

User manual

Strategies for grouping chemicals to fill data gaps to assess acute aquatic toxicity endpoints

Document history

Version	Comment
Version 1.0	April 2010, first version
Version 1.1	July 2013, second version, updated for version 3.1 of the Toolbox

Issue date: July 2013

Language: English

If you have questions or comments that relate to this document, please send them to ehscont@oecd.org or visit the QSAR Toolbox discussion forum at https://community.oecd.org/community/toolbox_forum

Table of Contents

1	How to use this guidance document	5
2	Acute aquatic toxicity endpoints	6
3	Primary profilers relevant to acute aquatic toxicity endpoints	8
3.1	Aquatic toxicity classification by ECOSAR	8
3.2	Acute aquatic toxicity MOA by OASIS	8
3.3	Acute aquatic toxicity classification by Verhaar.....	9
3.4	Protein binding by OASIS V1.1	9
3.5	Protein binding by OECD.....	9
4	Database relevant to acute aquatic toxicity.....	10
5	Profiling results: What they tell say about a grouping strategy.....	11
5.1	Potential profiling results	11
5.1.1	Propylbenzene - narcosis	12
5.1.2	Ethylcinnamaldehyde – reactive	12
5.2	General conclusions regarding the outcome of profiling strategy results	13
5.2.1	The use of multiple mechanistic profilers.....	13
5.2.2	The use of endpoint specific profilers.....	14
6	Secondary profilers relevant to acute aquatic toxicity endpoints	14
6.1	Defining the structural domain of a chemical category	15
7	Initial category formation using the primary profilers to define the mechanistic domain	16
7.1	Inert chemical (narcosis)	17
7.2	Reactive chemical acting via single covalent mechanism.....	18
7.3	General conclusions regarding the sub-categorisation with the primary profilers	19
8	Profiling examples which result in the ability to fill data-gaps	19
8.1	Profiling and data gap filling for inert chemical by trend analysis	20
8.1.1	Initial and secondary profiling	20
8.1.2	Data-gap filling via trend analysis.....	21
8.2	Profiling and data gap filling for reactive chemical using read across	21
8.2.1	Initial profiling using the primary profilers	22
8.2.2	Initial category formation and sub-categorisation using the primary profilers	22
8.2.3	Empiric sub-categorisation using the secondary profilers.....	23
8.2.4	Data-gap filling via read across.....	23
9	General approach for the development of categories for acute aquatic toxicity endpoints	24
10	References	26

Table of Tables

Table 3.1: Primary profilers for acute aquatic toxicity available in the OECD QSAR Toolbox..... 8
 Table 4.1: Summary of acute aquatic toxicity databases in the OECD QSAR Toolbox..... 10
 Table 6.1: Secondary profilers relevant to acute aquatic toxicity endpoints available in the OECD QSAR Toolbox V3.1..... 14

Table of Figures

Figure 2.1: Screenshot of the list of the acute aquatic toxicity effects available in the OECD QSAR Toolbox V3.1. 6
 Figure 2.2: Screenshot of the acute aquatic toxicity data classification for mortality effect. 7
 Figure 5.1: Result of primary profiling for inert chemical - propylbenzene using the profilers available for acute aquatic toxicity in the OECD QSAR Toolbox V3.1..... 12
 Figure 5.2: Result of primary profiling for 4-ethylcinnamaldehyde using the profilers available for acute aquatic toxicity in the OECD QSAR Toolbox V3.1..... 13
 Figure 7.1: Sub-categorisation window showing the additional structural alerts identified using the aquatic toxicity classification by ECOSAR profiler when sub-categorising the initial category formed for the target chemical propylbenzene..... 18
 Figure 7.2: Results of sub-categorisation using the five primary profilers for 2,3-dimethylvaleraldehyde. 19
 Figure 8.1: Data-gap present (shown in grey) for propylbenzene in the 96 hour *Pimephales promelas* assay. 20
 Figure 8.2: Trend analysis allowing an LC50 to be predicted for propylbenzene in the 96 hour *Pimephales promelas* assay. 21
 Figure 8.3: Results of primary profiling for formylcyclohexane using the profilers available for acute aquatic toxicity in the OECD QSAR Toolbox V3.1..... 22
 Figure 8.4: Read across prediction made for the *Pimephales promelas* 96hr LC50 endpoint for formylcyclohexane..... 24
 Figure 9.1: General scheme for category formation for acute aquatic toxicity endpoints 25

1 How to use this guidance document

It is the purpose of this document to provide guidance on the use of the profilers and databases within the OECD QSAR Toolbox (V3.1) with the express aim of providing non-prescriptive guidance to help the user build categories that are mechanistically and structurally robust. The aim being that such categories will allow the filling of data gaps for acute aquatic toxicity endpoints.

This document is for users having some experience with the workflow of the OECD QSAR Toolbox. OECD recommends that users first read the manual for getting started, which are available at www.oecd.org/env/existingchemicals/qsar.

The document is split into several sections, these being:

- Sections 2 – 4: Introductory material about the profilers and databases available for acute aquatic toxicity.
- Sections 5 – 8: Worked examples for profiling of target chemicals and this information can be used to form chemical categories for inert and reactive chemicals. These sections are intended to be used as examples that the reader can follow as illustrations of several recommended strategies.
- Section 9: Summary of a general strategy that can be used to generate chemical categories suitable for filling data gap. This summary was used to generate the example categories covered in this guidance document. The worked examples in sections 5 – 8 should be undertaken before attempting to use this summary information.

2 Acute aquatic toxicity endpoints

Acute aquatic toxicity means the intrinsic property of a substance to be injurious or fatal to an organism during a short-term exposure to that substance [1]. Acute aquatic toxicity is normally determined using a fish 96-hour LC50 (OECD Test Guideline 203), a crustacea species 48-hour EC50 (OECD Test Guideline 202) and/or an algal species 72- or 96-hour EC50 (OECD Test Guideline 201). These species cover a range of trophic levels and taxa and are considered as surrogate for all aquatic organisms.

Acute aquatic toxic effects are the most data-rich endpoints in the OECD QSAR Toolbox and are subcategorised into classes based on the observed adverse effects, such as: accumulation, avoidance, behaviour, biochemistry, cell(s), development, ecosystem process, enzyme(s), feeding behaviour, genetics, growth, histology, hormone(s), immobilisation, injury, intoxication, morphology, mortality, no effect coded, physiology, population, reproduction and undefined effect (Figure 2.1).

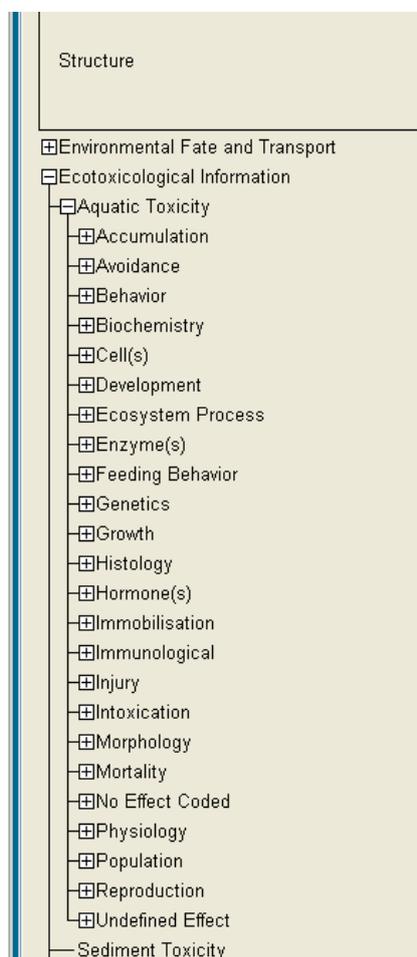


Figure 2.1: Screenshot of the list of the acute aquatic toxicity effects available in the OECD QSAR Toolbox V3.1.

These adverse effects are further divided into different classes based on the measured endpoint (e.g. LC50, EC50, LOEC) and duration and exposure regimes (e.g. 24h, 48h, 72h, 96h). Finally, the acute aquatic toxicity data are classified based on the organism upon which the test was carried out (Figure 2.2). As a result of three decades of testing there is an extensive depth and breadth of acute aquatic toxicity data. This testing has resulted in several important factors being identified:

- Acute aquatic toxicity has a water solubility-related minimal toxicity, which, while it may be superseded by other modes of action, forms a baseline for potency.
- The majority of the industrial organic chemicals, especially the most common ones, are baseline toxicants, which act via the non-polar narcosis mode of toxic action.

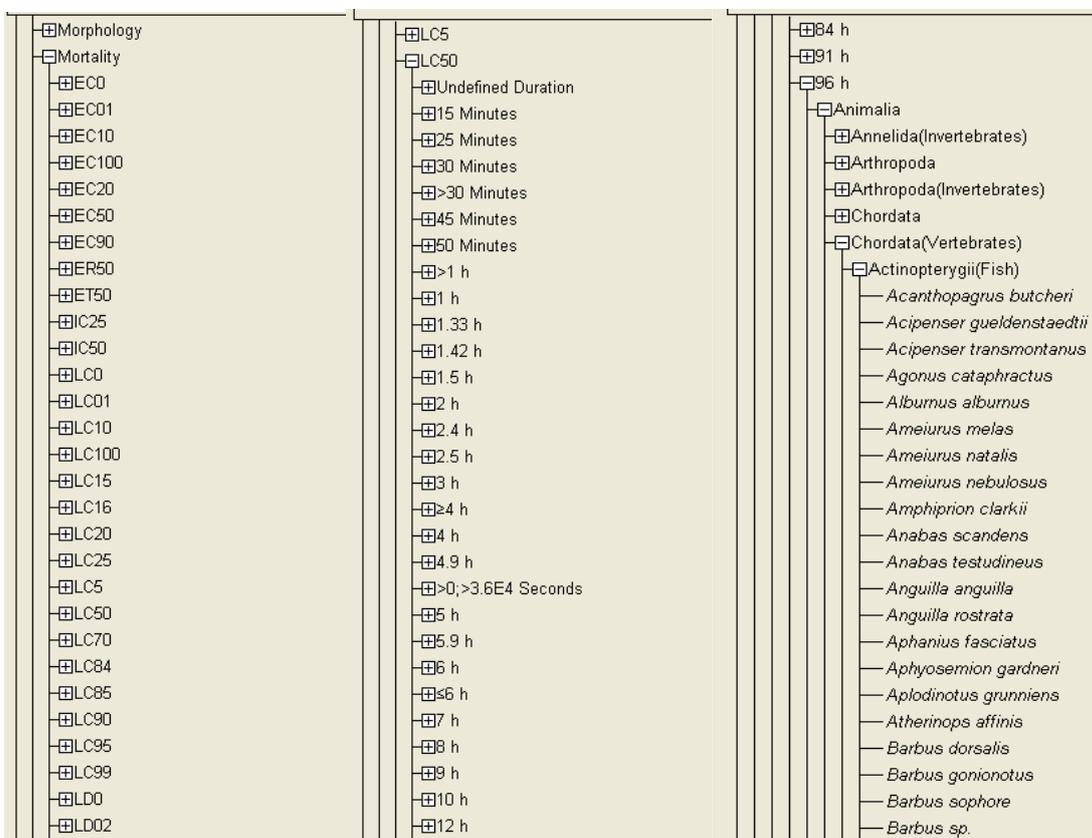


Figure 2.2: Screenshot of the acute aquatic toxicity data classification for mortality effect.

3 Primary profilers relevant to acute aquatic toxicity endpoints

The five primary profilers relevant to acute aquatic toxicity endpoints can be divided into two types: three endpoint-specific profilers and two mechanistic profilers (Table 3.1). The endpoint-specific profilers contain structural alerts that have been identified as being associated with toxicity from an analysis of aquatic toxicological data. The mechanistic profilers have been developed from knowledge of the organic chemistry related to the formation of a covalent bond between a chemical and a protein. These profilers contain structural alerts related to this organic chemistry; they are not however, necessarily supported by toxicological data.

Table 3.1: Primary profilers for acute aquatic toxicity available in the OECD QSAR Toolbox

Profiler name	Type	Number of alerts	Section
Aquatic toxicity classification by ECOSAR	Endpoint	118	2.1
Acute aquatic toxicity MOA by OASIS	Endpoint	6	2.2
Acute aquatic toxicity classification by Verhaar	Endpoint	5	2.3
Protein binding by OASIS	Mechanistic	87	2.4
Protein binding by OECD	Mechanistic	102	2.5

3.1 Aquatic toxicity classification by ECOSAR

The 'aquatic toxicity classification by ECOSAR' profiler is the most detailed profiler that is applicable to acute aquatic endpoints. It is based on 40 years of experience in the Office of Pollution Prevention and Toxics of the United States Environmental Protection Agency. This profiler has 118 structural alerts that have been linked to toxicity in excess of baseline. Many are related to specific modes of toxic action or specific molecular initiating events and thus have a sound mechanistic basis. Others, however, lack this mechanistic grounding. The major advantage to the 'aquatic toxicity classification by ECOSAR' profiler is the large number of alerts that are based on experimental evidence gathered from fish, aquatic invertebrates (*Daphnia*), and aquatic plants (green algae). The major disadvantage of the profiler is that it often lacks mechanistic transparency for the basis of the category. The neutral organic or basesurface narcotics are often arrived at because the target chemical does not fit in any other categories. Thus, there may be no experimental evidence for chemicals assigned to this category.

3.2 Acute aquatic toxicity MOA by OASIS

The 'acute aquatic toxicity MOA by OASIS' profiler was developed by the Laboratory of Mathematical Chemistry, Bourgas "Prof. As. Zlatarov" University, Bourgas, Bulgaria. It is based on a broader set of structural alerts gathered primarily from the fathead minnow toxicity testing and defined by Russom et al. [2]. The profiler classifies a chemical into one of seven categories: aldehydes; alpha, beta-unsaturated alcohols; phenols and anilines; esters; narcotic amines; basesurface narcotics. It also includes further rules based on simple metabolism. The major advantage to the OASIS acute toxicity mode of action profiler is it assigns categories based on modes of toxic action; thus, there is usually a clear mechanistic foundation to the category, which improves transparency and aids acceptability.

However, this profiler is based on fish toxicity data and its applicability to other organisms may be less relevant and should always be critically considered.

3.3 Acute aquatic toxicity classification by Verhaar

The 'acute aquatic toxicity classification by Verhaar' profiler was developed utilising acute toxicity data collection for guppies and fathead minnows [3]. This scheme is based on structural alerts that allow chemicals to be assigned to one of five classes. These classes being: class 1 or "inert" chemicals, which are nonpolar narcosis or baseline toxicity; class 2 or "less inert" chemicals, which are the polar narcotics; class 3 or "reactivity" chemicals, which are typically non-selectively, covalently reactive with protein moieties; class 4 or "specifically-acting" chemicals, which show reactivity to specific receptors and class 5 or "unclassified" chemicals. The structural alerts within this profiler allow chemicals to be readily assigned to one of the first three classes. This frequently results in the 'acute aquatic toxicity classification by Verhaar' profiler correctly assigned chemicals to class 1. However, a large number of industrial organic compounds get relegated to class 5 as the fail to trigger any structural alerts in the profiler.

3.4 Protein binding by OASIS V1.1

The 'protein binding profile developed by OASIS V1.1' includes 87 structural alerts. These structural alerts are related to mechanisms by which chemicals can covalently react with thiol (SH) and amino (NH₂) groups of proteins [4]. Briefly, covalent bonds between a substrate and a target molecule are formed by reactions between electron-rich nucleophiles and electron-poor electrophiles. Electron-rich groups usually contain heteroatoms (ones other than carbon or hydrogen), especially in nucleic acids and proteins. The preference of a chemical toward a specific molecular site of action can be explained by a classification of electrophiles and nucleophiles according to their polarisability, in other words, the chemical "hardness" and "softness" of the electrophilic or nucleophilic centre. Generally, soft electrophiles will react preferentially with thiol groups e.g. cysteine amino acids in proteins, while harder electrophiles will prefer to react with the amino groups of e.g. lysine amino acids in proteins. Thus, establishing whether a compound is electrophilic in nature and secondly the type of electrophile (and associate this with a reaction mechanism) can be of great benefit to predicting acute aquatic toxicity because electrophilic reactivity is often the molecular initiating event and potency-determining property for acute toxicity. The major advantage to the protein binding profiler is it that the resulting categories are based on well documented and well-understood chemical reactions. Therefore, there is a clear mechanistic foundation to the category, which improves transparency and acceptability.

3.5 Protein binding by OECD

The 'protein binding by OECD' profiler was developed by an analysis of direct acting structural alerts based on theoretical organic chemistry (the profiler does not contain metabolically/abiotically activated structural alerts) [5]. The alert compilations were analysed in order to place the information contained within the literature into a mechanistic chemistry framework. This mechanistic chemistry can be used as the basis for chemical category formation when utilising the protein binding by OECD profiler. Within each of the five mechanistic domains, related structural alerts have been grouped based on the presence of a common reactivity site into so-called

mechanistic alerts. Chemical category formation can be carried out at either the mechanistic alert or structural alert level using this profiler. The protein binding by OECD profiler contains 18 mechanistic alerts covering 102 structural alerts. These data are supported by mechanistic chemistry and references to the scientific literature (the meta data).

4 Database relevant to acute aquatic toxicity

The OECD QSAR Toolbox is well populated with experimental data for acute aquatic effects due to the long history of acute aquatic toxicity testing in fish, crustaceans, algae, and protozoa. These studies have resulted in four searchable databases within the Toolbox (Table 4.1). In order to collect the largest number of possible chemicals within a chemical category, it is recommended to use all four databases.

Table 4.1: Summary of acute aquatic toxicity databases in the OECD QSAR Toolbox

Database	Number of chemicals	Number of data points	Number of endpoints
Aquatic ECETOC	734	9487	33
Aquatic Japan MoE	464	2900	4
Aquatic OASIS	2390	4826	8
Aquatic US-EPA ECOTOX	7894	337204	147

5 Profiling results: What they tell say about a grouping strategy

A number of mechanisms have been identified that can lead to aquatic toxicity, with the majority of industrial chemicals exerting their toxic influence via non-covalent mechanisms. These mechanisms being: polar narcosis and non-polar narcosis [6]. In addition, a smaller, yet significant, proportion of industrial chemicals exert their toxicity via irreversible covalent bond formation. This involves electrophilic reactions between amino acid side chains such as cysteine and lysine and certain structural features of the chemical. Several other, less common, mechanisms have been also identified [7]. The profilers in the OECD QSAR Toolbox can be used to develop a chemical category suitable for data-gap filling that take this mechanistic knowledge into account. This can be achieved by following steps for category formation:

1. Profile the target chemical (for which a data gap exists) for potential mechanism\mode of action
2. Use the result of this profiling to select chemical analogues from the four aquatic toxicity databases
3. Define the mechanistic and structural domain of the resulting chemical category
4. Fill the data-gap using trend analysis and/or read across

As noted earlier defining the chemical category is the critical step in the workflow of the OECD QSAR Toolbox. The application of the initial profilers relevant to acute aquatic toxicity endpoints typically assigns the target chemical as either inert (defined as acting via non-polar or polar narcosis) or reactive (capable of forming a covalent bond with a protein).

5.1 Potential profiling results

The first of the two possible profiling results for aquatic toxicity occurs when one (or both) of the mechanistic profilers identifies a single mechanism related to covalent protein binding that is supported by a single structural alert identified by at least one of the endpoint specific profilers. Importantly, it does not matter whether only one of the mechanistic profilers or both identify the single mechanism (as long as if both do they identify the same mechanism). This is because there is significant overlap between the structural alerts contained within these profilers as they both outline the chemistry associated with covalent protein binding. Confidence in the profiling results is gained by the appropriate endpoint specific profiler identifying a complementary structural alert. The confidence is gained due to the fact that the endpoint specific profilers contain only structural alerts that have toxicological data associated with them (again only a single endpoint specific profilers needs to identify a complementary structural alert due to the overlap in the toxicological data from which they have been developed). In the case of reactive chemicals evidence from either the 'aquatic toxicity classification by ECOSAR' or 'acute aquatic toxicity MOA by OASIS' profilers are preferred as they are able to identify specific functional related to reactivity. In contrast, the 'acute aquatic toxicity classification by Verhaar' profiler only identifies a reactive chemical as belonging to class 3 and does not identify specific functional groups. This makes it less useful in support of the protein binding mechanism identified by the mechanistic profilers.

There is a second type of ideal profiling scenario that can occur when neither of the mechanistic profilers identifies a mechanism related to covalent protein binding suggesting that the target chemical is non-reactive (in this case it is important that the two mechanistic profilers are in

agreement to ensure there is, as far as the current knowledge in the Toolbox is concerned, there is no evidence of covalent protein binding). If these mechanistic profiling results are supported by one of the endpoint specific profilers identifying a chemical as being a narcotic then one has confidence in the profiling results. Consider the following examples:

5.1.1 Propylbenzene - narcosis

Profiling propylbenzene using the two mechanistic profilers shows neither of them to identify any mechanisms related to covalent protein binding. These results are supported by the endpoint specific profilers that suggest propylbenzene to exert its toxicity via hydrophobicity dependent narcosis [3]. It is important to realise that a chemical is classified as a narcotic by the endpoint specific profilers when it is profiled as a neutral organic by the ‘aquatic toxicity classification by ECOSAR’ profiler, as a basesurface narcotic by the ‘acute aquatic toxicity MOA by OASIS V1.1’ profiler and as belonging to class 1 or 2 by the ‘acute aquatic toxicity classification by Verhaar’ profiler (Figure 5.1).

filter endpoint tree...	1 [target]
Structure	
<ul style="list-style-type: none"> ⊕ Substance Identity ⊕ Physical Chemical Properties ⊕ Environmental Fate and Transport ⊕ Ecotoxicological Information ⊕ Human Health Hazards ⊖ Profile <ul style="list-style-type: none"> ⊖ General Mechanistic <ul style="list-style-type: none"> Protein binding by OASIS v1.1 Protein binding by OECD ⊖ Endpoint Specific <ul style="list-style-type: none"> Acute aquatic toxicity classification by Verhaar Acute aquatic toxicity MOA by OASIS Aquatic toxicity classification by ECOSAR 	<ul style="list-style-type: none"> No alert found No alert found Class 1 (narcosis or baseline toxicity) Basesurface narcotics Neutral Organics

Figure 5.1: Result of primary profiling for inert chemical - propylbenzene using the profilers available for acute aquatic toxicity in the OECD QSAR Toolbox V3.1.

5.1.2 Ethylcinnamaldehyde – reactive

Profiling 4-ethylcinnamaldehyde shows the ideal profiling situation for reactive chemicals in that at least one of the mechanistic profilers identifies a mechanism associated with covalent protein binding (in the example, the results indicate either Michael addition or Schiff base formation). These mechanistic profiling results are supported by the profiling results from both ‘aquatic toxicity classification by ECOSAR’ and ‘acute aquatic toxicity MOA by OASIS’ profilers which identify structural features related to the suggested protein binding mechanisms. These results are summarised in Figure 5.2.

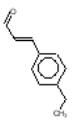
Filter endpoint tree...	1 [target]
Structure	
<input type="checkbox"/> Substance Identity	
CAS Number	N/A
Chemical IDs	NA
Chemical Name	
Structural Formula	<chem>c1(C=CC=O)ccc(CC)cc1</chem>
<input checked="" type="checkbox"/> Physical Chemical Properties	
<input checked="" type="checkbox"/> Environmental Fate and Transport	
<input checked="" type="checkbox"/> Ecotoxicological Information	
<input checked="" type="checkbox"/> Human Health Hazards	
<input type="checkbox"/> Profile	
<input type="checkbox"/> General Mechanistic	Michael addition Michael addition >> a,b-unsaturated carbonyl compounds Michael addition >> a,b-unsaturated carbonyl compounds >> a,b-unsaturated aldehydes Schiff base formation Schiff base formation >> Schiff base formation with carbonyl compounds Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes Michael addition Michael addition >> Polarised Alkenes Michael addition >> Polarised Alkenes >> Polarised alkene - aldehydes Schiff Base Formers Schiff Base Formers >> Direct Acting Schiff Base Formers Schiff Base Formers >> Direct Acting Schiff Base Formers >> Mono-carbonyls
Protein binding by OASIS v1.1	
Protein binding by OECD	
<input type="checkbox"/> Endpoint Specific	
Acute aquatic toxicity classification by Verhaar	Class 3 (unspecific reactivity)
Acute aquatic toxicity MOA by OASIS	Aldehydes
Aquatic toxicity classification by ECOSAR	Vinyl/Allyl Aldehydes

Figure 5.2: Result of primary profiling for 4-ethylcinnamaldehyde using the profilers available for acute aquatic toxicity in the OECD QSAR Toolbox V3.1.

5.2 General conclusions regarding the outcome of profiling strategy results

The above section outline the common examples of profiling results that one is likely to encounter when developing chemical categories for the endpoints discussed in this guidance document. The examples raise several important issues in terms of the confidence associated with profiling results and thus the subsequent category, these being:

5.2.1 The use of multiple mechanistic profilers

It is perhaps tempting to suggest that if both mechanistic profilers indicate the same mechanism that one should have more confidence in the profiling results. However, this is not the case as the mechanistic profilers have been developed from a range of toxicological data sources (many of which are the same for both profilers). Thus, in the case where both mechanistic profilers trigger the same alert it is likely that the underlying structural alert has been developed from the same (or similar) data. In addition, the situation where only one of the mechanistic profilers triggers an alert does not mean that the results are of lower confidence. All that can be stated in this scenario is that the target chemical contains an alert that is outside the domain of the second profiler. As discussed, the only time where one can make a decision about confidence based on the results from the two mechanistic profilers is in the situation where the results suggest multiple, competing mechanisms.

5.2.2 The use of endpoint specific profilers

The endpoint specific profilers contain structural alerts that have been shown to be associated with aquatic toxicity endpoints. In contrast, the mechanistic profilers have been developed from an analysis of a range of data sources (including general mechanistic chemistry knowledge related to covalent protein binding). Thus, not all of the structural alerts within them have been definitively associated with acute aquatic toxicity for which covalent protein binding is the molecular initiating event. As discussed, this means that the information in the mechanistic profilers can be supplemented with the information in the endpoint specific profilers allowing one to have increased confidence in the resulting category.

6 Secondary profilers relevant to acute aquatic toxicity endpoints

In addition to the primary profilers, a number of secondary profilers are also of use in category formation for acute aquatic toxicity endpoints. These profilers are summarised in Table 6.1. In contrast to the initial battery of profilers which are used in combination with one another, the secondary profilers are best utilised individually to help sub-categorise a chemical category. Such sub-categorisation is often needed to refine the structural domain of a chemical category allowing transparent structure-activity relationships to be developed. However, since the secondary profilers are based on imperfect structural similarity, it is important to review the list of structures provided with each sub-categorisation routine to assure one is not eliminating analogues for unknown reasons.

Table 6.1: Secondary profilers relevant to acute aquatic toxicity endpoints available in the OECD QSAR Toolbox V3.1

Profiler name	Type	Number of alerts
Organic functional groups	Empiric	489
Organic functional groups (nested)	Empiric	489
Organic functional groups (US EPA)	Empiric	467
Organic functional groups, Norbert Haider (checkmol)	Empiric	204
Chemical elements	Empiric	NA

The most commonly utilised and useful secondary profilers are the organic functional groups and chemical elements profilers. These profilers allow the user to develop sub-categories based on the presence or absence of common organic functional groups such as carbonyl, nitro or many others. In addition, the chemical elements profiler allows sub-categories to be developed based on the presence or absence of chemical elements. A combination of one of the organic functional group profilers and the chemical elements profiler can provide useful sub-categories depending on the makeup of the chemical category. The choice of which of the four organic functional group profilers to use is largely dependent on the data within the category one wishes to sub-categorise. However, as a general approach one is advised to use the organic functional group profiler as it relates to well established organic functional groups and thus is the most interpretable. The first two organic functional group profilers (general and nested) include the same functional groups. The difference is that the organic functional group (general) displays all functional groups present in the target

compounds, while nested one does not show the functional groups, which are only parts of larger ones. The remaining two organic functional group profilers should be used in cases where the organic functional group profiler does not provide a satisfactory sub-category. In addition to the organic functional group profiler, the chemical elements profiler is also a useful secondary profiler. This profiler encodes the chemical elements within a molecule allowing the user to exclude a given element or sets of elements. This would become useful during the fine-tuning of a chemical category as it allows the user to restrict the category members to those whose elements are the same as are present in the target chemical.

6.1 Defining the structural domain of a chemical category

One of the key functions of the secondary profilers is in the definition of the structural domain of the chemical category. It is important that chemicals containing (significantly) different elements and functional groups are removed from the category. Typically this is achieved using a combination of the organic functional group and chemical elements profilers (see Table 6.1 in section 6). Ideally, the category resulting from the primary profiling should contain only chemicals with the same elements and functional groups as the target chemical (those identified in the 'target menu' of the sub-categorisation window). However, this is not always possible and using such a tight structural domain results in the elimination of too many analogues from the category. In these instances, one can include more functional groups (by selecting them 'by hand' in the 'analogues menu' in the sub-categorisation window) to increase the number of analogues in the resulting category. A useful approach to ensure that the structural domain of the chemical category is suitable for subsequent data-gap filling is as follows (in usage order):

1. Profile the endpoint-specific category using the organic functional group profiler removing all chemicals that contain functional groups not present in the target chemical
2. Inspect the resulting chemical category – if it contains sufficient analogues (that one considers) suitable to fill the data gaps of interest then no further sub-categorisation is required (the absolute minimum for read across is a category containing the target chemical and a single analogue as this would allow for one-to-one read across. However, ideally one would like a category in which trend analysis and/or read across predictions could be made on a many-to-one basis. Thus, one wants a category containing at least two or three analogues if possible. For general guidance on grouping, chemical category formation and read across see [8]).
3. If step 1 results in insufficient chemicals considered suitable for data gap filling, then re-profile the endpoint-specific category using the organic functional group profiler. However, instead of removing all chemicals as before, additional simple non-ionisable organic functional groups not present in the target chemical should be included (simple alkyl groups for example). This increases the likelihood that there will be sufficient chemicals included in the resulting chemical category to allow for data gap filling.

When profiling for organic functional groups 'by hand' (as in step 3 above) it is extremely important to visually inspect the types (i.e. the chemical structures and associated functional groups) of chemicals that one is eliminating. The chemicals that will be eliminated can be visualised by right clicking on the 'sub-categorisation' window and selecting 'display selected'. One approach when sub-categorising organic functional groups in this way is to try to produce a chemical category in which

hydrophobicity is responsible for (the majority of) the trends in toxicity. Doing so will ensure that any subsequent predictions made by read across or trend analysis are as transparent and interpretable as possible.

7 Initial category formation using the primary profilers to define the mechanistic domain

In the OECD QSAR Toolbox, there are five primary profilers relevant to acute aquatic toxicity endpoints, which were discussed in sections 3.1 – 3.5. There are two mechanistic and three endpoint specific profilers; ‘protein binding by OASIS V1.1’, ‘protein binding by OECD’, ‘acute aquatic toxicity are aquatic toxicity classification by ECOSAR’, ‘acute aquatic toxicity MOA by OASIS’ and ‘acute aquatic toxicity classification by Verhaar’. As discussed in section 5.1 the ‘acute aquatic toxicity classification by Verhaar’ profiler is the least useful of these profilers due to its inability to identify specific structural alerts for reactive chemicals.

The formation of the initial chemical category is carried out by profiling the relevant databases to acute aquatic toxicity using either a mechanistic or endpoint specific profiler depending on whether the chemical is identified as being potentially reactive or not. The following general steps are recommended for the formation of the initial category:

1. Profile the target chemical using the primary profilers relevant to acute aquatic toxicity.
2. Using the results of the primary profilers profile the relevant databases¹ to acute aquatic toxicity for chemical analogues acting via the same mode of action as the target chemical. It is recommended that for baseline narcotics either the ‘aquatic toxicity classification by ECOSAR’ profiler or ‘acute aquatic toxicity MOA by OASIS’ profiler be used to develop the initial category. When the target chemical is profiled as reactive, then one of the mechanistic protein binding profilers should be chosen to form the initial category.
3. The resulting category is termed the ‘initial category’.

It is frequently necessary to perform a sub-categorisation of the initial category using one or more of the primary profilers relevant to acute aquatic toxicity endpoints (see sections 3.1 – 3.5). This is to ensure that the chemical category relates to a single mechanism of action. Sub-categorisation of a category is carried out as follows:

1. Profile the initial category with the primary profiler that was used to develop the initial category. This profiling will identify the mechanisms (if using a mechanistic profiler) or structural alerting groups (if using an endpoint-specific profiler) that are present in the initial category. These mechanisms (or structural alerting groups) are displayed in the sub-categorisation window.
2. Eliminate analogues from the initial category that contain additional mechanisms (or structural alerting groups if using an endpoint-specific profiler). Ensure that the ‘differ from target by’ option in the sub-categorisation window is set to ‘at least one category’.

¹ Profiling the relevant databases means searching the databases for analogues.

3. Repeat steps 1 and 2 using the other relevant primary profilers. The specific order in which these profilers are applied is not important, although it is recommended that the 'acute aquatic toxicity classification by Verhaar' profiler be used last.

Specific examples will now be discussed to show how the above process works for inert and reactive chemicals.

7.1 Inert chemical (narcosis)

This section relates to the development of a category suitable for data-gap filling for the target chemical propylbenzene. As outlined in section 5 the first step is to profile the target chemical using the five primary profilers shown in Table 3.1. The result of this profiling suggests that propylbenzene is an inert chemical that exerts its toxicity via non-polar narcosis. Therefore, the endpoint-specific 'aquatic toxicity classification by ECOSAR' profiler should be used to develop an initial category using the four applicable databases outlined in Table 4.1. This analysis results in an initial category consisting of 2109 chemicals (including target chemical).

The following sub-categorisations are then required to ensure the category contains only inert analogues that exert their toxicity via narcosis:

1. Sub-categorisation of the initial category of 2109 chemicals using the 'aquatic toxicity classification by ECOSAR' profiler (the profiler that was used to develop the initial category). Figure 7.1 highlights the additional structural alerts that are present in the initial category. This sub-categorisation results in a category of 1722 chemicals.
2. Sub-categorisation of the category of 1722 chemicals using the 'protein binding by OECD' profiler. This sub-categorisation results in a category of 1460 chemicals.
3. Sub-categorisation of the category of 1460 chemicals using the 'protein binding by OASIS V1.1' profiler. This sub-categorisation results in a category of 1427 chemicals.
4. Sub-categorisation of the category of 1427 chemicals using the 'acute aquatic toxicity MOA by OASIS' profiler. This sub-categorisation results in a category of 1160 chemicals.
5. Sub-categorisation of the category of 1160 chemicals using the 'acute aquatic toxicity classification by Verhaar' profiler. This sub-categorisation results in a category of 484 chemicals.

The resulting category is relatively large consisting of 484 chemicals. Therefore, it is recommended to apply the secondary profilers outlined in section 5 to define the structural domain of the category. The application of secondary profilers in this way is discussed in detail in section 8.1.1.

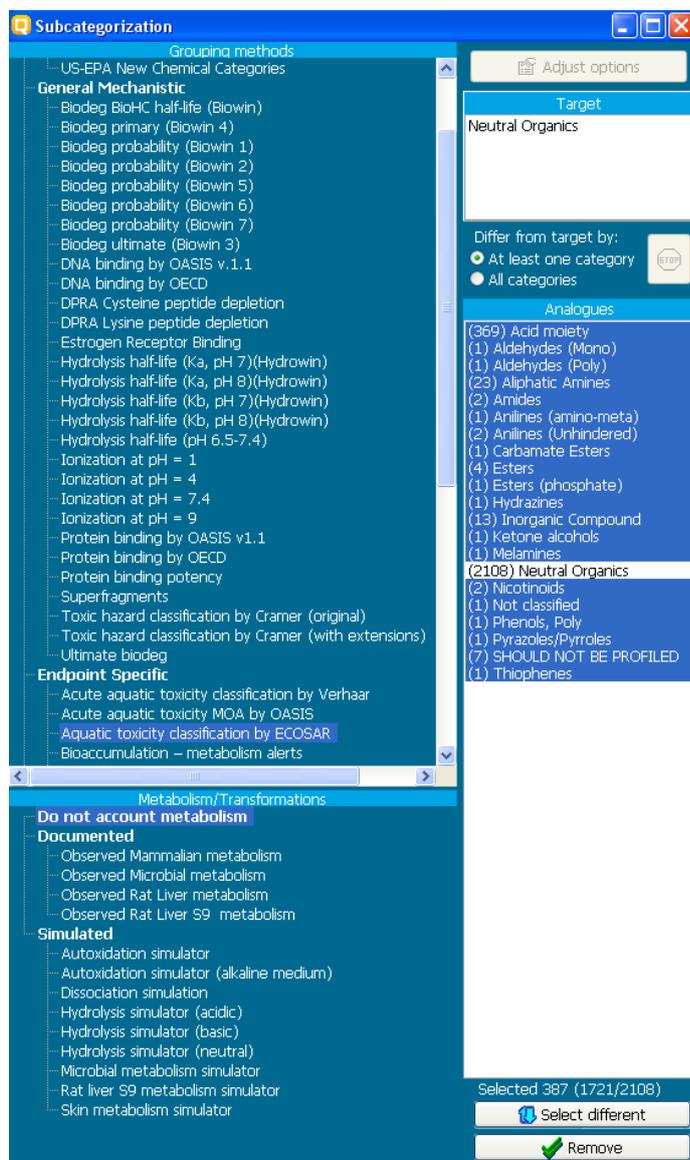


Figure 7.1: Sub-categorisation window showing the additional structural alerts identified using the aquatic toxicity classification by ECOSAR profiler when sub-categorising the initial category formed for the target chemical propylbenzene.

7.2 Reactive chemical acting via single covalent mechanism

An analogous category formation and sub-categorisation process using the five primary profilers can also be carried out for the target chemical 2,3-dimethylvaleraldehyde. The profiling results suggest this chemical is an electrophile that exerts its toxicity via Schiff base mechanism due to the presence of a carbonyl moiety. In cases such as this in which the initial profiling indicates the chemical to be reactive it is recommended that one of the mechanistic profilers be used to develop the initial chemical category (it doesn't matter which of the two one chooses). The 'protein binding by OECD' profiler identifies mechanistic analogues from the four applicable databases that results in an initial category of 120 chemicals (including target chemical).

The following sub-categorisations are then required to ensure the category contains analogues acting via a single electrophilic mechanism of action (Figure 7.2):

1. Sub-categorisation of the initial category of 120 chemicals using 'protein binding by OECD' profiler (the profiler that was used to develop the initial category). This sub-categorisation results in a category of 75 chemicals.
2. Sub-categorisation of the category of 75 chemicals using the 'protein binding by OASIS V1.1' profiler. This sub-categorisation results in a category of 75 chemicals.
3. Sub-categorisation of the category of 75 chemicals using the 'aquatic toxicity classification by ECOSAR' profiler. This sub-categorisation results in a category of 46 chemicals.
4. Sub-categorisation of the category of 46 chemicals using the 'acute aquatic toxicity MOA by OASIS' profiler. This sub-categorisation results in a category of 44 chemicals.
5. Sub-categorisation of the category of 44 chemicals using the 'acute aquatic toxicity classification by Verhaar' profiler. This sub-categorisation results in a category of 35 chemicals.

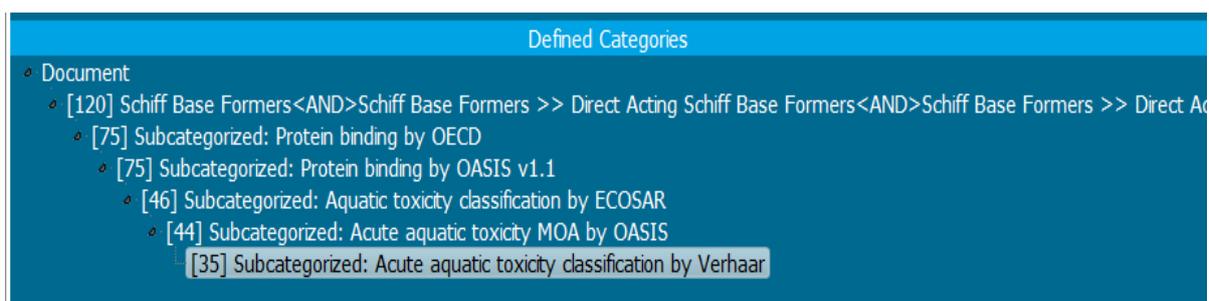


Figure 7.2: Results of sub-categorisation using the five primary profilers for 2,3-dimethylvaleraldehyde.

7.3 General conclusions regarding the sub-categorisation with the primary profilers

The examples discussed in this guidance document thus far highlight the importance of performing a series of sub-categorisations with each of the five primary profilers in turn. Such sub-categorisations are important to ensure that the resulting category consist of chemicals acting via a single mechanism of action related to acute aquatic toxicity. In addition, the sub-categorisations carried out using the endpoint-specific primary profilers ensure that only analogues that contain the same structural alerts present in the target chemical are included in the category. The choice of which order to apply the primary profilers is dependent on whether the initial profiling of the target chemical indicates it to be reactive or not.

8 Profiling examples which result in the ability to fill data-gaps

The OECD QSAR Toolbox is designed to enable data-gaps to be filled using the concept of chemical category formation. A prediction can be made for the target chemical using the existing data that is present in the various databases for the analogues identified as being part of the category. These predictions are typically made by either trend analysis or read across. This section provides examples

of data-gap filling using both of these methods for the mortality of *Pimephales promelas* as measured in a 96hr *in-vivo* assay.

8.1 Profiling and data gap filling for inert chemical by trend analysis

This section outlines how to profile propylbenzene in order to build a chemical category to allow a data-gap to be filled via trend analysis (Figure 8.1). The following example assumes the user is familiar with the workflow of the OECD QSAR Toolbox. Thus, multiple steps and keystrokes in the workflows are omitted with only key screenshots being included. All of the profiling steps detailed should be carried out with the 'differ from target by' option set to 'at least one category' unless otherwise stated. It is important that the user is familiar with the general approach to category formation within the OECD QSAR Toolbox [9]. It is recommended that the previous section of this guidance should have been attempted before attempting this section.

Species	Data Gap
Notemigonus crysoleucas	
Notropis atherinoides	
Notropis blennioides	
Notropis hudsonius	
Notropis ludibundis	
Oncorhynchus clarki	
Oncorhynchus gorbuscha	(2/7)
Oncorhynchus keta	
Oncorhynchus kisutch	(2/7)
Oncorhynchus mykiss	(9/21) M: 1.55 mg/L
Oncorhynchus nerka	(1/2)
Oncorhynchus tshawytscha	(1/1)
Onyzias latipes	(6/7)
Paracheirodon axelrodi	
Paralichthys olivaceus	
Parophrys vetulus	
Perca flavescens	
Perca sp.	
Phoxinus eos	
Phoxinus phoxinus	
Pimephales notatus	
Pimephales promelas	(12/63)
Platichthys flesus	(1/1)
Platichthys stellatus	
Pleuronectes americanus	
Pleuronectes platessa	
Pleuronectiformes	
Poecilia mexicana	
Poecilia reticulata	(7/16)
Poecilia vivipara	
Pomoxis nigromaculatus	

Figure 8.1: Data-gap present (shown in grey) for propylbenzene in the 96 hour *Pimephales promelas* assay.

8.1.1 Initial and secondary profiling

Initial category formation for the target chemical propylbenzene is as discussed in sections 5.1 and 45.2. This analysis suggested that target chemical was inert and thus exerted its toxicity via narcosis. The initial category consisted of 484 analogues. However, due to the large number of chemicals in the category further sub-categorisation to define the structural domain of the category is

recommended. This is achieved using two of secondary profilers listed in Table 6.1. As a general rule it is recommended to define the structural domain in this manner using the ‘organic functional group’ and ‘chemicals elements’ profilers as follows:

1. Sub-categorisation of the category of 484 chemicals using the ‘organic functional group’ profiler. This sub-categorisation results in a category of 36 chemicals.
2. Sub-categorisation of the category of 36 chemicals using the ‘chemical elements’ profiler. This sub-categorisation results in a category of 36 chemicals.

8.1.2 Data-gap filling via trend analysis

The sub-categorisation carried using the primary and secondary profilers result in a category of 36 chemicals. Trend analysis based on the 12 analogues in the category allows the data-gap in the 96 hour *Pimephales promelas* assay to be filled for propylbenzene. This analysis results in a predicted LC50 value for propylbenzene 7.43mg/l (Figure 8.2).

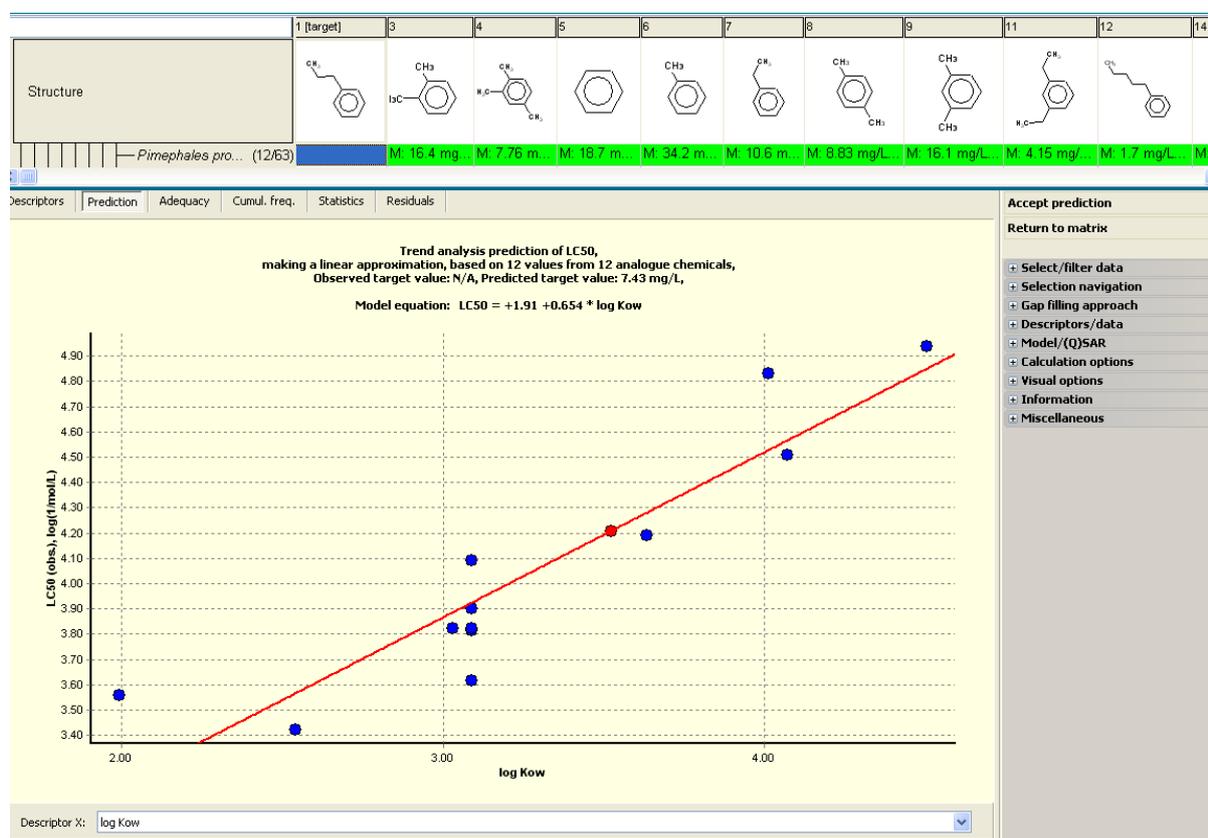


Figure 8.2: Trend analysis allowing an LC50 to be predicted for propylbenzene in the 96 hour *Pimephales promelas* assay.

8.2 Profiling and data gap filling for reactive chemical using read across

It is also possible to fill data-gaps using read across in the OECD QSAR Toolbox. This is particularly useful approach for predicting the toxicity of reactive target chemicals. In addition, read across is useful when the category contains only a small number of analogues. Consider the need to fill the data-gap present in the 96 hour *Pimephales promelas* assay for the target chemical formylcyclohexane.

8.2.1 Initial profiling using the primary profilers

The first step in the development of a chemical category for formylcyclohexane is to profile it using the primary profilers applicable to acute aquatic toxicity (Table 3.1 in section 3). The profiling results for formylcyclohexane suggest that toxicity for this chemical is likely to be due to Schiff base formation as a result of the aldehyde moiety (Figure 8.3).

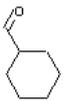
Filter endpoint tree...	1 [target]
Structure	
Substance Identity	
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	(1/1) M: 9.12 mg/L
Human Health Hazards	
Profile	
General Mechanistic	
Protein binding by OASIS v1.1	Schiff base formation Schiff base formation >> Schiff base formation with carbonyl compounds Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes
Protein binding by OECD	Schiff Base Formers Schiff Base Formers >> Direct Acting Schiff Base Formers Schiff Base Formers >> Direct Acting Schiff Base Formers >> Mono-carbonyls
Endpoint Specific	
Acute aquatic toxicity classification by Verhaar	Class 3 (unspecific reactivity)
Acute aquatic toxicity MOA by OASIS	Aldehydes
Aquatic toxicity classification by ECOSAR	Aldehydes (Mono)

Figure 8.3: Results of primary profiling for formylcyclohexane using the profilers available for acute aquatic toxicity in the OECD QSAR Toolbox V3.1.

8.2.2 Initial category formation and sub-categorisation using the primary profilers

The initial profiling results indicate that the most likely molecular initiating event for protein binding for formylcyclohexane is Schiff base formation due to the presence of aldehyde group. This information can be used to retrieve mechanistic analogues from the four applicable databases using the 'protein binding by OECD' profiler. This results in an initial category of 120 chemicals (including target chemical).

The following sub-categorisations are then required to ensure the category contains analogues acting via a single mechanism of action:

1. Sub-categorisation of the initial category of 120 chemicals using 'protein binding by OECD' profiler (the profiler that was used to develop the initial category). This sub-categorisation results in a category of 75 chemicals.
2. Sub-categorisation of the category of 75 chemicals using the 'protein binding by OASIS V1.1' profiler. This sub-categorisation results in a category of 75 chemicals.
3. Sub-categorisation of the category of 75 chemicals using the 'aquatic toxicity classification by ECOSAR' profiler. This sub-categorisation results in a category of 46 chemicals.
4. Sub-categorisation of the category of 46 chemicals using the 'acute aquatic toxicity MOA by OASIS' profiler. This sub-categorisation results in a category of 44 chemicals.

- Sub-categorisation of the category of 44 chemicals using the 'acute aquatic toxicity classification by Verhaar' profiler. This sub-categorisation results in a category of 35 chemicals.

8.2.3 Empiric sub-categorisation using the secondary profilers

The final stage in the development of a robust chemical category suitable for data gap filling is to ensure that the structural domain is well defined. One method to do this involves sub-categorising using a combination of the empiric profilers (removing all chemicals from the category that contain elements and functional groups not present in the target chemical). This sub-categorisation process is analogous to that carried using the primary profilers. The following sub-categorisations should be carried out on the category of 35 chemicals generated in section 8.2.2:

- Sub-categorisation of the category of 35 chemicals using the 'organic functional group' profiler. This sub-categorisation results in a category of 16 chemicals.
- Sub-categorisation of the category of 16 chemicals using the 'chemical elements' profiler. This sub-categorisation results in a category of 16 chemicals.

8.2.4 Data-gap filling via read across

The sub-categorisation carried using the primary and secondary profilers result in a category that has a well-defined mechanistic (defined as a result of the sub-categorisation in section 8.2.2) and structural (defined in section 8.2.3) domains. The utility of defining the structural domain can be seen by inspecting the analogues in the category, which are all simple aliphatic aldehydes. This category can now be used to fill the data-gap that is present in the 96 hour *Pimephales promelas* assay for the target chemical formylcyclohexane (Figure 8.4). A read across prediction based on the five closest analogues (in terms of hydrophobicity) results in LC50 prediction for *Pimephales promelas* of 17.5 mg/l.

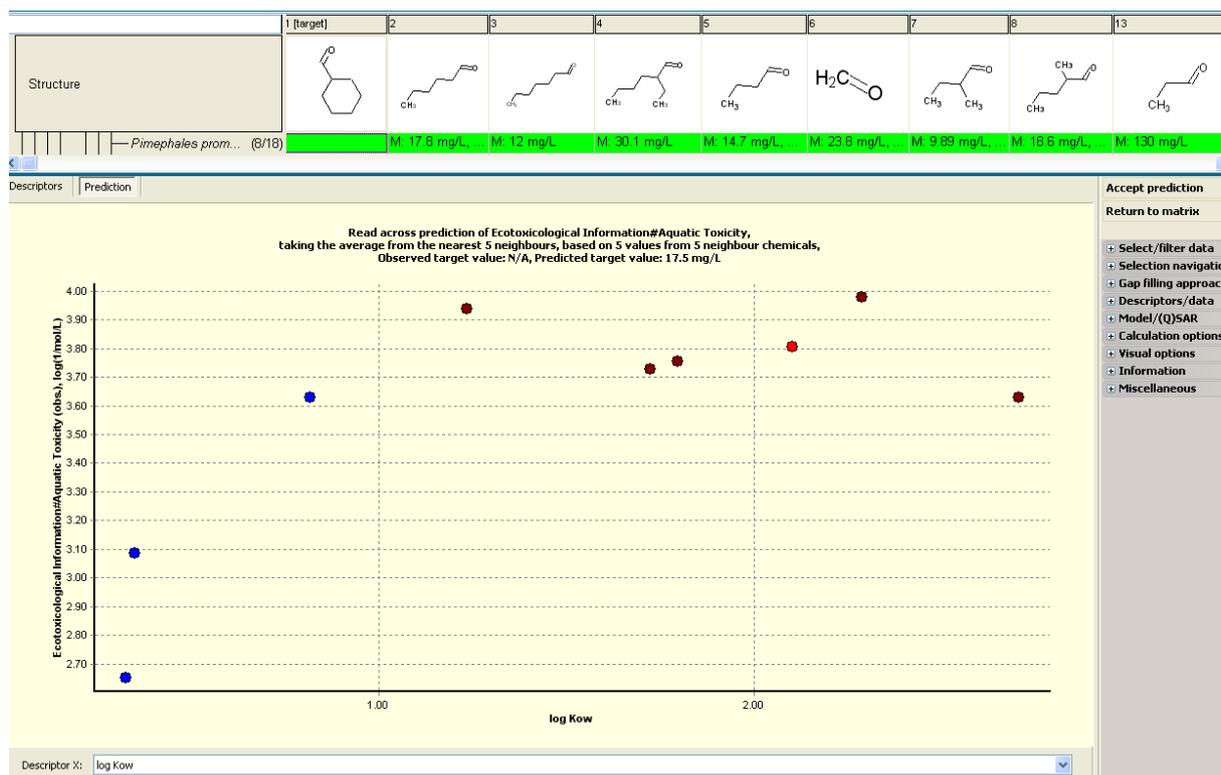


Figure 8.4: Read across prediction made for the *Pimephales promelas* 96hr LC50 endpoint for formylcyclohexane.

9 General approach for the development of categories for acute aquatic toxicity endpoints

The following outline can be considered a good general approach for the development of chemical categories for acute aquatic toxicity endpoints. These instructions are summarised in a flow chart (Figure 9.1).

1. Profile the target chemical using the five initial profilers relevant to acute aquatic toxicity endpoints.
2. Define the initial chemical category by profiling the four relevant databases to acute aquatic toxicity using an appropriate profiler depending on whether the chemical is reactive or not. It is recommended that for inert chemicals the endpoint-specific 'aquatic toxicity classification by ECOSAR' profiler or 'acute aquatic toxicity MOA by OASIS' profiler should be used for developing the initial category. In contrast, if the target chemical is profiled as reactive then one of the mechanistic protein binding profilers should be chosen to form the initial category (it doesn't matter which of the two are selected).
3. Sub-categorisation using a combination of the remaining primary profilers to eliminate chemicals that contain additional potential structural alerts related to the identified mechanism. The use of endpoint specific profilers in this step identifies structural alerts that have toxicological data associated with them.
4. Sub-categorisation using the secondary profilers in order to define the structural domain. One should use a combination of the empiric profilers (it is recommended to use the 'organic functional group' and 'chemical elements' profilers in the majority of cases) to restrict the structural domain of the category so that it is similar to that of the target chemical. The guiding principle should be towards the descriptor that one will use in any subsequent read across or trend analysis. This helps keep any predictions made using read across or trend analysis as transparent as possible. It is worth recalling that sometimes this profiling step requires the inclusion of analogous containing simple organic functional groups that are not present in the target chemical.
5. Always ensure that the data used in any read across or trend analysis predictions are quality checked and that unusual or outlying data within a category are investigated before use. Please remember that the OECD is not responsible for the quality of the data within the OECD QSAR Toolbox.
6. Create the appropriate reporting format in The Toolbox (see guidance [9]).

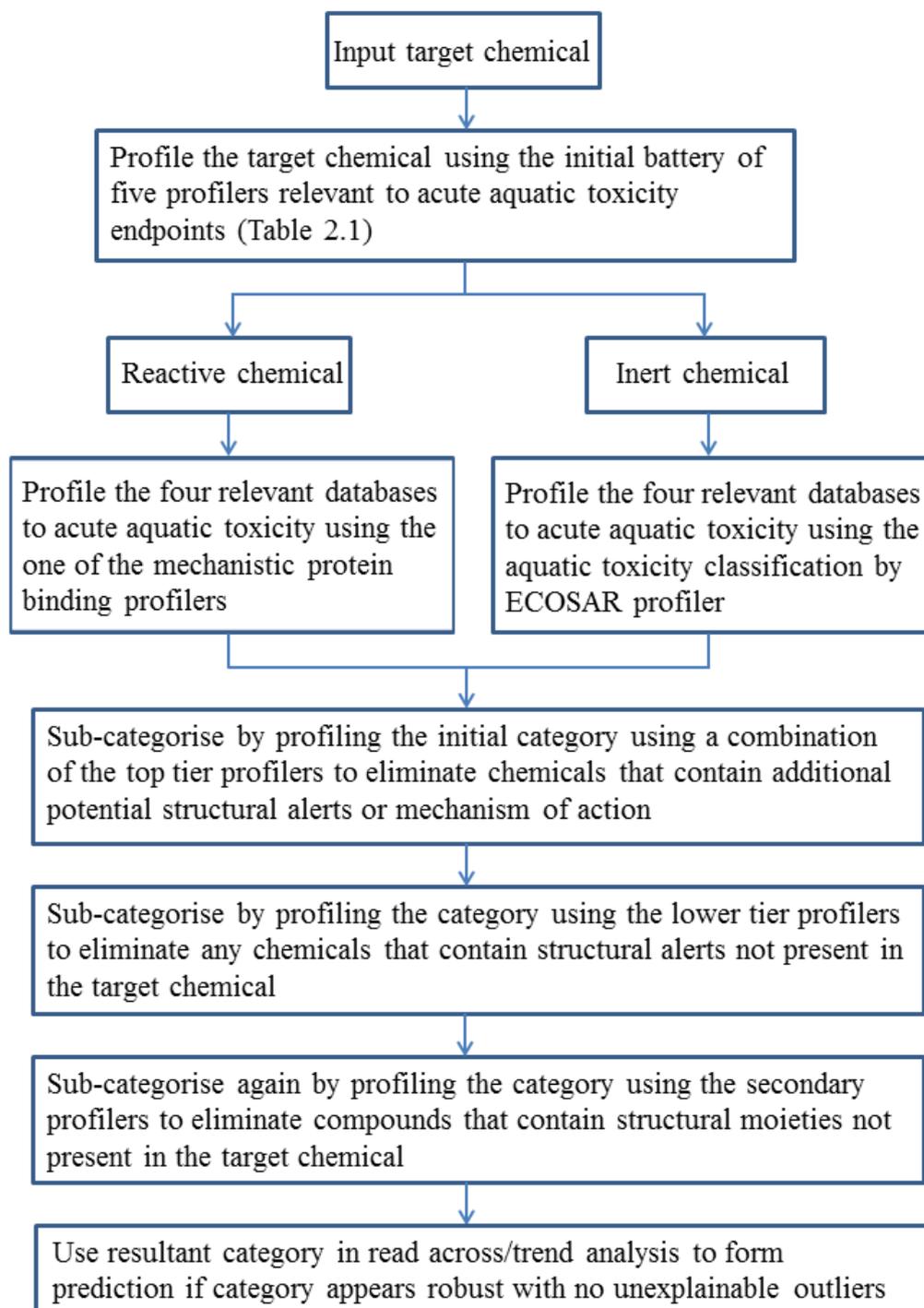


Figure 9.1: General scheme for category formation for acute aquatic toxicity endpoints

10 References

1. Globally Harmonised System of Classification and Labelling of Chemicals (GHS)
<http://www.hse.gov.uk/ghs/>
2. Russom, C.L. Bradbury, S.P. Broderius, S.J. Hammermeister, D.E. and Drummond, R.A. 1997. Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (*Pimephales promelas*), *Environ. Toxicol. Chem.* 16:948–967.
3. Verhaar, H. J. M. van Leeuwen, C. J. Hermens, J. L. M. 1992. Classifying environmental pollutants. 1: structure-activity relationships for prediction of aquatic toxicity. *Chemosphere* 25:471-491.
4. Dimitrov, S.D. Low, L.K. Patlewicz, G.Y. Kern, P.S. Dimitrova, G. D. Comber, M.H.I. Phillips, R.D. Niemela, J. Bailey, P.T. Mekenyan, O.G. 2005. Skin sensitization: Modeling based on skin metabolism simulation and formation of protein conjugates. *Internat. J. Toxicol.*, 24:189-204.
5. Enoch, S.J., Ellison, C.M., Schultz, T.W., Cronin, M.T. 2011. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit Rev Toxicol.* 41(9):783-802.
6. Enoch, S.J., Hewitt, M., Cronin, M.T.D., Azam, S., Madden, J.C. 2008. Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree, *Chemosphere* 73(3): 243–248.
7. Schultz, T.W., Sinks, G.D., Cronin, M.T.D. 1997. Structure–toxicity evaluation of phenols: a mechanism of action approach, in F.C. Fredenslund, G. Schüürmann (Eds.), *Quantitative Structure–Activity Relationships in Environmental Sciences – VII*, SETAC Press, Pensacola, FL, USA. 329–342
8. OECD. (2007) *Guidance on Grouping of Chemicals*. (available from: http://www.oecd.org/LongAbstract/0,3425,en_2649_34379_39404999_119829_1_1_37407,00.html).
9. OECD. (2010) *The OECD (Q)SAR Application Toolbox Guidance Documents* available from http://www.oecd.org/document/54/0,3343,en_2649_34379_42923638_1_1_1_1,00.html#Guidance_Documents_and_Training_Materials_for_Using_the_Toolbox.

OECD

2, rue André Pascal
75775 Paris Cedex 16
France

Tel.: +33 1 45 24 82 00

Fax: +33 1 45 24 85 00