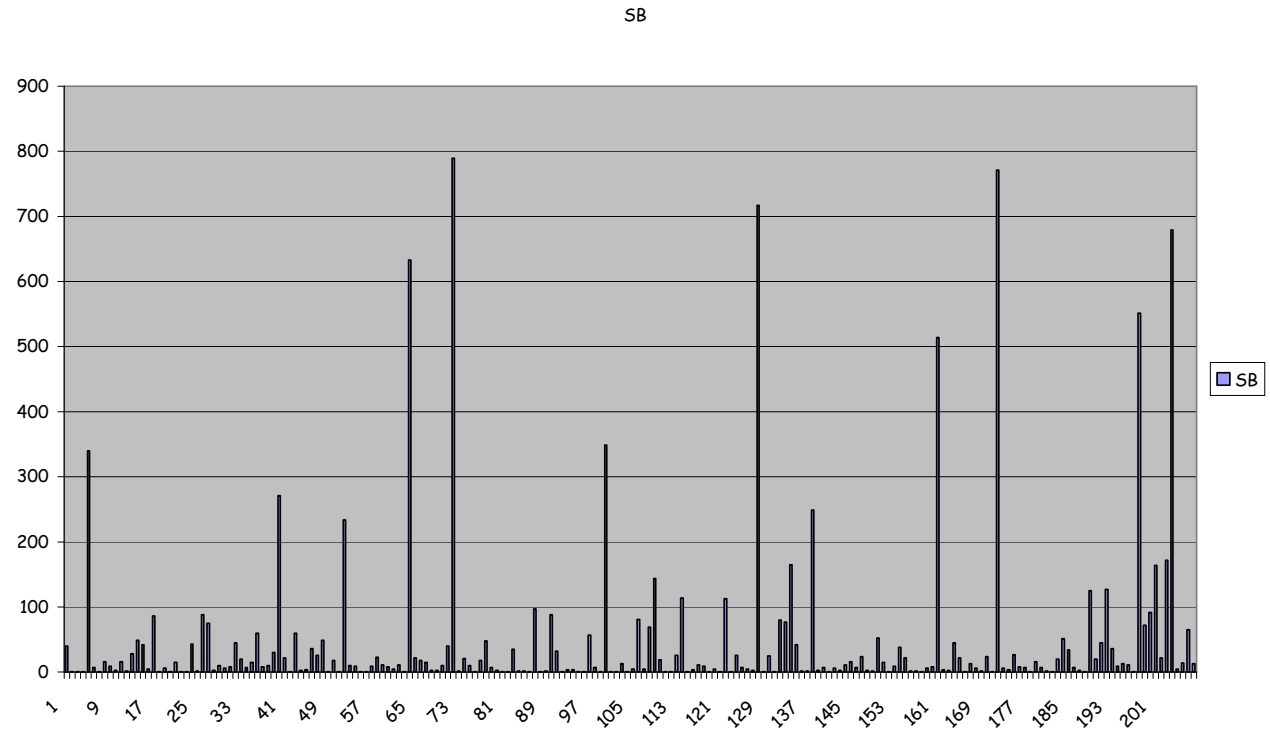


Figure 10. Distribution of clusters for the predicted S_N2 reactors

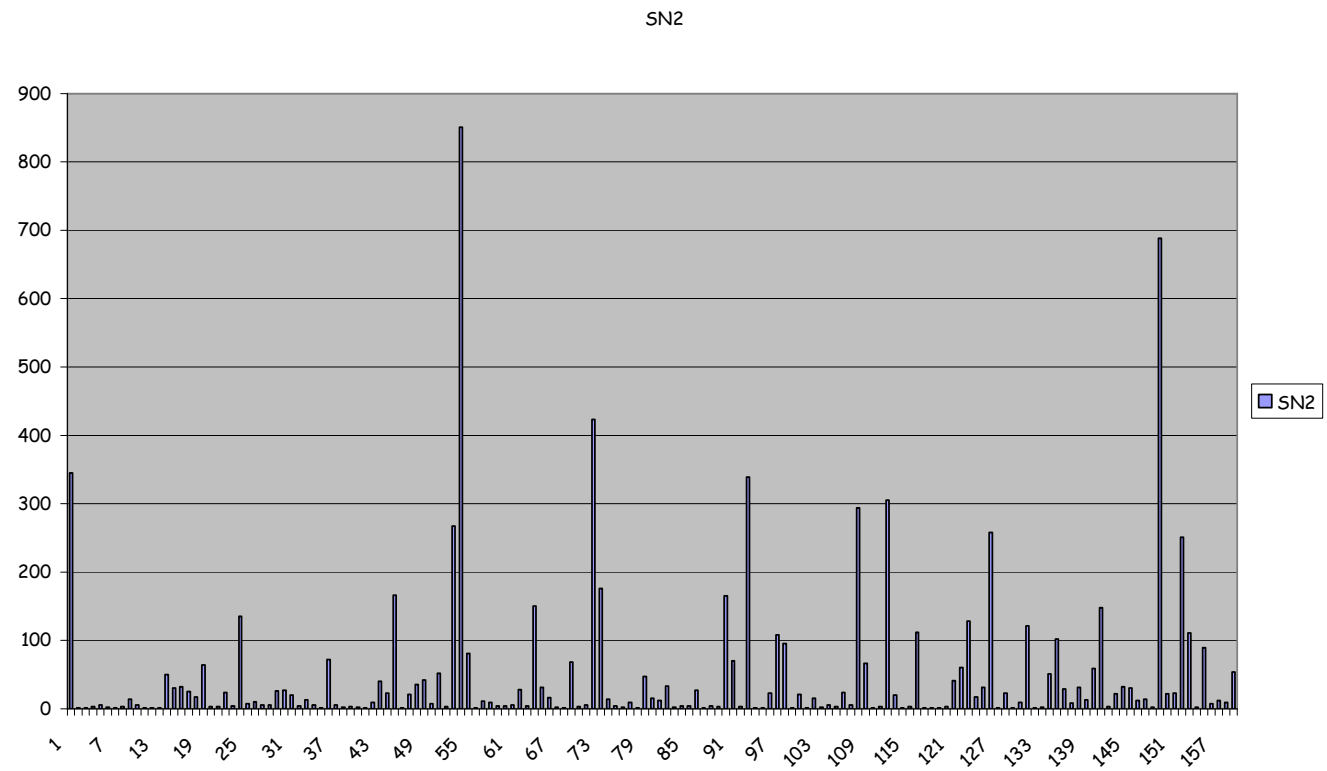


Figure 11. Distribution of clusters for the predicted S_NAR reactors

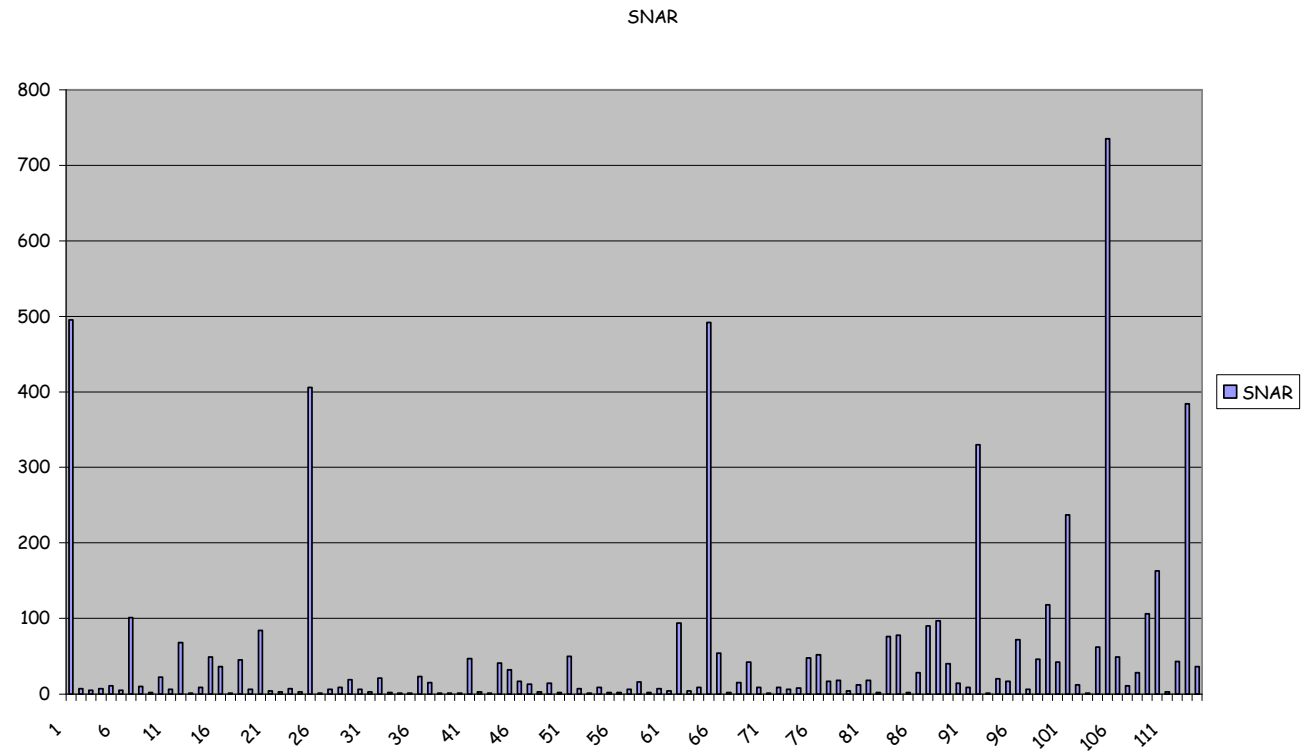


Figure 12. Distribution of clusters for the predicted special class

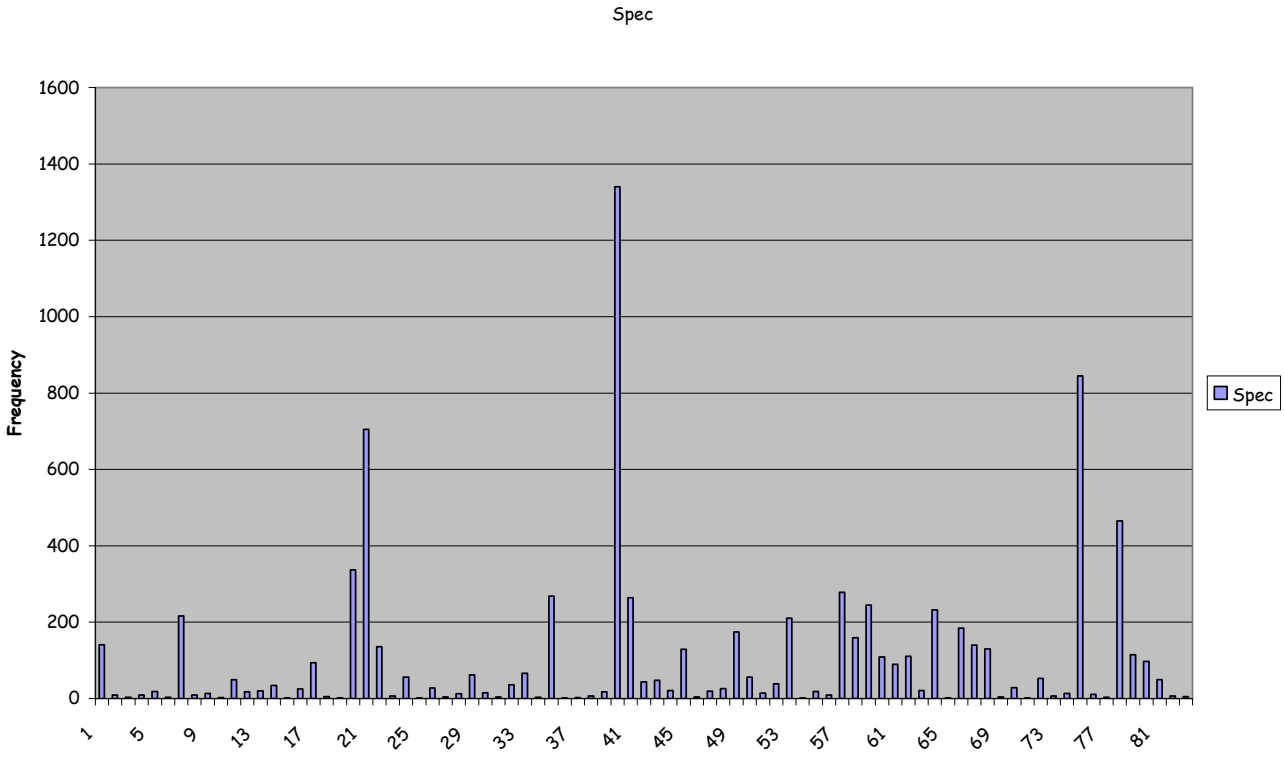
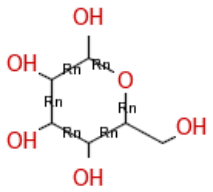
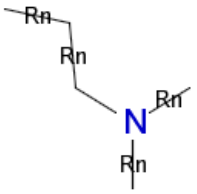
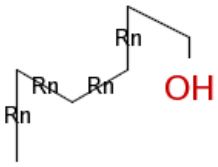
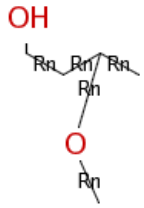
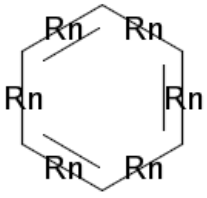
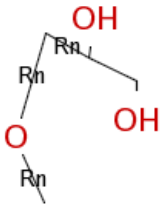
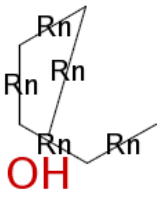
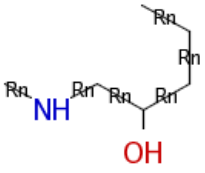
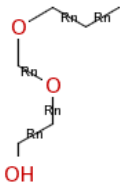
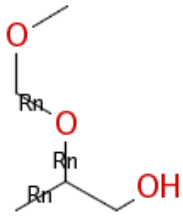


Figure 13. 20 Subgraphs for Cluster 14 (non-sensitisers)

Structure	
	
	
	
	
	

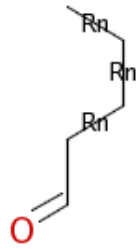
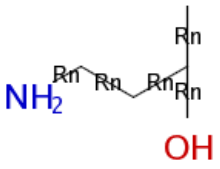
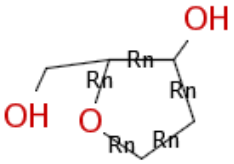
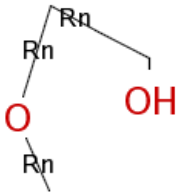
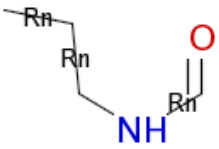
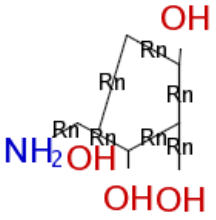
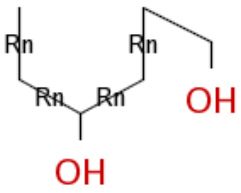
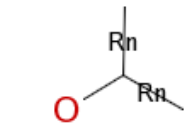
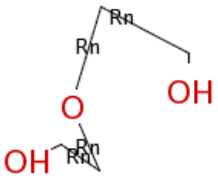
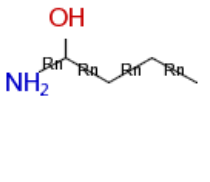
	
	
	
	
	

Figure 14. 7 Subgraphs for Cluster 33

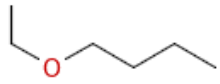
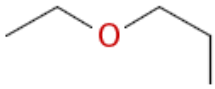
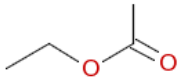
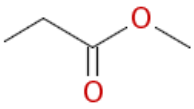
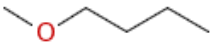
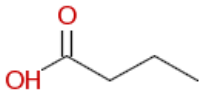
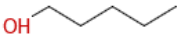
Structures	
	
	
	
	

Figure 15. Subgraph for Cluster 2

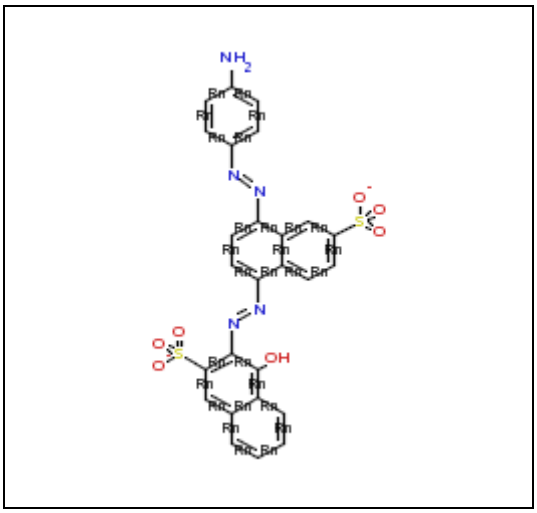


Figure 16. Subgraph for Cluster 93

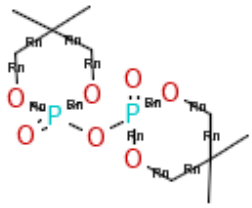


Figure 17. 11 Subgraphs for Cluster 32

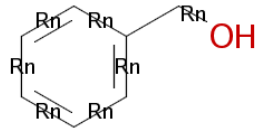
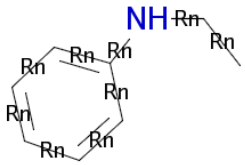
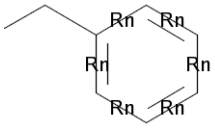
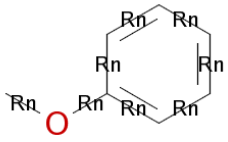
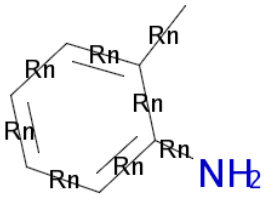
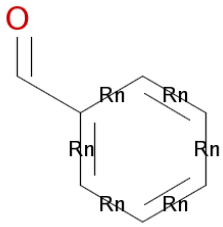
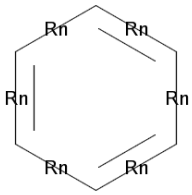
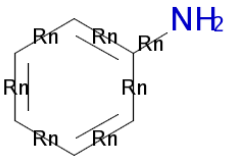
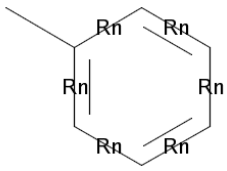
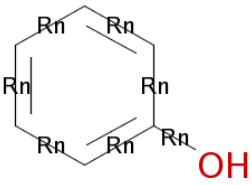
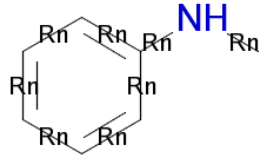
Structures	
	
	
	
	
	
	

Figure 18. MCS for Non-sensitisers

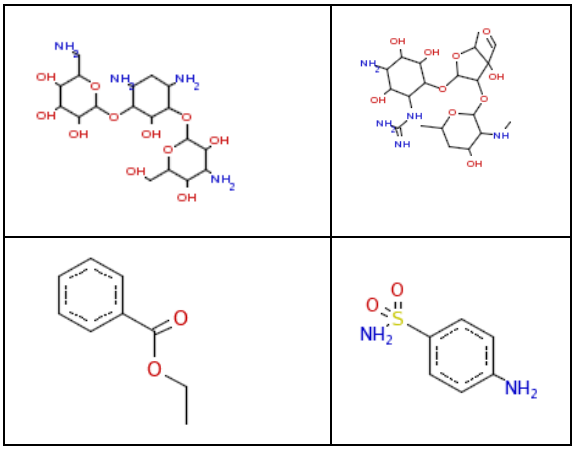


Figure 19. MCS for Acyl formers

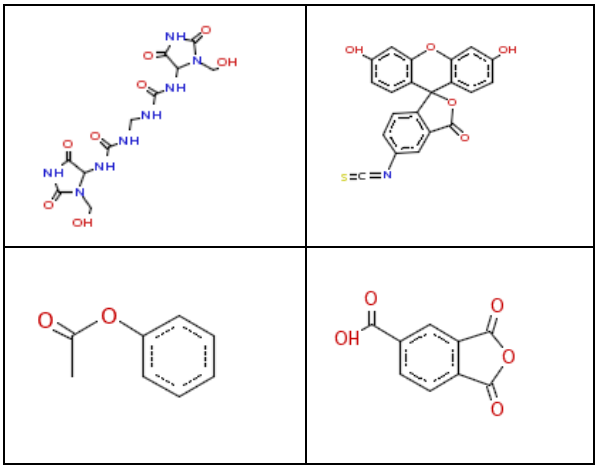
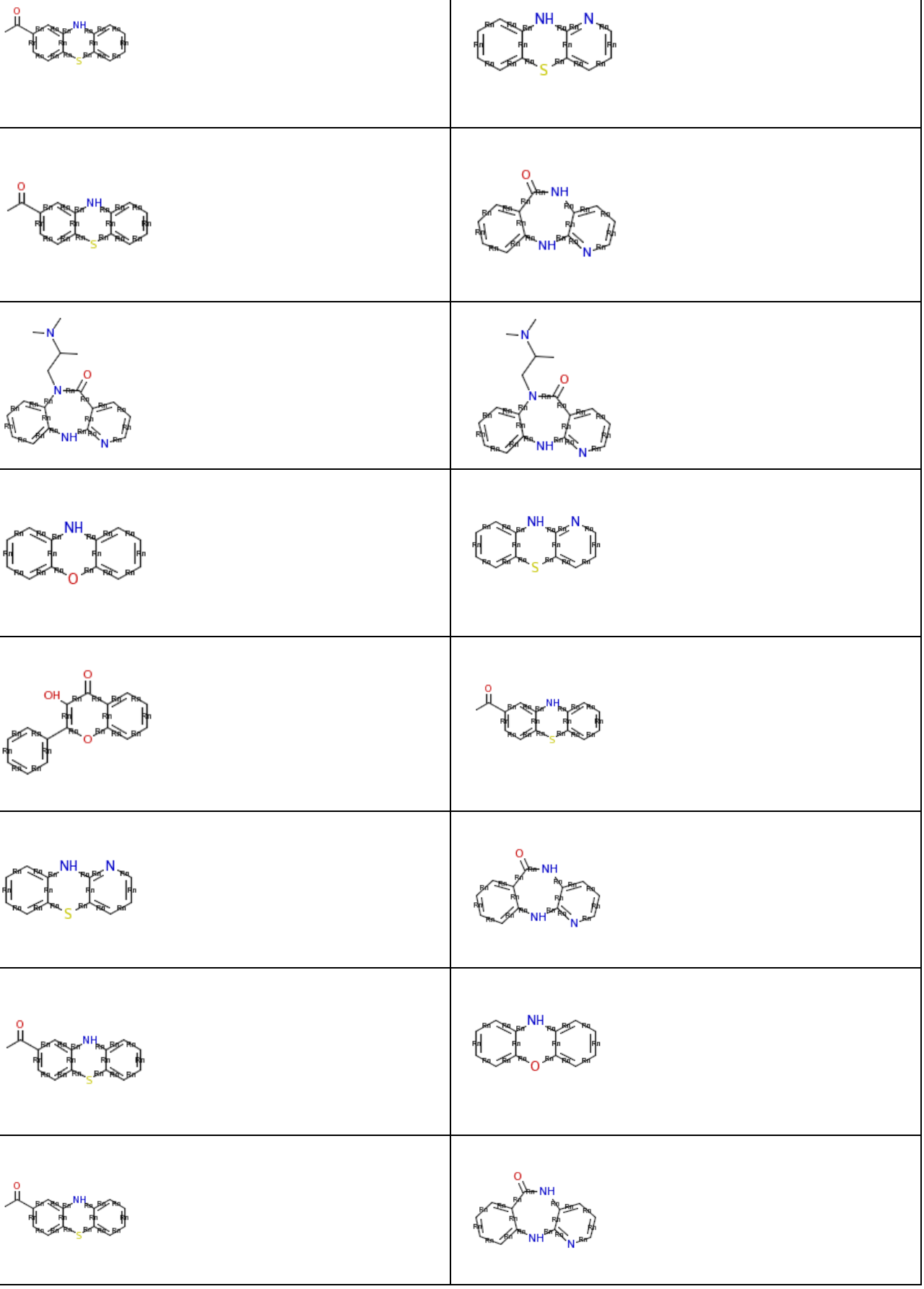


Figure 20. Parent MCS structures extracted from Cluster 32



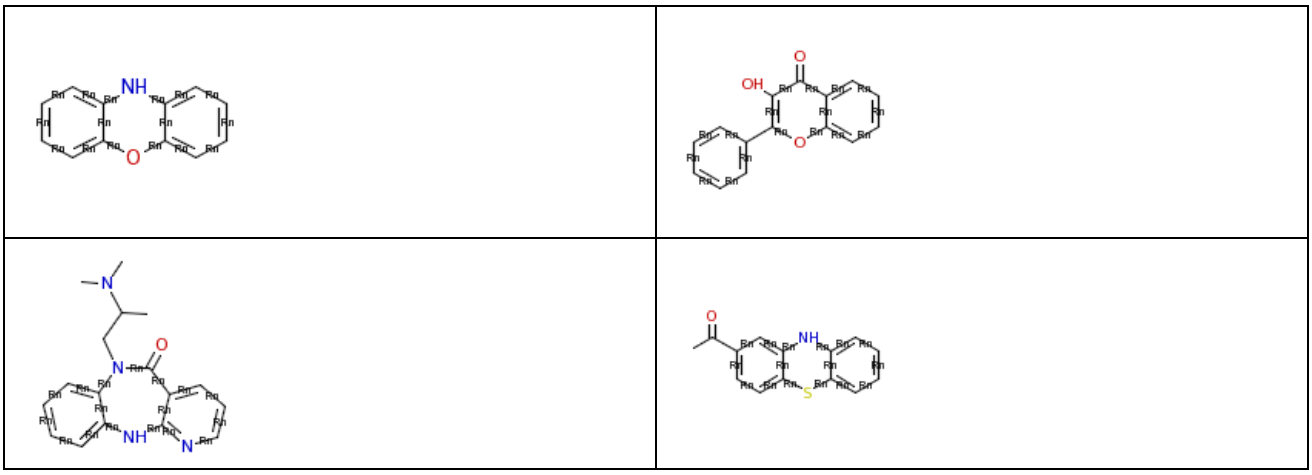
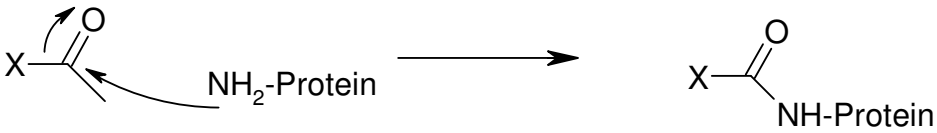


Figure 21. Rule taken from reference (3) for acylating agents



X = halogen, or other group such that XH is sufficiently acidic for X- to act as a good leaving group. Includes anhydrides, cyclic or non-cyclic. X = o-alkyl does not qualify apart from when part of a strained system. Analogous rxns can occur with attack at sulfonyl S and phosforyl P and thioacyl C.

European Commission

EUR 22941 EN – Joint Research Centre – Institute for Health and Consumer Protection

Title: The Use of Computational Methods in the Grouping and Assessment of Chemicals - Preliminary Investigations

Author(s): Worth A, Bassan A, Fabjan E, Gallegos Saliner A, Netzeva T, Patlewicz G, Pavan M and Tsakovska I

Luxembourg: Office for Official Publications of the European Communities

2007 – 123 pp. – x cm
EUR – Scientific and Technical Research series – ISSN 1018-5593

Abstract

This document presents a perspective of how computational approaches could potentially be used in the grouping and assessment of chemicals, and especially in the application of read-across and the development of chemical categories. The perspective is based on experience gained by the authors during 2006 and 2007, when the Joint Research Centre's European Chemicals Bureau was directly involved in the drafting of technical guidance on the applicability of computational methods under REACH. Some of the experience gained and ideas developed resulted from a number of research-based case studies conducted in-house during 2006 and the first half of 2007. The case studies were performed to explore the possible applications of computational methods in the assessment of chemicals and to contribute to the development of technical guidance. Not all of the methods explored and ideas developed are explicitly included in the final guidance documentation for REACH. Many of the methods are novel, and are still being refined and assessed by the scientific community. At present, many of the methods have not been tried and tested in the regulatory context. The authors therefore hope that the perspective and case studies compiled in this document, while not intended to serve as guidance, will nevertheless provide an input to further research efforts aimed at developing computational methods, and at exploring their potential applicability in regulatory assessment of chemicals.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.